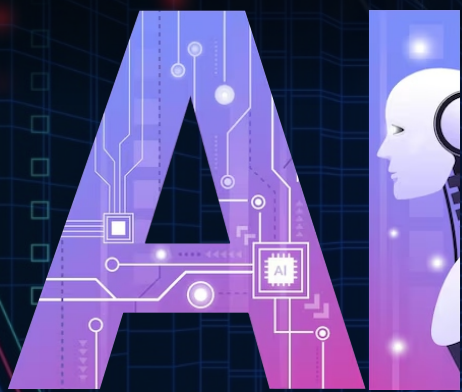


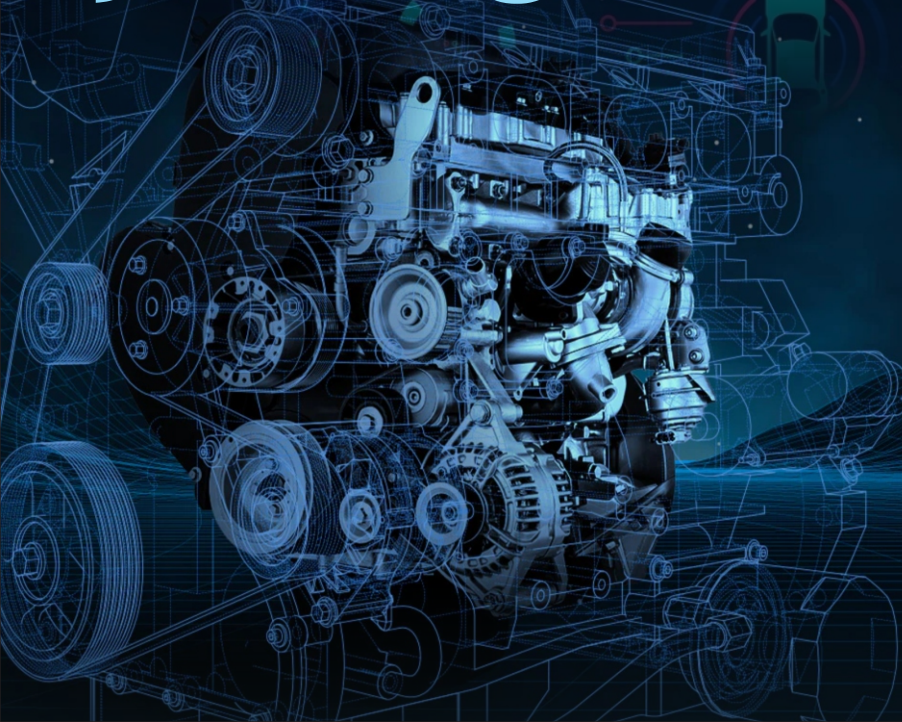
Dr. Ir. Agus Wibowo, M.Kom, M.Si, MM



Artificial Intelligence

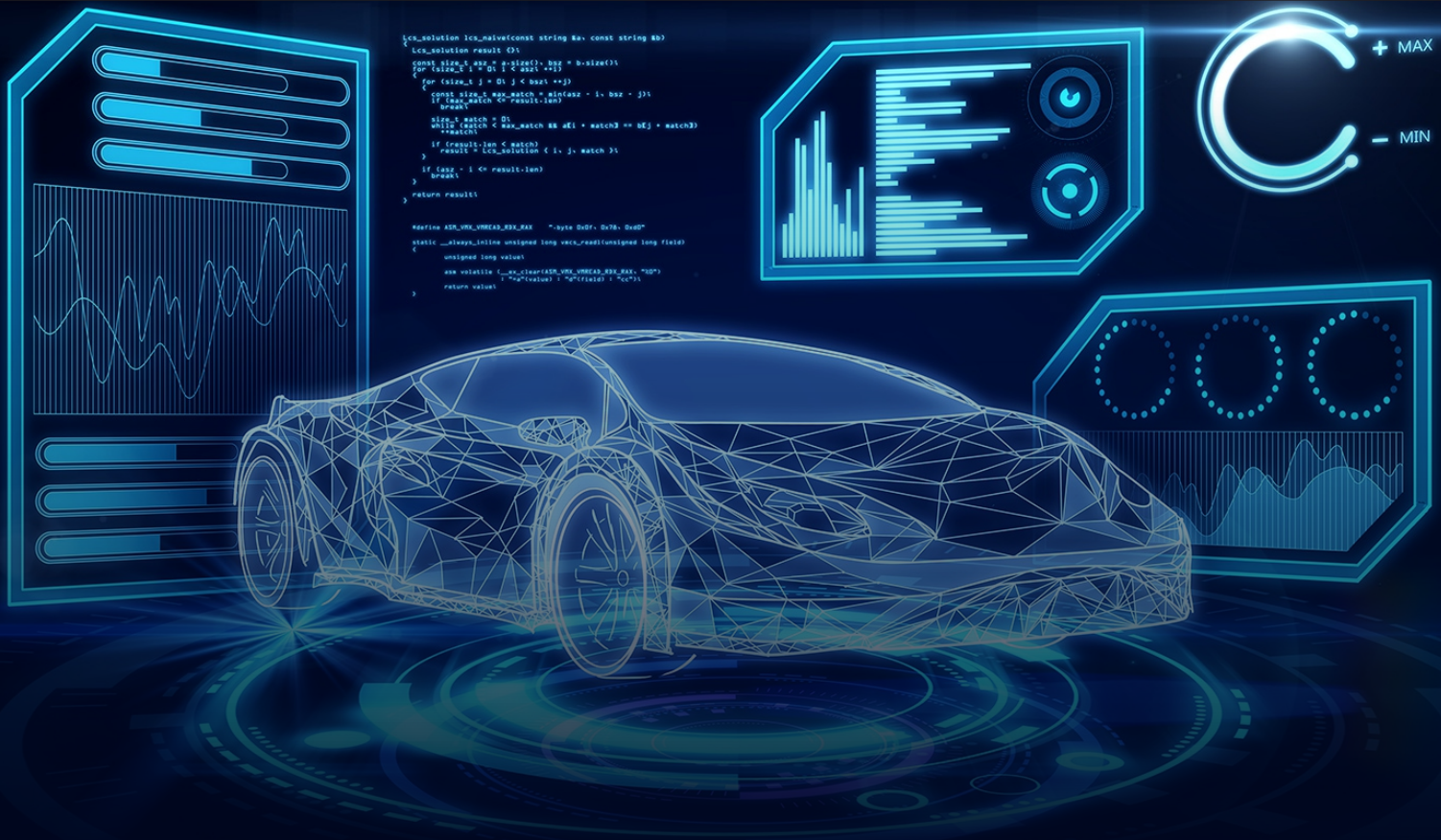
dalam

Jaringan Otonom

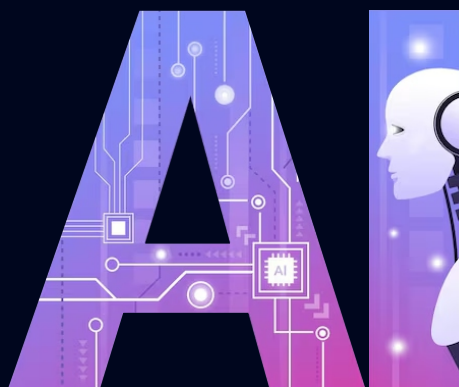


YAYASAN PRIMA AGUS TEKNIK





Dr. Ir. Agus Wibowo, M.Kom, M.Si, MM



Artificial Intelligence dalam Jaringan Otonom



YAYASAN PRIMA AGUS TEKNIK

PENERBIT :
YAYASAN PRIMA AGUS TEKNIK
 Jl. Majapahit No. 605 Semarang
 Telp. (024) 6723456. Fax. 024-6710144
 Email : penerbit_ypat@stekom.ac.id

ISBN 978-634-7227-48-5 (PDF)



9 786347 227485

Artificial Intelligence (AI) dalam Jaringan Otonom

Penulis :

Dr. Ir. Agus Wibowo, M.Kom, M.Si, MM

ISBN : 978-634-7227-48-5 (PDF)

Editor :

Dr. Joseph Teguh Santoso, S.Kom., M.Kom.

Penyunting :

Dr. Mars Caroline Wibowo. S.T., M.Mm.Tech

Desain Sampul dan Tata Letak :

Irdha Yuniarto, S.Ds., M.Kom

Penebit :

Yayasan Prima Agus Teknik Bekerja sama dengan
Universitas Sains & Teknologi Komputer (Universitas STEKOM)

Anggota IKAPI No: 279 / ALB / JTE / 2023

Redaksi :

Jl. Majapahit no 605 Semarang

Telp. 08122925000

Fax. 024-6710144

Email : penerbit_ypat@stekom.ac.id

Distributor Tunggal :

Universitas STEKOM

Jl. Majapahit no 605 Semarang

Telp. 08122925000

Fax. 024-6710144

Email : info@stekom.ac.id

Hak cipta dilindungi undang-undang

Dilarang memperbanyak karya tulis ini dalam bentuk dan dengan cara
apapun tanpa ijin dari penulis

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat-Nya sehingga buku “*Kecerdasan Buatan untuk Jaringan Otonom*” ini dapat terselesaikan dengan baik. Perkembangan teknologi informasi dan komunikasi yang pesat telah mendorong transformasi luar biasa dalam cara jaringan dibangun, dioperasikan, dan diamankan. Buku ini hadir sebagai sebuah karya yang menyeluruh dan mendalam, mengupas tuntas bagaimana kecerdasan buatan (*Artificial Intelligence/AI*) berkembang dari sebuah konsep menjadi komponen utama dalam menciptakan jaringan otonom yang cerdas, efisien, dan adaptif.

Buku ini membahas secara sistematis perjalanan evolusi jaringan, mulai dari era otomatisasi tradisional hingga penerapan AI dalam operasi dan keamanan jaringan. Dengan gaya penyampaian yang runtut dan kaya akan ilustrasi konseptual, pembaca diajak untuk memahami setiap lapisan teknologi dan proses yang membentuk jaringan masa depan. Bab demi bab dalam buku ini disusun untuk memberikan gambaran komprehensif sebagai berikut:

Pada Bab 1, pembaca dikenalkan dengan peran strategis AI dalam otomatisasi dan keamanan jaringan. Penjelasan dimulai dari tiga generasi transformasi jaringan, menggambarkan perubahan fundamental yang dialami industri komunikasi. Bab ini menekankan bagaimana AI menjadi kunci dalam mengubah jaringan menjadi lebih cerdas dengan pendekatan data-driven yang membuka peluang menciptakan “dunia yang sama sekali baru” dalam konteks jaringan otonom.

Bab 2 mengulas fondasi kecerdasan buatan termasuk pembelajaran mesin (*machine learning*) dan pembelajaran mendalam (*deep learning*). Sejarah singkat AI disajikan untuk memberi konteks perkembangan teknologi ini. Selain itu, bab ini membedah berbagai metode pembelajaran mesin dan mengupas janji serta tantangan yang dihadapi AI saat ini dan pada masa depan.

Kemudian, Bab 3 membawa pembaca memahami pergeseran paradigma jaringan dari perangkat keras kaku menuju jaringan berbasis perangkat lunak. Dengan pemaparan sejarah dan perkembangan jaringan yang ditentukan perangkat lunak (SDN), diikuti penjelasan virtualisasi fungsi jaringan, bab ini menggambarkan integrasi teknologi yang mendukung kecerdasan buatan dalam operasional jaringan.

Keamanan jaringan menjadi fokus utama di Bab 4, yang menjelaskan konsep blockchain dan kriptografi sebagai alat penting dalam memperkuat keamanan sistem jaringan otonom. Penjelasan mendalam mengenai mekanisme konsensus dan aplikasi blockchain untuk mendukung AI memberikan wawasan kritis bagi pembaca yang ingin memahami hubungan erat antara keamanan dan kecerdasan buatan. Di Bab 5, buku ini mengajak pembaca melihat secara futuristik pembangunan jaringan otonom masa depan. Karakteristik pengguna, operasi, tantangan, dan contoh aplikasi nyata jaringan otonom diuraikan secara jelas, sehingga pembaca dapat memahami gambaran besar dan detil teknisnya sekaligus.

Bab 6 membahas peran open source dan teknologi cloud yang menjadi tulang punggung jaringan otonom modern. Pembahasan teknologi, filosofi open source, serta bagaimana operator cloud memanfaatkan AI memperlihatkan sinergi antara berbagai teknologi yang mendukung ekosistem jaringan pintar. Terkait komputasi awan dan desain layanan, Bab 7 menjelaskan peran mikro layanan dalam membangun jaringan cloud native dan memperluas kapabilitas AI. Bab ini juga mengulas masa depan jaringan AI berbasis cloud dan bagaimana perusahaan dapat memanfaatkan teknologi ini untuk berada di garis depan inovasi.

Sementara itu, Bab 8 memberi sentuhan teknis mendalam tentang platform perangkat lunak dan perangkat keras AI. Penjabaran berbagai kerangka kerja, seni ilmu data, dan karakteristik beban kerja yang dihadapi dalam AI membuat bab ini sangat penting bagi pembaca yang tertarik dengan aspek pengembangan teknis AI di jaringan. Bab 9 dan Bab 10 membedah lebih jauh penerapan AI pada operasi jaringan dan keamanan siber. Dari konsep zero-touch operation hingga alat analitik ancaman yang mengandalkan AI, bab ini menguraikan bagaimana kecerdasan buatan merevolusi cara jaringan dikelola dan diamankan secara real-time.

Transformasi digital di dunia perusahaan menjadi fokus Bab 11, yang menjelaskan efek AI pada arsitektur jaringan perusahaan dan bagaimana jaringan otonom dapat mempercepat proses transformasi digital di berbagai sektor industri. Perkembangan teknologi akses terbaru seperti 5G dan IoT diulas pada Bab 12, dengan penekanan pada penerapan AI untuk meningkatkan performa dan otomatisasi jaringan generasi berikutnya. Kasus penggunaan yang konkret menambah nilai praktis bab ini.

Bab-bab berikutnya, Bab 13, Bab 14, dan Bab 15, melengkapi buku dengan pengupasan tren terkini dalam infrastruktur dan lalu lintas jaringan, dampak AI bagi layanan pelanggan, serta batas-batas baru penerapan AI dalam jaringan otonom yang melibatkan aspek teknologi, proses, dan pengaruhnya terhadap sumber daya manusia di industri telekomunikasi.

Dengan cakupan yang luas dan terperinci, buku ini diharapkan menjadi pegangan penting bagi praktisi TI, peneliti, akademisi, dan siapa saja yang ingin mendalami kecerdasan buatan khususnya dalam bidang jaringan otonom. Semoga buku ini mampu memberi wawasan, inspirasi, serta landasan ilmiah dan praktis yang kuat untuk menghadapi tantangan masa depan jaringan komunikasi. Terima kasih kepada semua pihak yang telah berkontribusi dalam penulisan dan penerbitan buku ini.

Selamat & Semangat Membaca...!!!

Semarang, Oktober 2025

Penulis

Dr. Agus Wibowo, M.Kom, M.Si, MM

DAFTAR ISI

KATA PENGANTAR.....	ii
DAFTAR ISI.....	iv
BAB 1 PERAN AI DALAM OTOMATISASI DAN KEAMANAN JARINGAN	1
1.1 Pendahuluan.....	1
1.2 Tiga Generasi Transformasi Jaringan	1
1.3 Dari Otomatisasi Ke Kecerdasan Buatan	7
1.4 Jaringan Otonom Berdasarkan Data.....	13
1.5 Dunia Yang Sama Sekali Baru	18
BAB 2 PEMBELAJARAN MESIN, DEEP LEARNING, DAN AI	23
2.1 Pendahuluan.....	23
2.2 Kecerdasan Buatan Sejarah Singkat	23
2.3 Cara-Cara Pembelajaran Mesin	26
2.4 Pembelajaran Mendalam Janji Di Balik Hype	33
2.5 Kecerdasan Buatan Masa Kini Dan Masa Depan.....	40
BAB 3 PERGESERAN KE JARINGAN BERBASIS PERANGKAT LUNAK.....	45
3.1 Pendahuluan.....	45
3.2 Sejarah Jaringan Yang Ditentukan Perangkat Lunak.....	45
3.3 Jalan Menuju Virtualisasi Fungsi Jaringan	52
3.4 Menggabungkan Semuanya	61
3.5 Dukungan Untuk Kecerdasan Buatan Jaringan.....	62
3.6 Ringkasan.....	63
BAB 4 BLOCKCHAIN DAN KRIPTOGRAFI UNTUK KEAMANAN	64
4.1 Pendahuluan.....	64
4.2 Tinjauan Umum Blockchain.....	64
4.3 Konsensus.....	69
4.4 Mendukung Kecerdasan Buatan Dengan Blockchain	73
4.5 Menggunakan Blockchain	75
BAB 5 MEMBANGUN JARINGAN OTONOM MASA DEPAN	78
5.1 Pendahuluan.....	78
5.2 Jaringan Masa Depan	78
5.3 Pengguna Jaringan Masa Depan	85
5.4 Operasi Jaringan Masa Depan	87
5.5 Tantangan Jaringan Masa Depan.....	90
5.6 Aplikasi Jaringan Otonom	93
5.7 Ringkasan.....	95
BAB 6 OPEN SOURCE DAN CLOUD DI JARINGAN OTONOM	96
6.1 Zen Of Open Source.....	96

6.2	Dari Potongan Ke Arsitektur	102
6.3	Teknologi.....	105
6.4	Operator Cloud, Kecerdasan Buatan, Dan Jaringan Otonom	126
6.5	Tantangan Ke Depan	132
6.6	Membangun Jaringan Otonom Dengan Open Source	135
BAB 7	MESIN DAN CLOUD NATIVE DENGAN MIKRO LAYANAN	137
7.1	Pendahuluan.....	137
7.2	Kembangkitan Komputasi Awan Dan Komunikasi.....	138
7.3	Jaringan Pembelajaran Layanan Mikro.....	148
7.4	Masa Depan Jaringan AI Berbasis Cloud	155
7.5	Menatap Ke Depan: Perusahaan Di Tepi	156
BAB 8	PLATFORM PERANGKAT LUNAK DAN KERAS AI.....	157
8.1	Pendahuluan.....	157
8.2	Kecerdasan Buatan	157
8.3	Seni Ilmu Data.....	162
8.4	Kerangka Kerja Perangkat Lunak Kecerdasan Buatan	164
8.5	Pilihan Perangkat Keras Kecerdasan Buatan	169
8.6	Karakteristik Beban Kerja	174
8.7	Ringkasan.....	178
BAB 9	KECERDASAN BUATAN UNTUK OPERASI JARINGAN	180
9.1	Pendahuluan.....	180
9.2	Pengertian Operasi Jaringan.....	181
9.3	Jaringan Zero-Touch: Otomatisasi & Kecerdasan	188
9.4	Mewujudkan Visi: Perjalanan Menuju Masa Depan	199
9.5	Dampak Pada Peran Dan Tanggung Jawab Operasi	220
9.6	Kesimpulan	223
BAB 10	KECERDASAN BUATAN UNTUK KEAMANAN SIBER	224
10.1	Pendahuluan.....	224
10.2	Pertimbangan AI Dan Otomatisasi Keamanan	228
10.3	Penerapan AI Dan MI Untuk Keamanan.....	239
10.4	Aplikasi Dalam Analitik Ancaman	243
10.5	Ringkasan.....	256
BAB 11	KECERDASAN BUATAN UNTUK JARINGAN PERUSAHAAN	257
11.1	Pendahuluan.....	257
11.2	Arsitektur Perusahaan Untuk Transformasi Digital	258
11.3	Transformasi Perusahaan Dan Kecerdasan Buatan	259
11.4	Transformasi Jaringan Perusahaan Dan Kecerdasan Buatan.....	268
11.5	Dampak AI Pada Jaringan Dan Arsitekturnya	274
11.6	Ringkasan.....	275
BAB 12	KECERDASAN BUATAN UNTUK 5G+ DAN IOT	277
12.1	Pendahuluan.....	277

12.2	Latar Belakang Dan Motivasi Evolusi Teknologi.....	277
12.3	Arsitektur Jaringan Dan Akses	286
12.4	Kecerdasan Buatan Untuk 5G+.....	299
12.5	Kecerdasan Buatan Dan Otomatisasi Jaringan Baru.....	300
12.6	Kerangka Kecerdasan Buatan Dan Otomatisasi.....	301
12.7	Kasus Penggunaan.....	302
BAB 13	AI UNTUK INFRASTRUKTUR DAN LALU LINTAS.....	306
13.1	Pendahuluan.....	306
13.2	Tren, Kesenjangan, Dan Peluang	306
13.3	Tren Teknologi Yang Memungkinkan Jaringan Cerdas	309
13.4	Penerapan Ai Dalam Pemantauan Dan Optimasi Jaringan.....	311
13.5	Menerapkan Kecerdasan Buatan Pada 5G	322
13.6	Ringkasan.....	333
BAB 14	KECERDASAN BUATAN UNTUK LAYANAN PELANGGAN.....	334
14.1	Pendahuluan.....	334
14.2	Dampak Jaringan Otonom Pada Pengalaman Pelanggan.....	334
14.3	Pengalaman Digital Yang Bertransformasi.....	336
14.4	Membangun Sistem AI Untuk Layanan Pelanggan.....	338
14.5	Meningkatkan Layanan Pelanggan Operator	345
14.6	Ringkasan.....	346
BAB 15	BATAS BARU AI UNTUK JARINGAN OTONOM	347
15.1	Pendahuluan.....	347
15.2	Fungsi Dasar Jaringan	348
15.3	Menyatukan Semuanya	354
15.4	Pilar Transformasi: Teknologi, Proses, Dan Manusia	359
15.5	Evolusi Proses Di Industri Telekomunikasi	359
15.6	Dampak Terhadap Karyawan Di Industri Ini	361
15.7	Kasus Penggunaan Jaringan Otonom	364
DAFTAR PUSTAKA		370

BAB 1

PERAN AI DALAM OTOMATISASI DAN KEAMANAN JARINGAN

1.1 PENDAHULUAN

Bab ini merupakan pengantar teknologi, peluang, dan tantangan jaringan otonom, sebuah produk dari integrasi kecerdasan buatan (AI) dengan jaringan yang ditentukan perangkat lunak (SDN). Bab ini memberikan gambaran umum tentang pentingnya otomatisasi untuk mendorong personalisasi, kecerdasan, dan elastisitas dalam jaringan. Karakteristik ini mendorong efisiensi operasional, peningkatan keamanan, dan memungkinkan skalabilitas spektrum baru layanan latensi ultra-rendah untuk mendukung *Internet of Things* (IoT), mobil otonom, realitas virtual (VR), dan realitas tertambah (AR). Bagian 1.2 menyajikan ikhtisar transformasi jaringan dari jaringan vendor-locked proprietary menjadi SDN terbuka yang berjalan di cloud, dan segera, menjadi jaringan otonom yang mengadopsi AI. Bagian 1.3 memperkenalkan peran AI dalam mendorong jaringan yang elastis, mampu memulihkan diri sendiri, dan aman. Bagian 1.4 menjelaskan pentingnya data yang mendorong jaringan otonom. Bagian 1.5 menjelaskan dunia aplikasi baru yang akan dimungkinkan oleh jaringan otonom.

1.2 TIGA GENERASI TRANSFORMASI JARINGAN

Kembali ke Dasar Jaringan Komunikasi

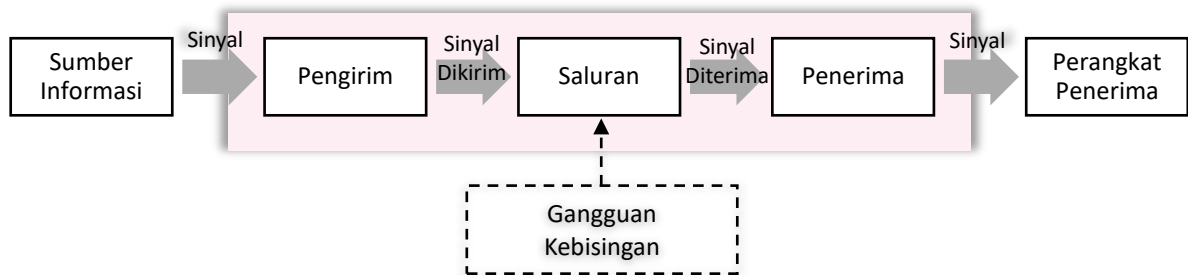
Meskipun bukti awal komunikasi diamati di Afrika, Amerika, dan Tiongkok melalui penggunaan sinyal asap dan drum, terobosan nyata terjadi pada abad ke-19 melalui penemuan telegraf listrik oleh Francis Ronalds. Telegraf adalah jaringan komunikasi pertama dari jenisnya yang memungkinkan transmisi sinyal jarak pendek. Pada tahun 1837, Samuel Morse, bersama asistennya Alfred Vail, menemukan kode Morse untuk transmisi informasi melalui telegraf listrik. Telegraf tersebut menggunakan pemancar, berupa perangkat pengirim atau enkoder, yang memungkinkan pengiriman serangkaian titik dan garis untuk mewakili huruf dan karakter. Di sisi penerima, dekoder digunakan untuk menerima informasi yang kemudian ditafsirkan oleh operator.

Jaringan komunikasi pada dasarnya adalah sistem yang saling terhubung yang memfasilitasi pertukaran informasi. Di balik jaringan komunikasi ini, terdapat beberapa komponen teknologi dasar. Sumber informasi menghasilkan sinyal, atau pesan, yang kemudian dikirim ke pemancar. Sinyal ini dapat berupa video, suara, atau teks, dan dihasilkan melalui perangkat, seperti kamera, telepon, atau papan ketik. Pemancar memproses sinyal ini dengan melakukan amplifikasi dan modulasi, dan kemudian menghasilkan keluaran yang sesuai untuk transmisi.

Sinyal dalam jaringan komunikasi awal ditransmisikan dalam bentuk analog. Berkat Claude Shannon dan teori informasinya, sebagian besar sinyal saat ini dikonversi ke bentuk digital digit biner satu dan nol. Sinyal yang dikodekan dapat ditransmisikan melalui saluran kabel, seperti kabel atau serat optik, atau saluran nirkabel, seperti Wi-Fi, BlueTooth, 4GLTE,



atau 5G. Saluran nirkabel melibatkan transmisi gelombang elektromagnetik dari satu antena ke antena lainnya (Gambar 1.1).



Gambar 1.1 Jaringan komunikasi dasar.

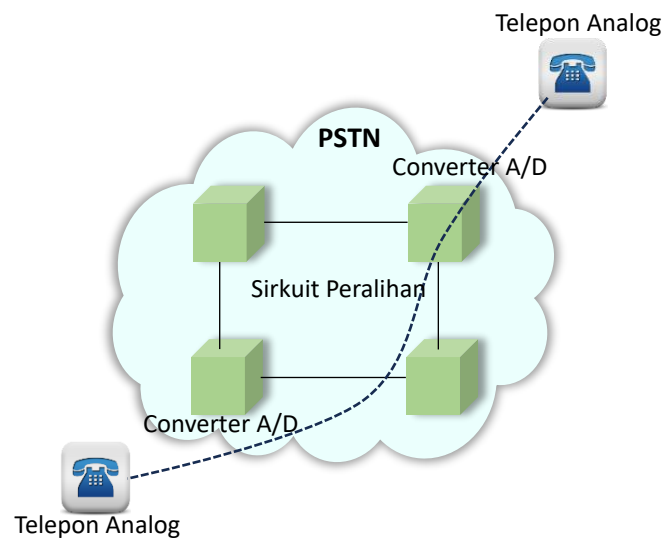
Mempertahankan sinyal berkualitas tinggi melalui saluran transmisi sangat penting untuk komunikasi yang efektif. Kualitas biasanya dipengaruhi oleh interferensi atau derau. Derau dapat berasal dari lingkungan, industri, atau elektronik dan disebabkan oleh digitalisasi atau sirkuit. Pengodean sinyal dan pemilihan saluran yang tepat dapat membantu memaksimalkan rasio sinyal terhadap derau seperti yang ditunjukkan oleh Shannon. Di sisi penerima, sinyal didemodulasi dan diubah menjadi bentuk yang sesuai untuk perangkat penerima, yang dapat berupa pengeras suara atau telepon.

Jaringan komunikasi modern didasarkan pada penemuan telepon pada tahun 1876 oleh Alexander Graham Bell dan Thomas Watson. Telepon digunakan di ujung pemancar dan penerima, dengan papan saklar terpusat yang terhubung ke masing-masing telepon. Penemuan ini melahirkan Jaringan Telepon Saklar Publik (PSTN).

Dari Lama ke Baru

Jaringan telekomunikasi, seperti yang dibayangkan oleh Alexander Bell, pada dasarnya merupakan bentuk telegraf dengan suara di kedua ujungnya. Jaringan ini mencakup transmisi arus listrik melalui kabel listrik yang menduplikasi gelombang suara. Penemuan circuit switching merupakan penyempurnaan dari penemuan Bell, yang memungkinkan suara dalam bentuk analog untuk menempuh jarak yang lebih jauh. Jaringan circuit switching (Gambar 1.2) mencakup simpul switching yang memfasilitasi transmisi suara dari pengirim ke penerima. Simpul switching tambahan biasanya dikembangkan untuk meningkatkan keandalan jaringan.





Gambar 1.2 PSTN dengan sakelar sirkuit.

Meskipun jaringan circuit switching cocok untuk komunikasi suara, jaringan ini menghadirkan beberapa tantangan. Tantangan pertama adalah inefisiensi sistem transmisi suara. Panggilan suara memerlukan kapasitas kanal khusus untuk keseluruhan panggilan. Kapasitas tersebut terbuang sia-sia meskipun tidak digunakan. Melakukan panggilan melalui simpul switching juga memerlukan penundaan yang lama.

Pada awal 1970an, PSTN mulai bertransformasi menuju jaringan packet-switched. Biasanya, pesan dibagi menjadi serangkaian paket yang lebih kecil beserta informasi kontrol, lalu dirutekan ke tujuan yang diinginkan. Dalam desain ini, paket dapat diantrekan dan ditransmisikan secepat mungkin berdasarkan prioritas. Koneksi antar node dapat dibagikan secara dinamis, sehingga memungkinkan data ditransmisikan bahkan saat lalu lintas padat.

Mengelola, memantau, dan mengoperasikan jaringan komunikasi merupakan pekerjaan yang sangat kompleks dan mahal. Fungsi jaringan, seperti router, sakelar, dan firewall, mencakup perangkat keras dan perangkat lunak khusus yang rentan terhadap kegagalan dan serangan keamanan. Perubahan atau penggantian peralatan ini memerlukan pekerjaan manual di lokasi tertentu. Namun demikian, dengan pertumbuhan lalu lintas yang linear dan dapat diprediksi hingga pertengahan 2000an, operator merasa puas dalam mengoperasikan dan memelihara jaringan yang berpusat pada perangkat keras ini.

Pada tahun 2007, jaringan komunikasi mengalami ledakan besar dengan lahirnya iPhone dan perangkat pintar. Tiba-tiba, lalu lintas mulai bergeser dari yang pada dasarnya suara dan beberapa data, menjadi sebagian besar data video. Antara tahun 2007 dan 2016, jaringan AT&T mengalami peningkatan lalu lintas sebesar 250.000%. Peningkatan permintaan eksponensial ini diperkirakan akan terus berlanjut seperti itu selama 10 tahun ke depan, yang memberikan beban yang signifikan pada jaringan komunikasi dan membutuhkan investasi modal yang substansial.

PSTN berbasis Protokol Internet (IP) dengan pemrosesan paket dan dengan peralatan jaringan yang berpusat pada perangkat keras "lama" tidak dapat lagi mendukung pertumbuhan lalu lintas yang eksponensial. Dengan tekanan pasar untuk menyediakan paket



tak terbatas kepada konsumen, menjadi jelas bahwa pendekatan lama tidak berkelanjutan, dan operator perlu menggunakan teknologi generasi baru yang memberikan fleksibilitas, elastisitas, dan otomatisasi saat mengelola fungsi jaringan. Beralih ke SDN dan virtualisasi fungsi jaringan (NFV) diperlukan untuk mengelola biaya operasi yang meningkat dan mendukung pertumbuhan lalu lintas yang eksponensial.

Perangkat Lunak Menguasai Dunia

Pada tahun 2011, Marc Andreessen menulis esainya yang terkenal, "Mengapa Perangkat Lunak Menguasai Dunia", di *The Wall Street Journal*. Perangkat lunak dan virtualisasi mungkin bukan hal baru bagi perusahaan web, termasuk Amazon, Google, dan Microsoft, tetapi merupakan cara berpikir baru bagi perusahaan jaringan dan komunikasi. Sejak artikel ini diterbitkan, operator telah menata ulang jaringan mereka sebagai SDN untuk mengelola peningkatan biaya operasional dan layanan mereka dengan lebih baik.

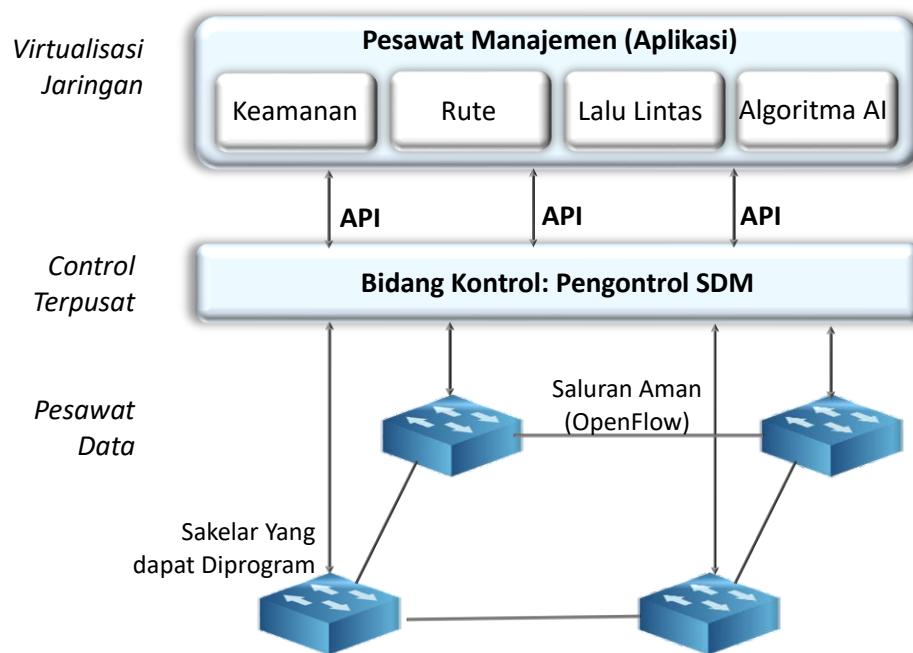
Premis utama di balik SDN adalah memisahkan bidang kendali, atau manajemen, jaringan, yang membuat keputusan tentang bagaimana paket mengalir, dari bidang data, atau bidang penerusan, yang memindahkan paket dari satu node ke node lainnya. Desain ini memungkinkan pemrograman bidang data, virtualisasi fungsi jaringan, dan pemanfaatan server siap pakai. Desain ini juga memungkinkan pemisahan sakelar perangkat keras dari pengontrol perangkat lunak melalui antarmuka selatan, atau protokol, seperti OpenFlow.

Berbeda dengan jaringan lama tradisional, yang menggunakan peralatan khusus yang berpusat pada vendor seperti firewall dan router, SDN memungkinkan aplikasi untuk berinteraksi dengan pengontrol melalui protokol northbound (Gambar 1.3). Dalam desain ini, operator dapat mengintegrasikan berbagai fungsi jaringan dan menerapkan kebijakan berbeda yang mungkin melibatkan prioritas, deprioritas, atau pemblokiran paket tertentu yang mengalir melalui switch tertentu. Hal ini juga memberikan fleksibilitas kepada operator untuk menjalankan beban kerja jaringan pada arsitektur multitenant komputasi awan. Hal ini menghasilkan fleksibilitas dan elastisitas yang lebih besar terhadap perubahan beban kerja melalui perangkat lunak sambil menggunakan switch komoditas yang lebih murah.

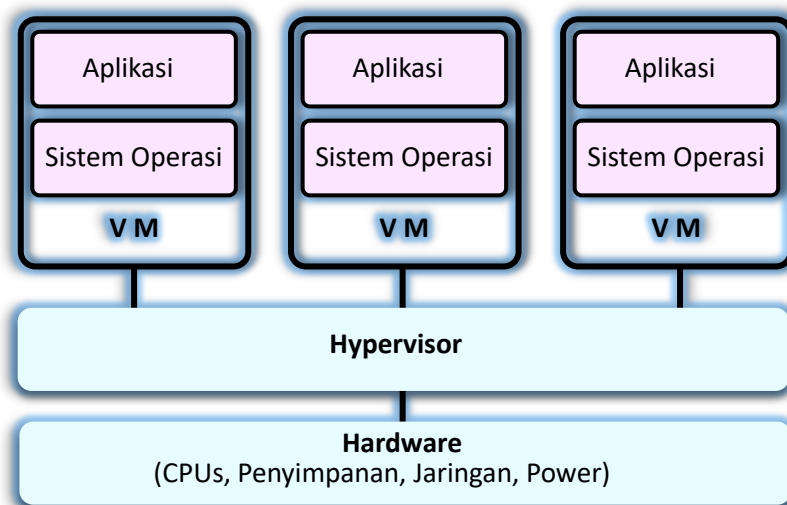
Kecerdasan Berada di Awan

Komputasi awan adalah salah satu terobosan paling luar biasa di abad ke-21. Konsep dasarnya adalah mampu menjalankan beban kerja dan mengakses serta menyimpan data melalui Internet, alih-alih komputer lokal. Keuntungan utamanya adalah mendapatkan kelincahan dengan kemampuan untuk mendapatkan akses sesuai permintaan ke komputasi, penyimpanan, dan jaringan ke beban kerja tanpa melakukan investasi modal yang besar. Awan publik, seperti Amazon dan Microsoft, menyediakan berbagai sumber daya komputasi yang berbeda.





Gambar 1.3 Arsitektur SDN dasar.



Gambar 1.4 Model cloud.

Cloud publik biasanya menawarkan tiga jenis bisnis cloud. Perangkat Lunak sebagai Layanan (SaaS), di mana bisnis berlangganan aplikasi melalui internet. Platform sebagai Layanan (PaaS), di mana bisnis mengembangkan aplikasi kustomnya sendiri yang dibagikan. Infrastruktur sebagai Layanan (IaaS), di mana bisnis "menyewakan" komputasi dan penyimpanan dengan model bayar per penggunaan. Netflix, misalnya, menyediakan layanan dengan menyewa IaaS dari Amazon.

Virtualisasi adalah fondasi di balik sebagian besar cloud berkinerja tinggi. Virtualisasi memungkinkan pemisahan perangkat keras fisik dari sistem operasi komputasi. Dengan kata lain, virtualisasi menyembunyikan sumber daya komputasi dari penggunaannya. Oleh karena itu, satu sumber daya fisik, seperti server, dapat tampak berfungsi sebagai beberapa sumber daya



virtual. Hypervisor adalah lapisan perangkat lunak yang berada di antara perangkat keras fisik dan sistem operasi (Gambar 1.4). Hypervisor menyediakan antarmuka untuk berbagi sumber daya yang tersedia melalui mesin virtual, atau VM. VM berperilaku seolah-olah berjalan pada satu mesin fisik.

Semua bisnis beralih ke komputasi awan dan virtualisasi untuk mengurangi biaya operasional dan modal, sekaligus menikmati kelincahan dan elastisitas komputasi dan penyimpanan. Transformasi ini membantu mereka memindahkan fungsi jaringan fisik, seperti router, switch, dan firewall, ke fungsi jaringan virtual (VNF). VNF berjalan pada VM dan, secara teori, dapat dikelola dan diorkestrasi tanpa sentuhan dalam SDN. VNF tersebut, yang mewakili "kecerdasan" di balik jaringan, dapat dirangkai bersama untuk membentuk layanan komunikasi jaringan yang lengkap. Detail lebih lanjut tentang VM, VNF, dan kontainerisasi disajikan di bab-bab selanjutnya.

Kisah CapEx dan OpEx

Penyedia layanan dengan cepat mengadopsi SDN dan NFV dengan harapan dapat menurunkan Belanja Modal (CapEx) dan Beban Operasional (OpEx) mereka. CapEx adalah dana yang digunakan untuk membeli, membangun, dan memelihara aset infrastruktur, perangkat lunak, dan perangkat keras, seperti membuat infrastruktur data, membangun dan merancang layanan baru, dan meningkatkan sistem TI. OpEx adalah dana yang digunakan untuk menjalankan fungsi bisnis sehari-hari, seperti penelitian dan pengembangan, perubahan perangkat lunak, dan perbaikan peralatan. CapEx dan OpEx memiliki implikasi pajak yang berbeda. OpEx dapat dikurangkan dari pajak pada tahun layanan, sementara CapEx harus disusutkan selama masa investasi.

SDN bersama dengan virtualisasi fungsi jaringan menjanjikan untuk menjadi pengubah permainan dalam menurunkan OpEx dan CapEx. Mereka berpotensi menurunkan biaya komputasi dan server, mengurangi kegagalan layanan dan biaya operasional pusat data, meningkatkan efisiensi melalui otomatisasi, dan mempercepat penerapan layanan. Ini adalah model yang sangat mirip dengan perusahaan web yang mengembangkan aplikasi mereka, misalnya, pencarian dan pemrosesan teks, sebagai perangkat lunak yang berjalan pada perangkat keras cloud komoditas. Dengan model ini, perusahaan web dapat meningkatkan dan menurunkan skala beban kerja mereka per permintaan.

Dari perspektif operator, SDN dapat menyediakan CapEx yang lebih rendah. Ini merupakan peralihan dari peralatan milik vendor, seperti switch dan router, ke semua perangkat lunak yang berjalan di lingkungan cloud. Sebagian besar operator besar merancang lingkungan cloud mereka untuk mendukung beban kerja jaringan. Model bisnis ini akan berubah seiring cloud publik mulai mendukung beban kerja jaringan. Operator kemudian akan mengadopsi strategi cloud hibrida di mana sebagian besar fungsi jaringan mereka berjalan di cloud privat mereka, sambil memanfaatkan cloud publik untuk beban kerja TI, spillover lalu lintas, dan redundansi geografis. Fleksibilitas ini memberi operator leverage saat menegosiasikan persyaratan pembelian dengan vendor jaringan dan komputasi.

Aspek lain dari penghematan dari SDN adalah biaya operasional. SDN memungkinkan operator untuk mengotomatiskan fungsi manajemen jaringan dan menemukan serta



memperbaiki kesalahan jaringan dengan lebih cepat. Sebuah studi terbaru oleh penelitian EMA menunjukkan bahwa rasio admin-ke-virtualisasi adalah 1:77, yaitu, setiap orang dapat mengelola sekitar 77 VM, dibandingkan dengan 1:65 untuk mengelola aset fisik. Virtualisasi melalui SDN, otomatisasi desain layanan, orkestrasi, dan manajemen siklus hidup VNF menyediakan model bisnis yang menarik bagi operator.

Sudah Amankah Kita?

Beralih ke jaringan yang berpusat pada perangkat lunak menghadirkan kemungkinan tantangan keamanan baru yang tidak begitu terlihat pada jaringan lama pada umumnya. Salah satu contoh tantangannya adalah potensi kerentanan yang mungkin disebabkan oleh protokol komunikasi antara bidang kontrol dan bidang data, atau bidang penerusan. Perlu diingat bahwa bidang data adalah tempat data, seperti paket, diteruskan melalui elemen jaringan. Bidang kontrol adalah logika yang menentukan bagaimana dan ke mana data dikirim dalam bidang data. Seperti yang telah kami uraikan sebelumnya, protokol komunikasi ini dapat dirancang dalam SDN melalui OpenFlow. OpenFlow merupakan target yang sempurna bagi penyerang untuk memenuhi saluran komunikasi yang menyebabkan serangan *Denial of Service* (DoS) dan kegagalan pengontrol SDN. Tantangan lainnya adalah mengingat desain pengontrol yang terpusat dalam SDN, hasilnya bisa sangat buruk, dan sangat penting untuk memiliki orkestrasi terpusat yang aman dan tangguh. Berbagai solusi telah diusulkan, termasuk memiliki bidang kontrol terdistribusi tetapi dengan pengontrol yang terpusat secara logis, dan memperkenalkan hierarki pengontrol.

Terlepas dari tantangan-tantangan ini, SDN dan otomatisasi dapat menghasilkan keamanan yang lebih baik dengan menawarkan jaringan virtual yang sepenuhnya terisolasi satu sama lain. Ini juga dikenal sebagai network slicing di mana beberapa jaringan virtual dapat berjalan pada infrastruktur jaringan fisik bersama. Pengguna yang berbeda, atau penyewa, dapat melakukan multipleks melalui satu infrastruktur fisik komputasi dan penyimpanan, memberikan kelincuhan, fleksibilitas, dan keamanan tambahan.

Desain OpenFlow dalam SDN juga menyediakan kemampuan terpusat untuk memfilter paket saat memasuki jaringan, yang secara efektif bertindak sebagai firewall di tepi jaringan. Transaksi yang mencurigakan dapat dialihkan ke kontrol keamanan tingkat yang lebih tinggi. Mengidentifikasi paket yang mencurigakan dan kerentanan dalam SDN sangat penting untuk mengembangkan jaringan otonom. Bagian selanjutnya akan menjelaskan bagaimana pola serangan potensial dapat dipelajari dan dideteksi melalui pembelajaran mesin (ML), dan bagaimana sistem keamanan dapat memperoleh manfaat dari kecanggihan tambahan sistem AI yang mendorong solusi tanpa sentuhan yang dapat diskalakan.

1.3 DARI OTOMATISASI KE KECERDASAN BUATAN

Kerangka Kerja untuk Kecerdasan Buatan

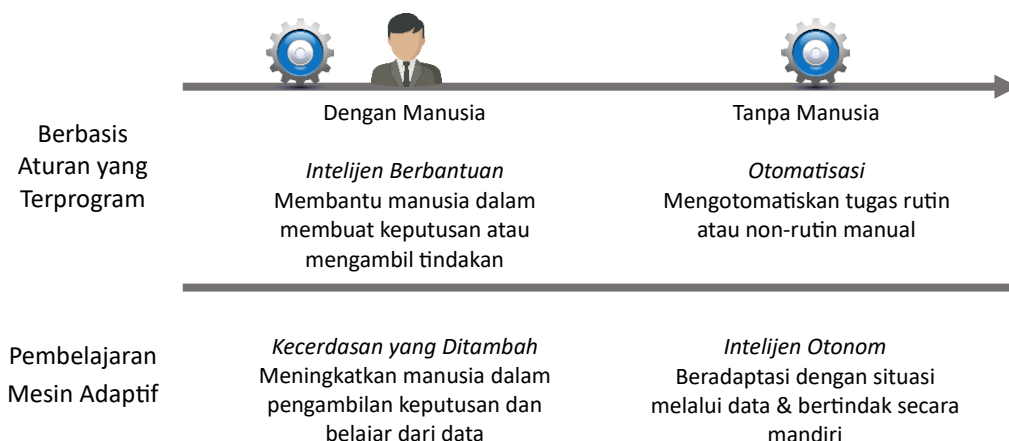
Pernahkah Anda membayangkan dunia di mana pasien dirawat sebelum mengalami gejala penyakit? Dunia di mana mobil Anda mengalami kerusakan, lalu ia mengemudi sendiri bolak-balik ke bengkel? Dunia di mana belanjaan dan makanan Anda diselesaikan dan disiapkan oleh robot rumah Anda? Di dunia tersebut, komunikasi menjadi hal yang lazim, dan



pengalaman tersedia di mana pun Anda membutuhkannya, kapan pun dan di mana pun. Tidak ada lagi ancaman keamanan siber atau roda berputar saat Anda mengunduh atau menonton konten. Perangkat terkecil Anda, seperti pancuran atau pembatas buku, memiliki kecerdasan ponsel pintar Anda saat ini. Inilah dunia yang dibayangkan dengan ML dan AI, yang mendorong jaringan otonom, mobil otonom, pengobatan otonom, dan robotika otonom.

Mari kita mulai dengan definisi dasar. Kami mendefinisikan ML sebagai program komputer (atau agen perangkat lunak) yang dapat belajar dari data untuk memprediksi keadaan atau kondisi di masa mendatang. Secara tradisional, ML, mirip dengan pemrosesan bahasa alami (NLP), dianggap sebagai subbidang AI. Kami mendefinisikan sistem AI sebagai bentuk sistem ML yang lebih canggih yang mampu bertindak dan terus mengoptimalkan tindakannya seiring waktu dan konteks. Salah satu analogi untuk mewujudkan hubungan antara ML dan AI adalah permainan catur. ML setara dengan memprediksi langkah terbaik berikutnya berdasarkan semua langkah historis. Ini adalah keputusan instan. AI setara dengan merumuskan strategi, atau proses, untuk memenangkan permainan. Ini biasanya melibatkan pengoptimalan urutan langkah terbaik berikutnya. Strateginya dinamis dan diperbarui setiap kali langkah dilakukan.

Gambar 1.5 menunjukkan kerangka kerja untuk AI dengan kisi 2x2. Sisi atas kisi mencakup sistem AI yang berbasis aturan dan terintegrasi. Kecerdasan Terbantu adalah salah satu jenis sistem ini yang melibatkan program perangkat lunak dalam bentuk sistem pakar yang menyediakan dukungan otomatisasi kepada manusia. Salah satu contohnya adalah mesin alur kerja yang digunakan di pusat kontak untuk mendukung agen langsung. Otomatisasi yang menggunakan sistem berbasis aturan dapat melibatkan bot yang melakukan tugas-tugas sederhana rutin atau non-rutin yang mungkin tidak melibatkan tingkat kecerdasan atau "berpikir" tingkat lanjut. ML adaptif memperluas AI ke dimensi baru dengan mengisi sistem ini dengan data.



Gambar 1.5 Kerangka kerja untuk sistem AI.

Kecerdasan tertambah mencakup sistem AI yang membantu manusia dengan memproses data dalam jumlah besar. Ini memiliki peluang signifikan dalam industri perawatan kesehatan, di mana pembelajar mesin dapat digunakan untuk memberikan rekomendasi



kepada dokter tentang diagnosis pasien. Kecerdasan otonom adalah bentuk sistem yang sepenuhnya otonom yang beradaptasi, bertindak, dan belajar secara mandiri. Sistem tersebut merupakan subjek penelitian yang signifikan, karena sangat penting dalam bidang-bidang seperti jaringan otonom dan mobil otonom, antara lain. Selain kisi AI 2×2 , para peneliti sering membahas bidang kecerdasan umum buatan. Ini termasuk sistem kecerdasan otonom yang dapat beroperasi dan menggeneralisasi di seluruh tugas dan masalah. Meskipun sistem seperti itu masih dalam tahap awal, mereka akan terus matang dan tersebar luas selama 10-20 tahun ke depan.

Kecerdasan Buatan Meninjau Kembali

Fondasi di balik pembelajaran mesin dan kecerdasan buatan (ML) diciptakan selama beberapa dekade oleh banyak pelopor, termasuk Alan Turing, Claude Shannon, Arthur Samuel, dan Frank Rosenblatt, di antara yang lainnya. Beberapa orang mungkin berpendapat bahwa momen penentu ini terinspirasi pada tahun 1950 oleh Alan Turing dan "Tes Turing"-nya. Tes ini menghitung lamanya waktu atau jumlah interaksi yang dapat dilakukan mesin untuk mengelabui manusia agar percaya bahwa ia adalah manusia. Saat berada di Bell Labs, Claude Shannon memajukan bidang kecerdasan buatan (AI) dengan menciptakan tikus memori, bernama "Theseus," untuk melintasi labirin. Tikus tersebut mampu mempelajari, tanpa pengawasan, jalur optimal dari sumber ke tujuan. Shannon, yang tertarik dengan eksperimen ini, mulai membangun upaya penelitian yang lebih luas di bidang AI. Pada tahun 1956, Shannon dan rekan-rekannya, Minsky, Rochester, dan McCarthy, menyelenggarakan lokakarya musim panas AI pertama di Dartmouth. Lokakarya tersebut menghasilkan terobosan mendasar di bidang AI dan, menurut para penulisnya, berupaya untuk "membuat mesin menggunakan bahasa, membentuk abstraksi dan konsep, memecahkan berbagai jenis masalah yang kini hanya dapat dipecahkan oleh manusia.

Pelopor lain di bidang AI termasuk Arthur Samuel dari IBM yang menulis program untuk permainan catur yang mengembangkan strategi dan langkah-langkah kemenangan. Frank Rosenblatt, pelopor AI lainnya, mengembangkan banyak algoritma dasar di balik jaringan saraf suatu bentuk pembelajaran mesin yang mencoba mensimulasikan operasi otak manusia. Teori dan eksperimen praktis yang dikembangkan oleh para pelopor ini dan banyak lainnya telah menghasilkan bidang-bidang ilmiah baru seperti pembelajaran mendalam, yang kini sedang merambah ke dalam otomatisasi cerdas untuk aplikasi dan layanan industri. ML dan AI tidak bersifat konseptual atau teoretis. Selama tiga dekade terakhir, teknologi ML telah diterapkan untuk berbagai layanan komersial, seperti Siri milik Apple dan Alexa milik Amazon. Algoritma ML digunakan untuk pengenalan wajah, pengenalan suara, dan NLP. Dalam pengenalan suara, model ML, yang dikenal sebagai deep learner, digunakan untuk memperkirakan probabilitas posterior unit linguistik yang bergantung pada konteks, seperti fonem dan kata.

Probabilitas tersebut diterapkan dalam model tersembunyi Markov untuk menemukan urutan unit terbaik. Melatih model ML ini pada berbagai macam dan volume data membantu membuat sistem pengenalan suara menjadi independen dari pembicara dan kuat terhadap lingkungan dan kondisi yang merugikan. Suatu bentuk ML, yang dikenal sebagai jaringan saraf tiruan, telah digunakan untuk memperkirakan parameter artikulasi model saluran vokal untuk



menghasilkan ucapan yang disintesis. Ini telah menjadi dasar di balik sistem teks-ke-ucapan modern. Jenis model ML lainnya, yang dikenal sebagai jaringan saraf konvolusional, telah digunakan untuk menemukan objek (misalnya, karakter, mobil, pakaian, sepatu) dalam gambar memungkinkan bisnis untuk melakukan periklanan yang ditargetkan. ML juga telah digunakan untuk memprediksi kegagalan jaringan. Tindakan proaktif ini dapat menghasilkan peningkatan penghematan biaya dan pengalaman pelanggan yang lebih baik. Demonstrasi aplikasi ML yang sukses ini, bersama dengan kemajuan dalam komputasi dan ketersediaan data besar-besaran, terus meningkatkan minat yang signifikan oleh perusahaan yang ingin mengembangkan sistem otomatis yang lebih cerdas.

Jaringan yang Ditentukan Perangkat Lunak dan Kecerdasan Buatan

SDN adalah tentang memvirtualisasikan jaringan dengan menjalankan fungsi jaringan sebagai perangkat lunak pada infrastruktur cloud komoditas, sedangkan AI berbasis ML, di sisi lain, adalah tentang mengumpulkan dan menganalisis data untuk meningkatkan otomatisasi dan mengekstraksi intelijen dan informasi. Perpaduan sempurna antara SDN dan AI menghasilkan jaringan otonom. Jenis jaringan ini tangguh, dapat menyembuhkan diri sendiri, dan belajar sendiri. AI dapat mengumpulkan dan menganalisis data, serta mengambil tindakan untuk memperbaiki kondisinya. Sebagaimana akan dijelaskan di bab-bab selanjutnya, tindakan dapat berkisar dari mengoptimalkan arus lalu lintas dan mengidentifikasi pola awal keamanan siber hingga memprediksi kegagalan VNF fisik atau VNF dan memindahkan lalu lintas sesuai kebutuhan.

Meskipun terdapat banyak publikasi tentang penggunaan teknologi ML dan AI untuk aplikasi multimoda dan multimedia, penelitian tentang penerapan ML dan AI pada SDN masih jauh lebih sedikit. Beberapa peluang utama yang telah kami identifikasi untuk penggunaan AI meliputi:

1. Optimalisasi jaringan mandiri
2. Analisis ancaman dan keamanan siber
3. Manajemen kesalahan
4. Peningkatan pengalaman pelanggan
5. Optimalisasi lalu lintas

AI menggerakkan jaringan otonom dengan memungkinkan peningkatan stabilitas, keamanan, skalabilitas, dan kinerja (juga dikenal sebagai S3P). Misalnya, AI dapat diterapkan untuk mengidentifikasi pola ancaman keamanan atau mendeteksi penurunan kinerja VNF. Sinyal memori dan CPU dapat dikumpulkan dari VM yang mendukung firewall virtual. Sekelompok pembelajar mesin kemudian dapat dilatih untuk memprediksi kesehatan firewall ini selama periode waktu tertentu. Kegagalan yang terdeteksi dapat diatasi dengan me-reboot firewall secara otomatis, memigrasikan lalu lintas ke firewall lain, atau menerbitkan tiket untuk intervensi manual.

Konsep pembelajaran mesin yang sama dapat diterapkan untuk interpretasi peristiwa pemicu atau tanda tangan ketika berhadapan dengan sistem loop kontrol. Pembelajar mendalam dapat dilatih untuk memetakan sinyal analitik dan konteks ke dalam serangkaian kebijakan atau tindakan yang telah ditentukan sebelumnya. Saat runtime, pembelajar dapat



memilih konfigurasi yang paling mungkin. Pembelajaran dapat memperbarui modelnya secara dinamis berdasarkan penguatan positif dan negatif. Algoritma back-propagation dapat digunakan untuk memperbarui probabilitas a posteriori agar paling mencerminkan tindakan atau kebijakan yang benar. Misalnya, jika pembelajar memprediksi kebijakan tanpa tindakan, dan suatu tindakan diperlukan, maka pembelajar akan menekankan secara negatif kelas "tanpa tindakan" dan secara positif menekankan tindakan yang benar.

Perangkat dengan pembelajaran mendalam adalah seseorang membuat keputusan instan tepat waktu dan tidak mengoptimalkan serangkaian keputusan menuju tujuan bersama. Hal ini baik-baik saja ketika mengenali objek dalam gambar, tetapi menimbulkan masalah serius ketika perlu merekomendasikan serangkaian tindakan tentang cara terbaik untuk memigrasikan lalu lintas dari satu simpul jaringan ke simpul jaringan lainnya, misalnya. Beberapa teknik yang menjanjikan telah menunjukkan kinerja yang unggul ketika diintegrasikan dengan pembelajaran mendalam. Ini termasuk pemrograman dinamis, suatu bentuk model Markov Tersembunyi, dan pembelajaran penguatan. AlphaGo dari DeepMind menerapkan pembelajaran mendalam dan pembelajaran penguatan untuk mengalahkan pemain Go terbaik dunia.

Perjalanan Menuju Jaringan Otonom yang Didefinisikan Perangkat Lunak

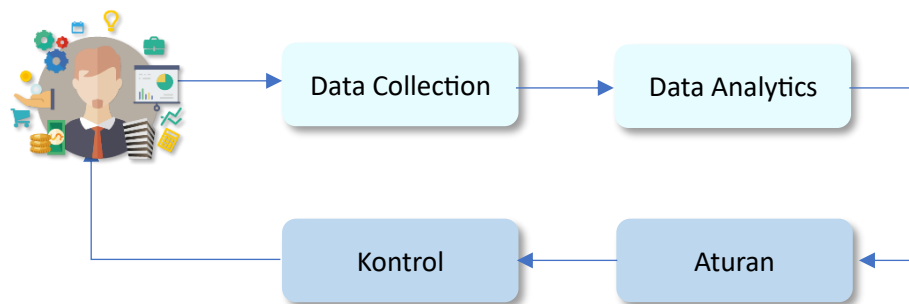
Layanan cerdas adalah layanan yang sadar diri, artinya, ia mengumpulkan informasi tentang sumber dayanya (pengguna, perangkat, jaringan, dan aplikasi) untuk memberikan layanan terbaik di lingkungan tempatnya beroperasi. Layanan kemudian membuat keputusan berdasarkan informasi ini dan pengetahuan domainnya untuk mempersonalisasi layanan bagi pengguna yang menggunakannya. Layanan cerdas menerima umpan balik tentang kinerjanya dan belajar. Ada tiga atribut utama yang mencirikan layanan cerdas, yang dikenal sebagai personalisasi, adaptif, dan dinamis (PAD). Layanan personalisasi prediktif adalah layanan yang mengantisipasi kebutuhan pengguna, secara proaktif mengambil tindakan cerdas, dan merekomendasikan informasi personalisasi yang berharga dan tepat waktu. Layanan adaptif belajar dari tindakan masa lalu semua pengguna dan menyesuaikan perilakunya untuk memberikan kualitas layanan yang unggul. Layanan yang dinamis adalah layanan yang tangguh dan dapat memperbaiki diri atau mengatur diri sendiri untuk menghindari gangguan layanan.

Mencapai jaringan otonom dengan karakteristik PAD dimungkinkan melalui pengenalan sistem loop tertutup (Gambar 1.6). Konsep dasarnya adalah menghitung sinyal kesalahan antara status target dan pengukuran aktual yang dihitung dari sensor. Loop kontrol umpan balik kemudian diterapkan untuk mengoptimalkan sistem dengan meminimalkan tingkat kesalahan yang terukur. Salah satu contohnya adalah termostat rumah yang secara terus-menerus mengendalikan sistem pemanas/pendingin untuk mempertahankan suhu ruangan yang diinginkan. Komponen dasar sistem tersebut, seperti yang ditunjukkan pada Gambar 1.6, meliputi pengumpulan data untuk mengukur suhu saat ini, analitik data untuk menghitung variasi antara pembacaan yang diinginkan dan aktual, aturan kebijakan untuk merekomendasikan tindakan terbaik berikutnya, dan pengontrol untuk menjalankan tindakan tersebut. Tindakan tersebut dapat mencakup menyalakan/mematikan pemanas, menyalakan/mematikan pendingin, atau tidak melakukan apa pun. Sistem kontrol loop



tertutup berbasis jaringan, secara umum, melibatkan integrasi backend ke sistem Sistem Pendukung Operasi (OSS) dan Sistem Pendukung Bisnis (BSS) lainnya, seperti pencatatan, penagihan, dan tiket.

Dalam jaringan otonom, sistem loop kontrol memainkan peran penting dalam menghasilkan penghematan biaya operasional. Otomatisasi loop kontrol dapat dikategorikan menjadi sistem loop terbuka dan loop tertutup. Sistem loop terbuka menangkap informasi telemetri dan diagnostik dari infrastruktur cloud yang mendasarinya (misalnya, peristiwa manajemen syslog, kesalahan, dan kinerja jaringan), melakukan serangkaian analitik, dan menyediakan pelaporan, tiket, atau alarm kepada tim operasi. Sistem loop tertutup terus memantau sistem untuk kesalahan dan S3P, serta menghitung serangkaian tanda tangan berdasarkan kondisi anomali yang terdeteksi. Mesin kebijakan kemudian menginterpretasikan tanda tangan ini dan tindakan korektif yang tepat direkomendasikan dan dilakukan. Setelah sistem diperbaiki, aplikasi pemantauan memeriksa status sistem untuk mempertahankan status target yang diinginkan. Tujuannya adalah mencapai nol waktu henti atau gangguan minimal.



Gambar 1.6 Konsep dasar sistem loop tertutup.

Prinsip Nol dan Satu

Dalam jaringan saat ini, setiap layanan baru yang diterapkan membutuhkan pengembangan perangkat lunak, pengujian, dan dukungan operasional baru. Penerapan dan operasi sebagian besar padat karya manusia, yang menciptakan tantangan selama kegagalan layanan atau ketika meningkatkan jumlah layanan. Merupakan visi setiap operator untuk merancang dan menerapkan layanan hanya sekali, dan berharap untuk mengeluarkan biaya operasional nol selama masa pakai layanan. Prinsip satu dan nol ini merupakan karakteristik utama dari jaringan otonom.

Mari kita tinjau terlebih dahulu desain layanan. Layanan adalah komposisi beberapa fungsi jaringan yang dirangkai bersama, seperti Lego. Bukan hal yang aneh jika fungsi-fungsi jaringan ini dikembangkan oleh beberapa vendor. Untuk mencapai "prinsip satu" dalam SDN, manajemen dan orkestrasi fungsi jaringan dirancang untuk digerakkan oleh model. Model tersebut dapat mencakup fungsi komputasi, konektivitas, dan fitur yang dapat disusun dan dioperasikan dalam lingkungan cloud. Dengan kata lain, model pada dasarnya merupakan resep manajemen yang mencakup parameter, kebijakan, dan proses fungsi jaringan. Merancang layanan mencakup pengembangan model tersebut, yang digunakan untuk



instansiasi dan penerapan layanan. Hal ini serupa dengan pencetakan 3D, di mana model dikembangkan untuk suatu objek yang kemudian digunakan untuk membuat produk yang sebenarnya. Desain berbasis model merupakan karakteristik yang kuat dalam SDN.

Di sisi lain, mendorong zero touch adalah di mana AI menjadi sentral. Sebagian besar biaya operasional untuk mendukung layanan dikaitkan dengan (a) pemutakhiran atau pemeliharaan perangkat lunak, (b) deteksi dan pemulihan kegagalan, dan (c) instansiasi, orkestrasi, dan penerapan fungsi jaringan. AI dapat mendorong otomatisasi manajemen perubahan untuk pemutakhiran dan pemeliharaan perangkat lunak. Penjadwalan kapan harus memutakhirkan, apa yang harus dimutakhirkan, apa saja dependensinya, bagaimana meminimalkan dampak layanan, kebijakan cadangan apa yang harus digunakan kembali jika terjadi kegagalan, dan bagaimana memantau pemutakhiran setelah selesai adalah fungsi-fungsi yang dapat dijalankan oleh sistem AI. Ini bukan sekadar masalah penjadwalan dan optimasi, tetapi juga masalah pembelajaran di mana data dari kegagalan dan keberhasilan masa lalu membantu meningkatkan hasil di masa mendatang.

Manajemen kesalahan dan kinerja juga merupakan kandidat ideal bagi AI untuk mencapai solusi tanpa sentuhan. Sebagaimana diuraikan di bagian sebelumnya, otomatisasi loop tertutup adalah sistem AI templat yang dapat diterapkan untuk memantau fungsi jaringan virtual dan fisik, VM, dan elemen infrastruktur, serta komponen fungsional untuk platform manajemen dan orkestrasi. Mengadopsi sistem otomatisasi loop tertutup saat menerapkan SDN dapat menghasilkan "prinsip nol".

1.4 JARINGAN OTONOM BERDASARKAN DATA

Data Adalah Mata Uang Baru

Selama dekade terakhir, data telah tumbuh secara eksponensial. Statistik terbaru dari Inside Big Data menunjukkan bahwa ukuran dunia digital berlipat ganda setiap dua tahun. Sementara data manusia-ke-mesin tumbuh 10 kali lebih cepat daripada data bisnis, data mesin-ke-mesin tumbuh lebih cepat, yaitu 50 kali lipat. Dikatakan bahwa perusahaan yang mengendalikan data dalam jumlah besar akan menjadi pemimpin baru di dunia AI. Memang, data adalah mata uang baru, dan kemampuan untuk mengumpulkan, mengelola, memindahkan, menyimpan, dan memproses data secara efisien dan dalam skala besar merupakan tantangan utama dalam membangun jaringan otonom yang andal yang menunjukkan prinsip nol dan satu.

Jaringan otonom perlu dirancang untuk beroperasi dengan big data. Selain hanya mempertimbangkan ukuran data, sebagai karakteristik penentu untuk big data, Doug Laney dari Gartner telah menawarkan beberapa atribut lebih lanjut dari big data dengan memperkenalkan "3V" dari volume, variasi, dan kecepatan. Sejak itu, yang lain (Livi13, Data14, Wikibig) telah memperluas 3V menjadi 5V atau 7V untuk memasukkan karakteristik lain seperti variabilitas, validitas/kebenaran, visibilitas/visualisasi, dan nilai. Atribut-atribut ini adalah bahan penting untuk mengembangkan sistem AI presisi tinggi, seperti yang akan dijelaskan di bab-bab selanjutnya.



Untuk jaringan otonom, big data dapat berasal dari aplikasi yang berjalan di perangkat, tepi pelanggan seperti kotak pengaturan, tepi jaringan seperti pusat data, atau cloud pribadi dan pihak ketiga yang lebih terpusat. Dari mana pun data berasal, mempertahankan data berkualitas tinggi merupakan sebuah tantangan. Atribut kualitas meliputi ketersediaan, kegunaan, keandalan, relevansi, dan presentasi. Tantangan lainnya adalah manajemen data. Terdapat berbagai isu manajemen dan kebijakan data yang harus ditentukan untuk sistem big data, seperti jenis data yang dikumpulkan, berapa lama data disimpan, apakah data diringkas atau dianonimkan, dan bagaimana data diproses dan diakses secara batch versus real-time. Memindahkan data antar situs sangat mahal. Oleh karena itu, sebagian besar sistem modern cenderung menghadirkan analitik dan ML ke dalam data, bukan sebaliknya. Data biasanya diproses terlebih dahulu untuk memungkinkan pemrosesan paralel. Ekosistem sumber terbuka Hadoop telah muncul sebagai yang memiliki kemampuan terdepan untuk memproses dan menyerap big data.

Terlepas dari semua tantangan dalam menangani, mengelola, dan memproses data, penting untuk merawat dan memberi makan data saat mengembangkan jaringan otonom. Data sangat penting untuk mengatasi manajemen kesalahan dan kinerja, keamanan siber, mengelola dan mengoptimalkan arus lalu lintas data, dan memprediksi lokasi untuk menempatkan edge cloud, sel mikro dan makro secara optimal. Data sama pentingnya dengan mata uang di dunia digital baru. Sama seperti uang yang dilacak dan ditangani dengan hati-hati, data perlu diperlakukan dengan cara yang sama. Sebagian besar perusahaan saat ini memperkenalkan tata kelola data untuk menetapkan kebijakan tentang cara terbaik mengelola, menyimpan, memproses, dan mengamankan data guna mengurangi biaya operasional, mendorong inovasi dalam ML dan AI, serta tetap mematuhi aturan privasi data.

Data adalah bahan bakar di balik model ML dan alasan keberadaannya saat ini. Semakin beragam atribut data pelatihan, semakin akurat pembelajaran mesin. Di sisi lain, data yang buruk dan bias dapat menyebabkan pembelajaran yang tidak akurat. Hal ini disebut sebagai bias algoritmik. Data berkualitas tinggi yang tidak bias penting untuk membangun sistem AI di SDN. Beberapa bab dalam buku ini dikhususkan untuk kualitas dan integritas data.

Layanan Mikro Berbasis Data

Salah satu kendala utama adopsi masif solusi AI berbasis data adalah tingginya hambatan masuk dalam penerapan teknologi ini. Model AI biasanya dibangun dengan data dalam jumlah besar, membutuhkan infrastruktur komputasi yang ekstensif, dan dilatih oleh para ahli yang memiliki gelar sarjana di bidang sains. Siklus waktu dari pembuatan model hingga penerapan layanan dengan AI juga panjang dan mahal. Sebagai contoh, peneliti di bidang teknologi ucapan dan video dapat menghabiskan waktu berbulan-bulan untuk membangun dan menyempurnakan model pembelajaran mesin (ML) untuk mengenali ucapan, atau mendeteksi wajah. Satu dari seratus model ini mungkin akan diterapkan dan digunakan dalam layanan komersial apa pun. Selain itu, model-model ini kurang dapat digunakan kembali, sehingga kemajuan dalam AI menjadi sangat lambat.

Membantu mendorong penggunaan kembali dalam AI dan menurunkan hambatan masuk adalah tujuan besar yang telah diperjuangkan komunitas selama bertahun-tahun. Salah



satu terobosan utama adalah munculnya layanan mikro. Layanan mikro, sebuah evolusi dari arsitektur berorientasi layanan, adalah gaya arsitektur perangkat lunak di mana aplikasi besar tersusun dari layanan-layanan kecil yang terhubung secara longgar dengan siklus hidup yang independen. Istilah ini dipopulerkan oleh para pemikir terkemuka seperti Martin Fowler yang menyatakan bahwa tujuan dari desain layanan semacam itu adalah untuk membangun sistem menggunakan kumpulan komponen inti umum yang lebih kecil yang memiliki API publik yang kuat. Hal ini menghasilkan beberapa manfaat, termasuk (a) kemudahan penerapan kapabilitas baru, (b) kemudahan pemeliharaan dan pemutakhiran, (c) kemudahan penghentian operasional, dan (d) tidak bergantung pada bahasa pemrograman yang digunakan.

Kontainer digunakan untuk membangun layanan mikro. Kontainer membantu mengisolasi, memindahkan, dan menjalankan perangkat lunak pada VM fisik dan VM. Kontainer berbeda dari VM karena aplikasi yang berjalan pada kontainer dapat berbagi sistem operasi yang sama. Layanan mikro biasanya diterapkan lebih cepat pada kontainer daripada VM. Hal ini dapat menjadi keuntungan ketika membutuhkan penskalaan horizontal. Penggunaan kontainer yang populer adalah melalui Docker. Docker adalah proyek sumber terbuka yang bertujuan untuk mengotomatiskan penerapan aplikasi di dalam kontainer portabel yang independen dari perangkat keras atau sistem operasi yang digunakan. Mengelola sekelompok kontainer Docker memerlukan teknologi dan alat baru. Yang paling populer adalah Kubernetes, yang merupakan proyek sumber terbuka untuk mengotomatiskan penerapan, penskalaan, dan pengelolaan aplikasi yang dikontainerisasi. Kubernetes, bersama dengan teknologi lain, seperti yang akan dijelaskan di bab-bab selanjutnya, sangat penting untuk mendorong layanan menjadi cloud native: ketahanan, skalabilitas, penyembuhan mandiri.

Meskipun layanan mikro yang memanfaatkan kontainer Docker telah menjadi sangat populer, penggunaan layanan mikro bertenaga data untuk AI tidak begitu dikenal luas. Hal ini sebagian disebabkan oleh kesenjangan keterampilan di antara arsitek yang menyediakan desain dan persyaratan, ilmuwan data yang membangun model, dan pengembang perangkat lunak yang mengembangkan dan menerapkan layanan akhir. Sejumlah kegiatan baru-baru ini diumumkan oleh Linux Foundation untuk mengatasi kesenjangan ini. Salah satunya adalah Proyek Acumos, yang bertujuan untuk menciptakan pasar layanan mikro yang berpusat pada AI, dan menyederhanakan proses pengembangan, mulai dari pembuatan model hingga penerapan layanan mikro di cloud hibrida.

Compute Mengikuti Anda

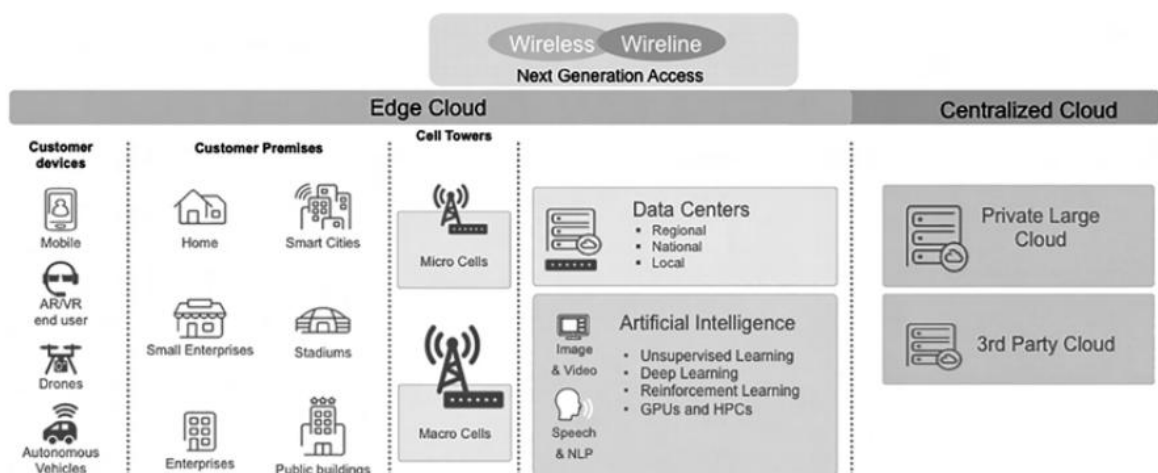
Selama dekade terakhir, telah ada upaya terfokus untuk memusatkan komputasi dan penyimpanan, serta mengembangkan infrastruktur virtualisasi yang memenuhi berbagai jenis beban kerja. Visi ini telah membantu perusahaan seperti Google, Amazon, dan Microsoft membangun infrastruktur cloud publik terpusat yang masif dan menyediakannya untuk digunakan oleh pihak ketiga. Virtualisasi melalui cloud publik telah mempercepat pertumbuhan banyak industri dan perusahaan rintisan baru, meningkatkan efisiensi operasional, dan mengurangi total biaya kepemilikan dengan meningkatkan pengembalian modal yang diinvestasikan. Mengadopsi virtualisasi juga menyiratkan bahwa aplikasi Anda



dapat berjalan ribuan mil jauhnya, yang memengaruhi kinerja dan pengalaman dalam banyak kasus.

Dengan munculnya pengalaman baru melalui mobil otonom, drone, dan realitas virtual, data kini dihasilkan di ujung jaringan pengguna, yang disebut "edge". Di sisi lain, aplikasi menuntut komunikasi real-time, sehingga membutuhkan pemrosesan data dinamis yang efisien di sisi pelanggan dan jaringan. Aplikasi tersebut menginginkan komputasi mengikuti perangkat, bukan sebaliknya. Hal ini berbeda dengan Jaringan Pengiriman Konten (CDN) tradisional, seperti video dan halaman web yang menyimpan cache relevan dan terlokalisasi lebih dekat dengan pengguna akhir.

Komputasi tepi adalah penempatan komputasi, penyimpanan, daya, dan jaringan di dekat aplikasi atau perangkat (Gambar 1.7). Hal ini dapat terjadi di "tepi" pelanggan, seperti rumah, kota pintar, usaha kecil, atau di "tepi" jaringan di pusat data regional, nasional, atau lokal. Komputasi tepi memberikan manfaat yang signifikan, termasuk (a) mengurangi lalu lintas backhaul, (b) mempertahankan kualitas layanan yang unggul, (c) mengurangi biaya pemindahan data, dan (d) meningkatkan latensi dan kinerja. Sumber daya cloud kini dapat didistribusikan secara optimal sesuai kebutuhan pengguna.



Gambar 1.7 Edge dan cloud terpusat.

Selama kurang lebih satu tahun terakhir, terdapat tren membangun komputasi cloud di "edge" pelanggan dalam bentuk "kotak putih". Kotak putih pada dasarnya adalah perangkat keras tanpa merek. Perangkat ini mendukung disagregasi atau pemisahan perangkat keras dan perangkat lunak melalui antarmuka non-eksklusif yang terbuka. Layanan jaringan, seperti router dan firewall, dapat secara otomatis diinstansiasi dan diorkestrasi secara lokal maupun jarak jauh. Perangkat keras tersebut pada dasarnya menjadi kotak putih komputer yang dapat menjalankan fungsi jaringan dan aplikasi sesuai permintaan. Pertimbangan untuk menempatkan kotak-kotak yang bertindak sebagai "cloud bayi" ini ke dalam lini manufaktur untuk membantu menjalankan berbagai beban kerja, seperti inspeksi dan pemantauan.

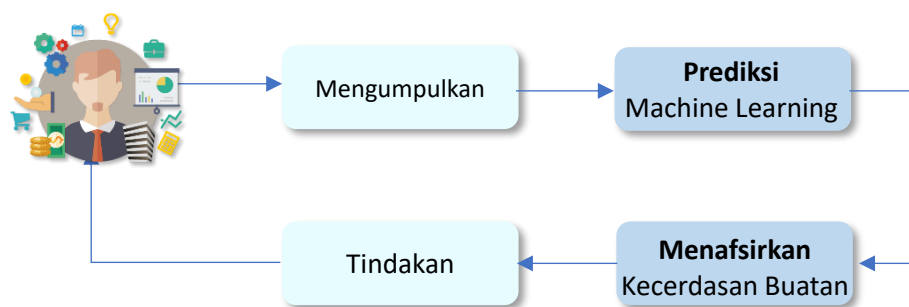
Memindahkan infrastruktur cloud ke edge memberikan peluang baru untuk mengalihkan pemrosesan dari perangkat dan menjadikan kecerdasan berada di edge. Hal ini



secara fundamental mengubah ekosistem perangkat dan memungkinkan pengembangan perangkat IoT yang sangat murah dan cerdas. Memindahkan kecerdasan tersebut ke edge membutuhkan pendistribusian teknologi AI di seluruh node edge dan menjalankannya pada GPU dan/atau server Higher Performance Compute. Beban kerja ML, seperti deep learning, unsupervised learning, atau reinforcement learning kemudian akan berpindah dari infrastruktur cloud terpusat ke node edge terdistribusi. Pergeseran paradigma ini, seiring dengan hadirnya 5G dan akses generasi berikutnya, akan menghasilkan pengalaman baru yang secara fundamental akan mentransformasi bisnis, bahkan masyarakat.

Menutup Loop Kecerdasan Buatan

Menangkap data di edge jaringan memberikan peluang baru untuk melakukan pemrosesan data primitif maupun kompleks, mulai dari agregasi data hingga ML tingkat lanjut dan analitik prediktif. Misalnya, dengan mengumpulkan data kinerja dan aplikasi dari VM yang mendukung firewall dan router virtual, kita dapat menjalankan analitik tingkat lanjut di edge untuk melakukan otomatisasi loop tertutup. Ini mungkin melibatkan identifikasi dan remediasi serangan siber secara otomatis, diikuti dengan pengalihan lalu lintas ke VM baru, penonaktifan alamat IP, atau boot ulang VM.



Gambar 1.8 Menutup Loop dengan AI.

Melakukan tugas-tugas tersebut secara semi-real-time di edge membutuhkan banyak tenaga manual dan hampir mustahil untuk diskalakan ketika Anda menangani lebih dari 10.000 node edge. Di sinilah AI memainkan peran penting dalam menutup loop. Alih-alih melakukan analitik reaktif, yang biasanya digunakan dalam model pembelajaran tanpa pengawasan, ML membantu memprediksi anomali atau peristiwa jauh sebelumnya, berdasarkan sinyal yang berubah seiring waktu. Demikian pula, alih-alih menerapkan kebijakan atau aturan yang terprogram untuk memandu sistem tentang tindakan terbaik selanjutnya, AI dapat membuat rekomendasi yang lebih cerdas berdasarkan data, konteks, pola sebelumnya, dan hasil. Lebih lanjut, seiring sistem terus beriterasi dan semakin banyak data yang dikumpulkan, ML dan AI dapat belajar dari keberhasilan dan kegagalan, serta menyesuaikan model mereka untuk meminimalkan kesalahan. ML dan AI membantu mengubah sistem dari statis menjadi dinamis dan dari reaktif menjadi proaktif. Hal ini digambarkan pada Gambar 1.8.

Menutup siklus dengan AI melibatkan pengumpulan data, memprediksi anomali, menginterpretasikan hasilnya, membuat keputusan, dan kemudian mengambil tindakan



korektif. Otomatisasi siklus tertutup dapat dianggap sebagai mikrokapabilitas fundamental untuk mencapai jaringan otonom dan mengembangkan layanan serta aplikasi tanpa sentuhan.

1.5 DUNIA YANG SAMA SEKALI BARU

Perpaduan antara SDN dan AI menghasilkan layanan baru yang dapat membuat masyarakat efisien, inovatif, dan lebih aman. Bagian ini akan mengulas sekilas masa depan.

Mengoperasikan Kota Virtual

Seberapa sering kita mengalami kemacetan lalu lintas? Peningkatan populasi seiring dengan pertumbuhan bisnis akan terus membebani sistem transportasi kita dan memberikan tantangan pada dunia fisik kita. Gambaran serupa berlaku untuk jaringan. Peningkatan eksponensial dalam data dan perangkat pintar akan terus menciptakan kemacetan lalu lintas jaringan dan menyebabkan penundaan saat mengunduh dan mengunduh data.

Meskipun penambahan spektrum dan lokasi seluler menambah kapasitas jaringan baru, pertumbuhan linier tersebut belum mampu mendukung peningkatan data secara eksponensial.

Bayangkan dunia fisik di mana jalan dapat berubah secara instan lebar, panjang, dan arahnya untuk mengakomodasi kemacetan dan menjamin mobil dapat mencapai tujuannya kapan saja tanpa penundaan. Inilah visi di balik jaringan otonom. Tujuannya adalah memberikan fleksibilitas dan elastisitas untuk rekayasa dan optimalisasi lalu lintas sehingga paket data mengalir dari satu simpul jaringan ke simpul lainnya tanpa penundaan kapan pun. Lebih lanjut, jaringan ini bersifat prediktif dan mengoptimalkan dirinya sendiri sepanjang waktu untuk mengantisipasi perubahan lalu lintas.

Seperti yang telah dibahas sebelumnya, SDN memungkinkan pemisahan bidang kontrol dan bidang data. Hal ini memungkinkan sentralisasi kontrol. Protokol seperti OpenFlow terkadang digunakan sebagai antarmuka komunikasi untuk memungkinkan manipulasi langsung perangkat jaringan di bidang data (Gambar 1.4). Aliran lalu lintas melibatkan perutean paket yang tiba di sakelar berdasarkan instruksi yang disimpan dalam tabel aliran. Kecerdasan dalam SDN berada di bidang manajemen yang mencakup aplikasi seperti router dan penyeimbang beban yang berinteraksi dengan model jaringan di bidang kontrol. Salah satu contoh utama platform manajemen adalah Linux Foundation *Open Network Automation Platform* (ONAP).

Memperkenalkan AI ke dalam ekosistem ini untuk memungkinkan optimalisasi lalu lintas dinamis dan peningkatan keamanan lalu lintas jaringan merupakan landasan yang kuat. Salah satu contohnya adalah karya Stampa dkk. tentang penggunaan pembelajaran penguatan mendalam untuk optimalisasi perutean SDN. Pembelajar mesin dilatih untuk menyediakan konfigurasi perutean dinamis guna meminimalkan penundaan jaringan. Berbagai fungsi penghargaan diterapkan dengan kebijakan target terkait. Penelitian ini merupakan pengembangan dari studi sebelumnya oleh Boyan dan Littman yang menerapkan Q-learning untuk menemukan kebijakan perutean optimal yang mengoptimalkan jumlah "hop" yang akan diambil sebuah paket untuk menghindari kemacetan.



Dari Internet of Things ke Internet of Everything

IoT adalah konsep yang menghubungkan perangkat apa pun, mulai dari ponsel dan mobil hingga kulkas dan mesin cuci, ke internet. Jaringan otonom ini menjanjikan untuk merevolusi ruang IoT dengan memungkinkan perangkat apa pun menjadi sangat "cerdas" dengan terhubung secara aman ke "tepi" jaringan. Contohnya adalah IoT untuk kota pintar. Sensor IoT jalanan memberikan informasi tentang tempat parkir kosong. Informasi tersebut kemudian digunakan oleh mobil otonom untuk mengidentifikasi lokasi parkir terdekat dan terbaik. Tren serupa diterapkan pada bisnis, di mana sensor IoT di lahan pertanian mengumpulkan data tentang suhu, jumlah curah hujan, dan kelembapan. Informasi ini kemudian digunakan untuk menentukan waktu dan jumlah air yang akan disemprotkan.

Banyak analis telah memberikan data tentang ukuran pasar IoT. Sebuah studi oleh IHS Markit memperkirakan jumlah perangkat IoT akan tumbuh dari 27 miliar pada tahun 2017 menjadi 125 miliar pada tahun 2030. Agar perangkat-perangkat ini dapat mengakses jaringan dengan aman dan melakukan tugas-tugas otomatisasi tipe loop tertutup (yaitu, mengumpulkan data, menganalisis, memprediksi, menerapkan logika, dan bertindak) menimbulkan tantangan yang luar biasa bagi jaringan otonom. Ukuran data yang sangat besar yang akan dihasilkan dari ekosistem IoT ini merupakan pekerjaan besar untuk dikelola dan diproses. Rantai pasokan data yang sukses dari ekosistem ini pasti akan menghasilkan sistem AI yang lebih cerdas.

Selama dekade mendatang, kemunculan jaringan otonom dan perluasan IoT akan melahirkan hal besar berikutnya *Internet of Everything* (IoE). IoE adalah konsep yang memungkinkan perangkat apa pun berinteraksi dengan perangkat apa pun tanpa harus terhubung langsung ke internet. Bayangkan sensor Pemanas, Ventilasi, dan Pendingin Udara (HVAC) Anda berinteraksi dengan sensor mobil Anda sehingga suhu rumah Anda dapat disesuaikan saat Anda berada dalam radius lima mil dari rumah Anda. Memiliki 125 miliar perangkat yang saling terhubung merupakan pergeseran paradigma yang fenomenal dalam masyarakat kita dan memunculkan sejumlah besar layanan inovatif yang menggunakan AI di "tepi" pelanggan atau jaringan.

Layanan Anda dalam Sekejap

Evolusi akses jaringan ke 5G, perpindahan aplikasi cloud ke tepi, dan pergeseran dari SDN ke jaringan otonom merupakan pendorong utama untuk memungkinkan layanan latensi ultra-rendah. Layanan ini dapat berjalan dengan latensi jaringan bolak-balik di bawah 20 milidetik jika beroperasi di "tepi" jaringan, dan latensi bolak-balik di bawah 5 milidetik jika beroperasi di "tepi" pelanggan. Satu-satunya tantangan yang tersisa adalah waktu komputasi, bukan latensi jaringan.

Daya komputasi untuk mendukung perkembangan AI dan IoT berkembang pesat. GPU modern, yang awalnya dikembangkan untuk aplikasi gim, kini mampu melakukan pemrosesan paralel berkinerja tinggi untuk algoritma pelatihan ML. GPU bersifat multi-threaded dan memungkinkan eksekusi tugas-tugas dalam memori secara paralel. Hal ini memberikan kecepatan dan efisiensi untuk memproses data dalam jumlah besar untuk pembelajaran dan adaptasi sistem AI.



Akselerasi komputasi dan jaringan yang pesat akan memungkinkan aplikasi AR dan VR berskala besar, serta layanan penting, seperti drone otonom dan mobil otonom. AR melapisi grafik 3D virtual ke dunia nyata kita, sementara VR membenamkan kita dalam dunia sintetis dengan pengalaman 360 derajat. Baik VR maupun AR memungkinkan kita mengubah bisnis dan masyarakat dengan mengubah dunia fisik mereka menjadi dunia virtual yang realistis. Hal ini berpotensi merevolusi pelatihan karyawan, konferensi, perbaikan produk dan layanan, serta diagnostik pasien.

Persyaratan utama untuk aplikasi latensi rendah, seperti AR dan VR, adalah mempertahankan pengalaman yang seragam tanpa jeda, tersendat, dan macet. Latensi di bawah 15 milidetik diperlukan untuk waktu gerak-ke-foton guna menghindari ketidaknyamanan dan untuk mempertahankan pengalaman yang imersif. Lebih lanjut, menjadi tanpa kabel dan mobile akan mendorong adopsi besar-besaran teknologi ini. Bayangkan saja mengalami pertandingan olahraga langsung dengan tampilan 360 derajat dari sudut mana pun Anda berpaling dan di mana pun di dunia. Setiap orang melihat sudut pandang yang berbeda dan menerima pengalaman yang berbeda. Di sinilah 5G dan edge cloud menjadi teknologi krusial untuk mencapai tingkat konektivitas nirkabel latensi rendah ini. Bersama AI dan jaringan otonom, kami mampu menangkap data dalam jumlah besar dan menyediakan layanan personal sesuai permintaan, seperti ringkasan acara, tayangan ulang, dan prediksi langkah selanjutnya, antara lain.

Drone Terbang Bersayap untuk Menyelamatkan

Dalam Badai Maria, salah satu bencana alam terbesar tahun 2017 yang melanda Puerto Riko, AT&T menerbangkan drone *Cell on Wing* (COW) untuk menyediakan layanan. Drone terbang COW tersebut melayang sekitar 60 meter di atas permukaan tanah untuk menyediakan LTE ke area seluas sekitar 120 kilometer persegi. Konektivitas nirkabel di daerah terpencil, bencana alam, dan kemacetan lalu lintas sangat penting untuk menyelamatkan nyawa dan membantu petugas tanggap darurat. SDN dan AI, yang bekerja sama, dapat mendorong dunia di mana konektivitas nirkabel tersedia di mana pun dan kapan pun dibutuhkan.

Memiliki konektivitas dan bandwidth yang memadai, sesuai kebutuhan, bukanlah masalah yang mudah. Tantangan pertama adalah mendeteksi kapan dan di mana konektivitas atau bandwidth tambahan dibutuhkan. Dalam kasus bencana alam, mudah untuk mengetahui kapan kebutuhan akan muncul dan lokasinya. Namun, dalam situasi normal sehari-hari, seperti kemacetan lalu lintas, kegagalan jaringan seluler, dan acara olahraga besar, mengetahui di mana terdapat kebutuhan bukanlah hal yang mudah. Dalam jaringan otonom, data dikumpulkan secara konstan dari semua sel mikro dan makro, dan kegagalan dideteksi hampir secara real-time. Dalam beberapa kasus, pembelajaran mesin memprediksi kegagalan dan kemacetan ini sebelum terjadi. Dalam waktu dekat, kita akan dapat melihat drone terbang secara otonom ke lokasi optimal di mana pun dan kapan pun konektivitas dan bandwidth lebih dibutuhkan. Ini adalah bentuk otomatisasi loop tertutup di mana aksinya adalah COW terbang yang ditempatkan secara optimal untuk sementara waktu guna meningkatkan jangkauan hingga perilaku normal kembali.



Meningkatkan konektivitas dan meningkatkan kapasitas adalah beberapa manfaat jaringan otonom di masa depan. Manfaat lainnya termasuk mendeteksi dan memperbaiki kegagalan fisik, baik di lokasi sel, tiang jalan, maupun perlengkapan utilitas. Operator saat ini menyewa atau memiliki jutaan tiang jalan yang mereka gunakan untuk memperpanjang kabel atau menempatkan peralatan seperti sel kecil. Beberapa di antaranya berada di lokasi yang sangat terpencil, sementara yang lain berada di kota-kota yang sangat padat. Perawatan atau perbaikan sederhana peralatan ini merupakan pekerjaan besar dan membahayakan keselamatan pekerja. Ini merupakan peluang besar lainnya bagi AI untuk memanfaatkan teknologi visi mesin guna mendeteksi dan melakukan perawatan secara mandiri. Dalam studi terbaru oleh AT&T, drone dilengkapi dengan kemampuan analitik video, dan model pembelajaran mendalam dilatih untuk mendeteksi ratusan jenis objek berbeda dalam sel kecil, seperti penyangga jumper dan ground bar (Gambar 1.9). Algoritma analitik video juga mampu mengidentifikasi puluhan jenis kategori perbaikan yang berbeda, seperti kotoran dan karat. Teknisi dikirim ke lokasi kejadian hanya jika perbaikan jarak jauh tidak dapat dilakukan atau robot drone tidak dapat memperbaiki masalah. Pergeseran ke arah adopsi drone seluler dan algoritma AI untuk membantu rekayasa dan infrastruktur akan berdampak luar biasa dalam menurunkan biaya perbaikan dan meningkatkan kualitas hidup teknisi dan insinyur.



Gambar 1.9 Video yang direkam oleh drone.

Bagaimana Saya Dapat Melayani Anda?

Salah satu aplikasi awal teknologi ML dan AI yang sukses adalah respons suara interaktif (IVR) untuk layanan pelanggan. Hampir dua dekade lalu, perusahaan mulai memperkenalkan pengenalan suara sebagai alternatif respons nada sentuh. Tujuannya saat itu adalah menyediakan antarmuka alami untuk memahami pertanyaan pelanggan dan mengarahkan pelanggan dengan tepat ke agen yang tepat. Perutean yang salah telah terbukti mengurangi skor kepuasan pelanggan dan meningkatkan biaya layanan perusahaan. Saat ini, pelanggan menghubungi dan berinteraksi dengan pusat kontak melalui bot berbasis suara atau obrolan dan menyelesaikan banyak kebutuhan penagihan, penjualan, dan perbaikan mereka tanpa harus berinteraksi dengan agen manusia.

Seiring kita beralih ke teknologi SDN dan cloud dengan kotak putih, layanan pelanggan juga akan bergeser dari reaktif menjadi proaktif. Dalam sebuah studi terbaru yang diselesaikan untuk layanan video broadband, sebuah algoritma ML digunakan untuk memprediksi dekoder mana yang akan mengalami masalah kualitas di masa mendatang. Pengalaman buruk sangat berkorelasi dengan ketidakpuasan pelanggan terhadap layanan mereka dan menghubungi pusat kontak mereka untuk perbaikan. Setelah kotak set-up dengan kinerja yang buruk



diidentifikasi, kemudian dijadwalkan untuk di-reboot selama waktu di luar jam sibuk. Algoritme tersebut terbukti mendeteksi 25% dari beberapa upaya perbaikan dengan penghematan tahunan sebesar Rp 10 miliar untuk 400 ribu pelanggan. Yang terpenting, ML terbukti meningkatkan skor promotor bersih pelanggan dan retensi mereka untuk tetap menjadi pelanggan setia. Studi juga menunjukkan bahwa algoritme ML semacam itu dianggap kurang bias daripada agen manusia.

Ketika masyarakat bergerak menuju transformasi digital, menerima revolusi AI, teknologi robotika akan mulai memainkan peran yang lebih besar dalam kehidupan kita. Tiongkok memimpin dalam hal ini, menyediakan robot percakapan untuk merawat populasi lansia yang besar dan terus bertambah. Robot-robot ini memiliki algoritme AI yang mengumpulkan data dan belajar beradaptasi dengan habitat dan preferensi pemiliknya. Mereka memastikan bahwa lansia menerima obat dan makan makanan mereka. Robot mengirimkan sinyal peringatan jika mereka melihat kejadian yang tidak biasa, seperti lansia jatuh atau tidak bernapas dengan baik. Penggunaan robot cerdas ini secara luas di dunia Barat akan terjadi dalam dekade mendatang seiring dengan semakin matangnya jaringan otonom dengan teknologi 5G dan edge cloud.

Adopsi besar-besaran perangkat, sensor, dan robotika yang hemat biaya dan cerdas hanya akan terjadi ketika jaringan otonom diterapkan secara luas. Bab 2 akan mengulas algoritma dan teknologi di balik AI. Sistem AI dapat berbasis aturan, digerakkan oleh data, atau gabungan keduanya. Sistem tersebut dapat diawasi, tanpa pengawasan, atau semi-diawasi. Bab ini menyoroti perbedaan-perbedaan tersebut dan mengulas algoritma AI yang relevan, mulai dari yang tradisional seperti regresi dan pengelompokan hingga teknik ML modern seperti NLP, pembelajaran mendalam, jaringan konvolusional, pembelajaran penguatan, dan lainnya. Bab ini juga menjelaskan beberapa aplikasi AI yang berhasil di bidang pengenalan suara, penerjemahan bahasa, pemrosesan video, serta pengenalan wajah dan objek.



BAB 2

PEMBELAJARAN MESIN, DEEP LEARNING, DAN AI

2.1 PENDAHULUAN

Kecerdasan buatan (AI) adalah bidang di mana satu ukuran pasti tidak cocok untuk semua, dan pemahaman tentang berbagai jenis pembelajaran mesin (ML) yang tersedia bagi kita adalah kunci untuk membangun solusi yang sukses untuk jaringan otonom. Sistem AI dapat berbasis aturan, digerakkan oleh data, atau gabungan keduanya. Sistem ini dapat diawasi, tanpa pengawasan, atau semi-diawasi. Dalam bab ini, kami membahas perbedaan-perbedaan tersebut dan pengaruhnya terhadap solusi yang kami ciptakan. Kami meninjau algoritma AI yang relevan, mulai dari yang tradisional, seperti regresi dan pengelompokan, hingga teknik ML modern, seperti pemrosesan bahasa alami, pembelajaran mendalam (DL), jaringan konvolusional, dan pembelajaran penguatan. Kami juga akan menjelaskan keberhasilan penerapan AI di bidang pengenalan suara, penerjemahan bahasa, pemrosesan video, serta pengenalan wajah dan objek.

Bagian 2.2 memberikan perspektif historis tentang evolusi kecerdasan mesin dan para peneliti inovatif yang terlibat, dari "perseptor" awal hingga sistem pakar, optimasi, dan peramalan, hingga DL modern. Bagian 2.3 memberikan gambaran umum tentang berbagai jenis masalah ML tradisional (misalnya, mendeteksi anomali, mengklasifikasikan spam, memprediksi kegagalan perangkat) dan cara penyelesaiannya. Bagian 2.4 membahas alasan di balik dan kemungkinan revolusi Pembelajaran Mesin (DL) dengan detail tentang di mana kehebohan tersebut memang diperlukan dan di mana kehebohan tersebut mungkin terlalu berlebihan. Kami mengakhiri Bab 2 dengan diskusi tentang terobosan AI saat ini dan apa yang akan terjadi di masa depan, baik tantangan maupun peluangnya.

2.2 KECERDASAN BUATAN SEJARAH SINGKAT

Apa itu AI? Menurut salah satu tokoh terkemuka di bidang ini, "begitu berhasil, tidak ada yang menyebutnya AI lagi." Definisi tentang apa yang menjadikan suatu sistem sebagai AI telah berubah seiring waktu, tetapi dengan kebangkitan baru-baru ini dalam ML, orang biasanya menggunakan AI untuk merujuk pada solusi yang menggunakan ML dan "menutup lingkaran" dengan mengambil tindakan, dan berpotensi belajar. Anda dapat membedakan lebih lanjut antara AI di atas dan kecerdasan umum buatan (AGI) dari mesin fiksi seperti HAL9000 atau Terminator. Dalam buku ini, kami akan berfokus pada model ML dan AI yang membantu jaringan otonom berfungsi, dan sistem AI yang menyatukan model menjadi sistem yang kohesif untuk pencegahan kesalahan, deteksi intrusi, dan perbaikan diri. Untuk memahami sistem ini, kami akan mengkaji sejarah ML dan AI, serta melihat berbagai jenis ML dan bagaimana keduanya digunakan.

Pembelajaran Mesin Awal

Pembelajaran mesin telah ada selama beberapa dekade. Beberapa bahkan mungkin mengatakan berabad-abad jika Anda memasukkan metode statistik seperti Teorema Bayes.



Untuk tujuan kita, kita akan mulai pada tahun 1950an ketika Marvin Minsky dan Dean Edmonds membangun mesin jaringan saraf pertama (Kalkulator Penguatan Analog Saraf Stokastik [SNARC]) dan kemudian Frank Rosenblatt menciptakan Perceptron. Alan Turing telah menulis makalah penting pada tahun 1950 yang memperkenalkan dunia pada "Tes Turing", dan dengan ukuran kecerdasan tersebut di benak publik dan keberhasilan baru SNARC dan Perceptron dalam memecahkan masalah nyata, "zaman keemasan" AI berkembang pesat dengan antusiasme dan dana penelitian yang mengalir deras.

Selama masa keemasan ini, kemajuan dalam metode AI menghasilkan pengembangan banyak "pemecah masalah" seperti "Pembukti Teorema Geometri" karya Herbert Gelertner dan "Pemecah Umum" karya Newell. Mereka mengabdikan diri untuk mencoba memecahkan masalah dengan cara yang mereka anggap "manusiawi", dengan mencari di ruang kemungkinan pilihan di setiap langkah, menelusuri kembali, dan mempertimbangkan pilihan baru (bayangkan kita memecahkan labirin). Yang lain mengabdikan penelitian untuk memecahkan masalah pemahaman dan komunikasi dalam bahasa alami seperti ELIZA karya Weizenbaum, yang dapat melakukan percakapan terapis sederhana. Kemajuan-kemajuan ini dan lainnya membuat Minsky memprediksi pada tahun 1970 bahwa "Dalam tiga hingga delapan tahun, kita akan memiliki mesin dengan kecerdasan umum manusia rata-rata." Kemajuan-kemajuan ini mendorong *Defense Advance Research Projects Agency* (DARPA) dan lembaga-lembaga lain, baik pemerintah maupun perusahaan, untuk berinvestasi dalam jumlah besar dalam penelitian AI, dengan Minsky dan John McCarthy membentuk laboratorium-laboratorium berpengaruh di MIT dan Stanford.

Namun, bahkan ketika Minsky memprediksi kecerdasan mesin "manusia rata-rata" akan muncul dalam waktu dekat, "Zaman Es AI" pertama sudah dimulai. Minsky sendiri menerbitkan sebuah kritik terhadap kemampuan Perceptron (jaringan saraf) yang menghentikan semua penelitian tentangnya selama bertahun-tahun, dengan dana penelitian mengering karena para peneliti menyadari keterbatasan komputasi dan ledakan kombinatorial dari masalah-masalah yang mereka coba pecahkan. Ilmuwan komputer mengadopsi komponen-komponen yang mudah digunakan dan dapat diprediksi seperti metode pencarian, yang kemudian tidak lagi dianggap sebagai "AI," dan penelitian AI itu sendiri terhenti.

Sistem Pakar dan Perjuangan Melawan Probabilitas

Pada pertengahan 1970an, Zaman Es AI pertama akhirnya mulai mencair dengan munculnya "Sistem Pakar". Sistem ini berbasis pengetahuan yang mengodekan keahlian domain dan proses bisnis untuk memungkinkan mesin membuat dan memandu keputusan di tempat kerja. Sistem pendukung keputusan berbasis aturan ini berbeda dari sistem AI generasi sebelumnya dalam dua hal utama: pertama, sistem ini diodekan secara eksplisit dengan pengetahuan detail tentang domain tempat mereka bekerja, alih-alih mencoba mempelajarinya dari prinsip-prinsip dasar, dan kedua, kemampuan prediktifnya terbatas pada satu masalah, alih-alih mencoba menjadi model umum yang mampu mempelajari "penalaran probabilistik" di seluruh domain.

"Zaman keemasan" kedua ini menyaksikan perkembangan program catur yang sangat terampil yang menggabungkan metode pencarian dari zaman keemasan pertama dan



pengetahuan domain dari ribuan rekaman permainan dari para master catur. Dengan cara yang sama, pencarian dikombinasikan dengan korpus pengetahuan di domain lain seperti kedokteran, dan bahkan area yang biasa tetapi berdampak seperti konfigurasi komputer (misalnya, XCON). Sekali lagi, janji itu membawa banjir uang dan gelombang baru prediksi muluk. Sekali lagi, prediksi tersebut gagal terwujud, dan gelembung AI kedua pecah, yang mengarah ke Zaman Es AI lainnya di akhir tahun 1980an.

Memprediksi Masa Depan, Mengoptimalkan Masa Kini

Selama Zaman Es AI baru ini di akhir tahun 1980an dan awal tahun 1990an, teknik-teknik yang berada dalam keluarga yang sama dengan AI tetapi telah menghindari stigmanya menggemparkan bisnis. Optimasi dan teknik-teknik lain dari komunitas riset operasi mulai berkembang melampaui industri aslinya. Metode peramalan mulai diterapkan lebih luas dan digunakan untuk memecahkan masalah di luar keuangan. Metode optimasi umum dan algoritma genetika merambah berbagai industri, menghasilkan kemajuan dalam desain sirkuit, bioteknologi, dan banyak domain lain di mana ruang pencarian kemungkinan jawaban terasa mustahil.

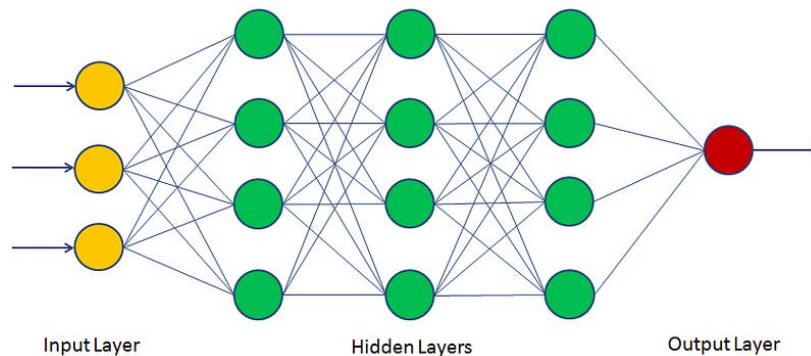
Pembelajaran Mesin Saat Ini dan Mendalaminya

Dengan maraknya internet dan e-commerce di akhir tahun 1990an dan 2000an, sejumlah besar data bisnis terakumulasi dan revolusi "big data" dimulai. Sejumlah besar data log dihasilkan oleh server komoditas dan penerapan layanan cloud. Platform periklanan daring yang besar pun berkembang. Semua pemilik sistem ini membutuhkan metode untuk memahami apa yang terjadi dan bagaimana mereka dapat secara efektif memanfaatkan informasi yang telah mereka kumpulkan. Teknik pembelajaran mesin tradisional mulai kembali menonjol, dan ledakan pembelajaran mesin baru pun dimulai. Tiba-tiba, para ahli statistik, yang beralih dari regresi data menjadi membangun model prediktif yang kompleks, berganti nama menjadi "ilmuwan data", dan mereka yang mampu mengelola data dalam skala besar mulai mendapatkan nilai yang sepadan. Memasuki tahun 2010, sistem seperti SparkML, H2O, dan lainnya berevolusi untuk menjalankan ML dalam skala besar, dan metode ML yang lebih canggih seperti *Support Vector Machines* dan *Boosted Decision Trees* segera mendominasi lanskap.

Sementara itu, sekelompok kecil orang di Kanada menolak untuk meninggalkan janji jaringan saraf, dan mereka akan segera meledak ke panggung. Geoffrey Hinton dan muridnya Alex Krizhevsky menemukan AlexNet, dengan mudah mengalahkan solusi terbaik untuk tantangan pengenalan gambar ImageNet pada tahun 2012 dan meluncurkan DL ke panggung. Jaringan saraf dalam ini dibangun di atas karya sebelumnya Hinton dkk. (back-propagation) dan Fukushima (jaringan saraf konvolusional), dan karya mereka disempurnakan dan diperluas oleh Yann LeCun dan Yoshua Bengio (di antara banyak lainnya) dengan lebih banyak lapisan tersembunyi dan neuron khusus. Sejak saat itu, jaringan dalam telah berkembang pesat dalam ukuran dan kemampuan, mengalahkan para grand master di Go (sebuah masalah yang kami pikir masih jauh dari solusi selama beberapa dekade) dan mendeteksi kanker kulit dan pneumonia lebih baik daripada pakar manusia. Pembelajaran mesin tradisional masih memiliki



tempat, tetapi model pembelajaran mesin berbasis jaringan saraf multilayer (Gambar 2.1) pasti akan memperluas jangkauannya, dan kita masih berada di tahap awal revolusinya.



Gambar 2.1 Jaringan saraf tiruan multilapis.

2.3 CARA-CARA PEMBELAJARAN MESIN

Sejarah singkat ML memang bermanfaat, tetapi hanya memberi kita sedikit wawasan tentang bagaimana ML beroperasi di alam bebas. Di bagian ini, kita akan membahas berbagai "cara" ML dan penerapannya, serta apa yang diperlukan untuk memanfaatkannya. Salah satu rahasia kotor di antara sebagian besar praktisi ML adalah bahwa 80% waktu kita dihabiskan untuk menangani data menemukannya, membersihkannya, dan memahami artinya. "Garbage in, garbage out" sangat relevan dalam ranah ML, jadi beberapa bagian pertama kita akan berfokus pada data. Kemudian, kita akan membahas bagaimana kita menggunakan data tersebut, mulai dari masalah "supervised" di mana kita tahu apa yang kita cari, hingga masalah "unsupervised" di mana kita mencoba menemukan struktur dari data yang kita miliki. Kemudian, kita akan menutup dengan melihat sistem yang belajar sendiri.

Belajar dari Data: Fitur, Label, dan Pemecahan Masalah yang Tepat

Inti dari pembelajaran mesin ditentukan oleh data yang dikandungnya. Agar dapat belajar dari data, sistem perlu diberikan data tersebut, dan dalam format yang dapat dipahami. Sebagian besar sistem pembelajaran mesin tidak menangani data mentah; data tersebut perlu diubah menjadi "fitur" (biasanya angka) melalui proses yang disebut "featurisasi", yang mengubah data yang masuk ke dalam format yang dapat digunakan sistem dan mengekstrak detail yang relevan dari yang tidak relevan. Jika Anda tahu apa yang Anda cari dalam data, data Anda mungkin juga memiliki "label", yang bisa berupa numerik (dalam hal ini Anda memiliki masalah regresi) atau kategorikal (dalam hal ini Anda memiliki masalah klasifikasi). Bayangkan kasus seperti pengenalan gambar; "Fitur" Anda adalah piksel gambar dan "label" Anda adalah objek utama dalam gambar (misalnya, "kucing") dari serangkaian "kelas" tetap seperti "kucing", "anjing", "perahu", dll.

Mengingat model ML dilatih dengan angka, salah satu masalah pertama yang kita hadapi sebagai ilmuwan data adalah mengubah data non-numerik kita menjadi angka yang berfungsi sebagai fitur untuk pelatihan. Ada banyak metode untuk melakukannya, tergantung pada jenis data yang kita tangani. Untuk data "kategorikal" (bayangkan kategori diskrit seperti warna atau jenis bunga), kita bisa saja mengodekannya sebagai angka (misalnya, "merah" = 0,



"hijau" = 1, "merah muda" = 2, ...), tetapi ini dapat membingungkan model, membuatnya salah mengira bahwa "merah" dan "hijau" lebih dekat daripada "merah" dan "merah muda". Kami sering menggunakan metode yang disebut "pengodean one-hot" untuk mengatasi hal ini, yang mengubah fitur kategoris menjadi vektor biner dengan hanya satu bit "hot" (misalnya, "merah" = [1, 0, 0]) yang memastikan setiap kategori bersifat independen.

Untuk data teks, kami sering menggunakan metode dari pemrosesan bahasa alami (NLP). Ini dapat melibatkan "stopwording", sebuah proses untuk menghapus kata-kata yang tidak relevan seperti "of", "and", dan "the"; dan "stemming" atau "lemmatization", yang keduanya dimaksudkan untuk mengubah kata-kata serupa dengan akar yang sama menjadi kata yang sama ("learn", "learns", dan "learning" semuanya dikonversi menjadi "learn" misalnya). Ini memungkinkan kami untuk menghapus kata-kata yang tidak prediktif, dan untuk menciutkan kata-kata yang secara semantik serupa. Kami kemudian mungkin memutuskan untuk hanya menginginkan 10.000 kata paling umum dalam kosakata kami, jadi kami mengubah setiap dokumen menjadi vektor jumlah kata dengan 10.000 elemen. Ini berpotensi mengubah dokumen besar menjadi vektor angka padat sepanjang 10.000, atau dokumen kecil seperti tweet menjadi vektor yang sangat jarang. Kami juga sering mempertimbangkan "hashing," yang memungkinkan kami untuk menggunakan kosakata lengkap tetapi masih membatasi jumlah fitur, dengan menempatkan setiap kata ke dalam "ember hash" yang mungkin bertabrakan dengan kata lain.

Featurisasi terkadang menggunakan teknik yang dikenal sebagai *Term Frequency/Inverse Document Frequency* (TF/IDF), yang memberi bobot lebih berat pada kata-kata yang tidak umum secara keseluruhan tetapi umum dalam dokumen tertentu (misalnya, "mesin" dan "pembelajaran" lebih banyak muncul di buku ini daripada di semua buku pada umumnya). Kita juga dapat menggunakan "embedding kata," atau jaringan saraf sederhana yang mempelajari vektor untuk setiap kata sehingga kata-kata yang secara kontekstual mirip menjadi "dekat" satu sama lain. Ini dapat mempelajari beberapa semantik yang cukup dalam, bahkan memungkinkan Anda untuk "mengerjakan matematika" pada vektor yang dipelajari dengan cara yang menarik. Pertimbangkan contoh terkenal di bidang ini: jika Anda menambahkan vektor "Raja" – "Pria" + "Wanita," hasil vektor terdekat akan menjadi "Ratu." Word2Vec menjadi terkenal di bidang ini, kemudian menjadi agak terkenal karena orang-orang menyadari bahwa Word2Vec tidak hanya mengkodekan kesamaan tetapi juga bias algoritmik (seksisme dan rasisme) dalam data yang mendasarinya.

Bagian penting dari featurisasi adalah "rekayasa fitur", di mana Anda menyederhanakan data menjadi hanya komponen-komponen yang relevan dengan permasalahan yang dihadapi. Sebuah contoh mungkin dapat membantu. Pertimbangkan kasus jaringan di mana Anda mencoba mendeteksi anomali. Data awal Anda mungkin berupa stempel waktu dan ukuran setiap paket masuk dan keluar di setiap switch. Ini adalah jumlah data yang sangat besar, yang sebagian besar kemungkinan tidak relevan. Anda dapat memilih untuk melakukan rekayasa fitur dengan menggabungkan jumlah permintaan dalam rentang waktu tertentu dengan ukuran permintaan rata-rata dan maksimum. Anda juga dapat memilih untuk melakukan ini pada beberapa tingkat agregasi. Misalnya, periode 1 milidetik untuk 30 milidetik terakhir,



kemudian periode 5 milidetik selama 30 milidetik, dst. Jika Anda memiliki anomali yang sudah diketahui, Anda dapat menghasilkan "set data berlabel" dan mulai melatih model.

Pelatihan (atau "pemasangan") model adalah proses mengambil algoritma ML yang ada dan menggunakan data untuk mengajarkannya memecahkan masalah tertentu. Untuk melakukannya dalam masalah pembelajaran terawasi (di mana kita memiliki data dan label yang ingin kita prediksi), kita biasanya membagi data kita ke dalam set yang berbeda "set pelatihan" adalah data yang kita tunjukkan pada model untuk melatihnya, dan "set validasi" digunakan untuk menguji model yang telah kita latih dan melihat seberapa baik kinerjanya pada data yang belum pernah dilihatnya. Seringkali set "uji" atau "tahanan" ketiga digunakan untuk memungkinkan kita membandingkan beberapa model satu sama lain pada data yang belum pernah ditemui oleh satu pun. Sebagai ilmuwan data, tugas kita adalah menentukan bagaimana kita membagi data kita ke dalam set-set tersebut, dengan mempertimbangkan ukuran data yang harus kita latih versus kemampuan untuk menguji kinerja kandidat model akhir kita. Jika kita melakukan pekerjaan kita dengan baik, kita akan mendapatkan model yang "menggeneralisasi" dan dapat bekerja dengan baik dalam sistem produksi pada data baru. Jika kita melakukan pekerjaan kita dengan buruk, kita bisa berakhir dengan "overfitting" dan model kita belajar untuk berkinerja baik pada data pelatihan tetapi berkinerja buruk pada data uji.

Mengevaluasi model selama pelatihan pada data validasi membantu meminimalkan overfitting. Kebutuhan akan data pelatihan, sebagian, mendorong revolusi "big data". Seiring perusahaan berupaya mendapatkan kecerdasan dari data mereka, mereka mulai menginstrumentasikan dan mencatat semuanya hingga mencapai titik di mana raksasa teknologi modern seperti Microsoft dan Google memiliki eksabita data yang digunakan untuk model pelatihan.

Kualitas Data: Big Data (Seringkali) tidak Cukup

Data saja tidak cukup untuk mendorong model yang berkualitas. ML, bahkan lebih dari bidang perangkat lunak lainnya, rentan terhadap kutukan "sampah masuk, sampah keluar". Mengumpulkan petabita data yang salah tidaklah berguna, dan bahkan dengan data yang valid, seringkali terdapat masalah yang harus diatasi sebelum Anda dapat menggunakannya dalam pembuatan model. Daftar lengkap semua masalah kualitas data berada di luar cakupan bagian ini, tetapi kami akan membahas beberapa secara detail untuk memberi Anda gambaran tentang teknik yang tersedia dan jenis masalah yang dihadapi praktisi ML.

Seringkali dalam masalah ML kita mengalami situasi di mana kita mencoba mengkategorikan data yang masuk ke dalam satu dari dua atau lebih kelas, dan satu kelas mungkin jauh lebih umum daripada yang lain. Pertimbangkan deteksi spam di mana email spam jauh lebih umum daripada non-spam, atau deteksi pelanggaran di mana sebagian besar lalu lintas jaringan bukan merupakan pelanggaran. Dalam kasus "tidak seimbang" atau "miring" seperti ini, Anda mungkin perlu membuang data di kelas mayoritas untuk menciptakan keseimbangan yang lebih besar, atau menggunakan cara buatan untuk memperluas kelas minoritas dengan menduplikasi data di kelas tersebut atau menggunakan teknik pembuatan data seperti *Synthetic Minority Over-Sampling Technique* (SMOTE). Fenomena yang sama dapat terjadi dengan fitur juga. Pikirkan skenario medis di mana



golongan darah adalah sebuah fitur; pasien AB- akan sangat jarang, yang dapat (antara lain) menyebabkan model ML mengabaikan fitur itu, bahkan jika itu menjadi sangat prediktif. Teknik augmentasi data yang serupa dapat membantu memecahkan masalah ini juga.

Masalah lain yang sering terjadi dalam ML adalah "kutukan dimensionalitas," yang sebenarnya adalah keluarga masalah yang terjadi ketika jumlah fitur dalam data Anda tumbuh melampaui segelintir menjadi ratusan, ribuan, atau bahkan ratusan ribu, yang berpotensi mengerdilkan jumlah sampel Anda. Ketika Anda mencoba melatih model dengan banyak fitur tersebut, Anda mungkin tidak memiliki cukup data untuk mengajarkan model Anda, pertimbangkan untuk mencoba mempelajari model untuk menggambar garis berkelok-kelok hanya dari dua atau tiga titik. Praktisi ML menangani masalah ini dengan mengurangi jumlah fitur dan/atau menambahkan lebih banyak data. Kami akan sering memeriksa fitur kami dan melakukan "pengujian signifikansi" untuk menentukan fitur mana yang memiliki daya prediktif paling besar untuk model kami dan membuang yang lain. Kami juga dapat menggunakan teknik seperti analisis komponen utama (PCA) untuk mengubah fitur asli menjadi ruang fitur komposit baru untuk memungkinkan kami memilih fitur konstruksi yang paling prediktif dari ruang baru ini. PCA sangat kuat dalam situasi di mana fitur input besar dan jarang, seperti dalam pemrosesan gambar. Hal ini dapat mengkonsolidasikan fitur-fitur tersebut menjadi fitur komposit yang jauh lebih sedikit dan terpisah secara rapi, sehingga pelatihan model menjadi jauh lebih mudah.

Sering kali dengan sistem baru, Anda akan menghadapi masalah yang berlawanan cold start. Bagaimana Anda bisa melakukan bootstrap pada sistem ML jika Anda tidak memiliki data yang ada? Sering kali, jawaban terbaik adalah mulai mengumpulkan data terlebih dahulu dan membangun komponen ML kemudian. Namun, jika Anda mengerjakan sistem baru di domain yang dikenal, Anda dapat menggunakan metode simulasi dan augmentasi data untuk membantu melakukan bootstrap pada model Anda. Misalnya, saat membangun jaringan baru, Anda mungkin memiliki wawasan tentang lalu lintas yang Anda harapkan berdasarkan pengetahuan domain Anda tentang jaringan, dan Anda dapat menggunakan ini untuk membangun simulasi dan menggunakan data yang dihasilkan untuk membuat model awal Anda. Saat Anda mengumpulkan data, Anda dapat menyesuaikan simulasi Anda untuk memodelkan jaringan Anda dengan lebih baik dan menggunakan teknik augmentasi untuk memperluas set data Anda. Putaran umpan balik ini bisa sangat efektif.

Ini hanyalah beberapa masalah yang dapat Anda hadapi dengan data, dan mengetahui cara memilih atau membangun fitur yang tepat dan memilih model yang paling tepat adalah seni sekaligus sains. Hal ini datang dengan pengalaman, keahlian di bidang tertentu, dan eksperimen. Teknik-teknik baru seperti auto-ML dan alat-alat baru seperti PROSE dapat membantu meringankan sebagian beban tersebut, tetapi dalam setiap proyek ML, kita tetap berharap untuk menghabiskan sebagian besar waktu pengembangan kita untuk memproses, membersihkan, dan bereksperimen dengan data.

Anomali dan Klaster

Kita umumnya membagi pembelajaran mesin menjadi dua jenis utama: pembelajaran terawasi di mana kita tahu apa yang kita cari dan memiliki pengetahuan tentang label target,



dan pembelajaran tanpa pengawasan di mana kita mencoba mempelajari kelas atau inferensi tanpa label target. Bagian ini dikhususkan untuk teknik pembelajaran tanpa pengawasan.

Terkadang kita berakhir dengan kumpulan data yang kita yakini mungkin memiliki beberapa struktur yang mendasarinya, tetapi kita tidak yakin dan perlu mencari tahu. Pengelompokan dapat membantu kita menemukan struktur ini dengan mengelompokkan item-item serupa dengan berbagai cara. Kita kemudian dapat memeriksa fitur-fitur item dalam klaster yang sama untuk menentukan karakteristik apa yang membuat item-item tersebut serupa dan apakah kita menganggapnya memiliki makna dunia nyata yang sebenarnya atau hanya artefak dari kumpulan data yang telah kita kumpulkan. Pengelompokan hadir dalam beberapa bentuk berbeda, tetapi metode yang umum adalah mengelompokkan item berdasarkan perhitungan jarak antar fiturnya. Ini berarti bahwa semua data kita harus diubah menjadi bentuk numerik di mana ukuran jarak tersebut memiliki beberapa makna, menggunakan berbagai cara seperti vektor one-hot atau embedding seperti yang dijelaskan sebelumnya. Salah satu contohnya adalah pengelompokan k-means di mana "k" klaster berbeda dihitung berdasarkan sentroid paling prediktif dari kumpulan data berdasarkan ukuran jarak yang Anda gunakan.

Pengelompokan juga dapat digunakan untuk mendeteksi titik data yang tampak abnormal anomali atau outlier. Setelah kita mengelompokkan data dan mengetahui tampilannya, titik yang terlalu jauh dari klaster mana pun dapat dianggap anomali. Ini hanyalah salah satu teknik dalam rangkaian metode deteksi anomali, yang berguna di berbagai bidang, tetapi dapat menjadi sangat penting dalam ruang jaringan untuk menemukan kegagalan, misalnya. Karena ini adalah metode tanpa pengawasan, kita tidak tahu secara apriori seperti apa anomali itu, dan kita bergantung pada model, eksperimen, dan keahlian domain kita untuk memilih ambang batas yang tepat (misalnya, apa yang "terlalu jauh" dari klaster). Alih-alih mengelompokkan, kita dapat menggunakan metode dari peramalan deret waktu untuk menentukan, misalnya, apakah pola lalu lintas terlihat tepat untuk waktu tertentu dalam sehari/minggu/bulan. Kita dapat melakukan ini dengan menggunakan rata-rata bergerak yang disesuaikan secara musiman atau metode *Auto Regressive Integrated Moving Average* (ARIMA) dan kemudian membandingkan data aktual kita dengan hasil peramalan kita. Seringkali dalam sistem produksi, ini dapat dikombinasikan dengan model lain (beberapa bahkan diawasi jika kita memiliki data berlabel) untuk memberikan jawaban yang lebih baik.

Dalam jaringan, seperti yang akan terlihat di bab berikutnya, beberapa sinyal akan bergerak ke arah yang sama pada waktu yang sama (atau serupa) dan dapat menjadi indikasi kegagalan kaskade atau masalah berkorelasi lainnya. Mencari korelasi di antara beberapa sinyal input dapat membantu Anda mengidentifikasi masalah yang muncul dan kronis, dan melacak ini dari waktu ke waktu dapat mengarahkan Anda ke pembuatan set data berlabel yang memungkinkan model yang diawasi untuk dilatih.

Pembelajaran di Bawah Supervisi

Pembelajaran tersupervisi sejauh ini merupakan bentuk pembelajaran mesin yang paling umum. Ketika Anda tahu apa yang ingin Anda prediksi dan memiliki contoh serta label target yang diketahui, maka Anda memiliki masalah pembelajaran tersupervisi. Pertimbangkan



contoh dari bagian sebelumnya tentang mendeteksi anomali dalam jaringan: salah satu cara umum yang digunakan orang untuk memecahkan masalah ini adalah dengan menemukan anomali jaringan sebelumnya, melabelinya, lalu melatih model untuk mengklasifikasikan data yang masuk. Perhatikan bahwa metode ini berbeda dari metode tanpa supervisi karena Anda memerlukan data berlabel, tetapi metode ini dapat berkinerja lebih baik dalam beberapa kasus dan berpotensi memberikan jawaban yang terarah tentang jenis masalah jaringan yang mungkin Anda alami.

Namun, satu kelemahannya adalah metode ini hanya dapat memprediksi masalah yang telah terlihat atau serupa dengan yang telah terlihat. Inilah sebabnya metode ini mungkin tidak berfungsi dengan baik untuk masalah seperti deteksi pelanggaran di mana pelaku kejahatan terus-menerus mengubah taktik. Dalam kasus seperti itu, kita dapat mempertimbangkan untuk menggabungkan model tersupervisi kita dengan metode tanpa supervisi, yang berpotensi lebih kuat daripada masing-masing metode secara terpisah. Teknik pembelajaran semi-supervised ini menggunakan data tak berlabel untuk lebih memahami distribusi data berlabel, sehingga model dapat digeneralisasi dengan lebih baik ketika data baru ditemukan.

Metode supervised biasanya terbagi menjadi dua jenis: model klasifikasi yang mencoba memprediksi kelas suatu item, seperti "itu kucing" (untuk model multikelas) atau "ya, ini masalah" (untuk model biner), dan model regresi yang memberikan prediksi numerik seperti "kami perkirakan 300 m rps." Model-model ini memiliki kompleksitas yang beragam, mulai dari pencocokan garis sederhana (regresi linier dan logistik) hingga keluarga Pohon Keputusan yang kompleks. Sebagian besar pekerjaan dalam ML sebelum DL muncul dikhususkan untuk menciptakan model yang lebih kompleks dan berkinerja tinggi. Diskusi yang kuat tentang model ML berada di luar cakupan bab ini (dan sebagian besar buku), tetapi ada beberapa teknik yang masih banyak digunakan dan mampu dalam domain tertentu untuk bersaing atau mengalahkan model DL, seringkali dengan sebagian kecil dari biaya dalam waktu pelatihan dan penilaian, seperti Pohon Keputusan yang Ditingkatkan yang dijelaskan di akhir bagian ini.

Peneliti pembelajaran mesin tahu bahwa sementara model yang mereka latih dapat berkinerja baik pada beberapa contoh, itu dapat berkinerja buruk pada banyak contoh lainnya. Sering kali, mencoba melatih model untuk bekerja dengan baik pada semua contoh akan menghasilkan kinerja agregat yang lebih rendah. Masalah ini membuat mereka menyadari bahwa mereka dapat mengelompokkan data mereka dan melatih model yang berbeda yang ditargetkan pada subset data tertentu, menyusun (atau menggabungkan) mereka bersama-sama untuk menghasilkan hasil akhir yang jauh lebih baik daripada hasil dari setiap model individu. Metode model-dari-model ensemble tertimbang ini mendominasi Kaggle (<https://www.kaggle.com/>) dan kompetisi ML publik lainnya dan menghasilkan model yang kuat dalam produksi; bahkan dapat memungkinkan model yang ada untuk beradaptasi dengan data baru tanpa memaksa orang untuk membuang semua pekerjaan mereka yang ada hanya dengan menggabungkan submodel baru yang ditargetkan pada segmen data baru.

Akhirnya, para peneliti menyadari bahwa mereka dapat lebih efektif menargetkan kelemahan dalam ansambel mereka yang ada dengan mengamati gradien kesalahan (gradien



kesalahan dibahas nanti di bagian "Pembelajaran Mendalam"). Sinyal ini memungkinkan mereka untuk membangun model baru yang menargetkan titik lemah, menambahkan model baru itu ke ansambel, dan beriterasi. Setelah beberapa iterasi, proses peningkatan gradien ini menghasilkan model dengan kinerja yang lebih tinggi daripada ansambel berukuran serupa, dan kinerja yang jauh lebih tinggi daripada model tunggal mana pun. Pohon peningkatan gradien ini merupakan yang tercanggih untuk teknik non-DL sekarang, tetapi tidak ada alasan untuk berpikir kita tidak akan melihat kemajuan di luar itu di masa depan, terutama seiring dengan kemajuan bidang auto-ML.

Pelatihan Ulang dan Pembelajaran Dinamis

Setelah Anda menerapkan sistem pembelajaran terawasi ke dalam produksi, bagaimana Anda memastikannya terus berkinerja? Mekanisme utamanya adalah pengambilan data berlabel baru dan pelatihan ulang model secara berkala. Sebagian besar model pembelajaran mesin (ML) meningkatkan kinerjanya saat terpapar data baru yang menjadi sumber pembelajarannya, dan proses ini dapat berulang seiring waktu, menghasilkan peningkatan bertahap. Dengan menerapkan proses iteratif ini secara ekstrem, kita akan mendapatkan proses yang disebut pembelajaran daring di mana model disusun sedemikian rupa sehingga mereka dilatih ulang secara terus-menerus, memperbarui estimasi terbaiknya di setiap paparan. Kita dapat menggunakan sinyal waktu nyata dari pengguna dan sistem untuk menyempurnakan model ini (misalnya, apakah pengguna mengklik iklan? Berapa latensi sebenarnya dari permintaan tersebut?). *Vowpal Wabbit* adalah salah satu contoh sistem semacam itu, yang bersumber terbuka oleh Yahoo dan saat ini dikelola oleh Microsoft, tetapi masih banyak lagi yang lain. Sistem-sistem ini cocok untuk masalah pembelajaran tanpa pengawasan di alam bebas, tetapi seringkali dapat mempelajari model yang bias dan rentan terhadap permainan dan oknum jahat.

Pembelajaran Penguatan

Meskipun model tanpa pengawasan dan dengan pengawasan sangat berguna, keduanya tidak termasuk dalam definisi konvensional "kecerdasan". Di mana para pemain catur? Di mana mobil tanpa pengemudi? Bagaimana kita mengajarkan mesin untuk belajar sendiri? Ini adalah bidang pembelajaran penguatan (RL) atau metode Pembelajaran Mesin (ML) di mana model yang dihasilkan menjadi lebih baik melalui penguatan positif dan negatif. Pembahasan lengkap tentang RL berada di luar cakupan bab ini, tetapi pemahaman tentang dasar-dasarnya penting dan akan membantu Anda memahami bagaimana Pembelajaran Mesin (DL) memengaruhi area ini juga.

Sistem pembelajaran penguatan bekerja melalui definisi konsep yang disebut fungsi-Q, yang merupakan representasi matematis praktis dari imbalan yang diharapkan untuk suatu tindakan tertentu. Misalnya, jika saya mengambil ratu Anda pada langkah ini, berapa kemungkinan saya akan memenangkan permainan? Fungsi-Q saya akan memberikan prediksinya, dan itu akan memengaruhi perilaku saya. Seiring saya terus memainkan beberapa permainan, saya terus melatih fungsi-Q saya untuk menghasilkan estimasi hadiah yang lebih baik, yang membuat saya bermain lebih baik. Teknik yang sama ini jauh melampaui catur, dan seperti yang akan kita lihat dengan pembelajaran penguatan mendalam (deep RL), teknik ini



dapat digunakan untuk menangani permainan yang menantang seperti Go dan masalah serumit mobil self-driving.

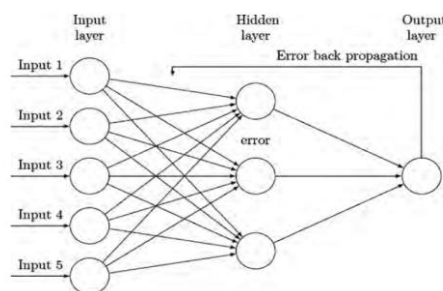
2.4 PEMBELAJARAN MENDALAM JANJI DI BALIK HYPE

Apa Itu Pembelajaran Mendalam dan Dari Mana Asalnya?

Saya telah membahas secara singkat DL di bagian sebelumnya dari bab ini, tetapi saya belum membahas secara detail apa itu DL, bagaimana cara kerjanya, dan mengapa tiba-tiba menjadi begitu menonjol. Hal itu berubah sekarang. Seperti yang disebutkan di bagian sejarah, jaringan saraf hampir ditinggalkan oleh sebagian besar di bidang ML kecuali beberapa tim terisolasi yang yakin bahwa janji yang ditunjukkan sebelumnya dapat direalisasikan dengan waktu dan upaya yang cukup. Dedikasi mereka terbayar awal dekade ini ketika pertemuan teknik semuanya diterapkan. "Backpropagation," "stochastic gradient descent," dan "fungsi aktivasi nonlinier" semuanya digabungkan untuk memungkinkan pelatihan jaringan yang lebih dalam dengan lebih sedikit sumber daya yang dapat mempelajari fungsi nonlinier yang semakin rumit. Di bagian berikut, saya akan memberi Anda sedikit latar belakang pada masing-masing, jenis jaringan baru yang mereka aktifkan, dan masalah yang dapat (dan tidak dapat) mereka pecahkan. Bersabarlah dengan saya karena saya akan sedikit "matematika," tetapi saya akan mencoba dan membuatnya seminimal mungkin.

Propagasi Balik

Propagasi balik, yang diperkenalkan oleh Rumelhart, Hinton, dan Williams pada tahun 1986, adalah teknik yang memungkinkan jaringan saraf tiruan untuk belajar dari setiap prediksi. Kita menciptakan jaringan kompleks ini, yang memodelkan fungsi nonlinier besar, tetapi jika kita memastikan bahwa jaringan tersebut terdiferensiasi, maka propagasi balik dapat berfungsi. Fungsi terdiferensiasi berarti kita dapat mendiferensiasikan fungsi-fungsi ini dan mendapatkan fungsi baru yang memungkinkan kita menghitung laju perubahan fungsi asli (anggaplah "lokasi" sebagai fungsi asli, dan "kecepatan" sebagai fungsi terdiferensiasi [atau laju perubahan lokasi]). Kita melakukan propagasi maju sinyal dari nilai masukan kita saat ini hingga mencapai akhir jaringan, lalu kita bandingkan dengan nilai harapan atau "label" kita. Setelah kita memiliki perbedaan tersebut di lapisan terakhir, kita dapat melakukan propagasi balik perbedaan tersebut melalui jaringan, dengan perbedaan fraksional (turunan parsial kesalahan terhadap setiap bobot) yang menyesuaikan bobot dan bias jaringan, secara bertahap mendorong jaringan ke arah yang benar (Gambar 2.2).



Gambar 2.2 Back-propagation.



Fungsi Optimasi dan Penurunan Gradien Stokastik

Salah satu masalah terbesar dalam mengoptimalkan jaringan menggunakan propagasi balik adalah proses optimasi itu sendiri. Optimasi umum menggunakan teknik yang disebut "penurunan gradien" (GD), yang melihat kesalahan pada hasil saat ini versus kesalahan pada hasil sebelumnya dan mendorong fungsi "menurun" (yaitu, ke arah gradien menurun). Dengan jaringan dalam yang berpotensi berisi ratusan juta bobot dan bias serta puluhan juta titik data masukan, mustahil untuk memproses seluruh masukan dan memperbarui semua bobot dalam satu langkah gradien. Penurunan gradien stokastik (SGD) memecahkan masalah ini secara sederhana: ia menarik satu sampel dari input, menjalankan jaringan, dan menyesuaikan gradien berdasarkan kesalahan dalam sampel tunggal tersebut.

Pelatihan sekali pada seluruh set input dikenal sebagai epoch, jadi GD bekerja dengan menjalankan seluruh epoch sebelum memperbarui bobot, melakukannya untuk beberapa epoch hingga mencapai ambang batas kesalahan yang telah ditentukan untuk berhenti atau menyadari bahwa kita telah mencapai batas (jaringan tidak membaik atau kita kehabisan waktu). Penurunan gradien stokastik berjalan melalui beberapa epoch dengan cara yang sama tetapi memperbarui bobot setelah setiap contoh dari input, mungkin mengacak set input di antara epoch bayangkan GD berjalan lurus menuruni bukit ke dasar, di mana SGD tersandung, tetapi secara bertahap menuruni bukit ke arah yang sama.

Sejak SGD, kita telah menciptakan banyak fungsi optimasi baru yang menyempurnakannya (konvergen lebih cepat). Bahasa Indonesia: Saya tidak akan masuk ke detail yang cukup bagi Anda untuk menerapkan salah satu dari ini, karena mereka sudah diterapkan di sebagian besar toolkit, tetapi berharap untuk menjelaskannya secara cukup rinci sehingga Anda dapat membuat pilihan yang tepat. Yang paling sederhana adalah SGD dengan momentum. Ini adalah jalan menurun yang acak, tetapi saat Anda berjalan ke arah yang sama, Anda tersandung lebih cepat.

Mini-batch gradient descent (MBGD) seperti SGD, tetapi dalam hal ini, Anda memproses beberapa contoh sekaligus. Ini membuatnya lebih mungkin Anda akan tersandung sedikit lebih sedikit secara acak dengan mengorbankan lebih banyak memori GPU Anda dan dapat dikombinasikan dengan momentum untuk keuntungan lebih lanjut. *Adaptive gradient descent* (AdaGrad) tersandung lebih agresif menuruni bukit ke arah yang tampak menjanjikan karena lebih stabil, untuk setiap contoh dari input atau mini-batch. *Root mean square propagation* (RMSProp) seperti AdaGrad, tetapi menggunakan rata-rata bergerak yang membusuk untuk bekerja melintasi beberapa gradien terkini. Estimasi Momen Adaptif (ADAM) adalah favorit saat ini dan mencoba menggabungkan manfaat AdaGrad dan RMSProp.

Sigmoid, Unit Linear Terserahkan, dan Selanjutnya

Saya telah menyebutkan sebelumnya bahwa jaringan dalam memodelkan rumus-rumus nonlinier yang besar. Hal ini berbeda dari upaya awal jaringan saraf tiruan yang memodelkan persamaan linear sederhana dengan menambahkan fungsi aktivasi nonlinier antar lapisan. Jaringan dalam mencapai hal ini dengan menambahkan "fungsi aktivasi" setelah sebuah neuron, yang mengontrol apakah sinyalnya merambat ke bawah jaringan dan



memungkinkannya untuk mati pada beberapa masukan. Jika fungsi aktivasi tersebut nonlinier, jaringan dapat belajar mengendalikan neuron dengan cara yang sangat kompleks.

Pada masa-masa awal DL, kami menggunakan nonlinieritas sederhana seperti Tanh dan Sigmoid yang sudah kami kenal dari aplikasi lain. Ini bekerja dengan baik (dan terkadang masih digunakan), tetapi membutuhkan komputasi yang mahal dan dapat secara signifikan meningkatkan waktu pelatihan dan evaluasi dalam jaringan besar. Kami menemukan bahwa dalam banyak kasus, nonlinieritas sederhana yang membiarkan nilai positif masuk dan menekan nilai negatif bekerja dengan cukup baik. Unit linier tersearahkan (ReLU) ini adalah fungsi linier sepotong-sepotong sederhana yang membiarkan positif tetap utuh dan menetapkan negatif ke nol, serta merupakan salah satu fungsi aktivasi yang paling umum dalam praktik.

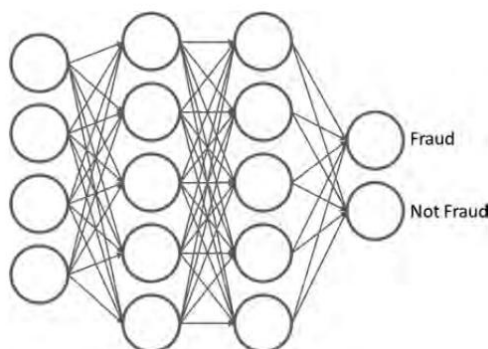
Beberapa modifikasi terbaru ReLU telah diusulkan dalam beberapa tahun terakhir dalam upaya meningkatkan akurasi dan generalisasi jaringan sambil mempertahankan karakteristik kinerja ReLU. Leaky ReLU memodifikasi ReLU dengan membiarkan sejumlah kecil sinyal negatif yang konstan masuk, sementara ReLU terparameterisasi mengubah konstanta tersebut menjadi parameter jaringan lain yang dipelajari selama pelatihan, yang meningkatkan waktu pelatihan dan ukuran jaringan tanpa memengaruhi beban evaluasi. Fungsi aktivasi masih menjadi bidang penelitian yang aktif, jadi saya tidak memiliki rekomendasi di sini karena kemungkinan besar sudah ketinggalan zaman pada saat artikel ini diterbitkan.

Jenis Jaringan yang Sepenuhnya Baru

Jaringan saraf dalam telah berevolusi dari akar jaringan saraf dasarnya, beranjak dari jaringan padat sederhana menjadi jaringan dengan jenis neuron khusus dan topologi rumit yang disesuaikan untuk membantu mereka menyelesaikan tugas-tugas tertentu. Di bagian ini, kita akan membahas dua jenis jaringan utama yang digunakan saat ini: jaringan saraf konvolusional (CNN) dan jaringan saraf rekuren (RNN) serta nenek moyang jaringan padatnya.

Jaringan Padat

Jaringan dalam berevolusi dari jaringan saraf yang lebih sederhana, dan seringkali untuk masalah klasifikasi yang sederhana, kita masih dapat melatih jaringan padat sederhana dari lapisan-lapisan yang terhubung penuh dengan fungsi aktivasi nonlinier dan mendapatkan hasil yang fantastis. Misalnya, jaringan pada Gambar 2.3 adalah jaringan padat sederhana yang memprediksi apakah suatu transaksi merupakan penipuan.



Gambar 2.3 Jaringan prediksi penipuan yang padat.



CNN

Jaringan saraf konvolusional adalah jaringan saraf dengan neuron konvolusional yang terspesialisasi. Ini sangat efektif dalam memproses data gambar karena konvolusi ini mampu belajar menarik fitur prediktif dari struktur lokal. Konvolusi adalah matriks kecil yang melintasi matriks masukan (misalnya, piksel dalam gambar) dan berkonvolusi dengannya untuk menghasilkan keluaran baru, yang kemudian dikumpulkan dalam lapisan berikutnya untuk difokuskan, sehingga meningkatkan medan reseptif lapisan berikutnya. Penumpukan lapisan-lapisan ini memungkinkan konvolusi berikutnya untuk mempelajari fitur tingkat yang lebih tinggi sehingga pada awalnya, jaringan mungkin mempelajari hal-hal seperti "ini adalah tepi," tetapi pada lapisan konvolusi terakhir, jaringan mungkin mempelajari "ini adalah telinga anjing."

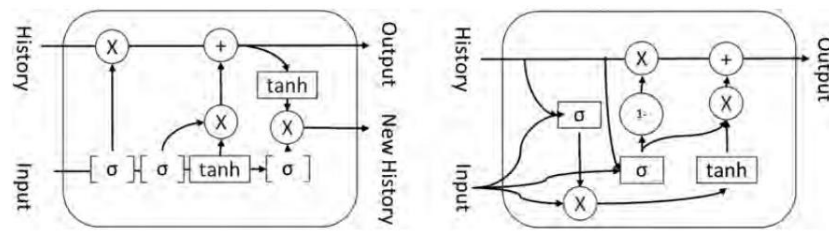
Biasanya, CNN ini digunakan pada data gambar, dan ledakan DL ke tempat kejadian sebenarnya disebabkan oleh CNN yang disebut AlexNet (dinamai menurut penciptanya Alex Krizhevsky, bermitra dengan penasihatnya G. Hinton dan I. Sutskever), yang mendominasi tantangan ImageNet untuk pengenalan gambar (misalnya, "ini adalah gambar kucing") pada tahun 2012. Sejak itu, CNN terus mendominasi tantangan ImageNet dalam berbagai bentuk tetapi juga telah berkembang melampaui gambar ke banyak bentuk data lain dengan struktur lokal. Mereka digunakan untuk pembuatan gambar dan video, dan dengan "konvolusi satu dimensi" mereka bahkan telah menggantikan struktur jaringan yang lebih rumit. Memahami CNN adalah inti untuk memahami DL, dan mereka kemungkinan akan menjadi inti solusi DL untuk tahun-tahun mendatang.

Jaringan Saraf Tiruan Rekuren: LSTM dan GRU

Tidak semua data memiliki struktur lokal. Data ucapan, teks, dan data sekuensial lainnya sering kali memiliki kombinasi struktur lokal dan tren jangka panjang atau fitur struktural, yang perlu ditangkap untuk menghasilkan prediksi yang efektif. Jaringan saraf tiruan rekuren (RNN) adalah salah satu solusi untuk masalah ini, dengan rekurensi internal yang memungkinkan nilai jaringan sebelumnya untuk dimasukkan ke dalam hasil saat ini dan membiarkan jaringan mempelajari struktur jangka panjang. RNN ini menunjukkan hasil yang menjanjikan pada data sekuensial (terjemahan bahasa, dll.), tetapi rentan terhadap masalah, yang sering menyebabkan kegagalan pelatihan jaringan. Karena mereka harus melakukan propagasi balik gradien melintasi ruang dan waktu (nilai sebelumnya) dengan membuka gulungan jaringan, gradien yang dihasilkan cenderung mengalami degenerasi, baik meledak atau menghilang dan menyebabkan jaringan gagal konvergen.

Para peneliti mengembangkan beberapa teknik berbeda untuk mengatasi masalah ini, dengan dua teknik yang saat ini mendominasi. Jaringan memori jangka pendek panjang (LSTM) menukar sel rekuren dasar dengan sel yang lebih rumit yang belajar melupakan menggunakan gerbang lupa (Gambar 2.4). Unit rekuren berpagar (GRU) merupakan alternatif pengganti sel rekuren dengan struktur yang lebih sederhana, yang dilatih lebih cepat, tetapi memiliki nilai ekspresif yang lebih rendah (Gambar 2.4). Bidang ini masih memperdebatkan mana yang lebih efektif dalam praktiknya, tetapi keduanya tampaknya efektif dalam mengatasi masalah gradien menghilang/meledak.





Gambar 2.4 LSTM versus GRU.

Jaringan Perhatian

Para peneliti menyadari ketika berfokus pada berbagai model RNN untuk penerjemahan mesin, misalnya, bahwa jaringan tersebut sedang belajar mengodekan seluruh konteks kalimat ke dalam keluaran pada lapisan tersembunyi terakhir. Mereka akan menjalankan sebuah kalimat melalui jaringan, dan ketika mencapai token akhir kalimat, pada titik tersebut jaringan tersebut harus menggunakan informasi yang telah dipelajarinya untuk menghasilkan terjemahan lengkap hanya berdasarkan informasi pada lapisan keluaran terakhir tersebut.

Bagaimana jika jaringan diizinkan untuk melihat masukan sebelumnya ketika membuat keputusan tentang setiap kata dalam terjemahan? Bagaimana jika kita membiarkannya belajar bagaimana memperhatikan berbagai bagian masukan untuk setiap bagian keluarannya? Mekanisme perhatian memodifikasi jaringan yang ada dengan menambahkan beberapa bentuk memori dari masukan sebelumnya, mirip dengan LSTM tetapi lebih eksplisit. Jaringan yang dimodifikasi ini membutuhkan lebih banyak RAM, karena Anda mungkin perlu menyimpan nilai perhatian untuk setiap kombinasi token masukan/keluaran, tetapi jaringan ini dapat jauh lebih akurat daripada jaringan sejenis LSTM untuk beberapa masalah. Mekanisme atensi merupakan bidang penelitian yang aktif, dan belum ada "praktik terbaik" di bidang ini, sehingga pembahasan mekanisme apa pun secara mendetail masih prematur.

Model Campuran dan Jaringan Adversarial Generatif

Beberapa tugas memiliki struktur lokal sekaligus persyaratan fokus jangka panjang pertimbangkan pemrosesan video, pemberian teks pada gambar, atau mengkorelasikan sinyal deret waktu yang masuk. Kita seringkali dapat menghasilkan model campuran (misalnya, CNN + LSTM) yang menghasilkan hasil yang jauh lebih baik daripada yang dapat dihasilkan masing-masing secara terpisah.

Kami juga menemukan bahwa model dapat dilatih bersama secara adversarial untuk menghasilkan data baru yang meniru masukan dunia nyata dalam proses yang disebut jaringan adversarial generatif (GAN). GAN ini menggabungkan jaringan generator yang dapat menghasilkan masukan palsu baru dan jaringan diskriminator yang dapat membedakan data palsu dari data nyata. Kami melatihnya bersama-sama hingga diskriminator benar-benar bingung, dan pada titik tersebut kami menentukan bahwa generator telah belajar membuat masukan palsu yang terlihat cukup nyata untuk mengelabui diskriminator (dan dalam beberapa kasus, mata manusia). Saat ini, GAN berguna untuk menghasilkan gambar palsu dan telah digunakan untuk meningkatkan resolusi gambar yang ada, menghasilkan keluaran yang



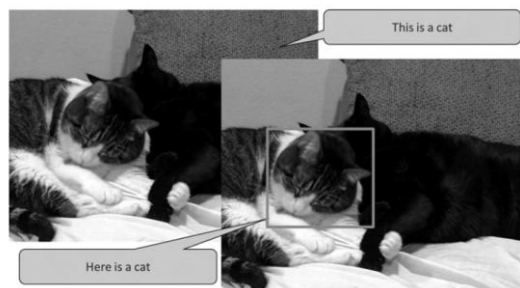
jauh lebih realistis daripada metode saat ini seperti pengambilan sampel bikubik, tetapi GAN paling menjanjikan dalam bidang pembuatan data pelatihan untuk domain di mana pengumpulan data pelatihan mahal atau sulit (misalnya, beberapa data medis).

Apa yang Dapat Dilakukan Pembelajaran Mendalam?

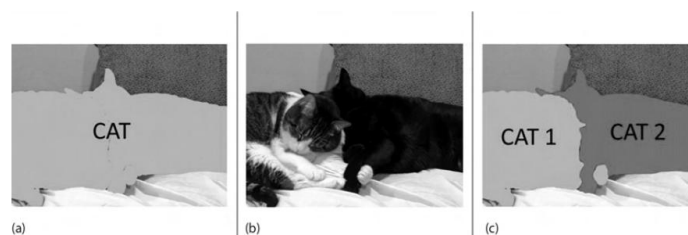
Gambar

Jaringan saraf konvolusional dan pengenalan gambar membawa pengenalan DL ke panggung dunia. Meskipun kami memulai dengan pengenalan gambar yang disederhanakan (misalnya, "ini kucing"), seiring berkembangnya metode DL, kami telah mampu memodelkan objek dan tindakan yang jauh lebih kompleks. Model deteksi objek tidak hanya mampu memberi tahu Anda bahwa "ini kucing", tetapi juga di mana kucing itu berada dalam gambar dengan mengidentifikasi kotak pembatas yang berisi objek tersebut. Model-model ini biasanya melakukan hal ini melalui mekanisme yang disebut wilayah minat, di mana beberapa wilayah diusulkan untuk setiap label dan prediksi yang dihasilkan memberikan probabilitas berdasarkan kombinasi label/wilayah (Gambar 2.5).

Kami bahkan telah melampaui model-model tersebut baru-baru ini dengan segmentasi semantik dan instans. Model-model ini menghasilkan prediksi per piksel yang memberi tahu Anda secara tepat kelas apa yang dimiliki setiap piksel dalam gambar. Segmentasi semantik memberi Anda kelas tetapi tidak dapat membedakan antara berbagai instans dari kelas tersebut misalnya, sepasang kucing hanyalah satu gumpalan "kucing". Segmentasi instans mengambil satu langkah ini dan mampu membedakan setiap instans domba dari yang lain (Gambar 2.6).



Gambar 2.5 Label pengenalan versus deteksi objek; mana yang merupakan pengenalan versus deteksi?



Gambar 2.6 Segmentasi semantik versus instan: (a) segmentasi semantik, (b) citra asli, dan (c) segmentasi instan.



Keduanya merupakan bidang penelitian yang sangat aktif, dan kami memperkirakan kinerja model-model ini akan meningkat drastis selama beberapa tahun ke depan seiring dengan dihasilkannya set data berlabel baru di bidang ini. Metode-metode ini saat ini sedang diterapkan dalam skenario "kota pintar" untuk penghitungan mobil dan pejalan kaki, memungkinkan estimasi lalu lintas yang lebih baik dan dapat memungkinkan jaringan pintar menjadi lebih adaptif terhadap pola lalu lintas manusia.

Sekuensial (Teks, Ucapan)

Metode jaringan saraf tiruan rekuren seperti LSTM dan GRU menjadi terkenal di bidang penerjemahan mesin dan pemrosesan ucapan-ke-teks. Upaya ini berlanjut dengan jaringan atensi, dan seiring munculnya kumpulan data yang lebih besar dan lebih tangguh di domain ini, kami berharap hasilnya akan semakin baik. Dalam praktiknya, kami menemukan bahwa setelah kami melatih jaringan untuk menerjemahkan antara dua bahasa, kami dapat menggunakan jaringan tersebut untuk melakukan bootstrap pelatihan untuk bahasa tambahan, dan jaringan yang dihasilkan seringkali dapat melakukan pekerjaan yang lebih baik dalam menerjemahkan kedua bahasa. Bentuk pembelajaran transfer ini umum di seluruh DL, tetapi dalam penerjemahan mesin terbukti sangat tangguh, memungkinkan jaringan penerjemahan baru untuk dilatih dari sangat sedikit contoh.

Jaringan atensi dan rekuren ini juga mengalami kemajuan dalam memproses data teks skala besar, dengan beberapa jaringan dilatih untuk menghasilkan teks ringkasan dari korpus besar, yang lain belajar untuk memprediksi sentimen, dan lebih banyak lagi mempelajari penyematan makna dan konteks yang lebih tangguh yang dapat digunakan untuk mendorong model prediktif lainnya. Model-model ini akan merevolusi pusat kontak dan panggilan karena konteks ditarik dari log panggilan dan manual sebelumnya, dibuat dapat dicari, dan panggilan masuk diurai dan diubah menjadi kueri cerdas. Ini adalah bidang penelitian yang aktif, dan salah satu area di mana kita sebagai industri perlu mewaspadaai bias bawaan dalam data pelatihan kita, karena kita telah melihat penyisipan kata yang dipelajari dari korpus teks yang ada seperti surat kabar yang menampilkan bias seksis dan rasis.

Campuran (Video)

Data video menggabungkan struktur lokal gambar dengan sifat sekuensial jangka panjang. Saat ini, sebagian besar model pemrosesan video biasanya menggunakan teknik keyframing untuk memproses gambar tunggal dengan model CNN yang dilatih pada data gambar, tetapi industri ini telah secara aktif meneliti metode untuk menggunakan model sekuensial seperti RNN dan mekanisme atensi untuk mempelajari struktur jangka panjang dari aliran video. Kita telah melihat penelitian yang menarik dalam peringkasan video dan bahkan pembuatan video, dan seiring dengan semakin tersedianya kumpulan data video, kami berharap bidang ini akan semakin aktif. Teknik-teknik ini dapat merevolusi layanan video-on-demand dan streaming dengan menyediakan ringkasan yang dihasilkan secara otomatis, cuplikan "sorotan" yang cerdas, dan bahkan model kompresi data yang lebih baik.

Pembelajaran Penguatan Mendalam



Pembelajaran penguatan mendalam (deep RL) mengambil konsep RL dan menggunakan DL untuk menyempurnakannya. Jika Anda ingat, fungsi-Q digunakan sebagai estimasi imbalan untuk setiap tindakan yang dilakukan terhadap tujuan kita (misalnya, jika kita mengambil ratu, seberapa besar kemungkinan kita memenangkan pertandingan?). Dengan RL yang mendalam, kita mengganti fungsi-Q sederhana dengan jaringan yang mendalam, yang memungkinkan fungsi-Q kita mempelajari strategi imbalan yang jauh lebih kompleks. Dalam kasus AlphaGo, Google menggunakan fungsi-Q kompleks dengan CNN 51-lapis yang mengodekan berbagai pengetahuan spesifik domain dari papan, dan selama ribuan pertandingan pelatihan, fungsi-Q ini mempelajari representasi yang cukup kompleks untuk memungkinkannya mengalahkan grandmaster Go. Teknik-teknik RL yang mendalam ini merupakan bidang penelitian yang sangat kuat saat ini, dan sistem pembelajaran mandiri berdasarkan RL yang mendalam pasti akan memberikan dampak yang besar di setiap bidang otomatisasi.

Generatif

Saya telah menyebutkan GAN sebelumnya, tetapi itu hanyalah salah satu keluarga jaringan generatif yang ada. Jaringan ini dapat dilatih untuk menghasilkan keluaran dalam berbagai modalitas, mulai dari transfer gaya neural menggunakan pencampuran CNN sederhana hingga pembangkitan teks menggunakan LSTM dan sistem modern seperti LyreBird, WaveNet milik Google, dan Deep Voice milik Baidu yang menggunakan model generatif untuk menciptakan ucapan dalam berbagai gaya. Seiring dengan semakin matangnya model-model ini, kita perlu mengembangkan metode untuk memastikan bahwa model-model tersebut digunakan untuk kebaikan (misalnya, aksesibilitas yang lebih baik bagi tunanetra) alih-alih kejahatan (serangan eksistensial terhadap kebenaran, karena "rekaman" palsu yang terdengar nyata dapat dihasilkan oleh siapa pun yang memiliki GPU).

2.5 KECERDASAN BUATAN MASA KINI DAN MASA DEPAN

Kecerdasan Buatan di Alam Liar

Pembelajaran mesin tidak hidup dalam ruang hampa. Di dunia nyata, ML perlu diintegrasikan ke dalam sistem yang ada dan diterapkan ke produksi, lalu digunakan oleh orang-orang, baik secara langsung maupun tidak langsung, untuk memberikan nilai. Dengan pesatnya penelitian ML, kami melihat kesenjangan yang semakin besar antara hasil yang dipublikasikan di situs akademis seperti arXiv.org dan model yang diterapkan di produksi. Hal ini sebagian disebabkan oleh tingkat perubahan yang fantastis, dan sebagian lagi karena kesenjangan dalam penggunaan alat antara konstruksi model dan produksi model, yang kami perkirakan akan segera tertutupi sehingga memungkinkan penggunaan kemajuan AI kami yang lebih cepat dalam skenario dunia nyata.

Hal ini tentu saja memunculkan pertanyaan, Apa yang diperlukan untuk menempatkan model ke produksi? Ada beberapa aspek antara pemodelan dan eksperimen serta penerapan produksi yang tangguh. Pertama, mayoritas proyek ML adalah persiapan data, jadi membangun alur data yang tangguh adalah kuncinya. Ini tidak hanya mencakup akuisisi data tetapi juga pembersihan (dan pengujian/regresi proses pembersihan Anda, serta pembersihan



ulang seiring berubahannya). Ini juga memperhitungkan asal data dari mana data Anda berasal? Hal ini semakin penting seiring dengan diberlakukannya Peraturan Perlindungan Data Umum (GDPR) dan undang-undang privasi lainnya.

Selanjutnya, model itu sendiri perlu berevolusi. Kita memerlukan pembuatan versi, dan kita perlu mengetahui data apa yang digunakan untuk melatih versi tersebut dan bagaimana kinerjanya dalam pengujian. Kita memerlukan kemampuan untuk "menerbangkan" model baru pengujian A/B untuk menentukan seberapa baik kinerjanya pada data nyata, dan peluncuran bertahap untuk memastikan kinerjanya baik saat dimuat. Kita memerlukan pemantauan yang kuat untuk kecepatan dan akurasi, dengan ukuran akurasi yang sesuai dengan sistem (misalnya, prediksi kanker tidak menilai positif palsu dan negatif palsu secara sama). Model juga "kedaluwarsa" seiring masuknya data baru dan sistem berevolusi, data yang sebelumnya diberi label mungkin menjadi kurang akurat dan prediksi model mungkin mulai gagal. Kegagalan ini menyoroti perlunya pelacakan akurasi dari waktu ke waktu, evaluasi prediksi yang terlewat, dan mekanisme yang kuat untuk melatih ulang dan menerapkan ulang model Anda.

Akhirnya, kita perlu mempertimbangkan penerapan itu sendiri apakah kita melakukan dockerisasi pada model dan menerapkan kontainer (instans aplikasi+OS mandiri) ke cloud? Apakah kita berjalan pada GPU atau CPU, atau apakah kita memiliki perangkat keras khusus *Application-Specific Integrated Circuit* (ASIC) atau *Field-Programmable Gate Array* (FPGA)? Apakah kita menerapkan pada perangkat? Jika demikian, bagaimana kita memastikan kita dapat mendorong pembaruan atau mendukung versi lama? Apakah masuk akal untuk memiliki model hierarkis, misalnya, di mana model yang lebih sederhana dan berpotensi terkompresi berjalan pada perangkat IoT dan menentukan apakah akan mengirim batch data ke model berbasis cloud yang lebih kompleks untuk analisis yang lebih detail?

Skenario produksi ini lebih kompleks daripada sistem perangkat lunak tradisional tetapi memiliki banyak kesamaan masalah lintas sektor (logging, pemantauan, deteksi dan pemulihan kesalahan, *Continuous Integration/Continuous Delivery* [CI/CD]), sehingga wawasan Anda yang ada akan berguna di sini. Di mana insting Anda rusak, di situlah asumsi determinisme Anda gagal. Ingatlah bahwa sistem pembelajaran mesin bersifat probabilistik, sehingga beberapa persentase kesalahan prediksi dan kegagalan tidak dapat dihindari, akurasi 100% dalam pelatihan adalah bug (overfitting) dan mustahil di dunia nyata. Alat-alat seperti Azure Machine Learning Workbench milik Microsoft, proyek Acumos sumber terbuka dari Linux Foundation yang baru-baru ini diumumkan, dan Pachyderm telah berevolusi untuk memenuhi beberapa kebutuhan ini, dan kami berharap lebih banyak alat akan muncul dan alat-alat ini akan terus berkembang.

Memahami, Berbicara, dan Menerjemahkan

Seperti yang telah kita lihat sebelumnya, model pembelajaran mesin yang ada telah membuat kemajuan luar biasa di bidang penerjemahan, sistem ucapan-ke-teks, dan teks-ke-ucapan. Kami berharap kemajuan ini akan berlanjut selama dekade berikutnya dan, dengan kompresi dan kuantisasi model, akan diterapkan di lebih banyak tempat (BabelFish dari *Hitchhiker's Guide to the Galaxy* hampir tiba). Selain itu, kami melihat kemajuan dalam



inferensi dan pemahaman kontekstual menjadi semakin penting seiring tersedianya kumpulan data dan korpora baru (melalui kompetisi Kaggle dan cara lainnya). Hal ini pada awalnya akan mengarah pada peningkatan analisis sentimen, tetapi kemudian akan berkembang menjadi peringkasan dokumen dan bahkan otomatisasi gudang data oleh agen cerdas. Analisis sentimen adalah kunci untuk skenario churn pelanggan dan layanan pelanggan dalam churn, kami dapat mengetahui kapan pelanggan marah dan kemungkinan akan pergi, dan selama sesi layanan pelanggan otomatis, mendeteksi frustrasi pelanggan dan mengarahkan mereka ke agen manusia dapat menyelamatkan bisnis mereka.

Penalaran tentang Gambar dan Video

Pemrosesan gambar merupakan salah satu domain pertama yang direvolusi oleh model pembelajaran mesin (DL), dan pada titik ini, pengenalan dasar hampir dianggap sebagai masalah yang terpecahkan. Para peneliti beralih ke tugas-tugas yang lebih rumit seperti segmentasi instans, pemrosesan video, dan ekstraksi semantik. Menggabungkan teknik-teknik ini dengan teknik pemrosesan ucapan sebelumnya akan memungkinkan solusi pemrosesan video yang lengkap (seperti Video Indexer milik Microsoft) untuk terus berkembang, tidak hanya memperkenalkan deteksi dan identifikasi objek dan orang di seluruh video, tetapi juga teks otomatis multibahasa, ringkasan otomatis, dan penandaan metadata yang kuat.

Kita kemudian dapat melihat masa depan model generatif di bidang ini kita akan melihat model generatif pertama kali berakar pada aplikasi "taman bermain" ("buatkan saya gambar lucu berdasarkan X"), tetapi kemudian meluas ke aplikasi untuk para kreator ("perbaiki mata agar lebih mirip Y"). Pada akhirnya, model generatif akan meluas ke pembuatan gambar dan video secara menyeluruh berdasarkan deskripsi belaka, atau modifikasi visual yang menarik dari karya yang sudah ada. Hal ini akan menimbulkan pertanyaan serius seputar penggunaan wajar, kepemilikan, dan hak cipta yang perlu diselesaikan.

Model generatif juga akan digunakan untuk menghasilkan data pelatihan secara otomatis dalam skenario yang sulit atau dipertanyakan untuk dikumpulkan (seperti yang telah disebutkan sebelumnya), dan hal ini, ditambah dengan meningkatnya daya komputasi GPU berbasis cloud dan ASIC khusus, akan menghasilkan kemajuan signifikan dalam kinerja model dalam diagnostik medis, penemuan obat, jaringan otonom, dan keamanan.

Musuh Masa Depan

Model generatif juga menyimpan potensi lain, terutama ketika kita mempertimbangkan penggunaannya dalam situasi yang bersifat permusuhan. Apa yang terjadi pada jejaring sosial kita ketika berita #FakeNews yang dihasilkan secara otomatis oleh model mendalam tidak dapat dibedakan (atau lebih baik) daripada berita yang dihasilkan oleh aktor jahat bayaran? Bagaimana hal ini berdampak pada sistem politik kita?

Setelah model generatif dapat, hanya dengan sedikit data suara atau gambar, menghasilkan rekaman audio atau video palsu yang terdengar atau tampak realistis dari siapa pun, apa pengaruhnya terhadap proses pembuktian? Bagaimana kita bisa menggugat perilaku buruk di ruang publik jika kita tidak memiliki cara untuk memverifikasi apakah itu benar-benar terjadi berdasarkan bukti yang terekam? Kita telah melihat model yang mampu menghilangkan tanda air dari gambar, sehingga menandai aliran video dengan tanda air akan



menjadi proses pascageneratif yang mudah. Kita perlu mengembangkan cara untuk mengidentifikasi materi palsu dalam menghadapi ancaman ini, dan para peneliti secara aktif mencari metode untuk mendeteksi keluaran model generatif.

Model mendalam saat ini juga rentan terhadap kesalahan pelabelan dalam menghadapi contoh-contoh yang bersifat adversarial, karena model tersebut menavigasi ruang pencarian yang begitu kompleks untuk menghasilkan prediksi mereka. Kita telah melihat contoh hal ini dalam model deteksi gambar di mana sejumlah kecil "derau pintar" dapat menyebabkan model salah mengidentifikasi siamang sebagai burung unta, dan contoh yang lebih signifikan di mana model penglihatan swakemudi salah mengidentifikasi rambu berhenti karena adanya derau (pita khusus) yang diterapkan pada rambu tersebut. Kita dapat memperkirakan akan melihat persaingan sengit di sini antara penipu dan peretas serta perusahaan, pemerintah, dan peneliti yang menghasilkan model.

Privasi juga bisa menjadi masalah dengan model mendalam, karena model tersebut mengodekan sejumlah besar informasi dalam parameternya. Kita telah melihat contoh model pembelajaran mendalam yang dihasilkan untuk deteksi wajah mampu mereproduksi gambar pelatihan dari model yang telah dilatih menggunakan proses yang disebut dekonvolusi. Dengan adanya undang-undang seperti GDPR, hal ini membawa kita pada diskusi tentang privasi data dalam menghadapi model mendalam jika seseorang ingin informasinya dihapus, apakah kita perlu membuang semua model yang dilatih menggunakannya? Adakah cara untuk melindungi kita dari keharusan melakukannya? Ini adalah area penelitian yang aktif, dengan model berantai yang tampaknya dapat memberikan solusi.

Pembelajaran Mesin untuk Membuat Mesin Pembelajaran

Apa yang terjadi ketika kita mengaktifkan teknik ML pada dirinya sendiri? Bagaimana kita dapat mengajarkan mesin untuk belajar lebih efektif? Area penelitian yang dikenal dengan nama-nama seperti meta-learning dan auto-ML ini merupakan area penelitian yang sedang berkembang pesat saat ini. Beberapa bagian tersulit dari penelitian pembelajaran mendalam saat ini adalah penyetelan hiperparameter dan pemilihan model. Karena terkadang butuh waktu sehari-hari atau berminggu-minggu untuk melatih model, memilih struktur model yang tepat, teknik optimasi, laju pembelajaran, dan hiperparameter lainnya dapat membuat perbedaan antara keberhasilan dan kegagalan. Kami telah sibuk merancang sistem yang secara otomatis akan melalui proses ini, menavigasi ruang pencarian hiperparameter, dan, dalam beberapa kasus, merancang topologi model baru berdasarkan suatu bentuk meta-pembelajaran.

Ketika GPU menjadi lebih umum, daya komputasi secara keseluruhan meningkat, dan lebih banyak set data berlabel muncul, bentuk meta-pembelajaran ini akan meningkat. Saat ini, para pemain besar berbagi temuan mereka (jenis neuron baru, dll.), tetapi kami berharap hal itu menjadi lebih jarang karena struktur neuron baru menjadi Kekayaan Intelektual (HKI) yang berharga. Tidak akan lama lagi sebelum struktur neuron dan jaringan baru menjadi lebih rumit daripada yang dapat dipahami para peneliti dengan mudah, dan mesin mulai menggantikan pengetahuan kita bahkan di ruang ML. Ketertelusuran, pemahaman, dan



kemampuan debug menjadi kunci dalam skenario ini, karena kita dipaksa untuk memercayai sistem yang bahkan pembuatnya sendiri pun belum tentu memahaminya.

Revolusi Industri Kecerdasan Buatan

Revolusi AI akan membantu mendorong jaringan yang mampu memulihkan diri, mengamankan diri sendiri, dan adaptif terhadap perubahan kebutuhan. Dengan revolusi sistem AI, jaringan dapat mendeteksi pola awal serangan siber, secara otomatis menyebarkan sumber daya di tempat yang dibutuhkan seiring perubahan dan evolusi pola lalu lintas, serta meningkatkan perutean otomatis untuk mengatasi kegagalan jaringan dan masalah lainnya. Sisa buku ini akan membahas skenario-skenario ini secara lebih rinci, memberi Anda gambaran tentang apa yang dapat dicapai setelah AI menjadi bagian yang merambah dalam jaringan.



BAB 3

PERGESERAN KE JARINGAN BERBASIS PERANGKAT LUNAK

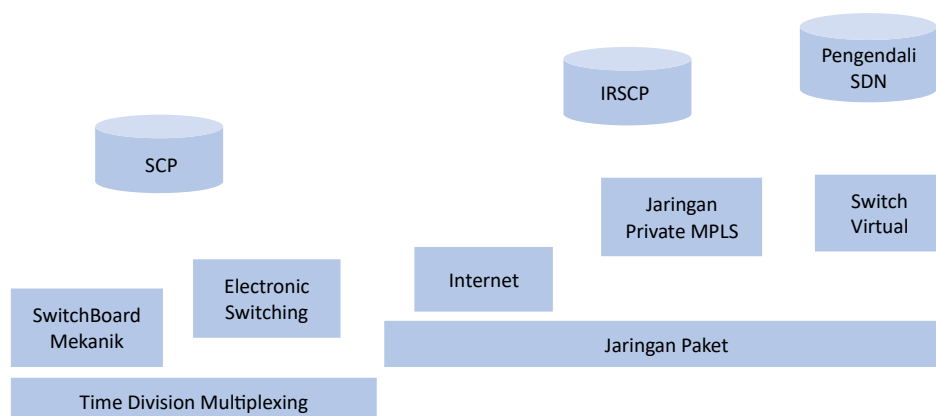
3.1 PENDAHULUAN

“Mr. Watson, kemarilah, saya ingin bertemu Anda” adalah frasa pertama yang membentuk fondasi jaringan 140 tahun yang lalu. Sejak saat itu, jaringan dan lalu lintasnya telah tumbuh secara eksponensial di seluruh pusat data nasional, regional, dan edge, serta dari kabel ke nirkabel. Dalam beberapa tahun terakhir, untuk menghadapi perubahan besar ini, industri telah bergerak cepat untuk merangkul virtualisasi fungsi jaringan (NFV) dan jaringan yang ditentukan perangkat lunak (SDN). Kedua upaya ini bergabung untuk menciptakan lingkungan baru untuk menyebarkan dan menskalakan layanan.

Kontrol dan pengelolaan lingkungan ini memerlukan ekosistem perangkat lunak baru untuk pembuatan layanan, orkestrasi, dan pengumpulan sumber daya dan layanan, konfigurasi, dan manajemen siklus hidup penuh. Bab ini menjelaskan bagaimana dunia teknologi cloud, NFV, dan SDN bergabung untuk menciptakan lingkungan jaringan baru ini dan selanjutnya membahas salah satu area inovasi dan disrupti utama berikutnya: kecerdasan buatan jaringan (AI), penerapan AI pada ekosistem ini untuk, misalnya, mencapai tingkat manajemen, pemantauan, pemulihan, dan kontrol loop tertutup prediktif yang baru dan lebih tinggi. Bab ini membahas bagaimana ekosistem NFV dan SDN menyediakan sarana untuk mengaktifkan AI jaringan dan menggambarkan arsitektur keseluruhan.

3.2 SEJARAH JARINGAN YANG DITENTUKAN PERANGKAT LUNAK

Jaringan yang ditentukan perangkat lunak telah digunakan selama beberapa dekade di bidang telekomunikasi. Bahkan, jaringan ini merupakan salah satu aplikasi komputasi komersial pertama. Menarik untuk melihat kesamaan masalah dan solusi yang muncul seiring perkembangan jaringan dari peralihan *time division multiplexing* (TDM) berorientasi sesi, menjadi peralihan paket, dan menjadi jaringan virtual saat ini.



Gambar 3.1 Evolusi jaringan telekomunikasi.



Jaringan telekomunikasi awal berawal dari sistem peralihan mekanis dengan papan colokan dan operator manusia yang mengatur dan memutuskan panggilan. Seiring dengan meningkatnya kompleksitas dan skala jaringan, sebuah cara untuk mengotomatiskan tugas melakukan panggilan telepon menggunakan mesin/komputer elektromekanis awal dirancang dan diterapkan untuk menggantikan operator papan sakelar.

Ketika peralihan elektronik menjadi mungkin melalui elektronik modern seperti transistor, konsep kendali program tersimpan (SPP) diterapkan pada desain jaringan. Sebuah program komputer akan mengumpulkan berbagai peristiwa, melakukan analisis yang disederhanakan, dan memicu tindakan. Loop kontrol fungsional sederhana akan menjadi dasar bagi AI tingkat lanjut dalam jaringan komunikasi, tetapi mari kita mulai dari dasar-dasar permasalahan yang sedang dipecahkan (Gambar 3.1).

Jaringan Terdefinisi Perangkat Lunak: Pemisahan Bidang Kendali.

Jaringan telekomunikasi awal memiliki tugas sederhana, menganalisis permintaan panggilan telepon dan membuat jalur melalui jaringan untuk menghubungkan pihak pemanggil A dengan pihak terpangil B. Tugas ini sederhana ketika terdapat opsi koneksi yang sederhana, tetapi seiring berkembangnya jaringan, menemukan jalur komunikasi "optimal" melalui jaringan menjadi sulit. Hal ini juga menjadi lebih rumit ketika pertanyaannya mencakup siapa yang harus membayar panggilan tersebut, pihak pemanggil (normal) atau pihak terpangil (panggilan kolektif).

Seiring dengan semakin canggihnya komputer, para insinyur juga menyadari bahwa layanan yang lebih canggih dapat diciptakan jika jalur sinyal dipisahkan dari jalur pembawa. Hal ini menjadi dasar standarisasi sistem sinyal nomor 6 sebagai pengenalan pensinyalan kanal umum, yang akhirnya digantikan dengan sistem sinyal nomor 7 (SS7). Pensinyalan antar komputer dapat dilakukan pada kanal bersama yang sudah ada sebelumnya sehingga tidak perlu diatur dan diputus untuk setiap panggilan. Pemisahan bidang kendali dan bidang pembawa memungkinkan serangkaian fitur luar biasa yang hanya didefinisikan dalam perangkat lunak sebagai serangkaian keputusan.

Keputusan rute dapat dilakukan melalui komunikasi peer-to-peer atau dengan algoritma terpusat berdasarkan digit yang dihubungi. Data tambahan dapat diterapkan untuk menentukan jalur, dan bahkan tujuan dapat diubah selama pemrosesan. Pada dasarnya, seiring komputer menjadi cukup cepat untuk melakukan analisis tambahan sebelum penelepon mengharapkan telepon berdering (keterlambatan pasca-panggilan pada masa itu sekitar dua detik), semakin banyak fitur yang tersedia untuk membuat keputusan tentang bagaimana dan ke mana harus mengarahkan panggilan. Pertukaran antara keterlambatan yang diharapkan pengguna dan waktu analisis komputer ini akan menjadi tema yang berulang bahkan dalam jaringan protokol internet (IP) saat ini.

Penerapan teknik-teknik ini menghasilkan cara yang lebih baik untuk melakukan panggilan otomatis tertagih. Penelepon menghubungi nomor khusus dan sistem akan menerjemahkannya ke nomor perutean jaringan (NRN) dan menghubungkan panggilan berdasarkan algoritma. Ini adalah jaringan overlay pertama di mana data asli ditulis ulang atau dienkapsulasi untuk memanfaatkan rencana perutean jaringan yang berbeda. Algoritmenya



bahkan dapat memutuskan berdasarkan kode layanan khusus bahwa pihak yang menelepon tidak akan dikenakan biaya untuk panggilan tersebut, melainkan akan menagih pihak yang dihubungi. Ini adalah layanan bebas pulsa 800 dalam bentuk aslinya. Teknologi ini menghasilkan standar "jaringan cerdas" (IN) yang memanfaatkan pemisahan bidang kontrol untuk menambahkan layanan yang lebih canggih, memungkinkan interoperabilitas peralatan dan interoperabilitas penyedia layanan.

Algoritme terpusat dalam bidang kontrol tidak hanya mencakup dasar-dasar untuk menghubungkan pihak yang menelepon dan yang dihubungi serta mencatat informasi untuk penagihan, tetapi juga fungsi-fungsi kompleks yang dapat digabungkan dalam pohon keputusan pemrosesan panggilan. Misalnya, algoritma terpusat dapat melihat nomor telepon pihak yang menelepon (Identifikasi Nomor Otomatis [ANI]) dan membuat keputusan berdasarkan kode area (tiga digit) atau NPA-NXX (enam digit) untuk mendistribusikan panggilan ke pusat panggilan terdekat. Algoritme ini dapat mendistribusikan panggilan di pagi hari ke Pantai Timur AS dan di malam hari ke Pantai Barat AS.

Ketika AT&T menambahkan peralatan di kantor pusat untuk memutar pengumuman dan mengumpulkan digit dari penelepon, algoritmanya dapat menambahkan lebih banyak input ke dalam keputusan tujuan. Hal ini juga memunculkan SDN berbasis TDM pertama, di mana paket panggilan pribadi dibuat agar pelanggan korporat dapat menggunakan paket panggilan khusus secara nasional tanpa jaringan khusus. Ini adalah virtualisasi jaringan TDM membuat jaringan pribadi kedua sebagai overlay pada jaringan publik nasional. Kombinasi algoritma, peralatan untuk pengumuman, dan pengumpulan serta pemetaan digit nomor yang dihubungi ke NRN adalah jaringan overlay yang ditentukan perangkat lunak pertama dalam panggilan telepon berorientasi sesi.

Jaringan circuit-switched juga mengalami transformasi dari peralatan switching monolitik yang terdapat dalam pusat switching bergerak (MSC dalam jaringan bergerak) atau sentral lokal (dalam jaringan tetap) yang membawa pemisahan bidang kontrol dan bidang pengguna satu langkah lebih jauh. Dalam transformasi ini, logika persinyalan dan layanan dipisahkan dari peralatan switching untuk membuat dua simpul khusus server dan switch yang dapat dipisahkan secara fisik dan memungkinkan sentralisasi server dan algoritmanya sambil mempertahankan penempatan peralatan switching yang optimal. Protokol untuk menghubungkannya merupakan upaya bersama antara *Internet Engineering Task Force* (IETF) dan ITU-T SG-16 dan didokumentasikan dalam spesifikasi H.248.

Implementasi algoritma ini secara real-time berada di *Network Control Points* atau *Signaling Control Points*. "Pengendali" ini akan menggunakan jaringan SS7 untuk menerima permintaan, memberikan instruksi ke switch, dan meningkatkan data melalui platform pengumuman dan pengumpulan digit. Mereka mengelola status jaringan privat khusus dan mewujudkan fungsi terpusat bagi pelanggan. Pengendali memiliki sistem pendukung operasi (OSS) yang memungkinkan pelanggan mengubah jaringan pribadi mereka hampir secara real-time mengubah rute yang disukai, menambahkan opsi panggilan tambahan, dll. Ini adalah lingkungan yang benar-benar dapat diprogram oleh pelanggan, meskipun pemrograman



biasanya dilakukan melalui Antarmuka Pengguna Grafis (GUI) pada OSS, alih-alih menulis kode C atau C++.

Manajemen Jaringan

Seiring munculnya SDN besar ini, serangkaian masalah lain muncul. Lonjakan lalu lintas dapat diamati karena berbagai hal di dunia nyata menyebabkan panggilan massal. Jaringan telah mengalami situasi kemacetan ketika bencana alam terjadi seperti banjir, tornado, atau badai, dan membutuhkan mekanisme untuk membatasi lalu lintas. Ide dasarnya adalah mengidentifikasi panggilan yang "kemungkinan gagal" dan memblokirnya di sumbernya sehingga tidak menghabiskan sumber daya yang terbatas di area terdampak. Maka, kontrol manajemen jaringan pun berlaku tindakan yang dapat diterapkan untuk menghentikan lalu lintas yang datang ke area bencana dan hanya mengizinkan panggilan keluar dari area tersebut. Hal ini memprioritaskan telepon yang berfungsi dan digunakan untuk melakukan panggilan keluar dari area terdampak dan memblokir panggilan masuk ke area yang mungkin tidak memiliki telepon yang masih terhubung ke jaringan.

Dengan kemajuan SDN, kebutuhan akan manajemen jaringan pun semakin meningkat. Ketika badai salju melanda, agen perjalanan dibombardir oleh pelanggan mereka yang mencoba merencanakan ulang penerbangan dan pengaturan perjalanan mereka. Dengan hanya sejumlah agen dan/atau koneksi di pusat data, jaringan perlu membatasi panggilan untuk mencegah permintaan tinggi dari satu pelanggan yang menghalangi lalu lintas pelanggan lain.

Deteksi kondisi ini (kelebihan beban tujuan terfokus) memerlukan pengumpulan telemetri pada upaya yang tidak efektif, menghitung peralihan sumber, dan kemudian menerapkan mekanisme gating berbasis waktu pada sumber yang dialihkan. Sistem yang melakukan perhitungan ini dan menentukan berdasarkan telemetri tingkat kontrol (celah panggilan) dan di mana menempatkan kontrol adalah Sistem Operasi Manajemen Jaringan (NEMOS). Ini adalah kasus algoritma manajemen jaringan eksternal yang menggunakan telemetri dari jaringan (laporan kinerja) untuk mengendalikan SDN. Dapat dikatakan bahwa ini adalah aplikasi pertama AI dasar untuk telekomunikasi, meskipun algoritmanya dikodekan secara keras.

Seiring dengan semakin besarnya jaringan privat, muncul pula kebutuhan untuk mengatasi penipuan. Catatan Detail Panggilan yang dibuat ketika setiap panggilan selesai dapat dianalisis, diringkas, dan dicocokkan polanya untuk menentukan bahwa pola panggilan tersebut kemungkinan merupakan penipuan dan kemudian memanggil tindakan di SDN untuk mengurangi penipuan. Sistem ujung ke ujung dari jaringan overlay privat yang ditentukan perangkat lunak dan elemen analisis dan kontrol yang membentuk sistem manajemen jaringan secara keseluruhan merupakan bagian penting dari SDN. Dalam istilah saat ini, ini adalah teknologi kuno, tetapi merupakan dasar bagi apa yang akan dibutuhkan untuk jaringan masa depan seperti Internet publik.

Internet Publik dan Pertumbuhan Jaringan Protokol Internet Pribadi Virtual (IP)

Ketika IP muncul sebagai protokol pemenang dalam jaringan paket, serangkaian pendorong serupa untuk evolusinya pun terjadi. Awalnya, IP sengaja dirancang secara



terdistribusi, tanpa kontrol terpusat, dan awalnya tidak ada pemisahan bidang kontrol dan penerusan. Masing-masing switch dan router membuat keputusan sederhana melihat paket, mencari tujuan, dan meneruskan paket ke tautan menuju tujuan. Jalur ditentukan berdasarkan keputusan lokal tentang status tautan dan jalur berbiaya terendah sebagaimana dilihat dari titik saat ini di jalur tersebut. Satu-satunya keputusan atau algoritma terpusat mungkin adalah layanan Sistem Nama Domain (DNS) yang menerjemahkan nama tujuan yang mudah digunakan *www.att.com* menjadi alamat IP 12.1.2.3. Terjemahan ini dapat bervariasi berdasarkan sumbernya, tetapi umumnya karena caching dan aspek teknisnya, kontrol yang dibutuhkan tidak terlalu detail.

Pemisahan bidang kontrol dari bidang pengguna (alias bidang pembawa) mulai terjadi dengan berbagai protokol tautan dan perutean seperti *Border Gateway Protocol* (BGP) untuk memungkinkan keputusan tentang ke mana paket harus diarahkan menjadi lebih dinamis, tetapi masih belum memberikan akibat langsung apa pun pada jaringan privat yang dimungkinkan dengan TDM SDN. Namun, dengan pemisahan bidang kontrol dan bidang pensinyalan, langkah selanjutnya menuju virtualisasi jaringan menjadi mungkin.

Multi-protocol label switching (MPLS), yang dimungkinkan oleh pensinyalan BGP, memungkinkan operator jaringan untuk membuat tabel perutean khusus pelanggan sehingga hanya titik akhir mereka yang ada di tabel perutean. Tabel perutean ini menciptakan jaringan privat virtual (VPN) di mana ruang alamat privat yang tumpang tindih (10.1.2.3) dapat digunakan di beberapa jaringan secara bersamaan seperti di TDM di mana ekstensi 10.1.2.3 dapat digunakan di beberapa paket panggilan privat. Jaringan privat tervirtualisasi ini sepenuhnya merupakan SDN dalam IP, Internet privat tujuan yang merupakan bagian dari perusahaan yang sama dan tidak berisiko berinteraksi dengan Internet publik. Internet privat ini merupakan overlay dengan pemisahan lapisan 2 yang dipasarkan AT&T sebagai Jaringan Privat Virtual (AVPN) AT&T; penyedia layanan IP lain menawarkan layanan serupa. SDN IP privat ini serupa dengan SDN TDM karena merupakan overlay, tetapi semuanya berupa paket IP tanpa koneksi, alih-alih panggilan telepon berorientasi koneksi.

Virtualisasi Jaringan Protokol Internet sebelum Cloud

Seiring berkembangnya MPLS, tabel-tabel routing tersebut juga dapat diubah dengan cara yang unik untuk menjembatani jaringan privat. Titik Kontrol Layanan Rute Terpadu (IRSCP), seperti halnya Titik Kontrol Jaringan (NCP), diciptakan untuk dapat memanipulasi tabel routing MPLS secara waktu nyata (real-time) berdasarkan kebijakan pelanggan dan jaringan. IRSCP digunakan untuk pembaruan routing antardomain privat yang tidak rentan terhadap masalah keamanan routing Internet dan sekali lagi memberdayakan pelanggan untuk mengontrol bagaimana mereka ingin jaringan overlay virtual mereka beroperasi.

Kira-kira pada masa inilah BGP Flowspec juga diciptakan yang memungkinkan pengontrol untuk mengomunikasikan pemrosesan aliran yang diinginkan melalui pesan BGP ke router. Flowspec dapat menentukan perutean yang diinginkan dalam pencocokan N-tuple data di header paket IP seperti alamat IP sumber, alamat IP tujuan, keluarga alamat (Internet atau VPN privat), atau protokol transport sehingga router tidak hanya dapat mengimplementasikan tabel perutean khusus tetapi juga keputusan perutean tingkat lanjut



berdasarkan pencocokan N-tuple yang secara efektif menyediakan kemampuan untuk memisahkan "alur" paket yang memenuhi kriteria yang sama. Alur-alur tersebut dapat dirutekan secara terpisah sehingga hal-hal seperti "scrubber penolakan layanan terdistribusi (DDOS)" dapat dirantai layanan ke dalam jaringan hanya untuk lalu lintas yang "kemungkinan buruk". Sekali lagi, ini adalah contoh overlay privat pada jaringan publik.

Semua jaringan IP ini digunakan oleh pengguna untuk mengakses server dan aplikasi server. Pertumbuhan teknologi server yang luar biasa dan skala ekonomi dengan server farm yang besar menghasilkan cloud privat pertama seperti Google, eBay, dan Amazon. Cloud ini perlu menangani masalah jaringan di dalam server farm mereka, di mana mesin virtual dapat dibuat di satu server fisik dan kemudian dipindahkan ke server fisik lain berdasarkan peningkatan beban atau pemeliharaan server fisik, tanpa memengaruhi jaringan eksternal. Mereka membutuhkan teknologi overlay agar aplikasi terlindungi dari pergerakan mesin virtual. Berbagai protokol tunnel overlay dikembangkan dan masih digunakan, dengan *Virtual Extended Local Area Network* (VXLAN) mungkin yang paling populer, tetapi semuanya membutuhkan manajemen penugasan dan pembaruan seiring perubahan terjadi.

Jaringan Berbasis Perangkat Lunak (Software-Defined Networking) Masa Kini

Kira-kira pada masa inilah, sekitar tahun 2009, dunia akademis mengusulkan paradigma baru untuk pemisahan logika kontrol dan bidang penerusan pada router, serta mengusulkan protokol baru yang disebut OpenFlow. Konsep dasarnya adalah agar pengontrol mengirimkan data (atau membalas kueri ketika aliran baru muncul di switch) dengan data tabel keputusan yang akan digunakan switch untuk penerusan paket. Tabel-tabel tersebut memiliki serangkaian fungsi dan keputusan yang relatif kecil, tetapi dapat disesuaikan untuk pelanggan berdasarkan serangkaian atribut N-tuple (pada prinsipnya mirip dengan flowingspec yang telah disebutkan).

Aspek kuncinya adalah penerusan paket berbasis aliran sehingga dapat mengatur perilaku di awal aliran dan kemudian mempertahankan keputusan pemrosesan paket yang sama selama aliran berlangsung. Dalam hal ini, penerusan paket sangat mirip dengan aliran TDM dan H.248 lama dalam evolusi jaringan circuit-switched. Namun, yang terpenting, ia masih menangani paket sehingga semua keunggulan jaringan paket dalam hal keputusan penerusan masih dapat diterapkan. Hal yang mungkin juga membedakan pengontrol aliran SDN modern adalah kemampuannya untuk menulis ulang header paket sehingga paket tidak hanya dapat dienkapsulasi tetapi juga dapat menggantikan informasi header paket yang mengubah pemrosesan langkah-langkah selanjutnya dalam algoritma perutean. Kemampuan untuk benar-benar mendefinisikan protokol perutean baru dalam jaringan yang sedang berjalan merupakan generasi berikutnya dari SDN.

Jaringan yang ditentukan perangkat lunak dalam jaringan OpenFlow sangat canggih dan beberapa pusat data cloud yang sangat besar menggunakannya dalam skala besar, tetapi untuk masalah Jaringan Area Luas (WAN) secara umum, kompleksitasnya telah membatasi penerapannya terutama pada lingkungan cloud lokal. Kombinasi jaringan WAN dan LAN-lah yang sangat dibutuhkan untuk memenuhi kebutuhan jaringan privat perusahaan (Gambar 3.2).



Aplikasi Jaringan yang Ditentukan Perangkat Lunak Protokol Internet

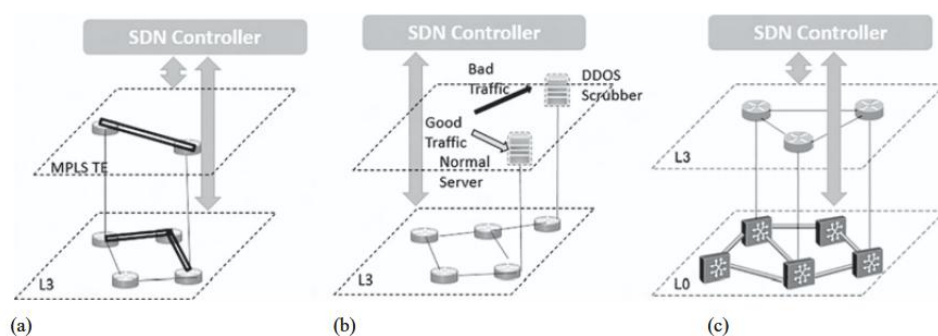
Dengan latar belakang SDN sebelumnya, mari kita bahas beberapa aplikasinya (Gambar 3.3).

Optimasi Jalur

Optimalisasi jalur didasarkan pada penggabungan topologi tautan dalam jaringan dengan telemetry dari jaringan terhadap arus lalu lintas melalui jalur saat ini, kapasitas tautan, dan okupansi tautan untuk menghitung fungsi biaya/optimasi jaringan. Dengan fungsi biaya/optimasi seperti utilisasi tautan secara keseluruhan, suatu algoritma dapat menghitung apakah jalur yang berbeda melalui jaringan akan menghasilkan utilisasi yang lebih baik atau ketahanan yang lebih baik saat terjadi kegagalan.

Fitur	Kontrol Program Tersimpan	MPLS VPN & Flowspec	OpenFlow
Kontrol Pembawa Pemisahan	CCIS6 SS7	BGP	OpenFlow
Hamparan	Nomor Perutean Jaringan	Pembeda Rute Target Rute	Header Paket Tulis Ulang
Logika Khusus Pelanggan	SCP	IRSCP	SDN Controller
Jenis Aliran	Berorientasi pada Koneksi	Tanpa koneksi	Aliran Tanpa Koneksi

Gambar 3.2 Perbandingan atribut jaringan yang ditentukan perangkat lunak.



Gambar 3.3 Aplikasi IP SDN: (a) optimasi jalur, (b) penulisan ulang tujuan, dan (c) SDN multi-lapis.

Saat ini, pengontrol SDN dapat menjadi bagian dari sistem loop tertutup yang menerima telemetry ini melalui data *BGP-Link State* (BGP-LS) dan *Path Computation Element Protocol* (PCEP), menerapkan analitik, lalu mengubah terowongan overlay *Multi-Protocol Label Switching – Traffic Engineering* (MPLS-TE) pada backbone untuk meningkatkan optimasi jalur.



Kombinasi telemetri, analitik, Antarmuka Program Aplikasi (API), dan tampilan jaringan yang lebih lengkap dari titik kontrol terpusat merupakan pendorong untuk jenis loop optimasi ini.

Penulisan Ulang Destinasi

Kami telah membahas secara singkat penggunaan pengontrol SDN untuk menyiapkan kriteria pencocokan N-tuple untuk alur perutean. Aplikasi kontrol jenis ini, jika dikombinasikan dengan telemetri dan algoritma, akan menjadi aplikasi pembersihan DDOS. Telemetri pada lalu lintas jaringan dan statistik server dapat digunakan untuk mengidentifikasi bahwa serangkaian alur yang cocok dengan suatu kriteria (sejumlah besar alur dari jaringan ini atau yang datang melalui tautan peering ini) adalah lalu lintas DDOS.

Algoritme dapat menentukan bahwa aturan alur pencocokan N-tuple ini dapat diterapkan pada pengontrol sehingga alur ini tidak akan menuju ke IP tujuan asli, melainkan akan dirutekan ke kompleks pembersih di mana lalu lintas yang buruk akan dihapus dan hanya lalu lintas yang valid yang akan diteruskan ke tujuan asli. Kemampuan pengontrol untuk mengklasifikasikan alur dan mengambil tindakan pada alur yang berbeda dari alur normal tidak akan berguna tanpa algoritma untuk menentukan tindakan yang harus dilakukan.

Jaringan Berbasis Perangkat Lunak Multilapis

Jaringan berbasis perangkat lunak bukan hanya fenomena Lapisan 2/Lapisan 3. Jaringan optik juga memiliki kemampuan kontrol perangkat lunak. Salah satu yang paling menarik adalah proyek OpenROADM di mana komponen yang dapat dioperasikan dapat dikonfigurasi melalui API standar industri. Kemampuan ini memungkinkan penambahan sirkuit bandwidth tinggi baru ke jaringan dalam waktu hampir nyata (menit satu digit). Kami baru-baru ini menggabungkan kemampuan ini dengan L3 SDN untuk memungkinkan kemampuan baru yang menarik untuk jaringan L3. Di bawah beban atau kegagalan, pengontrol Layer 3 dapat menentukan bahwa tidak ada lagi opsi untuk optimasi jalur dan alih-alih menyerah, ia dapat bertanya kepada pengontrol Layer 0 apakah ada sumber daya yang memungkinkan sirkuit baru antara dua titik akhir. Misalnya, jika ada kemacetan antara New York dan Los Angeles, jaringan Layer 3 dapat mengubah rute terowongan rekayasa lalu lintas tetapi masih memiliki kehilangan paket yang tinggi.

Pengontrol Layer 3 dapat bertanya kepada pengontrol Layer 0 apakah panjang gelombang tambahan dari NY ke LA dimungkinkan untuk delapan jam ke depan. Jika demikian, pengontrol Layer 0 dapat mengatur sirkuit itu dan memberi tahu pengontrol Layer 3 ketika selesai. Kontrol Layer 3 kemudian dapat menggunakan API yang ditentukan perangkat lunaknya untuk menghubungkan sirkuit tersebut ke antarmuka jaringan router Layer 3, mengubah sirkuit tersebut menjadi tautan IP. Terakhir, pengontrol Layer 3 dapat menempatkan terowongan MPLS-TE baru di atas tautan tersebut. Setiap kali, algoritma optimasi lalu lintas diaktifkan untuk memilih jalur yang tepat di Layer 0, rute yang tepat untuk tautan Layer 3, dan jalur serta bobot terowongan MPLS yang tepat di jaringan overlay terowongan.

3.3 JALAN MENUJU VIRTUALISASI FUNGSI JARINGAN

Sebagaimana disebutkan di bagian sebelumnya, hampir bersamaan dengan adopsi teknologi cloud dan virtualisasi jaringan, industri jaringan sedang mengembangkan arsitektur



SDN yang menghadirkan pendekatan baru terhadap arsitektur jaringan bidang kontrol dan bidang pengguna. Bagian ini membahas bagaimana teknologi cloud dan NFV bersama-sama memungkinkan pendekatan revolusioner terhadap jaringan.

Evolusi Pusat Data ke Teknologi Cloud

Teknologi cloud awal dan penerapannya pada pusat data didorong oleh kebutuhan untuk menjalankan dan mengelola pusat data tradisional dengan cara yang lebih hemat biaya. Hal ini mencakup penggunaan perangkat keras seefisien mungkin serta meminimalkan biaya siklus hidup pengelolaan dan pemeliharaan perangkat keras dan perangkat lunak. Teknologi cloud menyediakan kemampuan untuk membuat dan mengelola beban kerja aplikasi serta menerapkannya ke server di pusat data di bawah kendali program. Dengan memanfaatkan kapabilitas ini, perusahaan dapat membuat dan menerapkan strategi pemulihan bencana (DR), misalnya, dengan menggunakan kapabilitas cloud untuk mewujudkan beban kerja aplikasi yang konsisten dengan strategi DR mereka, seperti redundansi (geografis atau lainnya). Teknologi cloud dapat digunakan untuk mencapai ketersediaan dan ketahanan yang dibutuhkan, tetapi juga memungkinkan pengguna untuk mewujudkan geo-redundansi, distribusi, dan konektivitas yang diperlukan.

Perangkat keras server dapat digunakan secara lebih optimal dengan menciptakan beban kerja aplikasi yang membutuhkan kombinasi CPU, memori, penyimpanan, dan konektivitas untuk memanfaatkan sumber daya server secara lebih efisien. Selain itu, teknologi cloud menyediakan kemampuan untuk mengelola beban kerja guna menyediakan kapabilitas untuk:

- Menyesuaikan sumber daya yang disediakan untuk aplikasi secara dinamis (misalnya, CPU, penyimpanan, memori)
- Membuat instance tambahan dari aplikasi yang sama di pusat data yang sama
- Membuat instance tambahan dari aplikasi yang sama di pusat data di lokasi geografis lain

Peralihan ke Perangkat Keras White Box

Sebagaimana telah disebutkan sebelumnya, salah satu pendorong utama teknologi cloud adalah untuk memenuhi kebutuhan pengelolaan pusat data dengan cara yang lebih hemat biaya. Selain perangkat lunak cloud dan kemampuannya untuk menjalankan fungsi-fungsi yang disebutkan sebelumnya di bawah kendali program, komponen biaya signifikan lainnya, yaitu biaya perangkat keras itu sendiri, perlu ditangani. Hal ini ditangani oleh industri yang bersatu di sekitar konfigurasi perangkat keras standar untuk server yang disebut perangkat keras White Box. Perangkat keras White Box cenderung berupa teknologi x86 dalam berbagai konfigurasi standar. Kini, server standar ini dapat dibeli dalam jumlah besar sehingga memungkinkan tercapainya skala ekonomi.

Adopsi/Perpindahan Aplikasi Awal ke Teknologi Cloud

Perpindahan awal ke teknologi cloud dilakukan oleh perusahaan yang ingin meningkatkan ekonomi dan fleksibilitas pusat data. Di antara aplikasi pertama yang berpindah ke cloud terdapat beragam aplikasi perusahaan seperti keuangan, sumber daya manusia, penggajian, real estat, pemasaran, dan penjualan. Aplikasi-aplikasi ini diuntungkan oleh



kemampuan untuk menciptakan beban kerja yang mengoptimalkan penggunaan perangkat keras server dan kemampuan untuk mengalokasikan sumber daya secara dinamis (CPU, memori, penyimpanan), menyediakan keragaman geografis, pengaturan DR, dan penskalaan aplikasi. Hal ini dilakukan sambil memperkenalkan tingkat otomatisasi yang lebih tinggi, sehingga mengurangi biaya operasional dengan mengurangi operasi manual.

Virtualisasi Jaringan

Seiring dengan perkembangan SDN dan sekitar waktu yang sama dengan munculnya teknologi cloud, industri jaringan menyadari bahwa lalu lintas tumbuh pada tingkat yang tidak dapat ditopang secara hemat biaya dengan terus berinvestasi pada elemen jaringan fisik (NE) yang lebih tradisional seperti yang telah dilakukan di masa lalu; lalu lintas jaringan tumbuh lebih cepat daripada yang dapat diimbangi oleh penyedia jaringan dalam pembangunan peralatan jaringan tradisional. Pertumbuhan lalu lintas lebih cepat daripada pertumbuhan pendapatan, dan pembelian NE yang berkelanjutan tidak dapat dipertahankan secara finansial dalam lingkungan persaingan yang ketat dan tekanan margin.

Untuk mengatasi masalah ini dan menyediakan solusi yang layak secara finansial, industri jaringan sedang menyelidiki kelayakan transformasi Jaringan Jaringan (NE), yang dibangun dengan perangkat keras dan perangkat lunak khusus, menjadi aplikasi yang dapat berjalan pada perangkat keras standar (White Box). Idennya adalah jika fungsi NE dapat diubah menjadi aplikasi perangkat lunak yang dapat berjalan pada perangkat keras standar, maka hal itu akan mengubah persamaan ekonomi menjadi lebih baik. Aplikasi yang menyediakan fungsi jaringan yang dapat berjalan pada perangkat keras White Box standar dapat memanfaatkan biaya perangkat keras yang lebih rendah dibandingkan dengan perangkat keras khusus yang sebelumnya digunakan dalam NE tradisional.

Transformasi NE tradisional (perangkat keras dan perangkat lunak khusus) menjadi aplikasi yang dapat berjalan pada perangkat keras standar (misalnya, server x86) memang terbukti layak secara teknis. Penyedia layanan dengan antusias mengejar perubahan teknologi ini untuk mengubah ekonomi jaringan. Pada titik ini, terdapat konsensus yang kuat di antara penyedia bahwa kurva biaya dapat dibelokkan demi ekonomi yang lebih baik.

Virtualisasi Fungsi Jaringan

Virtualisasi fungsi jaringan dimulai dengan NE yang lebih kecil dan perangkat edge (misalnya, firewall dan switch) karena alasan kinerja. Kemudian, seiring dengan peningkatan kinerja perangkat lunak NE virtual, virtualisasi beralih ke elemen berkapasitas lebih besar seperti router edge penyedia dan DNS.

Peralihan ke Perangkat Keras White-Box

Peralihan ke perangkat keras standar merupakan tema utama virtualisasi jaringan. Perangkat keras standar berbiaya lebih rendah dan dapat diterapkan, diskalakan, dan dikelola dengan basis yang lebih hemat biaya dibandingkan perangkat keras khusus yang sebelumnya digunakan dalam NE tradisional. Personel operasi pusat data berpengalaman dalam penerapan dan pengelolaan, misalnya, melakukan operasi perbaikan kerusakan pada perangkat keras server standar.



Konvergensi Teknologi Cloud dan Virtualisasi Fungsi Jaringan

Sebagaimana telah dibahas sebelumnya, pada tahap awal virtualisasi jaringan, terdapat dua tema utama yang sedang dikaji: teknologi cloud untuk mendefinisikan, menerapkan, dan mengelola beban kerja aplikasi secara lebih efisien ke pusat data di bawah kendali program, memanfaatkan kapabilitas dinamis yang disediakan oleh teknologi, dan virtualisasi jaringan transformasi perangkat keras dan perangkat lunak jaringan khusus menjadi aplikasi yang berjalan di server perangkat keras standar. Kedua inisiatif ini menjadi landasan bagi salah satu pergeseran paradigma teknis paling dramatis dalam jaringan yang terjadi dalam beberapa dekade.

Pertemuan Teknologi

Seiring dengan kemajuan teknologi cloud dan virtualisasi jaringan, disadari bahwa jika fungsi jaringan virtual (vNFs) yang sedang berkembang (NEs yang difaktor menjadi aplikasi perangkat lunak) dijalankan sebagai aplikasi di pusat data cloud, maka semua kontrol dinamis yang disediakan oleh cloud dapat digunakan dalam membangun infrastruktur dan layanan jaringan penyedia berskala besar. Dan ini dapat dicapai dengan intervensi manual yang minimal. Pertemuan cloud dan virtualisasi jaringan ini mengarah pada arah industri saat ini untuk mengalihkan jaringan mereka dari tradisional ke VNF yang berjalan di pusat data cloud.

Kontrol Dinamis yang Dibawa ke Fungsi Jaringan

Dengan konvergensi cloud dan NFV, penyedia jaringan dan layanan kini memiliki kemampuan untuk menyebarkan ("spin up") NE—VNF (aplikasi) di pusat data, menetapkan sumber daya ke NE tersebut (CPU, memori, penyimpanan), dan mengelola VNF tersebut. Semua fungsi ini kini dapat dilakukan secara dinamis dan di bawah kendali program. Fleksibilitas yang kuat ini memiliki implikasi mendalam terhadap cara penyedia menerapkan dan mengelola jaringan dan layanan mereka. Berbagai situasi dapat diatasi dengan cara baru yang memanfaatkan cloud dan VNF. Misalnya:

Instansiasi dan Manajemen

- Kontrol sumber daya, Sumber daya seperti VNF dapat diterapkan ke pusat data (diinstansiasi) untuk membangun infrastruktur jaringan. VNF dapat diberi CPU, memori, dan penyimpanan yang sesuai dengan kebutuhannya dan kebutuhan jaringan serta layanan yang perlu didukungnya.
- Konfigurasi jaringan dan layanan, VNF dapat dikonfigurasi untuk infrastruktur atau kapabilitas layanan tertentu.
- VNF, jaringan, dan layanan dapat dipantau dan dikelola, semuanya secara dinamis dengan intervensi manual minimal. Misalnya, dalam kondisi beban lalu lintas jaringan yang tinggi, penyedia dapat mengaktifkan NE tambahan (misalnya, VNF seperti router) di pusat data yang terdampak atau di pusat data lain di lokasi geografis yang berbeda, tergantung pada jenis masalah dan metode penanganan yang diinginkan.
- DR dapat dilakukan dengan cara baru, misalnya, mengaktifkan VNF di pusat data bebas bencana dan lokasi di bawah kendali program.
- Dengan VNF yang berjalan pada perangkat keras standar, server dapat disebarkan ke pusat data dalam pod yang terdiri dari banyak server. Ketika jumlah server yang



berfungsi berkurang ke tingkat yang telah ditentukan, seluruh pod dapat diperbarui dengan server baru dari generasi yang sama atau lebih baru. Hal ini mengubah pemeliharaan server dari proses berdasarkan permintaan dan berdasarkan peristiwa (alarm memicu pengiriman personel) menjadi pembaruan perangkat keras yang terjadwal secara berkala. Hal ini menghasilkan lebih sedikit pengiriman personel dan kemungkinan menurunkan biaya pemeliharaan perangkat keras.

Orkestrasi dan Manajemen Layanan dalam Ekosistem

Seiring lanskap jaringan virtual ini mulai terbentuk, disadari bahwa seiring dengan lingkungan jaringan virtual baru yang sangat dinamis ini, pendekatan baru terhadap manajemen jaringan, dalam arti luas, diperlukan untuk menangani definisi layanan, instansiasi, dan manajemen di dunia di mana vNF dapat dijalankan, dihapus, atau dipindahkan secara dinamis dan di mana sumber daya untuk elemen-elemen ini dapat dialokasikan dan dimodifikasi di bawah kendali program semuanya dalam waktu hampir nyata. Ekosistem manajemen baru ini harus selincah lingkungan jaringan yang dikelolanya. Respons AT&T terhadap tantangan ini adalah pembuatan platform *Enhanced Control, Orchestration, Management and Policy* (ECOMP). Setelah pembuatan dan penerapan ECOMP, AT&T kemudian membuka kode platform ECOMP. Kode sumber terbuka tersebut disebut *Open Networking Automation Platform* (ONAP) dan dikelola di bawah naungan Linux Foundation. Hal ini dilakukan untuk memajukan perpindahan industri ke platform standar untuk pengelolaan infrastruktur jaringan virtual baru yang sedang dikejar oleh penyedia di seluruh dunia.

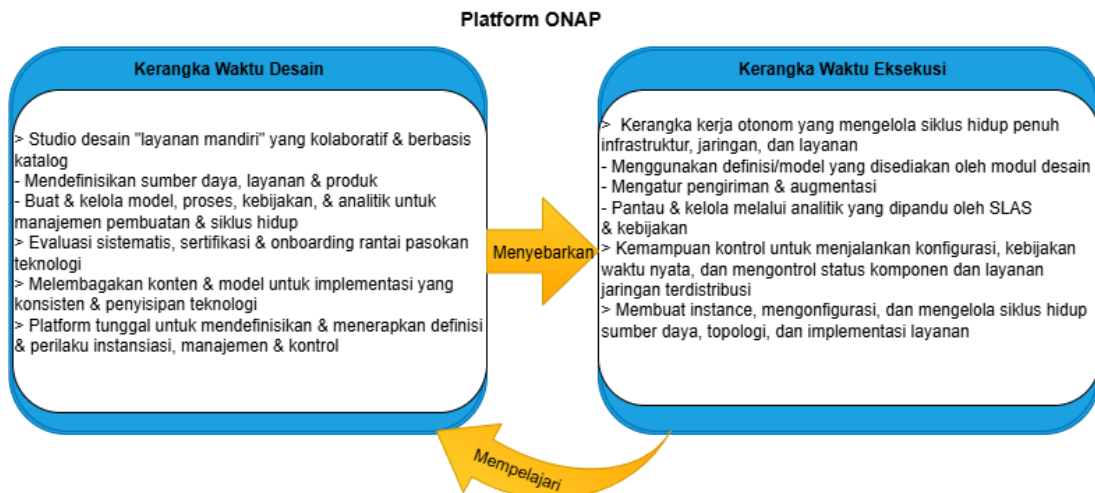
Arsitektur Platform Otomasi Jaringan Terbuka

Versi ONAP saat ini (rilis Amsterdam) mewarisi sebagian besar kemampuannya dari kode ECOMP yang disumbangkan dan inisiatif Open-O. Basis kode platform ini akan berkembang seiring waktu seiring semakin banyaknya pelaku industri yang berkontribusi kode untuk memajukan platform. Bagian berikut menjelaskan arsitektur ONAP.

Dua Kerangka Kerja Utama: Waktu Desain dan Waktu Eksekusi

Terdapat dua dimensi fundamental dan saling melengkapi, atau kerangka kerja, dalam platform ONAP yang berkaitan dengan waktu desain (misalnya, desain jaringan dan layanan) dan aspek waktu eksekusi (pengelompokan sumber daya yang diperlukan, konfigurasi jaringan dan layanan, pemantauan, dan manajemen) (Gambar 3.4).





Gambar 3.4 Platform ONAP Terdiri Dari Kerangka Kerja Waktu Desain Dan Waktu Eksekusi.

Kerangka Kerja Waktu Desain—Desain dan Pembuatan Layanan.

Penggerak strategis awal yang penting bagi ECOMP, dan selanjutnya ONAP, adalah penggunaan modal yang efisien dan pencapaian kelincahan yang lebih tinggi dalam pembuatan dan instansiasi layanan. Arsitektur ini memenuhi persyaratan ini dengan menggabungkan konsep berbasis model jika memungkinkan. Kerangka kerja waktu eksekusi digerakkan oleh model. Ini berarti bahwa perilaku platform waktu eksekusi ditentukan oleh metadata yang diberikan kepadanya, alih-alih dengan menulis atau memodifikasi perangkat lunak melalui proses dan siklus pengembangan perangkat lunak tradisional.

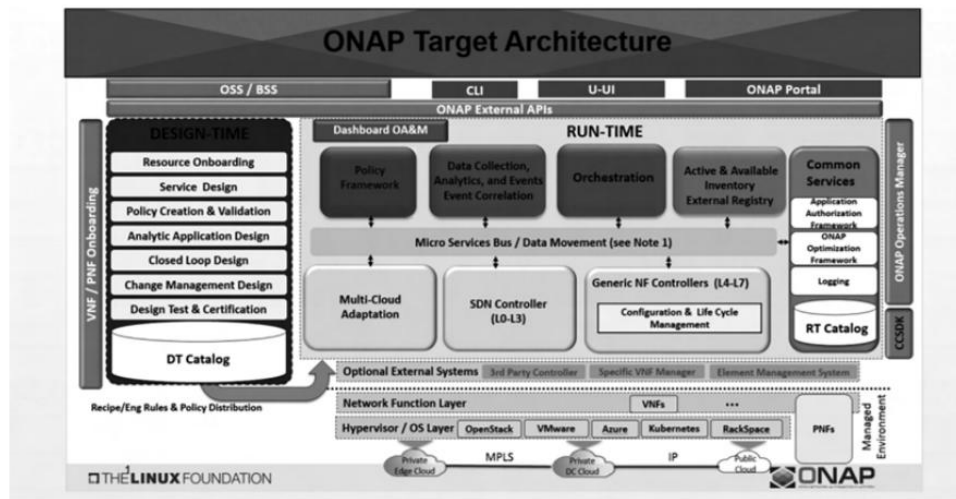
Hal ini memungkinkan seseorang untuk mengubah perilaku sistem dengan mengubah metadata yang diberikan, alih-alih dengan menulis kode untuk menentukan perilaku sistem. Platform desain dan pembuatan layanan (SDC), yang ditunjukkan pada Gambar 3.5, menyediakan paradigma studio desain bagi pengguna untuk memanipulasi VNF guna menentukan infrastruktur dan layanan. SDC menghasilkan metadata yang diperlukan untuk menentukan perilaku yang dibutuhkan kerangka kerja waktu eksekusi untuk membuat instance, mengonfigurasi, dan mengelola layanan berdasarkan permintaan pelanggan (atau peristiwa yang dihasilkan secara internal seperti alarm).

SDC memberi pengguna kemampuan untuk memasukkan VNF (yaitu, menyerap deskripsi metadata mereka sehingga dapat digunakan dalam merancang layanan) dan memanipulasi VNF tersebut untuk membuat layanan. Setelah ditentukan, platform SDC secara otomatis menghasilkan deskripsi metadata dari layanan yang ditentukan. Metadata ini, beserta permintaan layanan pelanggan, kemudian digunakan oleh platform eksekusi untuk membuat instans layanan tersebut, mengonfigurasinya untuk digunakan, dan melakukan pemantauan serta pengelolaan layanan tersebut.

Pada saat perancangan layanan, platform SDC juga digunakan untuk menentukan bagaimana layanan akan dikelola. Misalnya, pengguna dapat menentukan data yang akan dikumpulkan, dianalisis, dan ditindaklanjuti untuk mengelola jaringan atau layanan. Dengan



menggunakan contoh sederhana IP MPLS VPN, setelah VNF terpasang, pengguna SDC dapat menentukan layanan dengan mengorganisasikan VNF ke dalam layanan yang diinginkan menggunakan metafora studio desain. Setelah layanan dirancang, pengguna dapat menentukan bagaimana layanan akan dikelola.



Gambar 3.5 Arsitektur target ONAP.

Dalam contoh ini, pengguna (misalnya, bagian operasional) dapat memutuskan untuk mengelola berdasarkan throughput antar-situs. Dalam hal ini, pengguna akan menentukan data yang harus dikumpulkan untuk menghitung throughput antar-situs, analitik yang diperlukan untuk menghitung throughput, dan tindakan yang harus diambil jika throughput berada di luar ambang batas yang ditentukan. Platform SDC kemudian akan menghasilkan metadata yang diperlukan agar platform eksekusi dapat menjalankan fungsi-fungsi manajemen ini.

Sifat ONAP yang berbasis model menjawab tujuan strategis pemanfaatan modal yang efisien dan kelincahan penciptaan layanan dengan meminimalkan ketergantungan pada proses pengembangan perangkat lunak tradisional dan jangka waktu perancangan serta penciptaan layanan. Pendekatan ini juga memungkinkan layanan didefinisikan/diciptakan lebih cepat daripada proses pengembangan tradisional. Hal ini memungkinkan pemanfaatan modal yang lebih efisien dalam penciptaan dan modifikasi layanan dibandingkan metode pengembangan tradisional.

Kerangka Waktu Eksekusi

Kerangka waktu eksekusi ONAP bertanggung jawab untuk memproses permintaan layanan serta merespons peristiwa yang dihasilkan secara internal di lingkungan yang memerlukan tindakan. Bagian berikut menjelaskan kapabilitas kerangka kerja eksekusi.

Orkestrasi dan Kontrol

Orkestrasi bertanggung jawab untuk mengelola sumber daya seperti VNF di pusat data yang sesuai untuk memenuhi permintaan tertentu. Dalam contoh MPLS yang disebutkan sebelumnya, hal ini dapat berupa penerapan router tepi penyedia yang sesuai ke pusat yang



tepat untuk memenuhi permintaan layanan pelanggan. Misalnya, hal ini dapat dilakukan berdasarkan lokasi geografis lokasi pelanggan. Orkestrasi melakukan hal ini melalui interaksi dengan lingkungan cloud untuk meminta peningkatan sumber daya tertentu (misalnya, router tepi penyedia [VNF] di pusat data). Setelah sumber daya yang sesuai tersedia, kontrol beralih ke domain pengontrol.

Pengontrol bertanggung jawab atas konfigurasi VNF (misalnya, untuk memenuhi permintaan layanan, atau sebagai respons terhadap peristiwa dari lingkungan). Perlu dicatat bahwa orkestrasi berkaitan dengan pengorganisasian sumber daya yang sesuai; orkestrasi tidak sadar layanan. Misalnya, orkestrasi tidak tahu cara mengonfigurasi NE untuk layanan tertentu. Konfigurasi merupakan tanggung jawab pengontrol. Selain itu, ketika orkestrasi menyelesaikan satu unit kerja, orkestrasi tersebut tidak lagi berada dalam alur kontrol. Pengontrol, di sisi lain, mempertahankan status lingkungan yang mereka kendalikan.

Inventaris Aktif dan Tersedia

Sebagaimana telah disebutkan sebelumnya, infrastruktur jaringan virtual bersifat dinamis, di mana sumber daya seperti vNF dapat dijalankan, dihapus, dan dipindahkan di bawah kendali program secara hampir real-time. Agar ONAP dapat mengelola hal ini dengan sukses, perlu ada sumber informasi inventaris yang mutakhir pada horizon waktu yang sama dengan saat perubahan dinamis dapat dilakukan. Dalam ONAP, A&AI menyediakan kemampuan ini.

Pengumpulan Data, Analisis, dan Peristiwa

Komponen pengumpulan data, analisis, dan peristiwa (DCAE) dari ONAP adalah ekosistem yang memungkinkan pengumpulan dan analisis data tersebut serta penyimpanan dan distribusi data tersebut ke komponen platform lainnya. DCAE dirancang untuk menjadi platform yang sangat mudah didistribusikan. Hal ini dilakukan karena dua alasan utama: untuk lebih mendukung latensi rendah fungsi manajemen tertentu dengan menempatkan pengumpulan data dan analisis dekat dengan elemen yang dikelola, dan untuk memungkinkan pemrosesan data dekat dengan lingkungan yang dikelola sehingga data yang perlu dikirim

ke hulu dapat memiliki kualitas yang lebih tinggi (diproses di dekat sumber) dan volume yang lebih rendah. Dengan cara ini, data manajemen yang perlu dikirim ke sistem hulu dapat lebih mudah dikelola. Dengan demikian, tingkat pendistribusian instans DCAE umumnya bergantung pada penyedia berdasarkan persyaratan, seperti waktu respons manajemen, kinerja, dan volume data.

Kebijakan

ONAP menyediakan kemampuan untuk menentukan kebijakan, tindakan yang dapat dipanggil berdasarkan hasil penerimaan peristiwa jenis tertentu (tanda tangan). Tanda tangan ini dapat merupakan hasil analitik yang dilakukan pada data (jaringan) yang dikumpulkan oleh DCAE. Kemampuan untuk melakukan pengumpulan dan analitik data serta menentukan tindakan yang akan diambil (kebijakan) berdasarkan hasil analitik tersebut merupakan pola desain yang merepresentasikan kapabilitas fundamental ONAP.

Sebagaimana disebutkan sebelumnya, pengumpulan data, analitik, dan kebijakan ditentukan pada waktu desain menggunakan platform SDC. Berbagai contoh pola desain ini,



yang digerakkan oleh berbagai jenis metadata jaringan, cloud, atau layanan dapat digunakan berulang kali sehingga, baik saat mengelola jaringan, lingkungan cloud, atau layanan, contoh pola desain yang sama dapat digunakan.

Loop Kontrol Terbuka dan Tertutup

Dalam ONAP, loop kontrol adalah pola desain yang merepresentasikan alur proses di mana data dikumpulkan dari lingkungan jaringan (misalnya, vNF, PNF, lingkungan cloud), dianalisis, dan tindakan (sebagaimana didefinisikan oleh kebijakan) ditentukan, berdasarkan hasil analitik (tanda tangan). Selain tindakan yang akan diambil, kebijakan juga mendefinisikan aktor yang bertanggung jawab untuk mengambil tindakan tertentu. Biasanya, aktor dapat berupa orkestrasi layanan, pengontrol, atau vNF, tergantung pada cakupan masalah dan tindakan yang diperlukan untuk remediasi. Loop kontrol (yaitu, data yang akan dikumpulkan, analitik yang akan dilakukan, dan tindakan yang akan diambil) didefinisikan pada waktu desain di SDC.

Loop kontrol dapat terbuka atau tertutup. Loop terbuka adalah loop kontrol di mana tindakan yang akan diambil dilakukan melalui proses manual. Misalnya, penyedia layanan dapat memutuskan bahwa personel operasi adalah aktor yang tepat untuk mengambil tindakan tertentu. Loop terbuka juga dapat digunakan untuk memvalidasi tindakan yang ditentukan oleh kebijakan.

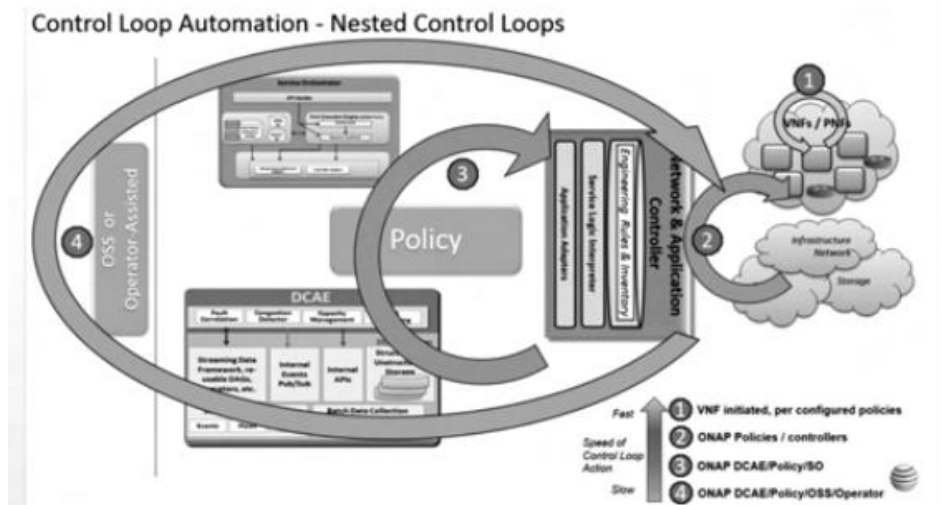
Loop tertutup adalah loop kontrol di mana tindakan yang akan diambil dilakukan secara otomatis di bawah kendali perangkat lunak. Kontrol loop tertutup dapat digunakan untuk menyediakan otomatisasi tingkat tinggi dalam mendukung pengurangan Biaya Operasional (OpEx) dengan menghilangkan proses manual dalam mengelola berbagai aspek jaringan (misalnya, manajemen kesalahan dan kinerja). Selain itu, otomatisasi loop tertutup dapat menganalisis dan mengambil tindakan pada skala waktu mulai dari submilidetik hingga milidetik, detik, menit, atau lebih lama.

Skala waktu loop kontrol bergantung pada aktor yang dibutuhkan untuk memperbaiki masalah. Loop kontrol dapat disusun dari skala waktu terpendek hingga terpanjang. Secara umum, jika aktor yang tepat untuk perbaikan tertentu adalah vNF, tindakan ini dapat diambil pada skala waktu terpendek. Loop kontrol jenis ini adalah loop kontrol yang "terketat". Jenis loop kontrol berikutnya dalam nesting adalah loop kontrol di mana aktor yang sesuai dalam loop tersebut adalah pengontrol (misalnya, pengontrol mungkin perlu berinteraksi dengan satu atau lebih vNF untuk memperbaiki masalah). Berikutnya dalam nesting adalah loop kontrol di mana aktor yang sesuai adalah orkestrasi layanan. Contoh loop ini adalah situasi di mana tindakan perbaikan adalah membuat instance vNF di pusat data lain. Dalam hal ini, orkestrasi diperlukan, seperti ketika vNF pertama kali dibuat. Terakhir, nesting terluar adalah ketika aktor yang sesuai adalah aplikasi atau sistem di luar ONAP dan lingkungan jaringan, seperti OSS. Loop kontrol jenis ini biasanya beroperasi pada skala waktu terpanjang (Gambar 3.6).

Khususnya dalam otomatisasi loop tertutup, loop yang beroperasi pada skala waktu terpendek terbukti penting dalam mendukung fungsi manajemen pada skala waktu yang dibutuhkan dalam mengelola aspek-aspek jaringan 5G, misalnya dalam mendukung aplikasi



seperti AR/VR atau mobil tanpa pengemudi, yang membutuhkan koneksi jaringan latensi rendah.



Gambar 3.6 Otomatisasi loop kontrol—loop kontrol bersarang.

3.4 MENGGABUNGKAN SEMUANYA

Jaringan yang ditentukan perangkat lunak dan NFV bersama-sama mengatasi tantangan peningkatan lalu lintas yang pesat, layanan sesuai permintaan, serta manajemen dan kontrol loop tertutup. Secara terpisah, keduanya merupakan alat yang ampuh, tetapi jika digabungkan, kemampuan untuk menciptakan sumber daya baru di jaringan berbasis cloud dan mengalirkan lalu lintas ke sumber daya tersebut secara dinamis merupakan kombinasi yang ampuh.

Penciptaan Bump Baru di Jaringan

Mari kita telaah kemungkinan menambahkan fungsi baru ke dalam lingkungan SDN/NFV yang dinamis. Salah satu cara untuk menangani serangan DDOS adalah dengan merutekan lalu lintas yang mencurigakan ke aplikasi scrubber yang terletak di pusat data yang strategis. Hal ini menambah penundaan dan meningkatkan pemanfaatan jaringan secara keseluruhan karena lalu lintas harus dialihkan dari jalur terpendek. Dalam dunia SDN/NFV, jaringan cerdas dapat memutuskan untuk membuat scrubber berbasis NFV baru di cloud yang sama dengan router virtual, stitch, atau rantai layanan scrubber ke jalur paket lokal. Hal ini tidak memerlukan sumber daya WAN tambahan dan meminimalkan dampak latensi seminimal mungkin untuk pembersihan lalu lintas. Setelah DDOS mereda, "benturan pada jaringan" dapat dihilangkan. Pendekatan penggabungan fungsi layanan ke dalam jalur ini bahkan dapat dilakukan untuk fitur yang lebih bermanfaat seperti aplikasi enkripsi, akselerasi web, dan firewall.

Penskalaan Elastis Titik Akhir

Aplikasi menarik lainnya dari SDN dan NFV adalah penskalaan elastis titik akhir. Seiring meningkatnya lalu lintas melalui atau menuju suatu aplikasi, telemetri dan kontrol loop tertutup dapat menentukan bahwa sumber daya tambahan akan memberikan layanan yang



lebih baik kepada pelanggan. Sistem SDN/NFV dapat mengorkestrasi sumber daya aplikasi NFV tambahan di lokasi cloud saat ini atau di lokasi cloud yang berbeda dan memodifikasi perutean paket untuk memindahkan lalu lintas dengan tepat ke sumber daya yang baru dibuat tersebut.

Bukan hanya pembuatan sumber daya (NFV) tetapi juga pergerakan lalu lintas ke sumber daya tersebut (SDN) yang benar-benar memaksimalkan potensinya. Penting juga untuk ditekankan bahwa ini bukan hanya penskalaan dalam satu pusat data, tetapi lintas pusat data. Hal ini dapat memungkinkan solusi unik seperti NFV berbasis private cloud, yang menggunakan sumber daya public cloud untuk lonjakan lalu lintas, mengubah model biaya karena penyedia layanan tidak perlu menyebarkan sumber daya lalu lintas burst di private edge cloud mereka untuk semua skenario. Jenis penskalaan hybrid cloud ini merupakan peningkatan langsung pada laba bersih.

Memindahkan Layanan Antar Lokasi

Seperti contoh penskalaan elastis, memindahkan layanan antar lokasi merupakan cara berbeda untuk menangani optimalisasi lalu lintas. Fungsi virtual dapat dipindahkan meskipun lalu lintas sedang kelebihan beban. Pelanggan pindah, menambah kantor, mengubah kebiasaan kerja, dan bahkan harus menghadapi badai salju yang membuat staf mereka tidak dapat masuk ke kantor. Di dunia lama fungsi jaringan fisik, latensi, dan kemacetan dapat diakibatkan oleh perubahan pola lalu lintas yang tidak praktis untuk diatasi. Dengan fungsi virtual, vNF dapat dipindahkan ke pusat data yang optimal untuk lalu lintas saat ini, baik secara semi-permanen maupun sementara. Sekali lagi, kombinasi kemampuan untuk membuat fungsi virtual di pusat data baru dan merutekan lalu lintas ke sana, ke mana pun ia dipindahkan, yang membuat jaringan lebih fleksibel.

3.5 DUKUNGAN UNTUK KECERDASAN BUATAN JARINGAN

Sama seperti SDN, cloud, NFV, ECOMP, dan ONAP mengantarkan teknologi dan pendekatan baru untuk jaringan dan layanan. Industri ini berada di ambang perubahan teknologi lainnya, AI dan pembelajaran mesin (ML). Bagian ini membahas bagaimana teknologi ini dapat digunakan dalam ekosistem seperti ONAP untuk menghadirkan tingkatan baru dalam kontrol dan manajemen otonom loop tertutup untuk jaringan dan layanan.

Telemetry

Penerapan AI/ML pada ekosistem manajemen seperti ONAP dimulai dengan data yang dikumpulkan dari lingkungan jaringan tempat analitik tingkat lanjut dilakukan, dan hasilnya digunakan untuk membuat keputusan kebijakan. Oleh karena itu, memiliki data yang benar dan berkualitas tinggi sangatlah penting. Terdapat peluang bagi AI untuk membersihkan data jaringan agar lebih mudah digunakan oleh aplikasi AI dan analitik lainnya dengan tujuan meningkatkan kualitas analitik dan keputusan yang dibuat menggunakan hasil analitik tersebut.

Tombol

Setelah data jaringan dibersihkan dan diagregasi sebagaimana dibahas sebelumnya, fase penting berikutnya adalah menerapkan AI untuk meningkatkan kebijakan yang digunakan untuk mengelola jaringan. Kebijakan adalah aturan yang menentukan tindakan apa yang harus



diambil berdasarkan tanda tangan, atau hasil analitik yang dilakukan pada data jaringan. Aturan yang menentukan tindakan yang akan diambil setelah menerima tanda tangan tertentu dapat ditingkatkan dengan menggunakan teknik AI. AI dapat menentukan apakah tindakan yang diambil, berdasarkan kebijakan, efektif dan apakah tindakan tersebut dapat ditingkatkan sehingga, misalnya, tindakan baru tersebut dapat memperbaiki masalah lebih cepat, meningkatkan kinerja jaringan dalam situasi tertentu, atau mungkin digunakan untuk menentukan cara yang lebih baik dalam memulihkan dari kegagalan vNF jenis tertentu.

Loop Terbuka dan Loop Tertutup

Kecerdasan buatan dan pembelajaran mesin (ML) dapat digunakan untuk meningkatkan dan menyempurnakan kontrol loop tertutup. Sebagaimana dibahas sebelumnya, kontrol loop tertutup terdiri dari pengumpulan data, analitik, dan tindakan yang didasarkan pada hasil analitik. Semua komponen loop tertutup dapat ditingkatkan dan disempurnakan melalui AI dan ML untuk meningkatkan kecepatan, akurasi, dan pada akhirnya efektivitas (nilai) kontrol loop tertutup. Khususnya, memprediksi kejadian dan kegagalan lebih awal serta menentukan tindakan perbaikan yang lebih efektif. Hal ini pada gilirannya akan berdampak positif pada ekonomi penyedia jaringan, OpEx, dan kualitas jaringan sebuah pembeda yang kompetitif.

Tiket Pintar

Penggunaan AI/ML akan memungkinkan tiket masalah "pintar" berkualitas lebih tinggi dengan memasukkan informasi yang relevan dalam tiket masalah (misalnya, hasil pengujian yang dilakukan) dan menghilangkan informasi yang asing atau bukan penyebab utama. Pemrosesan bahasa alami AI dapat membantu menafsirkan dan memproses catatan, jika digunakan, untuk memaksimalkan otomatisasi loop tertutup.

Menghindari Kebutuhan Tiket

Seiring dengan peningkatan kualitas kontrol loop tertutup dan keputusan kebijakan yang dimungkinkannya melalui penerapan AI/ML, kebutuhan akan tiket masalah seharusnya dapat dihindari.

3.6 RINGKASAN

Seperti yang Anda lihat, terdapat banyak kontrol yang disediakan oleh Jaringan Terdefinisi Perangkat Lunak modern bagi penyedia layanan. Penggunaan kontrol tersebut secara manual memang bermanfaat untuk menyelesaikan masalah lebih cepat, tetapi keuntungan sebenarnya berasal dari otomatisasi, yang, misalnya, tujuan aspirasionalnya adalah mengurangi secara drastis, jika tidak menghilangkan, tiket masalah, sehingga menghasilkan efisiensi operasional.

Bab-bab selanjutnya seperti Bab 9 tentang Kecerdasan Buatan untuk Operasi Jaringan akan memberikan detail lebih lanjut tentang penerapan AI pada Operasi Jaringan. Tiga area Penyediaan Jaringan, Pemeliharaan Jaringan, dan Administrasi Jaringan merupakan area yang kaya akan aplikasi AI. Bab-bab lain dalam buku ini juga akan relevan, karena tindakan yang dapat dilakukan oleh algoritma AI dan Pembelajaran Mesin hanya dibatasi oleh seberapa fleksibel API di lapisan Jaringan Terdefinisi Perangkat Lunak dan Virtualisasi Fungsi Jaringan.



BAB 4

BLOCKCHAIN DAN KRIPTOGRAFI UNTUK KEAMANAN

4.1 PENDAHULUAN

Bab ini memperkenalkan peran keamanan informasi dan mekanisme konsensus sebagaimana diterapkan pada kecerdasan buatan (AI) yang otonom. Bab ini memberikan gambaran umum tentang berbagai mekanisme konsensus dan menggambarkan perlunya catatan informasi yang tidak dapat rusak terkait dengan masukan keputusan AI, catatan keputusan, dan catatan umpan balik yang digunakan untuk mengoreksi dan melengkapi pengambilan keputusan AI. Pemanfaatan prinsip-prinsip Blockchain dan kriptografi memungkinkan fungsi AI untuk beroperasi dalam jaringan terdistribusi atau topologi geografis dan mendukung pengambilan keputusan kombinatorial menggunakan kumpulan informasi (atau keluaran keputusan AI lainnya) dari domain informasi yang berbeda.

Bagian 4.2 menyajikan gambaran umum tentang implikasi Blockchain dan kriptografi, yang dimanfaatkan baik secara umum maupun secara default di antara implementasi Blockchain terkemuka. Bagian 4.3 memperkenalkan algoritma konsensus yang umum digunakan dalam solusi Blockchain dan membahas trade-off di dalamnya. Bagian 4.4 membahas alasan utama untuk memanfaatkan fitur-fitur Blockchain guna melengkapi AI. Bagian 4.5 membahas bagaimana buku besar terdistribusi dapat mendukung AI terdistribusi, mekanisme umpan balik loop tertutup, dan memperkenalkan konsep AI parlementer yang relevan di masa mendatang untuk pengambilan keputusan tingkat tinggi menggunakan berbagai informasi yang aman.

4.2 TINJAUAN UMUM BLOCKCHAIN

Apa Itu Blockchain?

Pada intinya, Blockchain adalah buku besar digital. Layaknya buku besar fisik, Blockchain digunakan untuk mencatat serangkaian transaksi yang terurut. Transaksi ini dapat diperiksa kapan saja, dan dapat disesuaikan melalui entri baru ke buku besar. Namun, entri buku besar tidak boleh diubah atau ditulis ulang hanya diamandemen. Dalam hal ini, transaksi buku besar dianggap tidak dapat diubah dan buku besar, atau rantai, hanya ditambahkan.

Dalam buku besar fisik, urutan ditegakkan secara fisik. Setiap entri baru ditambahkan pada baris berikutnya di bawah entri saat ini. Jika diperlukan koreksi, baris baru ditambahkan untuk melakukan koreksi. Seorang pengamat yang mengamati buku besar yang dimodifikasi akan dengan mudah menemukan tindakan penipuan, seperti melewati baris kosong untuk menggunakannya untuk entri yang datang kemudian atau kembali dan mengubah entri yang sudah ditulis. Dalam lingkungan digital, verifikasi lebih sulit. Bit literal dari catatan digital dapat dimodifikasi oleh sistem atau aktor mana pun yang memiliki akses ke bit tersebut. Gagasan seorang pengamat lebih sulit segera setelah beberapa sistem terlibat, dan secara eksponensial sulit ketika beberapa jaringan terlibat.



Dalam buku besar digital, prinsip kriptografi seperti hashing, enkripsi, kunci, dan tanda tangan dimanfaatkan untuk membuat pesan dapat diverifikasi. Dalam mekanisme yang paling sederhana, setiap entri baru atau blok yang ditambahkan ke buku besar mencakup nilai yang sesuai dengan blok sebelumnya. Dengan cara ini, blok dirantai bersama oleh setiap blok yang berisi komponen sebelumnya. Merusak rantai menjadi sulit tanpa menulis rantai yang sama sekali baru dari titik tertentu dalam buku besar.

Sebagian besar implementasi Blockchain memanfaatkan komunitas peer sistem untuk mendistribusikan salinan buku besar di antara mereka. Setiap node Blockchain memiliki salinan buku besar, menerima dan memvalidasi entri baru saat dibuat, dan dapat mengusulkan entri. Risiko penulisan ulang rantai secara diam-diam dimitigasi dengan mewajibkan konsensus di antara node untuk menerima entri baru. Setiap node memiliki kemampuan untuk mendeteksi penulisan ulang yang terjadi dan memilih untuk menolaknya.

Dengan blok berantai, skema kriptografi untuk memeriksa integritas, dan sekumpulan peer sistem terdistribusi untuk memeriksa, memvalidasi, dan memasukkan transaksi, satu-satunya hal yang tersisa adalah rasa tujuan. Terdapat berbagai implementasi Blockchain, yang dapat diringkas menjadi beberapa kategori dasar:

Perdagangan mata uang dan aset: Termasuk Bitcoin yang populer, implementasi ini ada untuk menambang, mengelola, memperdagangkan, dan mentransfer komoditas digital dengan nilai tertentu di antara para peserta.

Verifikasi identitas dan atribut: Karena kemampuan verifikasi yang kuat, implementasi Blockchain seperti Sovrin digunakan untuk memvalidasi referensi ke identitas atau atribut suatu identitas tanpa menempatkan semua identitas dan semua atribut secara terpusat ke entitas umum.

Komputasi terdistribusi: Termasuk jaringan Ethereum yang populer, implementasi ini seringkali memiliki beberapa elemen mata uang, di mana mata uang tersebut digunakan untuk membayar eksekusi logika arbitrer di antara rekan-rekan Blockchain dan mengembalikan hasilnya kepada pengirim.

Tujuan umum/kerangka kerja: Upaya Blockchain lainnya seperti proyek Hyperledger dari Linux Foundation berupaya menciptakan implementasi Blockchain tingkat perusahaan, yang dapat digunakan untuk memecahkan kebutuhan bisnis arbitrer dengan menciptakan aplikasi yang mendukung kerangka kerja Blockchain. Jenis masalah yang dapat diatasi Blockchain lebih penting bagi implementasi ini daripada menyelesaikan contoh masalah tertentu.

Bagaimana Prinsip Kriptografi Digunakan dalam Blockchain?

Karena sifat digitalnya, implementasi Blockchain memerlukan mekanisme untuk memberikan jaminan integritas rantai. Masalah ini bukanlah hal baru bagi Blockchain, salah satu contoh pertama yang umum digunakan dari teknik ini adalah dalam email terenkripsi dan ditandatangani.

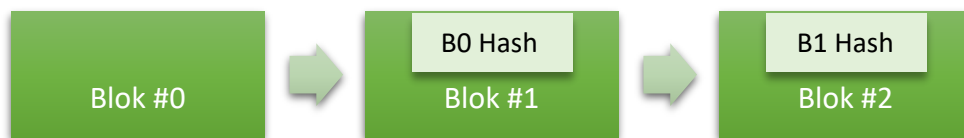
Setiap partisipan dalam percakapan pertama-tama membuat pasangan kunci. Pasangan ini terdiri dari kunci publik, yang dibagikan kepada semua orang, dan kunci privat yang disimpan oleh setiap partisipan. Ketika seorang penulis menulis pesan, pesan tersebut



dienkripsi menggunakan kunci publik penerima yang ditunjuk. Hal ini memastikan bahwa hanya penerima yang dapat membaca pesan tersebut, dengan mendekripsi menggunakan kunci privat mereka sendiri yang cocok.

Namun, ketika percakapan tidak perlu bersifat privat, seperti di papan pesan, penulis lebih peduli untuk memastikan bahwa pemikiran mereka yang terlihat publik adalah milik mereka sendiri bahwa pemikiran tersebut tidak diutak-atik. Untuk mencapai hal ini, setiap penulis menandatangani setiap pesan yang mereka tulis. Tanda tangan tersebut memperhitungkan isi pesan dan sertifikat serta kunci privat pengirim. Tanda tangan tersebut kemudian dimasukkan ke dalam pesan. Siapa pun yang membaca pesan dapat memvalidasi bahwa pesan tersebut utuh dengan melakukan komputasi terhadap isi pesan dan tanda tangan menggunakan kunci publik yang dipublikasikan oleh pengirim. Bila hasil nilai hash yang dihitung cocok, maka keaslian pesan terkonfirmasi.

Implementasi Blockchain memanfaatkan kedua prinsip kriptografi ini. Untuk mengautentikasi ke rantai sebagai node anggota, kunci privat dan publik digunakan untuk mengonfirmasi identitas partisipan rantai yang diketahui. Jika identitas tersebut dikenali sebagai anggota Blockchain, node yang berpartisipasi akan bergabung, menyinkronkan salinan rantai mereka, dan dapat mengusulkan blok baru, memvalidasi proposal lain, dan beroperasi pada rantai blok yang ada. Seperti yang ditunjukkan pada Gambar 4.1, ketika blok baru dibuat, setiap blok membawa hash dari blok sebelumnya, tanda tangan pembuat, dan informasi lain yang spesifik untuk tujuan implementasi Blockchain.



Gambar 4.1 Ilustrasi Blockchain Sederhana.

Dengan menggunakan teknik ini, di mana hash dari setiap blok berikutnya mencakup nilai hash dari blok sebelumnya, setiap blok terhubung secara permanen ke blok sebelumnya.

Meskipun konsep yang dijelaskan sebelumnya merupakan teknik inti, beberapa teknik khusus diperlukan untuk mencapai skala terdistribusi dalam validasi dan transaksi multipihak dengan blok. Metode yang paling umum memanfaatkan Pohon Merkle untuk melakukan Bukti Merkle. Merkle Proofs pertama kali diterapkan di Bitcoin Blockchain sebagai mekanisme untuk memasukkan informasi transaksi ke dalam blok sehingga riwayat transaksi dapat divalidasi dalam header blok, alih-alih harus mengunduh dan memverifikasi ulang konten setiap blok. Dengan kata lain, status buku besar dapat divalidasi hanya dengan mentransfer dan menyimpan sejumlah kecil data.

Ketika sebuah transaksi individual perlu diverifikasi, buktinya dapat ditegaskan untuk menunjukkan bahwa transaksi tersebut berada di dalam Merkle Tree dengan induk tingkat atasnya (simpul akar) yang ada di header blok dalam rantai. Karena Merkle Proofs memungkinkan metode validasi yang nyaman untuk atribut tertentu, Merkle Tree tambahan



dapat digunakan dalam implementasi Blockchain untuk memberikan bukti tentang status atribut lain seperti pemilik aset yang dilacak Blockchain saat ini, saldo akun, atau jika hasil tertentu dari suatu transaksi telah menciptakan beberapa informasi tambahan atau nilai, yang perlu dipahami dalam ekosistem partisipan rantai.

Contoh Implementasi Blockchain Terkemuka

Sebagaimana dijelaskan sebelumnya, terdapat beberapa jenis implementasi Blockchain, masing-masing menargetkan tujuan utama tertentu. Domain yang paling populer berdasarkan jumlah implementasinya adalah domain perdagangan mata uang dan aset, yang dipopulerkan oleh penetrasi ekonomi Bitcoin yang luas.

Bitcoin pertama kali daring pada Januari 2009 dengan Bitcoin pertama yang diciptakan dan klien sumber terbuka yang dirilis. Dikreditkan ke nama samaran daring Satoshi Nakamoto, Bitcoin digunakan terutama oleh mereka yang memiliki minat hobi dalam kriptografi dan konsep menggabungkan mata uang kripto, anonimitas, dan sistem nilai yang mengaktualisasikan diri secara mandiri dan independen dari lembaga mana pun. Selama dua tahun pertama, protokol ini dijalankan oleh semakin banyak penggemar, yang mengarah pada penemuan dan koreksi kerentanan sistem utama pertama yang memungkinkan penciptaan Bitcoin dalam jumlah tak terbatas. Ekosistem tersebut mengoreksi diri sendiri, memilih untuk mendasarkan kembali buku besar dari tepat sebelum titik eksploitasi dengan implementasi yang diperbarui yang menutup cacat tersebut.

Sejak 2011, para penghobi, penggemar, calon pesaing, dan lainnya telah memanfaatkan implementasi protokol Bitcoin sumber terbuka untuk menciptakan mata uang kripto butik lainnya, meskipun belum ada yang mencapai ukuran dan skala partisipasi Bitcoin. Perusahaan seperti Mt. Gox dan Coinbase muncul untuk menyediakan layanan jual beli Bitcoin menggunakan mata uang tradisional. Nilai dan permintaan Bitcoin, meskipun fluktuatif, terus meningkat.

Implementasi Blockchain terbesar berikutnya disebut Ethereum, yang dirilis untuk penggunaan publik pada tahun 2014. Proyek Ethereum bertujuan untuk mencapai tujuan teknologi dan ekosistem yang sangat berbeda. Alih-alih hanya berfokus pada mata uang, Ethereum menciptakan ekonomi komputasi yang difasilitasi oleh Blockchain dengan mata uang kripto asli yang mendukung pertukaran upaya-untuk-nilai. Pada intinya, Mesin Virtual Ethereum yang digunakan oleh node Ethereum mendukung bahasa Turing lengkap, yang berarti dapat digunakan untuk memecahkan masalah komputasi arbitrer. Ether, mata uang yang digunakan dalam Ethereum, digunakan untuk membeli "gas", yang mewakili biaya komputasi yang dikeluarkan jaringan Ethereum untuk mengeksekusi kode aplikasi dalam Mesin Virtual Ethereum yang terdistribusi, yang menghasilkan pembuatan blok. Pendekatan ini, meskipun lebih kompleks daripada Bitcoin, memungkinkan lingkungan tujuan umum menggunakan instans jaringan Ethereum dan jaringan umum yang tangguh untuk komputasi metode bisnis arbitrer yang diajukan oleh ekosistem pengguna Ethereum.

Dengan sistem yang kompleks muncul potensi kerentanan yang lebih rumit. Pada tahun 2016, implementasi Ethereum diperbarui sesuai dengan re-base buku besar tiga kali sebelum eksploitasi. Beberapa teknik telah dikembangkan untuk memberikan ketahanan dan



pemeriksaan kesalahan yang lebih baik pada kode aplikasi arbitrer, yang disebut kontrak pintar, sehingga kondisi eksploitasi dapat dideteksi dan diatasi. Untuk meningkatkan skala ekosistem komputasi Ethereum yang kini umum digunakan, proyek Ethereum beralih dari konsensus Proof-of-Work ke implementasi Proof-of-Stake, yang akan dijelaskan nanti.

Implementasi Sovrin Blockchain sepenuhnya mengabaikan gagasan ekosistem dan mata uang yang kompetitif. Sebaliknya, Sovrin berupaya menciptakan jaringan identitas manusia individual dengan kepemilikan individual namun beroperasi secara terdesentralisasi. Tujuan jaringan identitas Sovrin adalah untuk memfasilitasi kepercayaan yang dapat diverifikasi antar pihak, dengan menjalankan verifikasi Blockchain atas atribut identitas secara aman menghindari kebocoran atribut identitas pribadi ke pihak lain. Peserta dalam jaringan identitas Sovrin, yang dapat berupa individu atau organisasi, menyimpan atribut manusia individual seperti usia, nama, tanggal lahir, dan alamat. Atribut-atribut tersebut diverifikasi dengan cara yang menjaga privasi.

Contoh termudah dari verifikasi atribut identitas Sovrin adalah sebuah pernyataan. Suatu organisasi mungkin mewajibkan pengguna berusia di atas 18 tahun. Permintaan ke jaringan Sovrin akan memvalidasi bahwa "ya" identitas tersebut berusia di atas 18 tahun. Permintaan tersebut tidak akan mengungkapkan usia spesifik pengguna. Pemilik identitas memegang kendali atas atribut, bagaimana dan kapan atribut tersebut dibagikan, serta asal-usulnya.

Akhirnya, proyek Hyperledger oleh Linux Foundation bertujuan untuk menciptakan kerangka kerja Blockchain generalis untuk berbagai kebutuhan. Ditujukan untuk pengembang dan perusahaan, seseorang atau organisasi dapat memodifikasi, memanfaatkan, atau menghosting komponen platform Hyperledger untuk menciptakan mata uang kripto seperti Bitcoin, jaringan ekosistem seperti Ethereum, platform identitas seperti Sovrin, atau memecahkan berbagai kombinasi masalah arbitrer lainnya. Hyperledger paling tepat digambarkan sebagai implementasi buku besar sumber terbuka umum dan sejumlah kerangka kerja, alat, dan pustaka perangkat lunak untuk memungkinkan operasi kustom dikodekan pada instans buku besar umum oleh jaringan rekan yang menjalankan instans alat klien Hyperledger untuk tujuan tertentu.

Mendukung Berbagi Informasi yang Aman

Implementasi Blockchain Sovrin memungkinkan berbagi informasi yang aman, di mana informasi spesifiknya menggambarkan atribut individu manusia. Baik Ethereum maupun Hyperledger dapat mendukung berbagi informasi yang aman dengan kode aplikasi yang sesuai atau modifikasi khusus. Prinsip berbagi informasi yang aman sangat mirip dengan teknik email pribadi yang dijelaskan sebelumnya, menggunakan kunci publik dan pribadi penerima, serta tanda tangan pengirim yang terverifikasi.

Dengan memanfaatkan Blockchain, blok dapat dikodekan dengan informasi pribadi. Kunci publik penerima yang dituju dapat digunakan untuk mengkodekan pesan termasuk tanda tangan pengirim. Meskipun muatan terenkripsi dapat muncul di buku besar publik dan dilihat oleh semua anggota, muatan tersebut hanya dapat dibaca oleh pihak yang ditunjuk. Dengan cara ini, sistem terdistribusi dapat memungkinkan berbagi informasi yang aman dan



tangguh tanpa tanggung jawab otoritas terpusat. Dengan menggunakan Hyperledger, grup rantai dapat digunakan bersama dengan percakapan yang dinegosiasikan pada rantai publik utama dan dilakukan dalam rantai terpisah yang eksklusif untuk pihak yang bercakap-cakap. Rantai percakapan dapat dibuktikan telah dinegosiasikan dalam rantai publik dengan menemukan hash blok, dan integritasnya dapat diperiksa dengan menegaskan Bukti Merkle.

Dengan menggunakan metode sederhana ini, banyak kasus penggunaan menjadi mungkin seperti membuat aplikasi yang mengurangi kebutuhan untuk mendiktekan atau menampilkan kata sandi Wi-Fi tamu kepada pengunjung, atau kemampuan untuk membuat dinding pesan yang tahan terhadap gangguan. Penerapan Blockchain untuk berbagi informasi yang aman dalam Jaringan Berbasis Perangkat Lunak (*Software-Defined Network*) dapat mencakup pemberian informasi kepada pihak tujuan tentang ancaman jaringan yang terdeteksi tanpa perlu mengetahui terlebih dahulu cara merutekan informasi tersebut ke node tujuan. Penerima akan dapat memvalidasi keaslian pesan agen pengirim tanpa konektivitas (dan kepercayaan) dengan agen pengirim.

4.3 KONSENSUS

Apa Itu Konsensus dalam Blockchain?

Karena implementasi Blockchain memanfaatkan kelompok partisipan yang tersebar, mekanisme harus digunakan untuk menciptakan kepercayaan pada hasil komputasi. Meskipun metode Proof-of-Work adalah yang paling umum, terdapat metode lain untuk mencapai kesimpulan serupa hasil yang dikirimkan oleh node partisipan tertentu benar, dan node-node dalam jaringan partisipan menerima hasil tersebut ke dalam buku besar.

Masalah konsensus tidak hanya terjadi pada Blockchain, meskipun implementasi Blockchain memang membutuhkan beberapa bentuk konsensus. Setiap sistem terdistribusi atau multi-agen membutuhkan mekanisme konsensus untuk menghasilkan hasil komputasi. Dalam komputasi multi-utas, memori dibagi dari utas induk ke utas anak. Karena beberapa utas anak dapat memodifikasi satu ruang memori bersama, mekanisme penguncian digunakan. Pada suatu titik waktu, tidak ada utas yang memodifikasi ruang memori, atau hanya satu utas yang memodifikasi ruang tersebut. Jika utas konkuren ingin memodifikasi memori, utas tersebut harus menunggu hingga kunci dilepaskan. Bentuk konsensus ini menyatakan bahwa satu utas di antara beberapa utas bertanggung jawab atas status akhir transaksi pada ruang memori bersama.

Demikian pula, agar jaringan Blockchain dapat melanjutkan penambahan entri baru ke buku besar, hasil sebelumnya harus dikenali terlebih dahulu. Di dalam Blockchain, serangkaian komputasi, pembuatan data, dan modifikasi data dapat terjadi sebagai bagian dari satu transaksi. Beberapa peserta peer mungkin bekerja secara paralel, masing-masing berusaha mengusulkan blok berikutnya dengan hasil mereka. Ketika blok-blok ini diusulkan, implementasi Blockchain menjalankan algoritma konsensus aktif mereka untuk memutuskan blok mana yang diusulkan yang akan diterima ke dalam buku besar, dan dalam urutan apa. Karena sifat jaringan partisipan yang terdistribusi ini, kelompok node dapat "setuju" untuk



melanjutkan entri buku besar yang berbeda. Seperti halnya Bitcoin dan Ethereum, buku besar dianggap "bercabang" dari titik konsensus yang berbeda ini.

Metode dan Kompromi: Bukti Kerja

Dengan implementasi yang digunakan dalam Bitcoin, mata uang turunan, dan Ethereum pada saat penulisan ini, Bukti Kerja adalah metode konsensus yang paling umum digunakan. Konsep dasar mekanisme Bukti Kerja adalah sulit untuk melakukan pekerjaan tetapi mudah untuk memverifikasi bahwa pekerjaan telah dilakukan. Seperti halnya menambang permata atau logam mulia, tindakan menambang logam membutuhkan banyak upaya, tetapi tindakan memvalidasi keberadaan logam keluaran sangatlah mudah jika dibandingkan. Istilah "menambang" digunakan dalam implementasi Blockchain untuk menggambarkan tindakan melakukan langkah kerja.

Fungsi Proof-of-Work dapat bervariasi gayanya beberapa lebih menuntut aktivitas jaringan, komputasi prosesor sistem, atau alokasi memori sistem. Sementara algoritma Proof-of-Work yang digunakan dalam Bitcoin sangat intensif komputasi prosesor, algoritma yang dimanfaatkan dalam Ethereum intensif memori sistem. Dalam setiap kasus, algoritma Proof-of-Work merupakan fungsi biaya sumber daya tetap, semi-tetap, atau variabel yang menyebabkan node jaringan yang melakukan pekerjaan harus mengeluarkan upaya untuk melakukannya. Dalam kasus Bitcoin, peluang nilai potensial dalam mata uang kripto ini telah menyebabkan konsentrasi besar sumber daya komputasi khusus (dikenal sebagai Sirkuit Terpadu Khusus Aplikasi [ASIC]) yang dimiliki oleh organisasi. Konsentrasi kekuatan ini merupakan upaya untuk memaksimalkan kemampuan mereka mencapai konsensus dengan cepat pada blok yang ditambang secara lokal.

Kompromi untuk Proof-of-Work adalah bahwa energi dan biaya untuk menghasilkan pekerjaan biasanya merupakan pemborosan material upaya yang dikeluarkan untuk melakukan pekerjaan tersebut tidak dihabiskan untuk kebutuhan bisnis, sosial, atau ekonomi. Sebaliknya, pekerjaan tersebut merupakan beban biaya bagi pihak yang melakukan pekerjaan tersebut, untuk mendapatkan hak istimewa partisipasi aktif dalam implementasi Blockchain. Meskipun cara ini berfungsi dengan baik untuk kelompok peserta terdistribusi yang tidak memiliki hubungan kepercayaan yang melekat di antara mereka, cara ini merupakan pilihan yang buruk bagi organisasi yang memanfaatkan Blockchain internal atau berkolaborasi dengan organisasi lain di mana ukuran itikad baik telah ditetapkan, seperti perjanjian hukum atau kontrak layanan.

Metode dan Kompromi: Proof-of-Stake

Ketika pertama kali menyelidiki alternatif Proof-of-Stake, proyek Ethereum menganggapnya sebagai pekerjaan yang cukup rumit. Meskipun beberapa metode Proof-of-Stake sedang dimanfaatkan dalam implementasi Blockchain, implementasi Blockchain ini tidak memiliki ekosistem partisipan yang besar dan memiliki pertukaran nilai kompetitif yang jauh di bawah Ethereum atau Bitcoin. Ketika metode Proof-of-Stake tradisional, termasuk pengacakan, penunjukan berbasis usia, penunjukan berbasis kuantitas, atau sebagian besar lainnya, diperiksa dalam ekosistem dengan sistem pertukaran nilai, metode tersebut rentan terhadap kerentanan dalam teori permainan sejumlah nilai atau tekanan eksternal yang



diberikan kepada pemegang saham yang mendorong pemegang tersebut untuk melepaskan sahamnya kepada aktor lain.

Mengesampingkan algoritma Proof-of-Stake tradisional, proyek Ethereum sedang mempersiapkan implementasi Proof-of-Stake, Casper, yang memanfaatkan stake dalam bentuk escrow mata uang beserta hukuman (kehilangan stake escrow) dan imbalan (penemuan perilaku buruk pemangku kepentingan) untuk memitigasi insentif permainan di jaringan Ethereum. Metode Casper akan dimanfaatkan dalam 1% transaksi blok, pada awalnya, untuk membuktikan kelayakannya.

Dalam lingkungan di mana insentif, pengaruh, dan kompetisi permainan tidak ada, atau dikendalikan melalui cara lain, implementasi Proof-of-Stake lain menjadi layak. Protokol Tendermint mengimplementasikan *Byzantine Fault Tolerance* (BFT) untuk memastikan konsistensi blok baru dalam Blockchain. Menggunakan algoritma rotasi sekuensial rekan yang dipilih sebagai pengusul blok beserta serangkaian pemungutan suara pra-komitmen terbatas waktu yang membutuhkan mayoritas super dua pertiga untuk meratifikasi blok antar anggota sebelum pemilihan pengusul baru secara otomatis. Toleransi Kesalahan Bizantium (BFT) tidak hanya terbatas pada implementasi Blockchain dan memiliki aplikasi yang luas dalam permasalahan ilmu komputer. Sebaliknya, munculnya penggunaan Blockchain telah menjadi keuntungan bagi cara implementasi sistem BFT menggunakan Blockchain untuk mendukung pemilihan BFT merupakan pilihan yang tepat.

Metode dan Kompromi: Bukti Otoritas

Baik Proof of Work maupun Proof of Stake memiliki pertimbangan khusus untuk memitigasi kurangnya itikad baik yang melekat pada pengaruh eksternal dan perilaku node partisipan dalam jaringan peer terdistribusi. Ketika asumsi ini dapat diabaikan, seperti dalam jaringan privat, jaringan pemilik tunggal, atau jaringan yang menggunakan mekanisme alternatif untuk menegaskan kepercayaan dan kebijakan (seperti Jaringan yang Ditentukan Perangkat Lunak), optimasi konsensus dapat dicapai. Jaringan yang menegaskan kepercayaan dapat memanfaatkan Proof of Authority untuk mencapai konsensus pada skala yang jauh lebih besar, yang dapat dicapai dengan Proof of Work atau kompleksitas dalam Proof of Stake.

Konsensus Proof of Authority bergantung pada node yang ditunjuk, alih-alih peer yang dipilih atau dipilih secara algoritmik. Node otoritas ini mengusulkan dan memvalidasi blok baru di Blockchain. Node otoritas baru dikonfirmasi (atau dikecualikan) oleh kumpulan node otoritas yang ada. Karena tidak diperlukan fungsi biaya atau algoritma mahal untuk menerapkan imbalan dan hukuman ekonomi, node otoritas dapat membuat blok dengan kecepatan yang setara dengan waktu yang dibutuhkan untuk menghitung hash data blok dan menyinkronkan blok baru di antara jaringan node otoritas. Divergensi grup node tidak perlu terjadi tanpa alasan bisnis.

Meskipun umumnya tidak cocok untuk jaringan publik terdistribusi tanpa identitas yang kuat atau hubungan itikad baik yang telah terjalin, mekanisme konsensus Proof-of-Authority dapat berfungsi di domain publik untuk beberapa kasus penggunaan. Sinkronisasi server waktu, bukti verifikasi audit post-facto untuk peristiwa pesanan, atau infrastruktur penandatanganan tanpa kunci dapat dibuat dan dimanfaatkan secara publik atau privat



dengan mekanisme konsensus Proof-of-Authority. Secara privat, mekanisme konsensus Proof-of-Authority dapat digunakan untuk mengonversi akses sistem organisasi dan modifikasi informasi menjadi rangkaian bukti rangkaian peristiwa untuk mengotomatiskan audit sistem, memfasilitasi deteksi ancaman, melacak sinkronisasi berkas antar sistem di jaringan yang berbeda, atau menyediakan dimensi yang sesuai untuk pola data yang ada di antara sistem.

Dalam Jaringan yang Ditentukan Perangkat Lunak, konsensus Proof-of-Authority dapat dimanfaatkan untuk memelihara daftar rute *Border Gateway Protocol* (BGP) dan menegaskan penandaan verifikasi sumber untuk melacak dan membatasi lalu lintas palsu yang tampak berasal dari sumber yang tidak dapat secara sah mendukung perutean lalu lintas tersebut. Jejak audit penerapan dan perubahan pada layanan mikro jaringan dapat dilacak dan divalidasi, bahkan (dan terutama) ketika beberapa operator jaringan dan agen instrumentasi terlibat. Node otoritas dalam Jaringan yang Ditentukan Perangkat Lunak dapat berkorespondensi atau diterapkan bersama dengan elemen bidang kontrol untuk kebutuhan sertifikasi dan verifikasi kebijakan.

Jenis Konsensus Lainnya

Tidak hanya terbatas pada implementasi Blockchain, konsensus di antara kelompok sejawat terdistribusi tentang status sistem, informasi, atau transaksi ekonomi apa pun menghadirkan masalah ilmu komputer. Konsensus berlaku untuk banyak kasus penggunaan, mulai dari replikasi data back-office hingga gim multipemain daring yang serba cepat. Banyak algoritma konsensus yang ada, dan hanya sebagian kecil yang telah diterapkan pada implementasi Blockchain dalam skala besar.

Algoritma konsensus Ripple memfasilitasi ekosistem pencatatan keuangan yang terdistribusi. Ripple berasumsi bahwa node yang melakukan validasi tidak memiliki kepentingan pribadi yang melekat dalam mengonfirmasi transaksi yang curang atau tidak benar, dan menggunakan pendekatan banyak buku besar. Di Ripple, buku besar baru dapat dibuat setiap beberapa detik. Buku besar ini diurutkan dan mewakili urutan transaksi yang disepakati oleh rekan sejawat yang telah terjadi, dan dikomit bersama dengan buku besar yang telah dinumerasi yang berisi transaksi tersebut. Rantai buku besar tersebut bersama-sama sesuai dengan urutan kelompok transaksi keuangan yang dianggap telah dikomit. Transaksi yang tidak muncul di buku besar mana pun bersifat sementara untuk sementara waktu, sebelum menyelesaikan buku besar terbuka, tetapi transaksi tersebut akhirnya ditolak karena kurangnya dukungan konsensus di antara kelompok sejawat yang memvalidasi dan tidak muncul di buku besar yang telah difinalisasi yang ditambahkan ke daftar buku besar.

Keluarga protokol dalam Paxos memfasilitasi konsensus di mana, berbeda dengan Ripple, rekan jaringan dianggap tidak dapat diandalkan dan seringkali gagal atau tidak tersedia. Paxos, pertama kali diterbitkan pada tahun 1980an dan 1990an, mendahului implementasi Blockchain apa pun dan telah digunakan dalam replikasi data perusahaan. Karena Paxos berupaya menciptakan ketahanan dalam replikasi yang dapat diverifikasi, Paxos sangat cocok untuk situasi di mana konsistensi data dan toleransi kesalahan antara volume, sistem, atau konstruksi gudang data diinginkan.



Dalam permainan multipemain, algoritma konsensus diterapkan untuk merasionalisasi sisi klien dari status pemain dengan status pemain lain yang diusulkan untuk memutuskan peristiwa dalam permainan di mana sumber daya terbatas. Misalnya, jika dua pemain sama-sama melewati item dalam game, server harus merasionalisasi pemain mana yang menerima item tersebut dan mendistribusikannya kepada pemain tersebut. Untuk memberikan pengalaman pengguna yang paling lancar hampir sepanjang waktu, klien akan menerima tingkat pergerakan dan tindakan pemain tertentu tanpa sinkronisasi server. Dalam kasus kondisi jaringan yang tidak seragam atau kepadatan server, tindakan korektif klien untuk menyesuaikan status server dapat mengganggu pengalaman pemain.

4.4 Mendukung Kecerdasan Buatan dengan Blockchain

AI dalam Pembelajaran, Catatan, dan Nilai

Pembuatan layanan AI membutuhkan perhatian cermat terhadap masukan informasi dan paparan lintas domain dari masukan dan umpan balik pelatihan AI. Sistem kognitif yang tidak dimitigasi oleh naluri, hati nurani, atau bahkan rentan terhadap kebingungan sistematis yang tidak disengaja, mudah dirusak oleh bias dan kelemahan laten. Untuk menciptakan subsistem AI otonom yang dapat diinstrumentasikan dalam situasi tepercaya, perhatian cermat harus diberikan pada materi pembelajaran, kecepatan pembelajaran, serta kualitas umpan balik dan penilaian.

Di luar kebutuhan untuk menciptakan AI pada awalnya, kemampuan untuk menyempurnakan dan menciptakan kembali AI yang murni dan dapat diuji diperlukan untuk mencapai skala dan sertifikasi serta instrumentasi spesifik domain. Ketika rencana pembelajaran AI dimekanisasi dan diotomatisasi, integritas rencana pembelajaran harus dilindungi. Blockchain dapat digunakan untuk mengodekan rencana pembelajaran lintas dimensi waktu, pengetahuan, paparan, dan umpan balik.

Dikodekan ke Blockchain, rencana pembelajaran secara inheren dapat diverifikasi dalam keadaan aslinya. Penyesuaian pada rencana pembelajaran dapat ditambahkan dan buku besar didasarkan kembali atau dikembalikan secara kondisional berdasarkan hasil pengujian. Selama fase pengujian, praproduksi, dan pasca-penerapan, skor, klasifikasi, rekomendasi, atau tindakan AI lainnya dapat dicatat dalam Blockchain asli yang terkait dengan AI selama masa pakainya. Jika terjadi kesalahan atau diperlukan studi, snapshot status buku besar akan menyediakan analisis forensik yang detail, pemutaran ulang, dan simulasi tindakan AI. Dengan penerapan kepatuhan AI terhadap domain waktu, penyesuaian skalar pada periode waktu dapat digunakan untuk mempercepat atau memperlambat frekuensi periodik input AI ke tindakan, sehingga memfasilitasi pelatihan ulang dan simulasi yang cepat untuk sertifikasi. Dengan kata lain, proses otomatis yang menjalankan rencana pelatihan dapat mempercepat waktu pelatihan AI secara artifisial, namun tetap mencapai umpan balik tepat waktu yang dirancang dalam rencana pelatihan.

Pembelajaran Penguatan dan Koreksi

Biasanya, teknik pembelajaran penguatan memanfaatkan sistem penghargaan selama proses yang diamati. Ketika pengamat menginterpretasikan keadaan lingkungan baru setelah



tindakan AI sebagai positif, proses pembelajaran diberikan penghargaan. Proses ini biasanya beroperasi dengan informasi nyata dalam lingkungan lokal atau simulasi lengkap. Penghargaan bisa positif, negatif, atau hanya "kurang positif" untuk memberikan umpan balik pembelajaran berkelanjutan kepada subjek AI.

Namun, sistem insentif harus dibangun dengan cermat, pengamat harus waspada, dan lingkungan yang representatif agar tindakan AI dapat terkonvergensi secara optimal. Mengingat AI yang terintegrasi jaringan menginformasikan fungsi firewall dengan interpretasi pengamat yang buruk, ketika AI diberikan insentif penghargaan untuk memblokir port yang menjalani pemindaian kerentanan, AI akan terus melakukannya lebih sering. Tindakan ini dapat dimanipulasi oleh penyerang yang tidak tertarik menemukan kerentanan jaringan, padahal penyerang dapat dengan mudah meyakinkan AI untuk melakukan penolakan layanan yang diaktifkan firewall dengan menutup port layanan yang secara efektif mengganggu kontinuitas layanan untuk klien yang sah. Memberikan penghargaan kepada AI yang melakukan prosedur medis dengan tidak adanya komplikasi di masa mendatang pada anggota tubuh tertentu akan menghasilkan bias terhadap amputasi.

Untuk melatih AI agar bertindak lebih cerdas dalam situasi kompleks, metode baru telah muncul untuk menciptakan bentuk alternatif sistem penghargaan dan koreksi. Dengan menggunakan pembelajaran magang melalui penguatan terbalik (AIRP), sistem penghargaan AI didasarkan pada seberapa dekat ia meniru dorongan manusia. Keadaan sistem yang diinginkan diamati secara langsung sebagai hasil dari tindakan manusia yang ahli. Dengan mengkodifikasi urutan masukan dan tindakan manusia ke dalam Blockchain, inefisiensi bawaan tindakan manusia (seperti waktu pemahaman, waktu reaksi, dan kesalahan) dapat dikurangi. Varians dalam kecepatan bertindak, responsivitas terhadap pengambilan keputusan, presisi masukan, dan perhatian terhadap semua masukan informasi yang relevan sekaligus dapat disesuaikan secara offline untuk mencapai optimum teoretis. Selama pemutaran, AI dapat belajar meniru perilaku dari manusia yang "sempurna".

Mengomunikasikan Pembelajaran

Dengan mengodekan model pelatihan serta pengalaman runtime AI dalam Blockchain yang spesifik terhadap instans, urutan kejadian dan keputusan yang diinginkan dapat diintegrasikan ke dalam AI baru atau yang sudah ada melalui pemutaran ulang. Dengan memanfaatkan snapshot buku besar yang sesuai dengan satu instans, instans lain dapat dibuat awalnya sebagai klon dan memulai divergensi spesifik instans sesuai dengan masukan unik dan peluang untuk AIRP jika sesuai. Penggunaan kembali buku besar pelatihan induk menghadirkan peluang lisensi dan sertifikasi untuk domain informasi spesifik tertentu.

Di luar domain spesifik dengan instans AI yang sangat mirip, memanfaatkan prinsip-prinsip pembelajaran transfer mesin akan memungkinkan AI yang terlatih untuk memahami jenis tindakan dalam domain informasi tertentu untuk menciptakan kompetensi untuk jenis tindakan yang sama di domain informasi lain. Misalnya, AI yang dilatih untuk melakukan penilaian risiko keamanan untuk peristiwa autentikasi jaringan dapat dipindahkan ke domain asuransi untuk mengukur risiko sebagai komponen adjudikasi klaim.



Adaptasi domain dapat dipercepat dengan menerapkan AIRP pada instans AI yang dikomunikasikan dalam domain baru untuk segera mencapai tingkat kinerja yang dapat diterima. Penggunaan instans yang telah diadaptasi dalam mekanisme A/B memungkinkan instans baru dibuat menggunakan AIRP, yang kini dapat mengeksplor dan mengoptimalkan ledger. Ledger baru yang telah dioptimalkan kini menjadi representasi dari rencana pelatihan khusus domain.

Ketika sebuah organisasi membangun praktikum untuk konstruksi AI, pencampuran elemen-elemen dari domain yang berdekatan secara cermat ke dalam rencana ledger akan memungkinkan mutasi AI terbaik dengan kemampuan adaptasi yang tinggi dan sensitivitas informasi yang lebih rendah. Aset-aset ini dapat menghasilkan peluang pendapatan baru bagi organisasi yang dapat menciptakannya secara efektif.

4.5 MENGGUNAKAN BLOCKCHAIN

Menutup Loop

Menerapkan instans AI yang didukung Blockchain dalam sistem orkestrasi seperti Software-Defined Networks memungkinkan loop tertutup asli untuk input, tindakan, hasil, dan koreksi. Dalam fungsi bidang kontrol, instans AI harus menerima input analitik perilaku jaringan. Analitik ini merepresentasikan bentuk lalu lintas dan karakteristik penggunaan jaringan, yang disegmentasi oleh sejumlah variabel termasuk asal, tujuan, jenis, throughput, dan detail lainnya. Lihat Bab 3 untuk detail lebih lanjut tentang proses loop tertutup.

Saat AI memancarkan tindakan sebagai respons terhadap jaringan, tindakan ini beserta fitur analitik yang sesuai disimpan ke dalam struktur data untuk sebuah blok. Tindakan AI yang dipancarkan diterima oleh proses pengamat, yang secara kondisional memancarkan instans kebijakan sebagai respons terhadap tindakan AI, jika berlaku. Hasil implementasi untuk kebijakan tertentu, kapan pun terjadi, ditambahkan ke struktur blok dan blok tersebut difinalisasi.

Setelah kebijakan dikeluarkan, kemunculan berikutnya dari status lingkungan jaringan baru dapat diukur dan dibandingkan dengan status sebelumnya untuk divergensi. Jika divergensi terlokalisasi pada ketiadaan pola ancaman yang diidentifikasi dalam tindakan sebelumnya, AI menerima imbalan positif. Jika divergensi mencakup perubahan signifikan pada bagian non-ancaman dari karakteristik jaringan, AI akan menerima imbalan nol atau negatif. Jika memungkinkan, tindakan kebijakan awal dapat memicu simulasi jalur jaringan dinamis untuk menciptakan kondisi pengujian A/B. Hasil AI dapat diuji dan diperkuat dalam jalur jaringan yang disimulasikan sebelum sertifikasi dan penerapan di jalur layanan jaringan.

Kecerdasan Buatan Terdistribusi dengan Buku Besar Terdistribusi

Dalam elemen bidang kendali suatu ekosistem, baik dalam Jaringan yang Didefinisikan Perangkat Lunak (SDI) maupun domain lainnya, instans AI dapat didukung oleh satu buku besar umum atau beberapa buku besar mandiri. Untuk instans di mana instrumentasi waktu nyata kurang penting daripada konsistensi, buku besar umum dapat mencatat masukan, paparan, dan umpan balik dari beberapa identitas instans AI. Tanpa perlu fungsi sentralisasi, mekanisme



konsensus buku besar terdistribusi akan merekonsiliasi tindakan yang dipancarkan dari semua instans dan menyinkronkan hasil terbaik yang disepakati.

Namun, dalam beberapa kasus, lebih berguna untuk menerapkan konsensus alternatif seperti Bukti Otoritas untuk memastikan bahwa aturan yang diterapkan pertama kali oleh setiap instans AI secara otomatis direplikasi ke setiap instans, memastikan konsistensi publik. Dengan aturan konsensus yang tepat, otoritas dapat dicampur dan blok buku besar dipilih untuk dimasukkan untuk beberapa tindakan dan otoritas didelegasikan ke node pengamat untuk yang lain.

Selain instrumentasi jaringan, mekanisme konsensus yang dimanfaatkan di seluruh buku besar terdistribusi yang mendukung instans AI dapat memastikan kerja sama multipihak dan saling menguntungkan dari instans AI terdistribusi. Pembelajaran transfer dapat diotomatisasi, instans yang lebih matang dapat berfungsi untuk memodelkan perilaku untuk instans yang lebih baru, dan jejak audit untuk semua tindakan instans di antara instans AI terdistribusi dibagikan dan dapat diverifikasi.

AI untuk Parlemen: Konsensus dan Pemungutan Suara

Organisasi yang berkecimpung dalam konstruksi AI harus mencocokkan domain bisnis dengan beberapa implementasi AI terbaik yang matang melalui mutasi rencana pelatihan yang cermat, adaptasi domain selektif, dan penguatan pemagangan ahli. Instansi-instansi ini dapat diorkestrasi untuk mencatat hasil ke dalam buku besar khusus, yang dirancang untuk operasi otonom tingkat tinggi. Misalnya, perusahaan e-commerce dapat membuat implementasi domain AI untuk preferensi pelanggan, perilaku belanja daring, keamanan jaringan, karakteristik penggunaan perangkat seluler, dan logistik inventaris.

Menggunakan AI domain ini bersama-sama memberikan peluang untuk melatih AI agar memahami tindakan proses bisnis tingkat tinggi sesuai dengan ukuran Indikator Kinerja Utama (KPI) perusahaan. Jika banyak pesanan datang untuk produk lama, instansi parlementer akan memeriksa buku besar terkini dari perilaku belanja daring, keamanan jaringan, karakteristik penggunaan, inventaris, dan instansi preferensi untuk menentukan sebagian besar pesanan tersebut merupakan penipuan. Instansi parlementer akan mengirimkan tindakan untuk menunda inventaris sambil menunggu umpan balik manusia.

Dengan konsensus hasil dari beberapa instans AI spesifik domain, AI parlementer mampu merasionalisasi tindakan yang telah dipilih dari setiap domain menjadi pola proses bisnis. Fungsionalitas ini dapat menyediakan sistem informasi yang aman bagi organisasi atau ekosistem, serta sistem informasi untuk memantau kesehatan organisasi secara keseluruhan antar domain mengidentifikasi potensi risiko atau peluang untuk optimasi lintas batas domain organisasi tradisional.

Agar domain yang sebelumnya terputus dapat mengomunikasikan tindakan dan hasil melalui sistem parlementer, berbagi informasi yang aman harus dimanfaatkan untuk melindungi fungsi dan dampak tingkat tinggi yang penting dari pengaruh data input atau eksekusi output. Setiap instans AI domain akan diberikan kunci unik yang menyelesaikan instans parlementer. Saat mengirimkan hasil, buku besar parlementer mencatat muatan, dan instans parlementer bertanggung jawab untuk mendekripsi dan memproses instruksi. Berbagi



informasi yang aman memungkinkan organisasi yang tersebar di seluruh jaringan, geografi, atau batas konektivitas internal untuk menciptakan sistem kolaborasi AI, menjaga auditabilitas dan verifikasi, serta memitigasi risiko kebocoran informasi bisnis strategis saat transit.

Permeasi dalam Jaringan Otonom

Konstruksi AI parlementer, buku besar terdistribusi, berbagi informasi yang aman, validasi yang kuat, perekaman jalur keputusan yang ramah audit, dan pilihan yang tepat menggunakan konsensus Bukti Otoritas memungkinkan terciptanya jaringan otonom yang belajar dan berkembang. Baik menghadapi ancaman keamanan tradisional, penyerang teori permainan, audit lisensi vendor VNF, kebutuhan bisnis dan lalu lintas yang terus berubah, atau gangguan layanan regional yang selektif, teknik yang dijelaskan menyediakan jaringan yang dapat mengelola dan menyesuaikan diri secara mandiri.

Jaringan saat ini sudah mencakup mekanisme penyembuhan diri dan dibangun untuk tingkat redundansi dan lapisan fisik tertentu. Namun, seiring jaringan berevolusi dari lapisan fisik mode tunggal atau ganda dengan karakteristik lalu lintas umum dan bertujuan umum menjadi pola penggunaan paralel domain campuran, kemampuan untuk menentukan tindakan dalam perangkat lunak harus dilengkapi dengan konfigurasi, penyesuaian, dan kebijakan yang digerakkan oleh perangkat lunak.

Sesi jaringan harus menjangkau makhluk fisik untuk memberikan pengalaman yang berpindah antar perangkat, dengan mulus. Kapasitas harus bergerak, secara prediktif, dengan kebutuhan bisnis yang diantisipasi. Ekosistem layanan yang dikonsumsi individu dan organisasi harus selalu memiliki kedekatan jaringan dengan pelanggan mereka.

Sementara itu, jaringan otonom harus tetap menjadi tempat yang aman bagi data, transaksi, aset digital, dan melindungi hak milik individu. Jaringan otonom masa depan harus menggabungkan berbagai pendekatan canggih seperti yang dibahas di sini agar dapat mandiri, mengoptimalkan diri, dan menjaga diri. Jaringan otonom dijelaskan secara rinci, demikian pula kebutuhan untuk berevolusi dari jaringan yang dibangun khusus dan dirancang berdasarkan spesifikasi saat ini menjadi jaringan masa depan yang terus berkembang dan dinamis.



BAB 5

MEMBANGUN JARINGAN OTONOM MASA DEPAN

5.1 PENDAHULUAN

Bab ini menguraikan beberapa tantangan teknis mendasar dalam jaringan yang ditentukan perangkat lunak saat ini, yang sekaligus merupakan peluang untuk bergerak menuju jaringan otonom jaringan otonom adalah jaringan yang mandiri; ia mengelola sendiri, mengoptimalkan sendiri, dan melindungi dirinya dari segala bentuk serangan siber. Jaringan otonom bereaksi terhadap rangsangan eksternal dan internal dengan cara yang sama seperti tubuh manusia atau hewan. Contoh yang baik dari hal ini adalah mobil tanpa pengemudi, yang dapat menavigasi secara otonom ke tujuan terprogram.

Bagian 5.2 mengilustrasikan visi untuk jaringan masa depan, diikuti oleh Bagian 5.3 yang menggambarkan siapa/apa saja pengguna jaringan masa depan. Bagian 5.4 menjelaskan bagaimana kami membayangkan jaringan masa depan dikelola dan dikendalikan, diikuti oleh Bagian 5.5 yang membahas tantangannya. Terakhir, Bagian 5.6 dalam bab ini membahas beberapa kasus penggunaan untuk menggambarkan jaringan masa depan.

5.2 JARINGAN MASA DEPAN

Visi Jaringan Masa Depan

Jaringan komunikasi sebelumnya hanya untuk satu tujuan. Peralatan dalam jaringan difokuskan untuk menyediakan layanan tunggal yang berfungsi dengan baik (misalnya, layanan suara atau layanan data) dan cenderung terdiri dari peralatan fisik dan memiliki pola penggunaan yang umum. Layanan tersebut dispesifikasikan dengan baik melalui standar, dapat diprediksi, dan definisi layanan terintegrasi ke dalam peralatan. Peralatan jaringan agak terisolasi dari sistem pendukung bisnis, dan sistem pendukung operasional hanyalah itu sistem pendukung untuk pengoperasian peralatan jaringan.

Hal ini kini telah berubah dalam sejumlah hal mendasar. Jaringan komunikasi kini mendukung berbagai layanan, dan setiap layanan memiliki karakteristik dan persyaratan yang berbeda. Pola penggunaan tidak lagi dapat diprediksi dan dapat berubah dengan cepat, misalnya, karena pengenalan perangkat konsumen baru atau keberhasilan dan kegagalan aplikasi konsumen populer. Peralatan fisik menjadi virtual dan multiguna, dan jaringan menjadi dapat diprogram. Perpindahan dari fungsi fisik ke fungsi virtual, bersama dengan jaringan yang dapat diprogram, menciptakan kemungkinan bagi fungsi jaringan untuk direlokasi secara real-time.

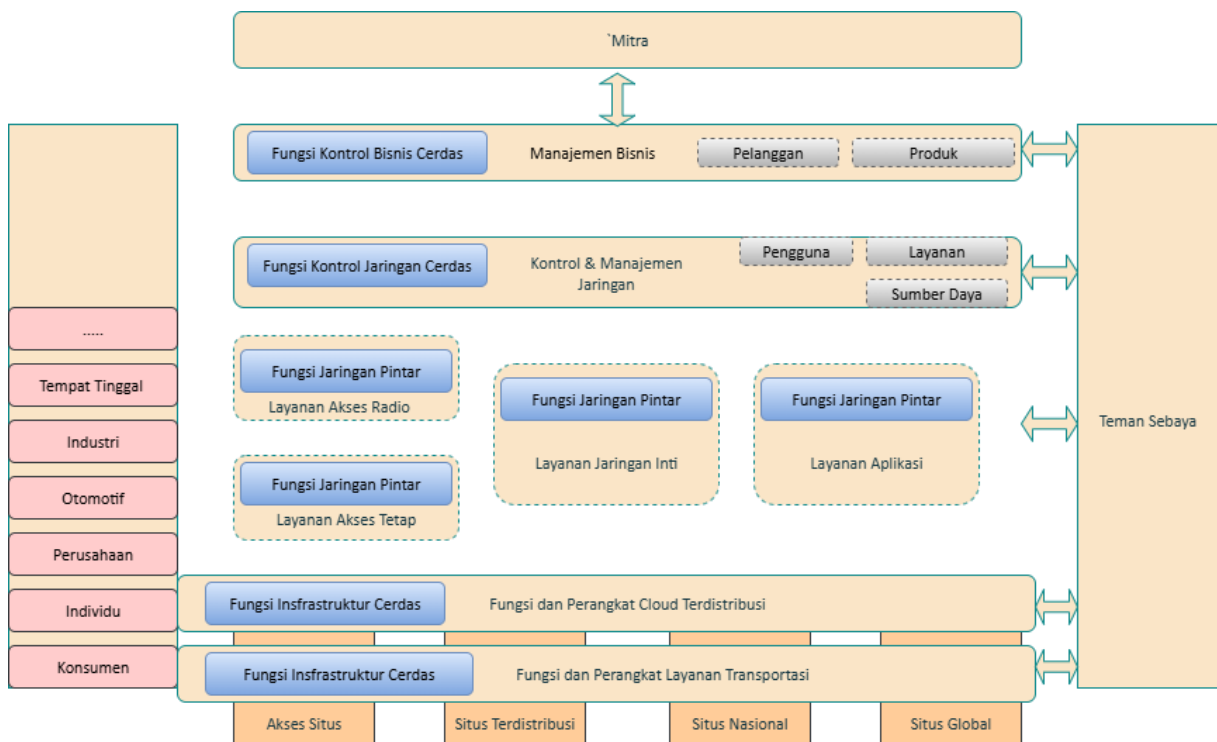
Hal ini sendiri memungkinkan kemungkinan pengoptimalan baru yang sebelumnya tidak mungkin. Lebih lanjut, peningkatan daya komputasi yang tersedia memfasilitasi bahwa fungsi jaringan tidak hanya multiguna tetapi juga menjadi lebih pintar, yang berarti bahwa mereka dapat mempelajari tugas yang diperlukan dalam parameter yang ditetapkan. Sistem pendukung operasional berevolusi dari mendukung jaringan dan peralatan menjadi



mendefinisikan dan mengelola layanan yang disediakan oleh jaringan, memanfaatkan kemampuan pemrograman peralatan jaringan yang terus berkembang.

Sistem pendukung operasional juga berevolusi untuk memiliki hubungan yang lebih erat dengan sistem pendukung bisnis untuk memungkinkan pengenalan layanan yang cepat dan layanan pelanggan yang lebih responsif. Dengan diperkenalkannya kecerdasan buatan (AI) dan meningkatnya kemampuan pemrograman serta kebutuhan untuk mendukung kasus penggunaan yang sangat berbeda, perubahan ini semakin cepat. Jaringan masa depan, yang semakin dekat, akan menyediakan beragam layanan, dan setiap layanan akan memiliki karakteristik yang sangat berbeda serta pola penggunaan yang tak terduga.

Konsumen layanan yang disediakan akan beragam, mulai dari manusia dengan perangkat tetap/bergerak hingga perangkat non-manusia (misalnya, *Internet of Things* [IoT]/komunikasi tipe mesin). Perangkat non-manusia mencakup barang-barang seperti kamera pengawas hingga bel pintu dan peralatan rumah tangga (misalnya, kulkas dan mesin cuci), serta kendaraan semi-otonom/otonom dan peralatan industri.



Gambar 5.1 Arsitektur jaringan masa depan.

Peralatan jaringan dan fungsi jaringan merupakan sumber daya untuk mendukung penyampaian layanan. Keduanya akan disesuaikan untuk mendukung satu layanan atau beberapa layanan dan menjadi fungsi jaringan cerdas (SNF). Menjadi cerdas menyiratkan kemampuan belajar mandiri yang disediakan oleh AI yang dipadukan dengan kebijakan berbasis maksud yang menangkap batasan. Ini menyiratkan bahwa mereka dapat belajar dalam konteks mereka sendiri cara terbaik untuk menjalankan tugas yang dialokasikan.

Untuk mengilustrasikan bagaimana suatu fungsi jaringan dapat menjadi cerdas, pertimbangkan bagaimana fungsi jaringan mengalokasikan sumber daya yang tersedia.



Beberapa informasi pengguna sebaiknya disimpan dalam memori cepat; namun, untuk pengguna lain, mungkin lebih baik untuk memindahkan informasi status pelanggan ke penyimpanan massal. Kemudian, mempelajari perubahan perilaku pengguna dapat memungkinkan fungsi tersebut untuk memindahkan data dari penyimpanan massal ke memori cepat sebagai persiapan untuk aktivitas pengguna.

Arsitektur jaringan masa depan diilustrasikan pada Gambar 5.1. Lapisan-lapisan utama dijelaskan sebagai berikut:

- Lapisan manajemen bisnis, yang berisi fungsi kontrol bisnis cerdas. Lapisan manajemen bisnis mendukung identifikasi produk, manajemen pelanggan, dan mitra.
- Lapisan kontrol dan manajemen jaringan, berisi fungsi kontrol jaringan pintar, yang mengelola dan mengoptimalkan jaringan dan fungsi-fungsi jaringan untuk mengoptimalkan layanan yang disediakan.
- Lapisan fungsi dan perangkat cloud terdistribusi, berisi fungsi infrastruktur pintar, dan menyediakan lingkungan eksekusi untuk SNF, fungsi kontrol bisnis pintar, dan fungsi kontrol jaringan pintar.
- Lapisan fungsi dan perangkat layanan transportasi, juga berisi fungsi infrastruktur pintar yang berfokus pada transportasi, menyediakan fungsionalitas untuk mendukung layanan transportasi.
- Lapisan fungsi jaringan yang mencakup layanan akses, inti, dan aplikasi, terdiri dari SNF dan menyediakan fungsionalitas aplikasi jaringan kepada pengguna akhir.

Perlu digarisbawahi bahwa lapisan cloud dan transportasi terdistribusi dapat dikatakan meluas ke peralatan konsumen. Fungsi infrastruktur pintar terutama menyediakan kemampuan untuk mendukung lingkungan eksekusi (komputasi, penyimpanan, dan jaringan) dan dapat berupa fisik maupun virtual. Mereka dapat mempelajari pola aplikasi dan mengoptimalkan layanan mereka. Jaringan masa depan terus mendukung mitra. Beberapa mitra akan menggunakan layanan yang disediakan oleh jaringan untuk mendukung kebutuhan bisnis mereka, dan beberapa mitra menyediakan layanan yang diperlukan untuk menjalankan layanan ujung ke ujung. Bagian-bagian yang berbeda dijelaskan lebih detail di bagian berikut. Fungsi jaringan dijelaskan sebagai pengelompokan:

- SNF untuk mendukung layanan akses radio
- SNF untuk mendukung layanan akses tetap
- SNF untuk mendukung layanan jaringan inti
- SNF untuk mendukung layanan aplikasi

SNF menyediakan fungsionalitas aplikasi kepada pengguna akhir jaringan dan dijalankan pada infrastruktur cloud terdistribusi yang terdiri dari fungsi-fungsi infrastruktur pintar. Lokasi pasti cloud terdistribusi bergantung pada jaringan operator. Namun, umumnya terdapat pusat data yang lebih kecil yang dapat ditempatkan di jaringan akses (di tepi jaringan), di lokasi pelanggan, hingga lokasi yang lebih global yang menghosting pusat data besar.

Sistem pendukung operasional jaringan masa depan disebut sebagai kontrol dan manajemen jaringan. Fungsi kontrol jaringan pintar (dari lapisan kontrol dan manajemen jaringan) pada dasarnya bekerja dengan konsep pengguna, layanan, dan sumber daya. Artinya,



ia mengelola sumber daya untuk menyediakan layanan (yang didefinisikan dan dikelola oleh lapisan kontrol dan manajemen jaringan) kepada pengguna akhir. Sumber daya tersebut adalah SNF dan fungsi infrastruktur pintar. Fungsi ini bertanggung jawab atas manajemen siklus hidup layanan dan sumber daya secara lengkap, mulai dari definisi, jaminan, optimalisasi, hingga penghentian layanan. Fungsi kontrol jaringan pintar menjadi otak jaringan, yang memprogram SNF untuk menyediakan layanan yang diperlukan. Fungsi kontrol jaringan pintar, bersama dengan SNF, mempelajari layanan umum dan pola penggunaan, sehingga memperoleh pemahaman tentang cara mengoptimalkan jaringan serta mempelajari cara menerapkan solusi baru untuk mendukung layanan yang disediakan dan mengoptimalkan jaringan.

Fungsi kendali jaringan pintar, dan sampai batas tertentu, SNF, akan mampu mengantisipasi kebutuhan pengguna di masa mendatang. Hal ini memungkinkan mereka mengalokasikan sumber daya yang diperlukan terlebih dahulu, memaksimalkan pengalaman pengguna dengan biaya tambahan minimal. Fungsi kendali jaringan pintar bagaikan seorang dalang yang menarik tali boneka untuk menciptakan pengalaman yang menyenangkan bagi penonton, yang merupakan konsumen pertunjukan. Fungsi kendali jaringan pintar bertugas menyederhanakan kompleksitas operasional jaringan bagi penyedia layanan.

Karena fungsi kendali jaringan pintar memungkinkan jaringan menjadi otonom dalam menyediakan layanan jaringan kepada pengguna, peran operator jaringan akan beralih dari mengelola jaringan menjadi mengelola otomatisasi.

Lokasi SNF dan fungsi kendali jaringan pintar akan dioptimalkan, menyeimbangkan pengalaman layanan dan biaya operasional, yang keduanya bergantung pada sejumlah faktor kompleks. Fungsi infrastruktur pintar dapat mendukung fungsi jaringan generik maupun khusus. Demikian pula, layanan dan fungsi transportasi menjadi pintar dan dapat belajar. Layanan transportasi dianggap sebagai fungsi infrastruktur pintar. Cara fungsi infrastruktur menjadi cerdas adalah dengan mempelajari cara mengoptimalkan layanan infrastruktur. Salah satu contohnya adalah merespons penyediaan layanan dan mengoptimalkan penggunaan fungsionalitas *Input/Output (I/O)* atau fungsionalitas penyimpanan berdasarkan pola aplikasi.

Mengoptimalkan jaringan ini akan menyeimbangkan ketegangan antara penyediaan pengalaman pengguna akhir yang optimal dan pemanfaatan sumber daya dan biaya yang efisien (biaya merupakan fungsi dari banyak faktor, seperti daya, transmisi, dan perizinan). Karena mayoritas SNF akan berupa fungsi virtual, SNF dapat ditempatkan di tempat yang paling hemat biaya dan optimal untuk layanan yang disediakan pada suatu waktu tertentu.

Jaringan masa depan akan tetap membutuhkan dukungan proses bisnis, seperti penagihan, manajemen hubungan pelanggan, mitra, dan pelanggan. Inilah peran manajemen bisnis. Hal ini, sebagaimana dijelaskan sebelumnya, terdiri dari fungsi kontrol bisnis cerdas yang memahami konsep pelanggan (pembayar), produk, dan mitra (mitra adalah aktor yang memiliki hubungan bisnis dengan operator dan menyediakan layanan kepada operator atau menggunakan layanan dari operator). Ini akan menyediakan fungsionalitas untuk berfokus pada kebutuhan pelanggan dan mitra, termasuk pemahaman produk yang diperlukan, serta



secara efisien mendukung kebutuhan bisnis, seperti penagihan, penagihan, dan manajemen hubungan.

Mungkin mudah untuk mendapatkan kesan bahwa kecerdasan tersebut tersentralisasi; namun, itu jauh dari kenyataan. Pengambilan keputusan akan didorong sedalam mungkin ke dalam jaringan, di mana lapisan yang lebih tinggi mendelegasikan keputusan sedalam mungkin (dengan batasan yang ditunjukkan), sambil tetap mempertahankan kemampuan untuk memengaruhi.

Akan mudah untuk menyatakan hari ini bahwa fungsionalitas sebelumnya diwujudkan melalui arsitektur layanan mikro; namun, teknologi realisasinya juga akan berkembang. Bahkan dengan teknologi realisasi baru, prinsip dasarnya akan tetap sama, yaitu komponen yang dapat digunakan kembali, yang kemampuannya akan diekspos melalui antarmuka berbasis model dan yang dapat dikombinasikan dengan layanan lain dari komponen lain untuk menciptakan aplikasi yang diperlukan. Sebagai pembaca, Anda mungkin membayangkan jaringan masa depan yang lain. Meskipun demikian, Anda mungkin dapat menghubungkan visi Anda dengan deskripsi sebelumnya. Apa pun visi Anda, satu hal yang jelas adalah bahwa kecerdasan tersebar di mana-mana.

Fleksibilitas Peluang dan Kutukan

Secara tradisional, banyak keputusan yang memengaruhi arsitektur jaringan penyedia layanan yang diterapkan dibuat selama fase perencanaan dan desain jaringan berdasarkan definisi layanan yang dipahami dengan baik. Lebih lanjut, perilaku peralatan jaringan dijadikan masukan.

Ini mencakup keputusan-keputusan seperti:

- Lokasi fungsionalitas
- Jumlah redundansi yang akan diterapkan, dan skema redundansi
- Topologi jaringan protokol internet (IP)
- Topologi fungsi jaringan
- Lokasi fungsi jaringan dan peralatan jaringan

Banyak waktu dan pertimbangan dicurahkan untuk membuat keputusan-keputusan ini, dan setelah keputusan dibuat dan jaringan diterapkan, upaya yang signifikan, baik dari segi waktu maupun biaya, diperlukan untuk mengubahnya. Misalnya, memindahkan node fisik, seperti Entitas Manajemen Mobilitas (*Mobility Management Entity*), Serving GW, atau Packet Data Network Gateway (berdasarkan standar 3GPP2) dengan titik terminasi jaringan statis, ke lokasi lain bukanlah hal yang mudah. Ada perangkat keras yang harus direlokasi, jaringan IP yang harus didesain ulang, daya yang harus dibangun, dan sebagainya. Hal yang sama berlaku untuk perluasan jaringan sederhana. Jika staf operasional jaringan ingin mengubah jaringan yang telah diterapkan (misalnya, dengan merelokasi fungsionalitas atau menerapkan fungsi baru dengan jenis yang sama), skala waktu yang dibutuhkan cukup panjang, sehingga secara umum staf operasional berfokus pada keputusan jangka pendek yang kurang berdampak.

Virtualisasi dan pemrograman jaringan mengubah hal ini. Tiba-tiba, dimungkinkan untuk dengan cepat membuat fungsi jaringan baru di berbagai lokasi, memprogram jaringan IP untuk mendukung desain jaringan baru, dan menyesuaikan fungsi jaringan untuk



mendukung berbagai kebutuhan seiring dengan adaptasi layanan. Hal ini memungkinkan jaringan virtual yang dirancang khusus untuk tujuan bisnis tertentu untuk diterapkan secara paralel (sering disebut sebagai network slicing, yaitu jaringan logis terpisah untuk tujuan atau serangkaian karakteristik tertentu). Seiring dengan perkembangan kapabilitas jaringan, kebutuhan untuk memanfaatkan fleksibilitas baru ini juga meningkat. Sifat layanan yang disediakan menjadi tidak teratur, dan terus ada layanan baru yang harus dikelola, beberapa di antaranya berumur pendek. Secara keseluruhan, virtualisasi dan pemrograman jaringan menghasilkan peningkatan kelincuhan yang sangat besar dalam jaringan. Hal ini tentu saja merupakan berkah.

Sisi lain dari kelincuhan jaringan adalah banyak keputusan yang sebelumnya dibuat selama fase perencanaan dan desain jaringan kini perlu dibuat selama fase operasi jaringan. Waktu untuk mempertimbangkan semua detail keputusan berkurang, dan tingkat pengambilan keputusan yang dibutuhkan jauh lebih tinggi. Lebih lanjut, waktu antar pemutakhiran fungsi jaringan berkurang drastis dari rencana pemutakhiran tahunan atau setengah tahunan ke metodologi Integrasi Berkelanjutan/Penyebaran Berkelanjutan (CI/CD) yang mengarah ke dan mencakup paradigma DevOps. Hal ini menghasilkan peningkatan tingkat pengambilan keputusan yang dibutuhkan dan tingkat pelaksanaan penyebaran yang diharapkan dari staf operasional. Mengharapkan hal ini dari staf operasi jaringan yang menggunakan metode dan teknologi manajemen jaringan tradisional, sementara pada saat yang sama menjaga biaya operasional tetap stabil, hampir mustahil. Inilah kutukannya.

Kutukan ini dapat diatasi dan kompleksitasnya dapat dikelola melalui kombinasi pemisahan tugas yang dipadukan dengan abstraksi jaringan dan penerapan teknologi pintar berupa AI. Kecerdasan buatan, sebagai sebuah teknologi, akan sangat penting untuk memungkinkan otomatisasi pengambilan keputusan cepat yang dibutuhkan dan memprediksi kebutuhan masa depan (misalnya, layanan baru atau perluasan sumber daya) yang dapat terwujud dalam tindakan yang telah disetujui sebelumnya untuk dieksekusi, serta mendeteksi dan memprediksi kegagalan, sehingga memastikan bahwa jaringan tersebut mandiri. Hal ini akan menyederhanakan operasi jaringan.

Jalan Menuju Jaringan Masa Depan

Ada beberapa transformasi yang sedang berlangsung ketika mempertimbangkan cara mencapai jaringan masa depan.

- ❖ Virtualisasi fungsi jaringan dan cloud.
 - Virtualisasi jaringan (fungsi jaringan, fungsi infrastruktur, manajemen dan kontrol, serta manajemen bisnis) memungkinkan fleksibilitas dalam cara jaringan, termasuk sistem pendukungnya, diterapkan dan dikelola. Terdapat skala ekonomi dalam fungsionalitas infrastruktur umum untuk mendukung virtualisasi fungsi jaringan secara efisien.
- ❖ Evolusi fungsi jaringan dan perangkat.
 - Fungsi dan perangkat jaringan sedang divirtualisasikan dan berevolusi menjadi fungsi jaringan virtual. Dengan demikian, mereka menjadi lebih serbaguna dan "diprogram" untuk berperilaku. Pada saat yang sama, mereka menjadi cerdas,



karena dikembangkan dengan kemampuan yang memungkinkan mereka mempelajari cara terbaik untuk melakukan tugas yang diberikan.

- Namun, beberapa fungsi dan perangkat jaringan, karena kebutuhan, akan tetap bersifat fisik (misalnya, sakelar, server, stasiun pangkalan radio). Hal ini akan tetap cenderung dapat diprogram dan cerdas, dan dari perspektif fungsi jaringan atau perangkat eksternal, mereka akan terlihat sangat mirip dengan fungsi virtual, dengan beberapa kemampuan yang lebih terbatas (misalnya, tidak dapat diskalakan atau direlokasi).
- ❖ Jaringan dapat diprogram (melalui jaringan yang ditentukan perangkat lunak).
 - Pergerakan menuju jaringan transportasi yang dapat diprogram, yaitu diprogram secara terpusat untuk tugas yang sedang dikerjakan, telah dimulai dan sedang dalam proses. Sentralisasi kendali jaringan transportasi memberikan gambaran yang lebih lengkap tentang pendekatan optimasi pendukung jaringan (memungkinkan pendekatan berbasis model dan kebijakan untuk manajemen jaringan) yang sebelumnya tidak mungkin dilakukan dalam pendekatan terdesentralisasi di mana semua keputusan diambil secara lokal. Lebih lanjut, kemampuan pemrograman jaringan transportasi memberikan kemampuan untuk mengkonfigurasi ulang jaringan dengan cepat guna mendukung situasi baru. Hal ini penting untuk mendukung alokasi dinamis fungsi jaringan virtual. Kemampuan pemrograman jaringan juga disebut sebagai jaringan yang ditentukan perangkat lunak.
- ❖ Jaringan menjadi cerdas.
 - Seiring dengan evolusi fungsi dan perangkat jaringan menjadi cerdas, jaringan menjadi cerdas. Jaringan menjadi mampu belajar dan beradaptasi dengan situasi terkini berdasarkan pengalaman masa lalu. Jaringan dapat dipengaruhi oleh lapisan yang lebih tinggi (orkestrasi, kebijakan, analitik, dan AI) melalui kemampuan pemrograman.
- ❖ Generalisasi sistem pendukung operasional.
 - Sistem pendukung operasional saat ini didedikasikan untuk domainnya (misalnya, sistem pendukung operasional untuk transportasi, jaringan akses radio, dan infrastruktur cloud). Pendekatan berbasis model terhadap sistem pendukung operasional menyiratkan bahwa sistem tersebut akan menjadi lebih horizontal dalam hal menjadi lebih multidomain (sistem pendukung yang sama mendukung beberapa domain), sekaligus menyediakan dukungan spesifik domain melalui model untuk domain tersebut (radio, transportasi, inti, infrastruktur cloud). Pada saat yang sama, sistem pendukung operasional mengambil peran yang lebih besar dalam definisi layanan dan menjadi cerdas.
- ❖ Dari otomatisasi ke jaringan otonom.
 - Otomatisasi sedang berlangsung saat ini. Titik awalnya adalah membuat skrip untuk tugas-tugas berulang, dan juga melalui orkestrasi. Dalam teknik saat ini, pengetahuan yang diketahui ditangkap dengan cara yang telah ditentukan



sebelumnya. Hal ini tidak cukup, karena tidak dapat menangani situasi yang tidak termasuk dalam pengetahuan yang diketahui, dan diperlukan pendekatan yang lebih otonom.

- Namun, otomatisasi akan bergerak menuju otonomi. Hal ini akan memungkinkan jaringan untuk menangani situasi yang tidak diketahui (misalnya, situasi yang sebelumnya tidak dipertimbangkan) dan tetap menanganinya dengan mendukung layanan yang telah ditentukan dan dalam batasan yang ditetapkan. Hal ini akan mengarah pada evolusi otomatisasi loop tertutup di mana AI akan memahami situasi jaringan saat ini dan yang diinginkan serta menyarankan tindakan yang tepat, dan kebijakan adaptif akan mengendalikan kondisi batas.

Singkatnya, fungsionalitas yang ada dalam jaringan akan menjadi pintar secara bertahap. Artinya, fungsionalitas tersebut akan beralih dari terprogram atau terskrip untuk mengikuti rutinitas yang telah ditentukan sebelumnya berdasarkan pengetahuan yang telah diketahui sebelumnya menjadi otonom dan mempelajari pendekatan terbaik dalam batasan yang telah dijelaskan (kendala).

5.3 PENGGUNA JARINGAN MASA DEPAN

Pengguna jaringan saat ini akan terus berlanjut, lalu ada pengguna masa depan yang sudah dapat kita identifikasi saat ini, dan ada pengguna masa depan lain yang belum kita bayangkan. Bayangkan jalan acak di kota besar mana pun, dan akan ada orang, mobil, taksi, bus, kereta api, lampu lalu lintas, toko, pintu, tempat sampah, sensor udara, dan segala macam hal. Bayangkan daerah pedesaan di mana peralatan pertanian, seperti traktor, mesin pemanen, sensor tanaman, dan label hewan pertanian akan tersedia. Bayangkan sebuah perusahaan; akan ada komputer, mesin kasir, peralatan komunikasi, printer, server, dan sebagainya. Bayangkan pabrik industri atau perusahaan industri apa pun; Akan ada mesin produksi, inventaris, barang produksi, peralatan transportasi, dan sebagainya. Bayangkan sebuah rumah keluarga; akan ada alarm, lampu, pemanas, peralatan hiburan, peralatan rumah tangga, sensor taman, pengontrol, dan sebagainya.

Ini adalah contoh sederhana tentang apa yang mungkin menjadi pengguna jaringan di masa depan. Daftarnya bisa terus bertambah, karena contoh selanjutnya mempertimbangkan rumah sakit, mal, pusat konferensi, pusat perbelanjaan, restoran, pusat olahraga, layanan darurat, dan sebagainya.

Kebutuhan pengguna di masa depan akan sangat bervariasi. Memahami kebutuhan mereka, menyediakan proposisi komersial yang sesuai bagi pengguna, dan menyediakan dukungan jaringan yang diperlukan perlu dilakukan, namun dilakukan dengan cara yang sederhana mengingat variasi yang begitu besar. Contoh-contoh berikut juga akan menggambarkan bahwa klasifikasi sederhana pengguna di masa depan seperti yang dijelaskan tidak cukup untuk memahami kebutuhan mereka.

Konsumen Swasta

Konsumen swasta merupakan gambaran umum dari banyak pengguna jaringan saat ini. Ini dapat mencakup rumah tangga dengan koneksi broadband hingga pengguna broadband



seluler dengan berbagai jenis perangkat seluler (ponsel, tablet, komputer, jam tangan, atau perangkat wearable lainnya).

Konsumen pribadi sadar akan biaya dan hanya ingin membayar untuk nilai dan penggunaan yang mereka terima. Hal ini menyiratkan penyesuaian layanan yang ditawarkan kepada mereka dengan memperhatikan nilai yang mereka rasakan. Selain peningkatan bandwidth dan ekspektasi untuk selalu terhubung, konsumen pribadi juga akan memiliki ekspektasi yang lebih tinggi terhadap keamanan. Lebih lanjut, konsumen pribadi juga akan memiliki perangkat lain di balik perangkat yang terhubung ke jaringan (peningkatan mesin yang terhubung, keamanan dan pengawasan, peralatan yang mungkin memerlukan dukungan keamanan tambahan karena tekanan biaya produsen perangkat, dan waktu pemasaran yang menyebabkan perangkat tersebut mungkin tidak aman secara inheren).

Internet of Things

IoT biasanya mengacu pada mesin dan komunikasi jenis mesin. Hal ini sendiri dapat bervariasi dan berkisar dari perangkat berdaya rendah dengan bandwidth rendah (misalnya, sensor air taman dan bel pintu yang terhubung) hingga perangkat berdaya tinggi dan bandwidth tinggi, seperti peralatan pengawasan atau perangkat yang dikendalikan dari jarak jauh. Ini mencakup berbagai aspek lain, seperti mobilitas, kecepatan pengiriman data, dan penjadwalan data tersebut. Pengguna ini akan membutuhkan prediktabilitas jaringan dan keamanan yang lebih baik. Beberapa kelas perangkat IoT juga akan mendapatkan manfaat dari dukungan jaringan lain, seperti menganalisis data, memberi tahu pemilik tentang masalah, dan peningkatan perangkat.

Meskipun saat ini pengguna yang membeli perangkat IoT bergantung pada kapabilitas broadband seluler atau tetap yang ada, bersama dengan definisi layanan yang ada, jaringan masa depan akan menyediakan dukungan untuk menyesuaikan jaringan dengan kebutuhan perangkat IoT melalui jaringan otonom. Kemampuan virtualisasi jaringan akan memungkinkan jaringan logis terpisah yang dikhususkan untuk kebutuhan perangkat, dan bahkan mungkin menjadi bagian dari kontrak penggunaan perangkat tersebut.

Perusahaan

Perusahaan adalah bisnis yang kecil atau besar dan membutuhkan layanan komunikasi, seperti konektivitas, mobilitas, dan komunikasi antarpribadi. Selain kantor, perusahaan dapat berupa bisnis khusus: apotek, dokter, rumah sakit, dan toko komersial. Persyaratan perusahaan akan berkembang untuk menuntut hubungan yang lebih fleksibel dan disesuaikan dengan kebutuhan mereka, serta dukungan jaringan lebih lanjut berupa aplikasi khusus yang dapat disediakan bagi perusahaan melalui sistem masuk tunggal dan layanan terpadu. Keamanan juga akan menjadi kunci, begitu pula dengan hosting infrastruktur komputasi perusahaan yang aman. Ambil contoh rumah sakit sebagai sebuah perusahaan.

Rumah sakit mungkin memerlukan interoperabilitas sesuai permintaan dengan rumah sakit lain atau spesialis medis, baik untuk berkonsultasi mengenai suatu situasi dengan pasien, melakukan prosedur jarak jauh, atau berbagi akses ke peralatan khusus. Hal ini akan membutuhkan solusi komunikasi yang fleksibel dan khusus yang tidak dapat dikompromikan atau menjadi tidak tersedia. Rumah sakit juga akan membutuhkan dukungan untuk privasi dan



komunikasi yang aman. Mentransfer data analisis khusus dapat memerlukan bandwidth transmisi yang ekstrem dan tidak dapat diprediksi. Hal ini tidak mungkin dilakukan tanpa dukungan jaringan untuk menyediakan prioritas dan komunikasi yang aman dan andal.

Industri

Industri dapat dianggap sebagai pertambangan, manufaktur, atau utilitas infrastruktur yang menyediakan layanan, seperti listrik, gas, atau bahkan transportasi. Kebutuhan masing-masing industri akan sangat berbeda dan akan mencakup hal-hal ekstrem dalam hal cakupan, ketersediaan, dan bandwidth. Misalnya, operasi pertambangan atau manufaktur tidak dapat terpapar pada pemadaman jaringan, sehingga memerlukan ketersediaan tinggi, sementara pelaporan penggunaan dari meteran dan manajemen perangkat memiliki persyaratan yang sama sekali berbeda. Menerapkan mekanisme ketersediaan yang sama yang diperlukan untuk pertambangan dan manufaktur pada meteran akan menjadi sangat mahal.

Saat ini, beberapa industri ini membuat dan mengoperasikan jaringan khusus mereka sendiri. Ini merupakan cara untuk menyediakan ketersediaan dan keamanan yang dibutuhkan yang tidak dianggap tersedia dalam jaringan pita lebar atau pita lebar seluler pada umumnya saat ini. Jaringan masa depan, untuk memenuhi kebutuhan industri, akan menyediakan ketersediaan dan keamanan yang dibutuhkan industri dengan menyediakan jaringan yang secara logis terisolasi dari jaringan yang mendukung pengguna pribadi dan perusahaan. Tanpa dukungan yang disesuaikan untuk setiap industri, biaya jaringan bagi konsumen IoT, konsumen pribadi, atau konsumen perusahaan tidak akan ekonomis, atau karakteristik jaringan untuk industri lain akan terganggu.

Yang tidak Diketahui

Sebagaimana dijelaskan sebelumnya, jaringan masa depan perlu diperlengkapi untuk menangani kebutuhan pengguna masa depan yang tidak dapat kita prediksi saat ini, dalam hal menangani karakteristik mereka, menyesuaikan penawaran, dan memberikan penawaran kepada mereka dengan cukup cepat. Singkatnya, jaringan masa depan perlu mendukung kelincahan bisnis dan kelincahan jaringan.

5.4 OPERASI JARINGAN MASA DEPAN

Pergeseran Paradigma: Dari Mengelola Jaringan ke Mengelola Otomatisasi

Sebagaimana dijelaskan di bagian sebelumnya, jaringan dulunya berpusat pada satu layanan, dan jaringan tersebut sengaja dibangun untuk mendukung layanan ini; hal yang sama berlaku untuk pengelolaan jaringan tersebut. Semua operasi dan manajemen difokuskan pada penyediaan satu layanan ini bagi pengguna akhir. Seiring berkembangnya jaringan, kompleksitas pengelolaan jaringan ini pun meningkat. Setiap layanan baru yang ditambahkan juga disertai fungsi manajemen dan pemantauan tambahan. Seiring waktu, fungsi-fungsi ini mulai saling mengganggu. Lebih lanjut, mengorkestrasi pengelolaan jaringan secara keseluruhan yang terdiri dari berbagai layanan dan komponen telah menjadi prioritas dalam mengelola keseluruhan jaringan.

Saat ini, kita masih mengelola jaringan dan semua fungsi jaringan secara detail. Artinya, staf operasional penyedia layanan mengelola dan mengonfigurasi fungsi jaringan, jaringan,



dan perangkat secara langsung. Jika kita bandingkan bagaimana staf operasional mengelola jaringan saat ini dengan bagaimana pengaturan suhu di dalam mobil, akan terasa seperti operator mobil yang menyalakan dan mematikan komponen pemanas dan pendingin di dalam mobil, alih-alih mengatur suhu yang diinginkan.

Salah satu langkah untuk menjadikan jaringan lebih cerdas adalah dengan memperkenalkan pembelajaran mesin dan AI ke dalam jaringan untuk menciptakan fungsi-fungsi cerdas yang menangani bagian-bagian jaringan. Fungsi-fungsi berbasis data untuk jaringan akan menyempurnakan solusi yang ada dengan memanfaatkan informasi terkini yang semakin banyak daripada yang mungkin sebelumnya. Hal ini akan memungkinkan pelatihan dan adaptasi fungsi-fungsi ini secara berkelanjutan, sehingga memungkinkan fleksibilitas dalam perilaku jaringan yang belum pernah terlihat sebelumnya. Bahkan, pemecahan masalah yang sebelumnya tidak mungkin dipecahkan pun akan dimungkinkan. Teknologi pembelajaran mesin, seperti hutan acak atau algoritma pembelajaran mendalam dengan jaringan saraf tiruan, dapat memanfaatkan lebih banyak data dan parameter daripada yang dapat dilakukan oleh pakar manusia untuk menguji dan menemukan solusi baru dengan cepat. Hal ini berlaku baik untuk masalah yang sudah ada maupun masalah baru yang akan muncul seiring jaringan menjadi lebih kompleks untuk memenuhi kebutuhan kapabilitas baru. Setiap fungsi yang diperkenalkan dalam jaringan akan beroperasi pada kumpulan data tertentu dan menyelesaikan tugas tertentu.

Seiring meningkatnya otomatisasi untuk mengelola detail jaringan, manajer jaringan akan berperan untuk mengelola tujuan dan target jaringan. Fungsi pintar dalam jaringan akan menyesuaikan diri berdasarkan tugas yang diberikan untuk setiap fungsi. Untuk memelihara jaringan yang semuanya otomatis, penting untuk memahami bagaimana hal ini akan berubah seiring perkembangan jaringan. Seiring semakin banyaknya fungsi pintar yang dibangun ke dalam jaringan, parameter konfigurasi diabstraksikan dari manajemen jaringan. Setiap fungsi dapat dipandang sebagai layanan untuk jaringan Anda; layanan ini akan memiliki peran khusus dari sudut pandang manajemen. Contoh layanan tersebut antara lain optimasi latensi, pemanfaatan bandwidth, pemantauan SLA, prediksi penggunaan, atau pemantauan tautan.

Keluaran dari fungsi-fungsi seperti ini akan memberikan tampilan jaringan terkini. Dengan tampilan jaringan secara real-time ini, aplikasi dapat berinteraksi dengan fungsi pintar untuk membuat prediksi dan menetapkan ekspektasi sebelum eksekusi. Selain itu, setiap layanan akan beroperasi dalam konteks tertentu; konteks ini diberikan dari perspektif global untuk menghindari optimasi lokal yang dapat memengaruhi jaringan dalam skala yang lebih besar. Jika kita kembali membandingkannya dengan kontrol iklim di mobil, ini setara dengan mengatur suhu yang diinginkan di dalam mobil dan membiarkan mobil tersebut mengetahui cara menggunakan komponen pemanas dan pendingin untuk mencapai suhu tersebut, alih-alih menyalakan dan mematikannya secara langsung.

Ketika sebuah aplikasi ingin dijalankan di jaringan, jaminan tertentu dapat diminta dari jaringan, dan fungsi infrastruktur pintar kemudian akan menyediakan sumber daya yang diperlukan agar aplikasi tersebut dapat dijalankan. Jaringan juga bertugas memastikan bahwa aplikasi beroperasi dalam parameter yang telah disepakati sebelumnya. Perilaku ini juga



dipelajari dari waktu ke waktu dan dapat menjadi mekanisme umpan balik yang penting bagi layanan dan krusial untuk memahami dampak interaksi aplikasi terhadap fungsionalitas jaringan. Abstraksi juga merupakan bagian dari peralihan dari pengelolaan jaringan ke pengelolaan otomatisasi. Pada berbagai tingkatan dalam aplikasi jaringan, kebutuhan dijelaskan dengan berbagai tingkatan detail.

Ini bisa berupa persyaratan, seperti latensi, bandwidth, atau parameter lain yang relevan untuk perjanjian tingkat layanan (SLA) layanan. Semakin akurat deskripsinya, semakin akurat pula jaringan akan mampu beradaptasi dengan layanan dan aplikasi. Karena layanan dan aplikasi juga akan lebih mudah diamati, maka perilaku ini dan persyaratan selanjutnya dapat dipelajari seiring waktu. Ini juga akan memengaruhi bagaimana jaringan dibangun, kemungkinan pengaturan atau parameter yang dapat dimanipulasi oleh operator jaringan akan berkurang, atau lebih tepatnya ini akan diabstraksikan juga. Karena lebih banyak informasi tersedia untuk layanan yang berjalan di atas jaringan, akan lebih mudah untuk menyetel jaringan agar sesuai dengan layanan ini. Dengan cara yang sama, akan memungkinkan juga untuk mempelajari bagaimana pengaturan dan parameter bekerja bersama dengan campuran layanan tertentu yang perlu didukung. Setelah ini diketahui, akan menjadi langkah alami selanjutnya untuk mengabstraksikan metode untuk menyetel jaringan untuk layanan yang berbeda dengan mengelompokkan fitur-fitur ini dan menawarkan pandangan abstrak tentang cara mengoptimalkan.

Dalam contoh sederhana, Layanan A memerlukan sumber daya tambahan. Untuk meningkatkan sumber daya, parameter X, Y, dan Z perlu diubah. Operator hanya perlu memutuskan untuk meningkatkan sumber daya untuk A, dan jaringan akan mencari cara untuk mengubah X, Y, dan Z agar dapat memenuhi kebutuhan tersebut. Di sini, layanan dan aplikasi juga dapat dipantau untuk memastikan mereka mematuhi aturan. Ketika suatu layanan berperilaku tidak semestinya, layanan tersebut dapat dideteksi dan dilaporkan kepada penyedia layanan sebagai masalah. Penyelesaiannya akan bergantung pada layanan dan jenis masalah yang muncul.

Perspektif Pengguna

Pengguna saat ini mulai terbiasa mengakses layanan apa pun dari perangkat apa pun, kapan pun. Perspektif pengguna sederhana: semuanya harus berfungsi sepanjang waktu. Hal ini juga berlaku untuk jaringan otonom; namun, layanan dan aplikasi yang akan digunakan pengguna pasti akan memberikan beban lebih pada jaringan yang bertugas menyediakan semua ini kepada pengguna, baik konsumen pribadi maupun pelanggan perusahaan. Aplikasi seperti video 3D, aplikasi realitas virtual dan campuran, serta telepresensi yang tersedia setiap saat dan dari mana saja, menghadirkan persyaratan baru.

Dikombinasikan dengan segmen perangkat IoT yang terus berkembang, jaringan otonom harus mampu menyediakan layanan kepada semua pengguna. Pengguna juga bergantung pada penyedia akses mereka untuk mengamankan layanan mereka. Hal ini juga akan menimbulkan persyaratan baru bagi operator jaringan. Serupa dengan perbankan, nasabah tidak ingin mengubah perilaku dan bergantung pada bank untuk melindungi mereka dari penipuan dan ancaman keamanan lainnya. Persyaratan yang sama akan diberlakukan



pada penyedia jaringan untuk memastikan jaringan yang tepercaya dan aman secara lancar. Pelanggan perusahaan cenderung menerapkan SLA yang lebih ketat saat berlangganan layanan jaringan, tetapi ekspektasinya akan sama untuk layanan apa pun, kapan pun, dan di mana pun.

Mengelola Layanan

Dengan transformasi ke jaringan otonom, staf operasional akan beralih dari mengelola elemen jaringan ke mengelola otomatisasi. Aspek penting dari pergeseran ini adalah memfokuskan layanan yang disediakan dan juga bagaimana hal itu memengaruhi pengelolaan layanan tersebut. Layanan memberikan nilai kepada konsumen dan dapat dijelaskan dengan sejumlah atribut. Atribut-atribut ini dapat diterjemahkan menjadi indikator kinerja, seperti latensi atau penundaan, waktu pembentukan layanan, ketersediaan, konsistensi data, dan pengaturan layanan. Tugas staf operasional penyedia layanan akan bergerak untuk memastikan bahwa indikator kinerja yang dijelaskan dapat dipenuhi sekaligus mengonsumsi sumber daya seminimal mungkin, seperti energi dan transmisi.

Hal ini harus dicapai melalui seluruh siklus hidup sumber daya yang digunakan dalam jaringan untuk menyediakan layanan. Lebih lanjut, AI memiliki kemampuan prediktif, yang memungkinkan antisipasi kebutuhan layanan di masa mendatang. Hal ini memungkinkan operator menggunakannya untuk merancang dan menyetujui terlebih dahulu layanan yang mungkin diperlukan di masa mendatang. Saat ini, fokusnya adalah mengelola jaringan, yang berarti mungkin menggunakan serangkaian indikator kinerja utama (KPI) yang berbeda untuk memastikan SLA pelanggan terpenuhi. Dengan perubahan ke manajemen yang berfokus pada layanan, serangkaian KPI yang berbeda akan dipantau, dan dengan fungsi kontrol jaringan pintar yang berinteraksi lebih baik dengan fungsi pintar dalam jaringan, teknologi pembelajaran mesin akan memberikan wawasan yang bermanfaat tentang bagaimana data dan lalu lintas dapat ditangani untuk mencapai kepuasan pelanggan terbaik.

5.5 TANTANGAN JARINGAN MASA DEPAN

Semuanya tentang Data

Jaringan saat ini dibangun dari banyak komponen, mulai dari transportasi hingga cloud, hingga jaringan, manajemen jaringan, dan manajemen bisnis. Baik jaringan tersebut berupa entitas fisik maupun fungsi virtual, jaringan perlu beroperasi dengan data. Untuk menciptakan jaringan cerdas yang benar-benar otomatis, data ini harus tersedia agar fungsi cerdas dapat beroperasi. Data akan berada di berbagai tingkatan dalam jaringan. Data yang kami maksud adalah segala jenis data yang dapat diproduksi dan diekspos, seperti (namun tidak terbatas pada) telemetri perangkat, instrumentasi, statistik *Simple Network Management Protocol* (SNMP), log, status, konfigurasi terkini, alarm, data pengguna, Administrasi dan Manajemen Operasi (OAM), statistik pengguna, informasi perangkat, jenis langganan, kecepatan data, dan data cuaca. Dengan meningkatnya jumlah fungsi cerdas dalam jaringan, data apa pun yang dapat digunakan untuk mengoptimalkan fungsi jaringan akan digunakan sebelum diteruskan atau dibuang. Pada akhirnya, sebagian besar data ini mungkin berakhir di pusat data cloud atau semacam danau data.



Di sini, data dapat digunakan untuk pembelajaran offline guna menciptakan fungsi baru, tugas pembelajaran jangka panjang, pelaporan tren, dll. Hal ini mungkin tampak jelas bagi pembaca buku ini atau siapa pun yang memiliki latar belakang ilmu data, tetapi tanpa data, tidak akan ada pembelajaran dari data tersebut. Secara tradisional, data dari node dan fungsi dalam jaringan tetap berada di dalam node ini karena biaya pengambilan sampel, pencatatan, dan pemaparan data memerlukan sumber daya tambahan yang utamanya digunakan untuk pemecahan masalah ketika terjadi masalah dalam jaringan. Hal ini seringkali membuat operator jaringan tidak memiliki informasi ketika terjadi masalah karena tidak ada log, tidak ada data yang tercatat tentang apa yang terjadi di jaringan saat masalah terjadi. Saat merancang fungsi baru, ketersediaan data harus menjadi salah satu fungsi utama untuk setiap node ini. Jaringan tidak hanya harus dirancang untuk memungkinkan pemaparan data, tetapi juga perlu untuk dapat menginstrumentasikan bagaimana data dihasilkan untuk setiap fungsi ini. Hal ini akan membuat komponen lebih tahan terhadap masa depan daripada sebelumnya. Untuk mencapai hal ini, paparan data harus terintegrasi ke dalam setiap fungsi yang dirancang untuk jaringan dan dialirkan secara real-time. Membatasi pergerakan data diperlukan untuk menghindari kelebihan beban jaringan.

Dengan mengizinkan operator jaringan mengelola data, fokus pada parameter, fungsi, atau layanan tertentu pada waktu yang berbeda akan dimungkinkan. Biasanya, hal ini akan membantu dalam skenario pemecahan masalah atau saat melatih ulang model data. Data yang dihasilkan juga harus diproduksi tepat satu kali dan dibaca oleh banyak layanan atau fungsi sesuai kebutuhan. Dengan kemampuan untuk menginstrumentasikan pengumpulan data sebelum data dipublikasikan, manipulasi data selama pembuatannya juga dimungkinkan. Misalnya, seseorang mungkin ingin mengagregasi jenis data tertentu sedekat mungkin dengan sumbernya. Fungsi lain berpotensi ditambahkan langsung ke sumber data itu sendiri, seperti filter pintar, fungsi, atau model keputusan.

Sebagaimana telah disebutkan, saat mendesain ulang atau membuat elemen jaringan baru, instrumen tentang bagaimana data dibuat dan diekspos di dalam dan di luar fungsi jaringan harus dimungkinkan. Hal ini memiliki manfaat tambahan yaitu kumpulan data baru dapat dibuat sesuai permintaan. Karena data dikaitkan dengan biaya, hanya data yang bermanfaat yang boleh dibuat dan dipelihara. Namun, untuk mencoba fungsi baru, memecahkan masalah, atau meningkatkan pemahaman tentang peristiwa dalam jaringan, data tersebut akan menjadi alat yang ampuh untuk dapat secara fleksibel menghasilkan data untuk penelitian, diagnostik, atau peristiwa pemicu lainnya. Data juga dapat menjadi parameter siklus hidup untuk fungsi pintar di mana data diaktifkan dan dikumpulkan untuk pembelajaran selama periode tertentu, hanya jika kinerja model saat ini mulai menurun.

Lanskap Teknologi yang Berubah

Kecerdasan buatan, yang saat ini diwujudkan melalui ilmu data dan pembelajaran mesin, merupakan salah satu bidang teknologi yang paling pesat perkembangannya, memungkinkan metode, kerangka kerja, dan produk baru dirilis setiap minggu. Banyak dari metode ini yang diabaikan secepat proses produksinya. Namun, setelah beberapa kali diguncang dan dimatangkan, beberapa di antaranya menemukan jalannya ke dalam kerangka



kerja yang menjadi perangkat sehari-hari para ilmuwan data. Di jaringan masa depan, kita akan melihat evolusi fungsi-fungsi cerdas yang setara, yang entah akan dibantah atau berevolusi menjadi bagian integral dari operasi harian jaringan. Sebagai hasil dari perubahan ini, jaringan yang lebih kuat akan dihasilkan.

Salah satu aspek krusial untuk meningkatkan kekuatan jaringan adalah kemampuan untuk memanfaatkan data yang dihasilkan jaringan. Ini berarti lebih banyak sumber daya komputasi dan lebih banyak memori di seluruh jaringan. Meskipun hukum Moore tampaknya telah melambat dalam hal siklus clock mentah per inti, jumlah inti meningkat dengan laju yang spektakuler. Hukum Moore masih berlaku untuk memori, yang berarti harga memori berkurang setengahnya setiap delapan belas bulan atau lebih. Kombinasi peningkatan inti dan memori pada akhirnya menghasilkan daya komputasi yang lebih besar, meskipun kompleksitas komputasi meningkat dengan komputasi terdistribusi, algoritma yang lebih canggih, dan peningkatan kompleksitas masalah yang harus dipecahkan.

Selain komputasi tradisional dengan peningkatan inti, siklus clock, dan memori, perangkat keras yang lebih khusus digunakan untuk jenis komputasi ini. Unit Pemrosesan Grafis (GPU) menjadi semakin populer untuk aplikasi pembelajaran mendalam (deep-learning). Unit Pemrosesan Tensor (*Tensor Processing Unit*) yang dibuat khusus dari Google kini tersedia, IBM merilis rangkaian chipset pemrosesan jaringan saraf (*neural network*) SyNAPSE mereka, dan Qualcomm telah menempatkan *Neural Processing Engine* (NEE) di platform seluler Snapdragon terbaru mereka. Ada biaya yang terkait dengan peningkatan kapabilitas dalam jaringan. Seiring dengan penurunan biaya perangkat keras ini, lebih banyak layanan dapat ditangani dengan lebih baik, dan ekosistem pada akhirnya akan menyeimbangkan peningkatan investasi yang dibutuhkan.

Tantangan Manusia

Ada sejumlah tantangan di bidang jaringan otonom dan AI, baik yang bersifat teknis maupun non-teknis. Insinyur jaringan tidak mempercayai kotak ajaib yang akan menyelesaikan sebagian besar masalah mereka dengan lebih baik, lebih cepat, dan lebih murah daripada yang pernah mereka bisa. Upaya untuk menerima metode dan model tersebut mencakup meyakinkan insinyur berpengalaman bahwa mesin dan algoritma dapat mempelajari dan mengembangkan solusi lebih cepat dan lebih baik daripada yang mereka bisa. Lebih lanjut, kompetensi insinyur yang tersisa harus bergeser dari pemahaman jaringan menjadi pemahaman algoritma dan alat yang digunakan untuk mencapai pemeliharaan tanpa sentuhan.

Selain itu, generasi baru insinyur jaringan membutuhkan keahlian dalam kerangka kerja Antarmuka Pemrograman Aplikasi (API), platform orkestrasi, dan bagaimana penerapan layanan dan aplikasi operasi sistem berinteraksi dengan komponen fisik dan virtual aktual dalam jaringan. Area lain yang akan semakin sulit adalah pemecahan masalah. Meskipun penggunaan teknik pembelajaran mesin tingkat lanjut, seperti jaringan saraf, dapat sangat membantu dalam mengotomatiskan fungsi jaringan tertentu, memahami keluaran dan interpretasi antara dari fungsi ini serta mengapa fungsi ini gagal dalam kasus tertentu bisa sangat sulit dipahami.



Untuk mengatasi tantangan memahami cara-cara baru dalam menyajikan hasil, prediksi dan model perlu dikembangkan. Serupa dengan kemudahan pemahaman di sektor lain, informasi perlu disajikan sedemikian rupa sehingga pengguna sistem dapat dengan mudah memahaminya. Sebagaimana dijelaskan sebelumnya, pengguna harus memiliki pemahaman yang memadai tentang alat dan metode yang digunakan. Orang-orang yang bekerja di bidang manajemen jaringan saat ini harus mempelajari teknologi-teknologi baru ini agar tetap relevan. Orang-orang berpengalaman di bidang manajemen jaringan dapat memainkan peran penting dalam membantu mengembangkan dan menyempurnakan metode-metode otomatis untuk operasi jaringan otonom. Tentu saja, tidak semua orang akan mampu mengembangkan fungsionalitas baru dalam lingkungan seperti ini; hal ini membutuhkan orang-orang yang sangat terampil di bidang ilmu data. Akan ada pula spesialisasi tingkat tinggi di berbagai bidang dan aspek ilmu data, yang dengan sendirinya akan menambah jalur spesialisasi baru untuk menghubungkan beberapa metode yang beragam ini. Ilmuwan data harus bekerja sama dengan para ahli di bidangnya (misalnya, operator jaringan berpengalaman) dan pakar bisnis untuk menciptakan nilai maksimal bagi fungsi-fungsi baru bagi jaringan.

5.6 APLIKASI JARINGAN OTONOM

Berikut ini adalah contoh aplikasi yang dibayangkan oleh jaringan otonom.

Operasi Tanpa Sentuhan

Istilah operasi tanpa sentuhan cukup deskriptif. Artinya, jaringan dioperasikan tanpa campur tangan manusia. Namun, ekspektasi terhadap jaringan dapat bervariasi seiring waktu (misalnya, mengoptimalkan kapasitas atau mengoptimalkan jangkauan), dan layanan dapat diperkenalkan; oleh karena itu, staf operasional penyedia layanan mengelola jaringan melalui pengelolaan KPI yang diperlukan, alih-alih mengelola peralatan jaringan itu sendiri.



Gambar 5.2 Operasi zero-touch.



Di dalam jaringan (mencakup sistem pendukung operasional serta fungsi dan perangkat jaringan), kontrol loop tertutup beroperasi. Kontrol loop tertutup melibatkan pengambilan data dari jaringan, menganalisis data untuk memberikan wawasan tentang status jaringan guna mendukung layanan, menggunakan kebijakan untuk memutuskan tindakan yang diperlukan, dan menjalankan tindakan yang diperlukan melalui orkestrasi. Kecerdasan buatan dapat mendukung seluruh kontrol loop tertutup. Kontrol loop tertutup terus memastikan terpenuhinya perjanjian layanan sambil mengoptimalkan jaringan. Jaringan dapat mengidentifikasi beberapa cara untuk mengoptimalkan konfigurasinya agar memenuhi harapan pengguna akhir, bahkan jika terjadi insiden jaringan. Keputusan tentang alternatif terbaik yang telah diidentifikasi untuk diikuti akan didasarkan pada batasan (apa yang diizinkan) dan riwayat yang dipelajari (yaitu, tindakan mana yang paling mungkin menyelesaikan situasi tertentu). Hal ini diilustrasikan pada Gambar 5.2.

Dalam kasus insiden jaringan, suboptimasi, atau pelanggaran SLA, staf operasional diinformasikan dan diberikan saran tentang cara menanganinya. Ini dapat mencakup penyesuaian kebijakan dalam jaringan atau pembuatan perintah kerja untuk memperbaiki insiden fisik. Salah satu contohnya adalah skenario berikut:

Perusahaan 1 telah memesan konektivitas antara dua gedung, gedung A1 dan gedung A2. Demikian pula, Perusahaan 2 telah memesan konektivitas antara dua gedung, gedung B1 dan gedung B2. Sifat perjanjiannya adalah bahwa setelah waktu pemadaman yang disepakati, penalti finansial akan berlaku. Jika terjadi insiden jaringan yang menyiratkan bahwa kapasitas transportasi tidak mencukupi untuk melayani Perusahaan 1 dan Perusahaan 2, pendekatan klasiknya adalah memprogram strategi melalui kebijakan ke dalam jaringan. Strategi ini dapat berupa, misalnya, memprioritaskan lalu lintas untuk perusahaan yang paling dekat dengan pelanggaran SLA. Hal ini juga dapat berimplikasi jika dilakukan secara tidak benar karena konfigurasi jaringan yang terus-menerus berubah. Namun, dengan jaringan otonom, jaringan mempelajari bahwa pada saat insiden jaringan terjadi, Perusahaan A biasanya menggunakan kapasitas kurang dari yang dikontrak, sehingga memprioritaskan kapasitas untuk Perusahaan 2, sehingga kedua pelanggan tetap puas.

Mengoptimalkan Pengalaman Pengguna Akhir

Kasus khusus dari operasi zero-touch adalah dukungan pengenalan suara atau kapasitas pemrosesan video untuk pengguna akhir. Karena komputasinya intensif, ponsel umumnya bergantung pada dukungan jaringan, dan waktu respons dapat memengaruhi pengalaman pengguna secara keseluruhan. Mengalokasikan kapasitas pemrosesan tersebut di tepi jaringan sepanjang waktu dapat menjadi mahal, khususnya untuk jaringan seluler. Namun, jaringan otonom dapat mempelajari perilaku pengguna. Jaringan ini dapat mempelajari kapan dan di mana pengguna mengandalkan dukungan jaringan tersebut dan mengalokasikan fungsi serta kapasitas di lokasi tepi yang tepat yang paling membutuhkannya. Jaringan otonom juga dapat memprediksi kapan dan di mana pengguna akan membutuhkan dukungan ini.

Deteksi Intrusi

Keamanan jaringan masa depan akan terus menjadi perhatian besar. Jaringan otonom akan memahami perilaku umum pengguna dan jaringan pada waktu, minggu, bulan, tahun,



lokasi, dan bagaimana perilaku tersebut dipengaruhi oleh faktor eksternal seperti cuaca. Perubahan mendadak pada perilaku pengguna atau jaringan dapat menjadi indikasi penipuan atau intrusi. Sebagaimana dijelaskan sebelumnya, kemampuan observasi jaringan di masa mendatang akan memungkinkan operator jaringan untuk mendeteksi kapan perubahan tersebut terjadi. Jaringan kemudian dapat mengidentifikasi perubahan perilaku dan menganalisisnya untuk deteksi intrusi, mengambil tindakan untuk memberi tahu pengguna, staf operasional, serta mengidentifikasi dan mengisolasi perangkat atau peralatan jaringan yang mencurigakan.

5.7 RINGKASAN

Jaringan di masa mendatang akan cerdas. Artinya, jaringan dapat belajar dan beradaptasi dengan situasi yang tidak diketahui untuk membentuk sistem otonom. Hal ini membutuhkan fungsi cerdas di semua tingkat jaringan dengan pengumpulan data sesuai permintaan untuk mendukung diagnostik dan operasi loop tertutup. Hal ini akan mengubah operasi jaringan dari mengelola jaringan menjadi mengelola otomatisasi. Potensinya sangat menarik, dan terdapat sejumlah tantangan yang harus diatasi, seperti akses data, pelatihan yang optimal, dan transisi manusia. Kecerdasan buatan merupakan aspek fundamental dari kecerdasan.



BAB 6

OPEN SOURCE DAN CLOUD DI JARINGAN OTONOM

6.1 ZEN OF OPEN SOURCE

“Kami mengeksplorasi pendekatan yang berbeda dan lebih ‘terbuka’: sistem ini dianggap menawarkan berbagai fasilitas, yang mana pun dapat ditolak, diterima, dimodifikasi, atau diperluas oleh pengguna....

Seperti yang disebutkan dalam *“A Brief History of the Internet”* yang diterbitkan oleh Internet Society pada tahun 2003, Robert Kahn memperkenalkan gagasan jaringan arsitektur terbuka pada tahun 1972: “Jaringan individual dapat dirancang dan dikembangkan secara terpisah dan masing-masing dapat memiliki antarmuka uniknya sendiri yang dapat ditawarkan kepada pengguna dan/atau penyedia lain, termasuk penyedia Internet lainnya....

Dengan arah yang jelas menuju jaringan otonom yang digerakkan oleh kecerdasan buatan (AI), paradigma terbuka wajib diadopsi. Agar berbagai tantangan dapat diatasi, setiap inovasi disambut baik.

Definisi dan Motivasi

Istilah "perangkat lunak sumber terbuka" (OSS) berasal dari Pedoman Perangkat Lunak Bebas Debian. Pedoman ini disusun sebagai seperangkat prinsip yang mengatur perizinan dan distribusinya. Kode sumbernya disertakan, dan siapa pun bebas meretasnya. OSS tidak bergantung pada afiliasi vendor atau produk dan bersifat independen terhadap teknologi. Etos OSS adalah meretas kode dan membagikannya. Perangkat lunak sumber terbuka mewarisi cita-cita yang awalnya digagas oleh Lampson, Kahn, dkk. bertahun-tahun yang lalu.

Dari cita-cita altruistik awal tersebut, OSS yang ada saat ini telah menjadi lebih terorganisir. Awal mulanya yang bebas telah diorganisasikan ke dalam proyek-proyek, dengan spektrum struktur tata kelola yang mendasarinya. Dalam konteks ini, kesuksesan diukur dari ukuran dan keragaman pembuat dan pengelola kode sumber (disebut sebagai komiter) dan peserta (korporasi, akademisi, individu, dan lainnya) yang secara kolektif menggemballakan sebuah proyek dari awal hingga rilis. Dalam kasus terbaik, hasil proyek muncul dalam bentuk produk dan layanan serta mendorong pembentukan ekosistem yang lebih besar.

Perangkat lunak sumber terbuka juga telah menjadi alat strategis perusahaan dalam beberapa cara, yang digunakan untuk menggerakkan industri atau meningkatkan posisi kompetitif. Dalam hal ini, perangkat lunak sumber terbuka sering kali mendemokratisasi komponen bernilai rendah dari suatu produk atau ruang solusi untuk menciptakan peluang baru atau mengisolasi kekayaan intelektual dan memungkinkan partisipasi yang lebih luas.

Sumber Terbuka Versus Sumber Privat

Perangkat lunak sumber privat didistribusikan berdasarkan ketentuan kontrak bilateral antara penyedia perangkat lunak dan pengguna perangkat lunak. Pengguna akan membayar penyedia untuk lisensi penggunaan perangkat lunak sesuai ketentuan yang tercantum dalam kontrak tersebut. Perangkat lunak sumber privat tidak dibagikan kepada publik (kecuali versi



uji coba dengan fungsi terbatas), dan kode sumbernya tidak dapat diakses. Komitmen peta jalan untuk pembaruan kode fitur baru sepenuhnya merupakan wewenang satu penyedia.

Perangkat lunak sumber privat sebagian besar memiliki keunggulan kegunaan, termasuk dokumentasi daring dan antarmuka pengguna (UI) yang dirancang dengan baik dan intuitif untuk interaksi manusia. Dalam banyak kasus, perangkat lunak ini menjalani pengujian yang ketat dan biasanya dirilis setelah pengguna beta mengatasi sebagian besar bug.

Fungsi atau algoritma proprietary yang dimiliki oleh suatu organisasi, pemegang paten, atau institusi akademis umumnya tidak dimasukkan ke dalam rilis sumber terbuka. Badan keamanan militer dan pemerintah yang mengoperasikan platform keamanan siber memiliki justifikasi untuk menjaga privasi perangkat lunak. Penyedia perangkat lunak privat dapat menawarkan perangkat pengembangan perangkat lunak (SDK) atau antarmuka pemrograman aplikasi (API) terdokumentasi yang mengekspos platform mereka ke aplikasi atau pengembang pihak ketiga. Meskipun pengguna/ekosistem tidak dapat berkontribusi pada Protokol Internet (IP) inti, mereka dapat memberikan nilai tambah di sekitarnya.

Kontributor dan Pengguna

Kami mendefinisikan peran komiter kode sumber terbuka. Komiter membuat, memperbaiki, dan memelihara kode sumber. Salah satu ukuran kesuksesan adalah memiliki komunitas yang besar, aktif, dan dinamis. Namun, profil komiter kode sumber terbuka berbeda-beda, tergantung pada aspirasi karier dan posisi di perusahaan mereka.

"Koder bebas" adalah individu atau kelompok pengembang, sukarelawan, penghobi, mahasiswa, dan sejenisnya yang membuat potongan kode lalu mengunggahnya ke repositori Github agar tersedia bagi pengguna dan komiter. Kesuksesan mereka dapat mencakup kepuasan diri, nilai bagus di kelas, penghargaan hackathon, dan bahkan penyebutan dalam salah satu dari banyak daftar "kode sumber terbuka terbaik untuk" yang sering muncul. Berdasarkan pengalaman dan pengetahuan yang mereka peroleh, bukan hal yang aneh bagi koder bebas untuk bergabung dengan jajaran insinyur perangkat lunak berbakat yang direkrut oleh organisasi besar yang mengerjakan proyek sumber terbuka.

"Pelatun terorganisir" adalah kategori lain dari komiter kode sumber terbuka. Mereka berspesialisasi dalam satu atau lebih disiplin ilmu yang didukung oleh organisasi dengan sumber daya (misalnya, vendor, penyedia layanan (SP), perusahaan). Platoon terdiri dari beberapa atau semua hal berikut: pembuat kode, arsitek perangkat lunak, arsitek jaringan, insinyur perangkat keras, sysadmin, personel lab, insinyur pengujian, pekerja magang, desainer, manajemen, dan pemimpin dalam organisasi. Platoon berkembang pesat dalam lingkungan di mana pengambilan risiko, kegagalan, mencoba lagi, dan menghasilkan inovasi dihargai.

Pleton biasanya berada di perusahaan besar yang merupakan katalisator besar proyek sumber terbuka. Buktinya terlihat dalam daftar keanggotaan yayasan, statistik kontributor GitHub, dan persyaratan keterampilan papan pekerjaan. Banyak yang berbicara terbuka tentang penggunaan sumber terbuka mereka. Wajar untuk menyebutkan bahwa sebagian besar kode perusahaan besar untuk alat internal dan layanan eksternal terdiri dari sejumlah perangkat lunak sumber terbuka.



Agar jaringan otonom yang digerakkan oleh AI berhasil, kontribusi dan inovasi dari komunitas riset akan dibutuhkan. Penyedia layanan akan diminta untuk menerapkan infrastruktur dan layanan. Perusahaan yang ingin memperluas kehadiran pasar mereka akan didorong untuk menjadi kontributor aktif dalam pengembangan sumber terbuka. Lembaga pemerintah akan berperan, salah satu contohnya adalah situs Portal Pengembangan Aplikasi Sumber Terbuka (OSADP) milik Departemen Perhubungan AS (USDOT) dan Administrasi Jalan Raya Federal. OSADP adalah portal berbasis web yang menyediakan akses ke kode sumber dan aplikasi Sistem Transportasi Cerdas dengan tujuan memfasilitasi kolaborasi komunitas dalam proyek-proyek yang disponsori oleh USDOT.

Peran Konsorsium dan Ekosistem

Konsorsium, yayasan, dan ekosistem telah dan akan terus memainkan peran penting dalam kesuksesan perangkat lunak sumber terbuka. Mereka menyelenggarakan situs web, menjadi sponsor bersama konferensi, menyediakan edukasi, dan mendorong partisipasi komunitas maupun individu. Yayasan dapat membantu memulai dan memvalidasi proyek serta menjadi tonggak penting dalam perjalanan menuju penerimaan industri yang lebih luas. Proyek yang berafiliasi dengan yayasan memiliki tingkat prestise tertentu.

Apache Software Foundation telah berdiri selama hampir 20 tahun dan identik dengan perangkat lunak sumber terbuka. "Apache Way" didasarkan pada tiga prinsip: partisipasi individu, pengembangan proyek yang transparan dan terbuka, serta komunitas. Prinsip ini telah menjadi model bagi upaya kolaboratif di seluruh industri selama bertahun-tahun. Apache menawarkan lisensi, dan ekosistem perangkat lunak sumber terbukanya telah diterapkan di seluruh internet. Pengaruh dan kontribusi mereka tidak dapat dilebih-lebihkan. *Linux Foundation* (LF) didirikan pada tahun 2000 untuk menjadi rumah bagi OS Linux (termasuk Linus Torvalds dan kepala pengelola Greg Kroah-Hartman). Yayasan ini telah berkembang dengan baik, pertama karena asosiasi nama dan dalam beberapa tahun terakhir menjadi situs "tujuan" untuk semua hal yang berkaitan dengan sumber terbuka. Memang, dilihat dari jumlah proyek yang dihostingnya, jelas bahwa industri memandang LF sebagai lokasi yang diinginkan untuk inkubasi, pertumbuhan, akreditasi, dan rilis proyek sumber terbuka.

Namun, pertumbuhan tersebut menimbulkan tantangan, khususnya bagaimana LF dapat berinteraksi sebaik-baiknya dengan organisasi pengembangan standar (SDO). LF menerbitkan sebuah white paper yang membahas perlunya harmonisasi antara keduanya. White paper tersebut menunjukkan bahwa industri sekarang digerakkan oleh perangkat lunak operator merancang dan menerapkan layanan baru tanpa ketergantungan pada perangkat keras khusus. Adopsi teknologi bergeser dari siklus persyaratan, arsitektur, standar, dan implementasi klasik dan lebih mengarah pada platform referensi, arsitektur, dan kasus penggunaan. Tantangannya bermuara pada fakta bahwa proses pengembangan standar dan arsitektur berbasis konsensus memakan waktu. Definisi keterbukaan sedang berubah alih alih multivendor, kita memiliki vendor netral dengan nilai tambah yang diwujudkan melalui orkestrasi dan kontrol perangkat lunak. Lebih banyak komunikasi, baik informal maupun formal, mengorganisir proyek multi-SDO/OSS dan metodologi adopsi teknologi yang direvisi yang memadukan standar, sumber terbuka, kasus penggunaan, dan kontribusi vendor.



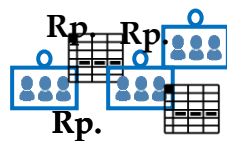
Penyempurnaan tanggung jawab dan peran yang dijalankan oleh pihak-pihak yang berpartisipasi dapat membantu. Menggabungkan kontribusi kode, alokasi pengeluaran iuran, dan mengelola pemasaran proyek serta aktivitas dukungan ke dalam satu wadah besar dapat menjadi tantangan. Idealnya, proyek dapat diatur sedemikian rupa sehingga anggota teknis berfokus pada kontribusi kode, dan pakar bisnis menangani pengeluaran dan dukungan pemasaran. Solusi yang ditunjukkan pada Gambar 6.1 menggabungkan sekelompok proyek terkait di bawah satu proyek payung jaringan yang seragam. Tata kelola teknis masing-masing komunitas proyek tidak akan berubah. Keputusan bisnis akan ditangani oleh yayasan dan dewannya di belakang layar. Untuk mencapai hal ini, LF membentuk sebuah proyek bernama Linux Foundation Networking Fund. Anggota awal proyek ini adalah *Fast Data Project* (FD.io), *OpenDaylight* (ODL), *Open Network Automation Platform* (ONAP), *Open Network Function Virtualization* (OPNFV), *Platform for Network Data Analytics* (PNDA), dan *Streaming Network Analytics System* (SNAS).

Standar

Organisasi pengembangan standar seperti *Internet Engineering Task Force* (IETF), *Third Generation Partnership Project*, *Institute of Electrical and Electronics Engineers*, dan *European Telecommunications Standards Institute* (ETSI) telah ada selama beberapa dekade. Tidak seperti proyek OSS, SDO tidak selalu menghasilkan kode.

MASALAH SAAT INI

- Banyak proyek jaringan
- Overhead tinggi
- Total pengeluaran yang besar
- Banyak "peristiwa" tahunan



SOLUSI: Dana Jaringan Yayasan Linux

- Tata kelola teknis komunitas-komunitas tersebut tetap berubah
- Satu yayasan dan dewan yang membuat keputusan bisnis di belakang layar
- Biaya operasi & pemasaran yang lebih rendah



Gambar 6.1 Solusi dana jaringan Linux Foundation.

Kontribusi mereka meliputi arsitektur, definisi protokol, interoperabilitas, kepatuhan produk, dan membangun koalisi vendor/operator. Bukti keberhasilan SDO terlihat jelas dari umur panjang mereka dan DNA internet serta komunikasi data. Terdapat ketidaksesuaian kecepatan antara komunitas SDO dan OSS. OSS bergerak dengan kecepatan yang didasarkan pada pengembangan kode dan keterlibatan komiter proyek. Sebaliknya, kerangka waktu SDO (misalnya *Internet Engineering Task Force* [IETF]) diperpanjang. Prosesnya terdiri dari draf teks tertulis, diikuti dengan diskusi milis dan pertemuan tiga kali setahun.

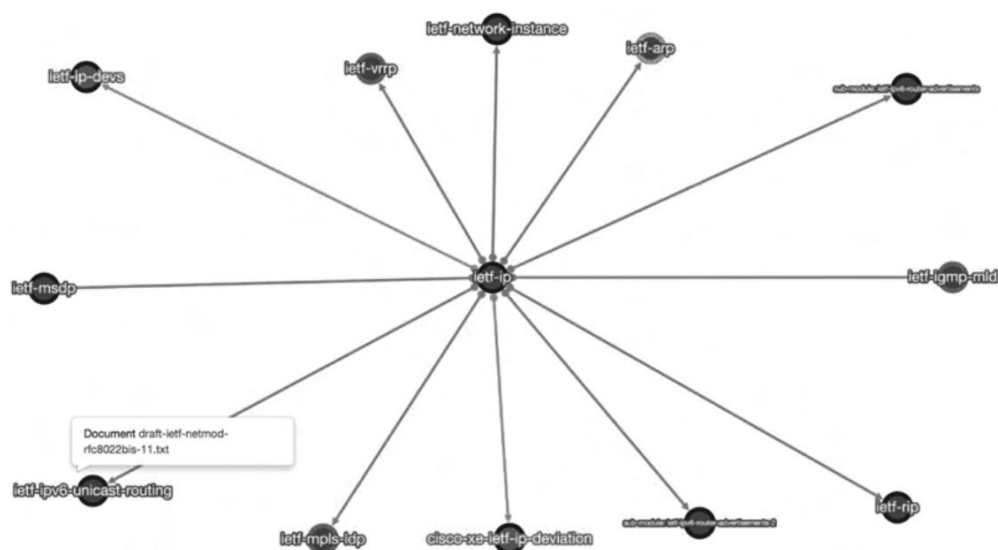
Birds of a Feather adalah pertemuan untuk berdebat dan memutuskan apakah perlu ada pertemuan kelompok formal untuk berdebat dan memutuskan. Hasil akhirnya adalah



spesifikasi yang dipublikasikan yang berpotensi mengarah pada penerapan dan operasionalisasi produk. Pada gilirannya, spesifikasi baru atau yang dimodifikasi dihasilkan dan prosesnya diulang. Siklus umpan balik ini (contoh yang baik adalah Transport Area Working Group) dapat berlangsung selama bertahun-tahun. Di sisi lain, otoritas di balik OSS berada di komunitas proyek dan kode yang berjalan bukan spesifikasi yang dipublikasikan atau struktur kelompok kerja formal.

Telah ada kemajuan tertentu yang dibuat di bidang SDO untuk merujuk, memanfaatkan, dan mempromosikan sumber terbuka dengan lebih baik. Referensi normatif disediakan dalam Permintaan Komentar (RFC) IETF yang merujuk pada standar lain. Dalam skenario baru ini, RFC dapat memberikan referensi normatif ke proyek OSS. Dokumentasi OSS termasuk kualitas dan keterbacaan kode akan bervariasi antar proyek dan mungkin tidak memenuhi standar referensi normatif yang digunakan oleh IETF.

Perkembangan terbaru adalah rilis YANGCATALOG. Ini adalah katalog dan registri model yang memungkinkan pengguna untuk menemukan dan menganalisis penerapan model YANG pada kasus penggunaan spesifik mereka. Aplikasi ini ditargetkan untuk beberapa audiens. Yang pertama adalah pengguna modul YANG yang ingin mencari model, memeriksa ulang metadata (misalnya, organisasi, kematangan, jumlah impor, implementasi, layanan atau elemen jaringan modul YANG, tipe pohon IETF atau OpenConfig), melakukan analisis dampak model dengan dependensi, dan menghasilkan kode Python dari modul YANG. Target kedua adalah perancang modul YANG dengan fungsi yang menawarkan validator regex YANG untuk bereksperimen dengan pernyataan pola W3C. Ini juga mencakup server *Network Configuration Protocol* (NETCONF) dan *Representational State Transfer* (REST) untuk mengambil model untuk berbagai tujuan, termasuk tetapi tidak terbatas pada inspeksi, pengujian, dan pengembangan pengetahuan.



Gambar 6.2 Contoh keluaran dari alat analisis dampak YANG.

Gambar 6.2 adalah contoh keluaran dari fitur Analisis Dampak YANG. Gambar tersebut menunjukkan pembuatan pohon langsung modul dependen berdasarkan modul target (ietf-ip



adalah contohnya). Dengan mengklik modul dependen tertentu (ietf-ipv6-unicast-routing adalah contohnya), sebuah popup dengan URL ke dokumen tempat modul tersebut dijelaskan akan muncul. Model YANG dari berbagai SDO dan vendor didukung. Alat ini menyediakan status implementasi modul YANG yang lengkap (dan tidak lengkap) di berbagai SDO. Industri ini bergerak cepat menuju pendekatan berbasis model, dan aplikasi ini dapat memfasilitasi kerja sama SDO dan OSS di area penting ini.

Area kerja sama lain yang sukses adalah IETF Hackathons. Peserta (siapa pun diundang) terlibat dalam pengkodean aplikasi eksperimental menggunakan hampir 100% OSS. Hadiah diberikan dan semua upaya dipublikasikan dan tersedia. Ini adalah cara yang bagus untuk menarik generasi baru programmer, yang mendalami perkakas dan pengembangan OSS, ke pertemuan IETF dan mulai berkontribusi pada proses spesifikasi. Lebih baik lagi, individu SDO dan IETF dapat bertemu dan berbicara tatap muka. Dan akhirnya, SDO dapat meniru model tempat transformasi dan penemuan kembali yang sukses sebagai proyek sumber terbuka terjadi.

Sebagai titik pembuktian, *Metro Ethernet Forum* (MEF) telah berputar dari badan standar layanan ethernet operator menjadi orientasi konsorsium sumber terbuka melalui arsitektur *Lifecycle Service Orchestration* dan proyek sumber terbuka mereka sendiri (OpenLSO dan OpenCS). MEF hadir dalam alat YANGCATALOG (salah satu organisasi yang dapat dipilih), bermitra dengan proyek ONAP dan menyelenggarakan acara hackathon mereka sendiri. Mereka juga telah membangun infrastruktur (yaitu, komputasi, jaringan, penyimpanan) tempat eksperimen dan pengujian "standar langsung" dapat dilakukan.

Apa yang tidak Dilakukannya

Apakah perangkat lunak sumber terbuka merupakan aplikasi mandiri yang siap pakai? Dalam semua kasus, kode lain (pribadi, buatan sendiri, sumber terbuka, atau lainnya) dan beberapa atau semua anggota tim terorganisir yang disebutkan di atas akan diperlukan untuk mengembangkan aplikasi yang bermanfaat, bahkan dalam bentuk demonstrasi/Bukti Konsep. Perangkat lunak sumber terbuka bukanlah sebuah standar. Perangkat lunak ini tidak mewajibkan kesesuaian fungsional dengan stempel persetujuan yang terdokumentasi. Ada beberapa perangkat lunak yang menerapkan standar industri. Namun, perangkat lunak tersebut sendiri hanya digunakan sebagai implementasi referensi atau eksperimen pengembangan pengetahuan.

Perangkat lunak sumber terbuka bukanlah perangkat lunak untuk dijual. Namun, perusahaan dapat mengomersialkan versi tingkat perusahaan dan menjual produknya. Tidaklah berlebihan untuk mengklaim bahwa hampir semua perangkat lunak (perusahaan, konsumen) memiliki fondasi sumber terbuka, dengan Android sebagai salah satu contohnya, dan pada tingkat yang lebih rendah, Apple iOS. Kita juga dapat melakukan apa yang telah berhasil dilakukan Red Hat dengan menciptakan versi OS Linux yang dikustomisasi (Red Hat Enterprise Linux), menggabungkannya dengan dukungan dan layanan, dan memasarkannya ke perusahaan. Perlu dicatat juga bahwa generasi muda insinyur perangkat lunak memiliki pengalaman dengan OSS, dan mereka akan memasuki sektor swasta dan menggunakan keterampilan ini untuk mengembangkan produk.



Siapa yang Mendapat Manfaat dari Open Source

Semua orang dapat memperoleh manfaat dari open source. Perusahaan dan operator dapat melibatkan diri dalam proses OSS dan menggunakannya untuk pembelajaran, pengembangan, dan pengujian. Mereka dapat berkolaborasi dengan pihak lain di industri (bahkan pesaing) untuk memengaruhi prioritas dan arah proyek OSS. Lagipula, hasil yang diinginkan dari setiap proyek OSS adalah seseorang akan menggunakan perangkat lunak tersebut, dan berpotensi menjadi kontributor kode.

Vendor peralatan menyukai open source karena mereka dapat memberikan solusi tanpa harus menempatkan fitur bernilai tambah di dalam perangkat keras. Sebaliknya, mereka dapat mengembangkan platform perangkat keras menggunakan komponen komoditas (misalnya, CPU Intel) dan mengoptimalkan interaksi dengan solusi penyediaan dan orkestrasi layanan open source.

Mengantisipasi Sumber Terbuka untuk Kecerdasan Buatan/Jaringan Otonom.

Yang telah kita pelajari sejauh ini adalah bahwa sumber terbuka mendorong kolaborasi komunitas dan inovasi cepat dalam bentuk kode yang dijalankan secara konsisten. Meskipun lebih bertahap, arsitektur dan standar SDO telah dibangun menjadi produk-produk interoperabel yang diterapkan di seluruh dunia dalam internet global. Munculnya AI dalam jaringan yang ditentukan perangkat lunak (SDN) bersama dengan kekuatan OSS dan SDO melahirkan jaringan otonom.

6.2 DARI POTONGAN KE ARSITEKTUR

Blok Bangunan

Dekonstruksi jaringan otonom berbasis AI menghasilkan node dengan berbagai ukuran, bentuk, lokasi, dan perilaku. Secara umum, node akan beroperasi secara otonom, artinya mereka tidak bergantung pada interaksi data bidang kontrol atau bidang data eksternal (tidak seperti router IP yang bertukar informasi topologi satu sama lain untuk menentukan ke mana paket akan diteruskan).

Kemungkinan besar mereka akan menjalankan perangkat lunak cerdas atau setidaknya melalui proksi atau server. Mereka dapat membentuk jaringan ad-hoc sementara dan berkomunikasi dengan entitas eksternal (misalnya, rekan, cloud, manusia) untuk pertukaran data aplikasi, unggahan telemetri, dan unduhan pembaruan perangkat lunak. Salah satu contohnya adalah ponsel pintar kita yang terhubung ke tubuh kita. Contoh yang lebih menarik adalah kendaraan terhubung (KV). Kendaraan terhubung membentuk jaringan ad-hoc bergerak yang bertukar data keselamatan dan non-keselamatan yang dikemas dalam pesan singkat. Kendaraan terhubung mengoperasikan sensor canggih untuk menerima data dari berbagai sumber data, termasuk di dalam mobil, di luar jangkauan penglihatan, GPS, infrastruktur jalan, dan cloud.

Kendaraan terhubung berisi sumber daya komputasi untuk menjalankan aplikasi, sangat mobile, dan memerlukan komunikasi bandwidth tinggi yang berkelanjutan dengan internet melalui tautan nirkabel (LTE/5G). Kendaraan terhubung menghasilkan data dalam jumlah besar untuk operasi runtime internal dan untuk ekspor telematika ke cloud. Kendaraan



terhubung dapat dianggap sebagai cloud di atas roda. Platform pengumpul data/analitik menyerap data yang dihasilkan node dalam bentuk statistik, status, objek, peristiwa, telemetri, dan sebagainya, dan menyajikannya ke aplikasi yang menggunakannya. Dalam beberapa kasus penggunaan, data yang dikumpulkan dialirkan langsung ke aplikasi. Dalam kasus penggunaan lainnya, data disimpan dalam basis data untuk analisis offline. Platform data berisi plug-in arah selatan untuk berkomunikasi dengan berbagai jaringan dan API arah utara untuk akses aplikasi eksternal. Platform orkestrasi/pengendali mendukung API arah utara untuk berkomunikasi dengan aplikasi dan plug-in arah selatan untuk menyampaikan data program dan kebijakan ke jaringan. Aplikasi membentuk otak sistem. Aplikasi mengonsumsi data dari berbagai sumber dan menjalankan fungsi yang menghasilkan keluaran berupa maksud alur kerja yang dikomunikasikan lintas API ke platform orkestrasi/pengendali.

Lingkaran Kebajikan

Pola membaca informasi dari jaringan, menjalankan aplikasi, dan (mem)program ulang jaringan telah menjadi topik berbagai makalah sejak awal tahun 2000an. Ide ini telah diterapkan pada manajemen jaringan otonom yang disebut MAPE-K (Monitor, Analyze, Plan, Execute, and Knowledge) yang muncul pada tahun 2003. Beberapa teknologi internet ternama menggambarkan bidang pengetahuan (KP) internet sebagai, "sistem yang tersebar luas di dalam jaringan yang membangun dan memelihara model tingkat tinggi tentang apa yang seharusnya dilakukan jaringan, untuk menyediakan layanan dan saran kepada elemen-elemen lain dalam jaringan... bersifat baru karena mengandalkan perangkat AI dan sistem kognitif...

Jaringan yang ditentukan oleh pengetahuan (KDN) menerapkan KP ke paradigma SDN. Lingkaran operasional KDN menggambarkan aliran melingkar dengan AI sebagai komponen dalam KP. RFC7575 menguraikan definisi dan tujuan jaringan otonom. Kelompok kerja Model dan Pendekatan Terpadu Jaringan Otonom IETF dibentuk untuk mendokumentasikan spesifikasi solusi secara keseluruhan.

Arsitektur Gambaran Besar

Setelah blok-blok penyusunnya disusun dan konsep lingkaran kebajikan dijelaskan, kita dapat menyusun arsitektur gambaran besar. Jaringan dan simpul-simpul penyusunnya dapat memiliki beragam karakter. Kendaraan terhubung merupakan pilihan yang tepat untuk diskusi ini, mengingat skala jumlah kendaraan yang sangat besar serta keberadaan dan ketergantungannya pada sumber daya komputasi dan jaringan di dalam kendaraan dan eksternal (untuk mengoperasikan berbagai aplikasi terkait keselamatan, infotainment, unggahan telemetri ke cloud, dll.). Kendaraan terhubung memiliki API, dapat diprogram ulang secara dinamis, dan dapat menghasilkan data telemetri dalam jumlah besar yang sebagian akan tetap berada di dalam kendaraan dan sebagian lagi akan ditujukan untuk platform cloud CV.

Ada satu hal lagi yang perlu diperhatikan, yaitu kita dapat memiliki jaringan yang terdiri dari beberapa CV dengan masing-masing menjadi simpul; sebut saja jaringan kendaraan eksternal. Mungkin juga terdapat jaringan "di dalam" kendaraan yang terdiri dari sakelar yang menghubungkan berbagai komponen (simpul), termasuk unit kontrol elektronik, unit utama, dan sistem bantuan pengemudi canggih; sebut saja jaringan internal kendaraan. Platform



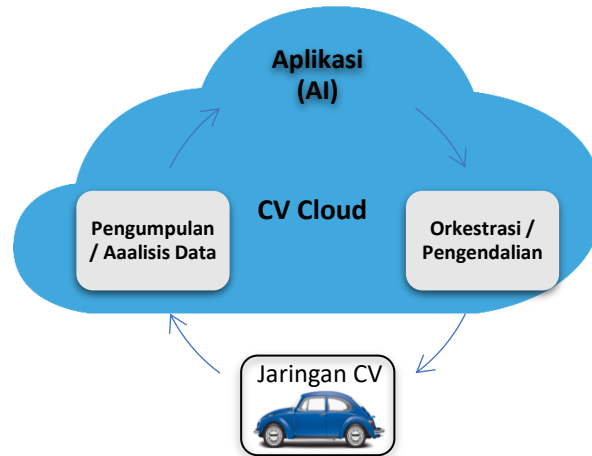
pengumpul data/analitik menyerap dan mungkin memproses data kendaraan. Jika data dihasilkan dari ribuan atau bahkan jutaan CV dan ditransmisikan ke cloud, maka itu merupakan masalah big data yang membutuhkan solusi big data. Sebuah platform yang dikustomisasi untuk menangani data internal kendaraan hanya ada di dalam kendaraan dan hanya memperhatikan lalu lintas jaringan internal kendaraan dan mungkin lalu lintas yang dihasilkan dari kendaraan atau infrastruktur di dekatnya (seperti pesan siaran keselamatan dan non-keselamatan). Platform pengumpul data/analitik ini, baik berbasis cloud maupun di dalam kendaraan, harus diskalakan secara proporsional dengan beban data dan persyaratan pemrosesan aplikasi.

Platform orkestrasi/pengendali menerima perintah alur kerja, intent, operasi prediktif yang diinginkan atau operasi reaktif yang diperlukan, pembaruan perangkat lunak, dan sebagainya dari aplikasi di seluruh API menerjemahkan data tersebut menjadi perintah konfigurasi dan kebijakan lalu mengomunikasikan kebijakan tersebut ke jaringan. Awan kendaraan yang terhubung akan berisi komponen platform orkestrasi/pengendali berskala besar; komponen di dalam kendaraan berukuran lebih kecil dan spesifik untuk operasi internal kendaraan.

Berbagai aplikasi, termasuk yang menggunakan AI, dapat hadir di awan. Sebuah contoh yang menggabungkan platform analitik data besar, aplikasi pembelajaran mesin (ML), dan pengontrol jaringan untuk mengotomatiskan deteksi anomali dalam jaringan CV dikembangkan dan diprototipekan oleh para insinyur di Cisco Systems. Kendaraan yang terhubung mengalirkan telemetri ke awan tempat aplikasi ML memeriksa data untuk potensi anomali. Jika kandidat anomali ditandai, aplikasi ML meminta pengontrol agar CV meningkatkan frekuensi transmisi telemetrinya. Dengan data tambahan yang datang untuk diperiksa, komponen ML dapat memverifikasi dan mengumumkan keberadaan anomali. Alur aplikasi otomatisasi loop tertutup untuk deteksi anomali CV ini mengikuti alur lingkaran kebijakan.

Perhatikan bahwa arsitektur gambaran besar biasanya ada di lingkungan kampus hingga penyedia awan berskala besar. Namun, ia juga dapat berada di dalam node individual seperti halnya dengan CV yang diilustrasikan pada Gambar 6.3. Kendaraan yang terhubung sebenarnya merupakan contoh node jaringan otonom, dan gagasan lingkaran kebijakan bersarang ini merupakan paradigma yang akan ditiru di berbagai domain. Arsitektur gambaran besar yang dijelaskan di bagian ini hanyalah abstraksi tingkat tinggi. Arsitektur ini dapat diselaraskan secara longgar dengan arsitektur dan fungsi untuk implementasi sumber terbuka yang ada dengan runtime ONAP.





Gambar 6.3 Lingkaran positif bersarang dalam jaringan kendaraan terhubung.

Kesesuaian Open Source

Pada prinsipnya, perangkat lunak open source dapat dimasukkan ke dalam blok penyusun mana pun dalam aplikasi deteksi anomali CV yang dijelaskan sebelumnya, tetapi jumlah dan perannya akan bervariasi tergantung domainnya. Dalam ranah CV, keakraban dan kenyamanan dengan platform yang relatif tertutup, siklus pengembangan yang panjang, dan kehati-hatian yang diperlukan (yaitu, tantangan keselamatan dan keamanan CV) akan menunda penetrasi open source. Hal ini kemungkinan besar tidak terjadi di cloud. Penerapan aplikasi yang disesuaikan seperti deteksi anomali CV ML di atas platform cloud open source akan menjadi praktik standar.

Kesesuaian Kecerdasan Buatan/Jaringan Otonom

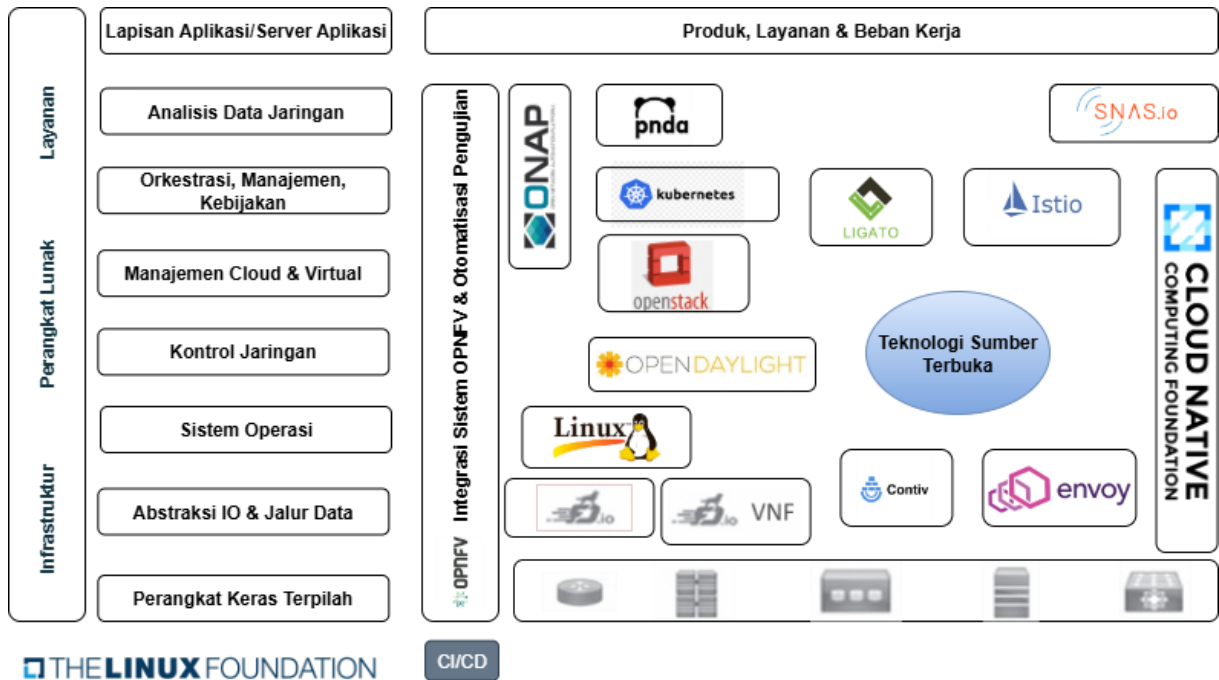
Kompleksitas, skala, dan volume data, baik untuk aplikasi maupun operasi, akan bergantung pada otomatisasi loop tertutup yang dicontohkan oleh lingkaran positif. Aplikasi kecerdasan buatan akan menjadi hal yang lumrah untuk dijalankan dengan, pada, dan di dalam jaringan otonom besar, mulai dari layanan telepon seluler nasional hingga cloud milik produsen mobil yang berkomunikasi dengan armada kendaraan komersial di jalan.

6.3 TEKNOLOGI

Terdapat dunia yang penuh dengan paket dan platform sumber terbuka. Bagian ini akan mensurvei serangkaian teknologi sumber terbuka yang secara individual, atau kolektif, dapat digunakan untuk membangun aplikasi jaringan otonom berbasis AI. Dengan semangat diagram arsitektur berlapis klasik yang kita kenal (yaitu, OSI), versi modifikasi dari tumpukan arsitektur LF (infrastruktur, perangkat lunak, dan layanan) ditunjukkan pada Gambar 6.4 dengan platform dan komponen sumber terbuka diposisikan pada lapisan yang sesuai dalam tumpukan tersebut.

Hal ini dapat berfungsi sebagai peta jalan untuk diskusi selanjutnya dan membantu dalam rekam di akhir bagian ini.





Gambar 6.4 Tumpukan teknologi sumber terbuka.

Definisi Umum

Akronim, nama, dan deskripsi teknologi sumber terbuka umum ditunjukkan pada Tabel 6.1.

Tabel 6.1 Nama dan Deskripsi Teknologi Open Source Umum

Akronim	Nama	Deskripsi Singkat
CN	Cloud native	Mencakup Kubernetes, container, dan microservices
CNF	Cloud native network function	VNF sebagai microservice
Container	Container	Paket perangkat lunak yang ringan, portabel, dan dapat dieksekusi secara mandiri
Contiv VPP	Contiv VPP	Plugin jaringan Kubernetes menggunakan FD.io yang dikontainerisasi sebagai data-plane
Envoy	Envoy	Diciptakan oleh Lyft, mendukung fungsi data-plane untuk layanan mesh Istio
etcd	etcd	Basis data k-v untuk konfigurasi dan data operasional yang digunakan dalam lingkungan CN
FD.io	Proyek data cepat	Nama sumber terbuka untuk VPP
Istio	Istio	Perangkat lunak service mesh yang dibuat oleh Google
k8s	Kubernetes	Platform orkestrasi container Cloud Native
Kafka	Kafka	Message bus menggunakan pola pub/sub
Ligato	Ligato	Platform yang berisi pustaka dan dokumentasi untuk mengembangkan CNF, agen VPP, dan pengendali rantai fungsi layanan CNF
Microservice	Microservice	Bagian dari aplikasi yang berjalan di dalam container
ODL	OpenDaylight	Platform pengembangan dan penyampaian aplikasi SDN
ONAP	Open Network Automation Platform	Arsitektur referensi, kerangka kerja, dan komponen perangkat lunak untuk membangun platform penyebaran cloud skala besar



OpenStack	Openstack	Mengelola komputasi, jaringan, penyimpanan, dll. pada komponen dalam sistem cloud
OPNFV	Open NFV	Platform untuk merancang, mengintegrasikan, dan menguji platform cloud
Plug-in	Plug-in	Perangkat lunak yang ditambahkan ke sistem yang sudah ada untuk memberikan fungsi tambahan
PNDA	Platform for Network Analytics	Platform Big Data sumber terbuka
Pub/sub	Publish and subscribe	Satu aplikasi menerbitkan data, satu atau lebih pelanggan mendengarkan dan menggunakannya saat data tiba
Service mesh	Service mesh	Lapisan overlay L7 yang memungkinkan microservices berkomunikasi satu sama lain menggunakan sidecars
Sidecar	Sidecar	Container atau proses yang 'dibautkan' ke aplikasi microservice, menyediakan fungsi tambahan termasuk jaringan dan keamanan
SNAS	Streaming Network Analytics System	Platform berkinerja tinggi untuk mengumpulkan, menyimpan, dan mengakses informasi sistem routing internet
VNF	Virtual network function	Switch, router, perangkat layanan virtual, dll.
VPP	Vector packet path	Perangkat lunak pemrosesan paket berkinerja tinggi yang dapat berjalan pada sistem fisik atau virtual
ZK	Zookeeper	Basis data konfigurasi dan operasi, biasanya digunakan dengan Kafka

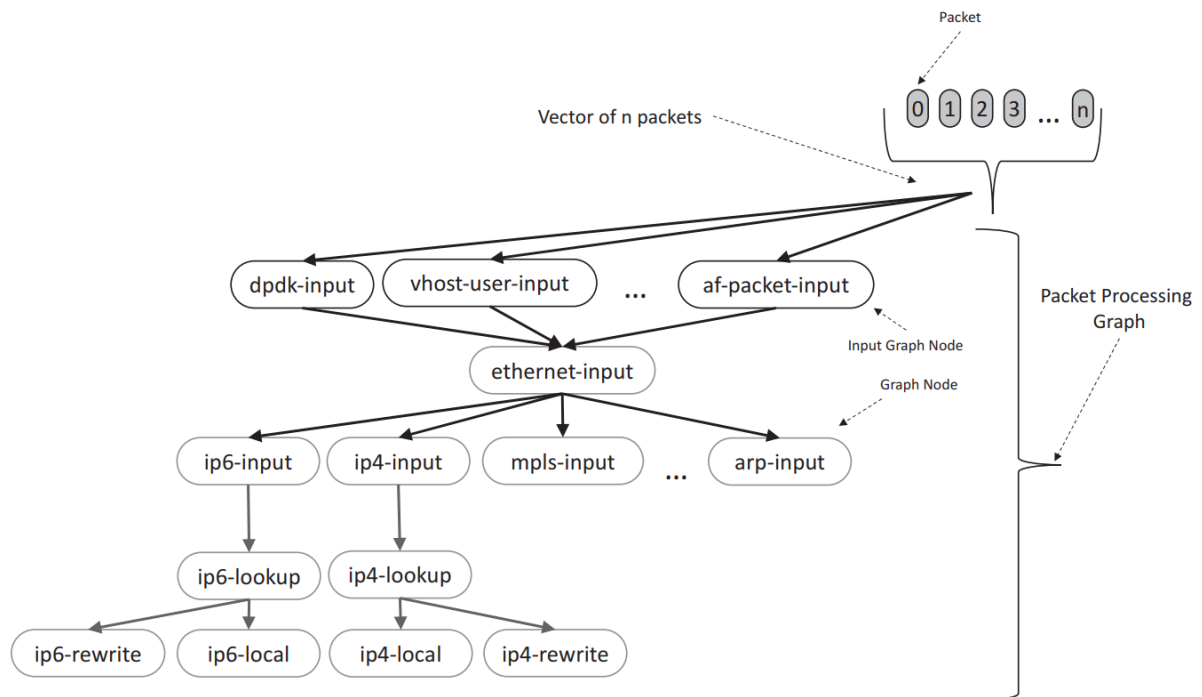
Jaringan

Proyek Data Cepat

FD.io adalah platform aplikasi terprogram dan bidang data yang sangat skalabel. FD.io dapat beroperasi sebagai sakelar virtual, router, dan perangkat layanan di berbagai lingkungan, termasuk bare metal, cloud, kontainer, fungsi jaringan virtual (VNF), dan sistem tertanam. FD.io berjalan pada rangkaian CPU standar dan menggunakan Intel DPDK (Data Plane Development Kit).

"Saus spesial" dalam FD.io disebut pemrosesan paket vektor (VPP). Gambar 6.5 mengilustrasikan proses ini. Pada tingkat tinggi, VPP membaca vektor paket dari I/O jaringan dan menjalankannya melalui grafik terarah pemrosesan paket (misalnya, satu simpul dalam grafik adalah input IPv4, simpul lain mungkin pencarian IPv6, dan seterusnya) dan kemudian beralih ke vektor paket berikutnya. Simpul yang mewakili fungsi yang berbeda dapat ditambahkan, diubah, atau dihapus sehingga menjadikannya cepat dan fleksibel. Proyek Data Cepat adalah proyek LF, mendapat dukungan industri yang luas, memiliki komunitas yang aktif dan terlibat, dan mengirimkan produk.





Gambar 6.5 Grafik pemrosesan paket vektor.

Kontainer

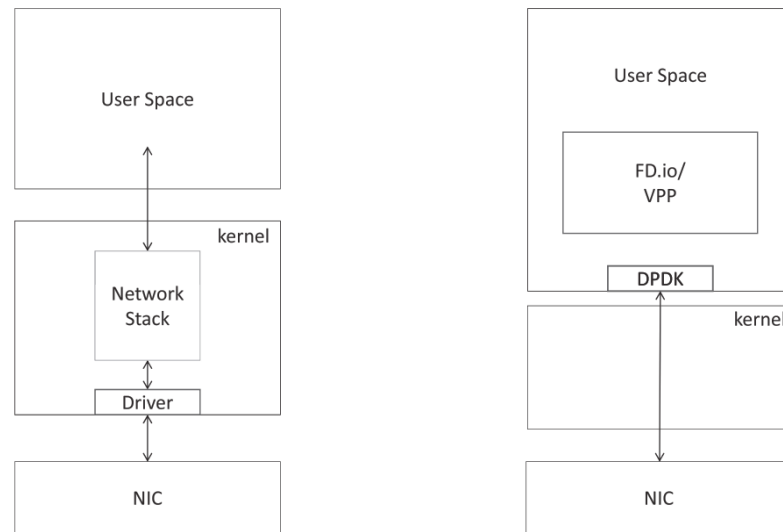
Kontainer dapat dianggap sebagai "VM ringan" yang menjalankan aplikasi atau bagian individual dari aplikasi yang disebut sebagai layanan mikro. Pembahasan mendetail tentang kontainer, layanan mikro, dan prinsip-prinsip *cloud native* (CN) dibahas dalam Bab 7. Di bagian ini, kita akan membahas aspek jaringan kontainer.

Kontainer diterapkan di seluruh jaringan server host. Kontainer dapat dihubungkan bersama dalam berbagai cara yang didukung oleh plugin jaringan native atau pihak ketiga. Jaringan kontainer adalah jaringan overlay virtual yang berjalan di atas jaringan underlay fisik. Jaringan layanan mikro yang ditempatkan di dalam kontainer merupakan persyaratan dasar untuk pemberdayaan, penskalaan, dan ketahanan aplikasi. Jaringan kontainer berkinerja tinggi akan berdampak positif pada kinerja aplikasi secara keseluruhan.

Antarmuka jaringan kontainer (CNI) adalah metode untuk jaringan kontainer. CNI mendefinisikan API (spesifikasi) antara kontainer dan jaringan. CNI menyederhanakan jaringan kontainer dengan mengelompokkan satu atau lebih kontainer di dalam sebuah pod (istilah Kubernetes ini akan dibahas di bagian selanjutnya). Setiap pod dikonfigurasi dengan alamat IP yang unik. Kubernetes menggunakan plug-in CNI untuk mengatur konfigurasi dan pembentukan jaringan kontainer. Ketika sebuah pod dibuat, plug-in CNI diinstal bersama dengan informasi konfigurasi jaringan yang diperlukan, dengan contoh-contoh termasuk konektivitas antar-pod, jenis enkapsulasi, dan kebijakan jaringan.

Seiring dengan semakin banyaknya aplikasi yang mengikuti prinsip-prinsip CN yang dikembangkan dan diterapkan, akan ada kebutuhan yang lebih besar untuk memenuhi persyaratan kinerja aplikasi. Peningkatan infrastruktur jaringan kontainer CN untuk memenuhi persyaratan tersebut meliputi:





Gambar 6.6 Jaringan kernel di sebelah kiri dan jaringan ruang pengguna di sebelah kanan.

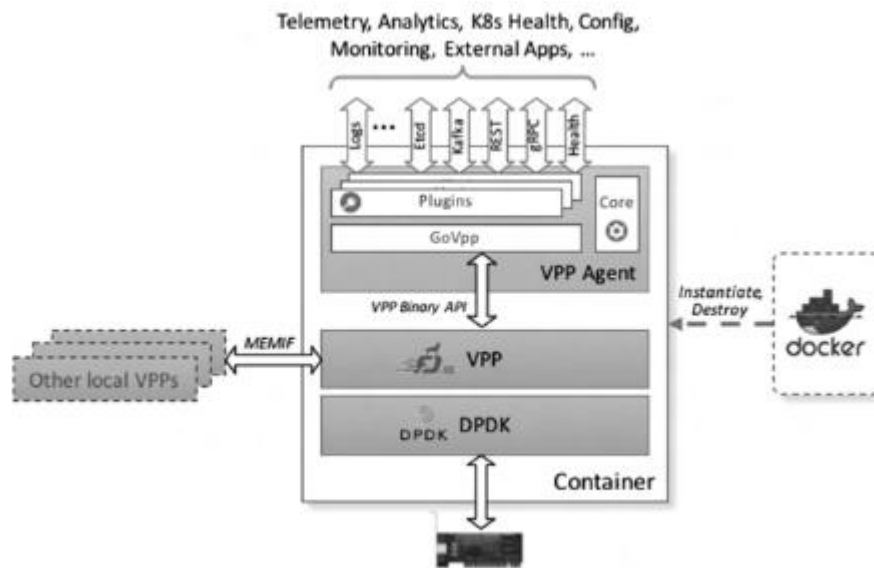
- Memindahkan fungsi jaringan kontainer dari kernel ke ruang pengguna. Hal ini menghilangkan overhead panggilan sistem (ke kernel). Tidak ada ketergantungan pada komunitas jaringan kernel untuk mengimplementasikan fitur-fitur baru. Inovasi dan penambahan fitur baru menjadi mudah tanpa perlu menyentuh kernel. Ketersediaan juga meningkat karena masalah ruang pengguna tidak akan melumpuhkan node. Gambar 6.6 mengilustrasikan perbedaan antara keduanya. Di sebelah kiri, tumpukan jaringan diimplementasikan di kernel. Di sebelah kanan, fungsi jaringan yang disediakan oleh FD.io/VPP berada di ruang pengguna dan melewati kernel sepenuhnya.
- Membangun fungsi jaringan sebagai fungsi jaringan *cloud native* (CNF). CNF adalah VNF yang diimplementasikan sebagai layanan mikro terkontainerisasi. Sistem perangkat, orkestrasi, dan manajemen yang sama yang digunakan untuk siklus hidup aplikasi CN dapat digunakan untuk CNF. Intinya, CNF menjadi warga negara kelas satu dalam topologi layanan aplikasi CN.

Agen VPP menyediakan bidang kontrol dan manajemen untuk CNF berbasis VPP. Terdapat beberapa plug-in khusus VPP (L2, L3, manajemen siklus hidup CN-Infra) beserta akses API northbound ke aplikasi dan plug-in eksternal. Pemrosesan paket vektor menyediakan bagian bidang data. Kombinasi agen VPP dan bidang data VPP membentuk CNF berbasis VPP dengan arsitektur yang digambarkan pada Gambar 6.7. Agen VPP merupakan komponen dari proyek sumber terbuka Ligato.

Satu contoh lagi CNF berbasis VPP adalah Contiv-VPP. Ini adalah plug-in jaringan Kubernetes yang mengabstraksi konektivitas dan jaringan kontainer. Plug-in ini diimplementasikan di ruang pengguna dan menyediakan jaringan berkinerja tinggi untuk CNF serta aplikasi lama menggunakan jaringan kernel. Selain itu, aplikasi dapat melewati kernel dengan menggunakan tumpukan Protokol Kontrol Transmisi (TCP) VPP host khusus. Jaringan otonom akan menuntut skala dan ketahanan kinerja yang lebih besar pada aplikasi



terdistribusi berbasis CN yang beroperasi di setiap penerapan cloud. Sangat penting untuk menyediakan kinerja bidang data setinggi mungkin untuk jaringan kontainer.



Gambar 6.7 CNF berbasis VPP yang diimplementasikan menggunakan agen Ligato VPP.

Alur Kerja dan Orkestrasi Layanan

Alur kerja adalah otomatisasi proses atau fungsi yang, ketika berhasil dijalankan, menghasilkan hasil bisnis yang diinginkan. Orkestrasi layanan mengubah alur kerja menjadi tugas-tugas yang menyediakan dan menyebarkan sumber daya ke platform jaringan komputasi yang memungkinkan aktivasi dan kesiapan layanan. Definisi ini dapat langsung diterapkan ke platform yang menjalankan tugas-tugas tersebut. Salah satu platform tersebut adalah ONAP, yang dibahas dalam Bab 3.

OpenDaylight

OpenDaylight adalah API dan platform pengembangan yang mendukung interaksi aplikasi-jaringan. Fondasi arsitektur ODL didasarkan pada hal-hal berikut:

- Perangkat, fungsi, topologi, dan layanan direpresentasikan oleh model YANG.
- Prinsip-prinsip Rekayasa Perangkat Lunak Berbasis Model berlaku. Model YANG secara otomatis menghasilkan REST API. RESTCONF adalah contoh di mana fungsi dan model NETCONF/YANG dikonversi menjadi REST API.
- Mendukung beberapa plug-in protokol southbound. Contohnya termasuk *Border Gateway Protocol—Link State* (BGP-LS), NETCONF, Openflow, dan Path Computation Element Protocol.

OpenDaylight memiliki komunitas yang aktif di dalam LF. OpenDaylight dengan cepat menjadi komponen utama dalam banyak proyek sumber terbuka seperti OpenStack dan ONAP.

Platform Otomasi Jaringan Terbuka

Penyedia layanan komunikasi (istilah lain operator cloud) menghadapi banyak tantangan, termasuk pembuatan dan peluncuran layanan, optimasi sumber daya virtual,



skalabilitas, kinerja, efisiensi operasional, dan biaya. Kurangnya standar atau arsitektur referensi membuat tugas-tugas ini sangat kompleks.

Operator menghargai layanan dan diferensiasi. Pelanggan tertarik pada layanan, bukan infrastruktur yang digunakan untuk menyediakan layanan tersebut. Dengan demikian, akan lebih hemat biaya bagi industri untuk menyelesaikan masalah ini sekarang dengan cara yang umum daripada terus-menerus menggunakan solusi yang terisolasi.

Demikian pula, vendor melihat perbedaan utama pada perangkat virtual dan fisik yang mereka sediakan untuk menciptakan layanan, dan kurang begitu pada cara layanan ini dioperasikan. Vendor akan lebih cenderung mendukung peralihan menuju tumpukan solusi layanan lengkap yang umum.

Operator dan vendor dapat berkolaborasi untuk mengembangkan dan mengoperasikan tumpukan jaringan hemat biaya yang menerapkan prinsip-prinsip berikut:

- Tumpukan solusi layanan lengkap yang mencakup waktu desain layanan dan waktu proses layanan untuk sumber daya virtual dan fisik. Solusi tumpukan parsial mungkin mencakup manajemen dan orkestrasi NFV tetapi tidak memiliki orkestrasi dan desain layanan.
- Tumpukan solusi layanan lengkap yang dibangun dan dirancang sebagai platform referensi menggunakan standar yang ada jika sesuai. Hal ini menawarkan "kelengkapan" tumpukan dan menghindari solusi operator independen yang mahal.
- Otomatisasi untuk mengurangi pengeluaran operasional (opex) dan memberikan waktu pemasaran layanan baru yang lebih cepat.
- Arsitektur dan platform referensi.

Prinsip-prinsip ini diadopsi oleh ONAP. *Open Network Automation Platform* lahir dari penggabungan antara *Open-Orchestrator* (Open-O) dan *Enhanced Control, Orchestration, Management, and Policy* (ECOMP). ECOMP merupakan kontribusi dari AT&T dan sebagian besar definisi di sana telah ditempatkan di bawah naungan proyek ONAP. Detail arsitekturnya dibahas dalam Bab 3, dan terdapat referensi pendukung yang tersedia dari repositori proyek ONAP. Dua komponen utama yang perlu diperhatikan adalah layanan, desain, dan pembuatan (SDC) dan kerangka kerja runtime.

SDC menyediakan lingkungan, perangkat, dan sistem untuk mendefinisikan, mensimulasikan, dan mensertifikasi aset dan fungsi, termasuk pengalaman pengguna, sumber daya, layanan, dan otomatisasi loop tertutup. Komponen manajemen kebijakan memungkinkan pembuatan dan penerapan kebijakan yang mengatur kondisi dan perilaku yang diharapkan untuk diterapkan selama penyediaan aset. Tersedia API dan perangkat pengembangan perangkat lunak (SDK) untuk mendorong partisipasi ekosistem. SDC menggunakan pendekatan berbasis model untuk definisi dan orkestrasi layanan menggunakan standar Spesifikasi Topologi dan Orkestrasi untuk Aplikasi Cloud. Standar ini menawarkan serangkaian templat yang memungkinkan definisi aplikasi (misalnya, node komputasi, basis data, dan aplikasi web) serta siklus hidup penerapan secara langsung.

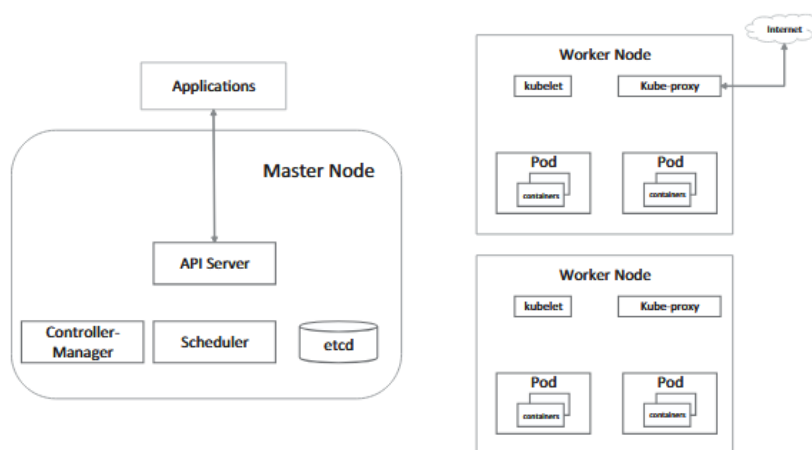


Kerangka kerja runtime mengeksekusi fungsi, aturan, dan kebijakan yang didefinisikan dalam SDC. Termasuk di dalamnya orkestrasi layanan, pengontrol, dan komponen pengumpulan data, analitik, dan peristiwa (DCAE). ONAP memiliki keistimewaan sebagai platform sumber terbuka pertama yang mewujudkan visi siklus hidup tertutup otomatis yang dibahas di bagian sebelumnya. ONAP merupakan sebuah platform, dan karenanya terdapat fleksibilitas untuk menggabungkan teknologi sumber terbuka lainnya, termasuk ODL dan FD.io. ONAP adalah proyek LF dan didukung oleh operator dan vendor, yang agak tidak umum untuk proyek sumber terbuka tradisional.

Kubernetes

Penggunaan kontainer dalam jaringan cloud terus berkembang. Otomatisasi manajemen siklus hidup kontainer dan penjadwalan kluster, mulai dari penggunaannya di lingkungan pengembangan hingga penerapan aplikasi dan layanan di cloud produksi, diperlukan. Kubernetes (k8s) adalah platform sumber terbuka yang menyediakan solusi otomatis untuk mengelola penerapan kontainer terdistribusi skala besar. Arsitektur Kubernetes terdiri dari beberapa komponen berbeda seperti yang ditunjukkan pada Gambar 6.8. Pod adalah tempat aplikasi yang dikontainerisasi berada, dan merupakan objek terkecil yang digunakan Kubernetes. Setiap pod dikonfigurasi dengan satu alamat IP.

Pod terdapat dalam node (juga disebut sebagai node pekerja). Node terdiri dari host fisik atau virtual yang memungkinkan jaringan kontainer dan manajemen sumber daya. Node berkomunikasi dengan node master untuk keperluan bidang kontrol. Terdapat agen kubelet yang mengelola node dan kube-proxy yang menawarkan fungsi proksi layanan dan penyeimbang beban untuk komunikasi eksternal ke Internet, sebagai salah satu contoh.



Gambar 6.8 Arsitektur Kubernetes tingkat tinggi.

Beberapa komponen berbeda berada di dalam node master. Server API menawarkan dukungan API arah utara ke aplikasi eksternal. Dijelaskan di bagian selanjutnya, etcd digunakan untuk menyimpan data kluster di dalam penyimpanan data. Fungsi penjadwal bertanggung jawab untuk men-deploy pod dan layanan pada node yang tersedia. Controller-manager memungkinkan dukungan untuk beberapa pengontrol, termasuk node (memantau ketersediaan node) dan replikasi (mengelola jumlah pod dalam sistem).



Kluster Kubernetes didefinisikan sebagai sebuah node master dan satu atau lebih node pekerja yang berada di bawah kendalinya. Nilai namespace Kubernetes dapat ditetapkan ke kluster virtual, sehingga menyediakan bentuk multitenancy. Misalnya, grup pengujian dapat berada di satu namespace dan grup deployment produksi di namespace lain. Namun, di dalam namespace, setiap grup dapat menjalankan aplikasi, layanan, konfigurasi pod, skema pengalamatan IP yang sama persis, dan sebagainya tanpa saling mengetahui atau mengganggu satu sama lain.

Dan terakhir, layanan Kubernetes adalah pengelompokan logis pod yang mendukung entitas layanan tertentu. Analoginya adalah penyeimbang beban front-end yang mendistribusikan koneksi HTTP ke beberapa server back-end. Dengan layanan Kubernetes, terdapat satu pod yang mewakili layanan (misalnya, URL) yang dapat dijangkau oleh aplikasi dan pengguna eksternal, dan satu atau lebih pod back-end yang hanya diketahui oleh pod front-end. Perlu dicatat bahwa pod back-end dapat muncul atau menghilang (misalnya, untuk alasan performa atau ketahanan) tanpa memengaruhi layanan. Kubernetes didukung oleh Linux Foundation dan Cloud Native Computing Foundation.

Kafka dan Zookeeper

Menurut halaman berandanya: “Apache Kafka digunakan untuk membangun jalur data waktu nyata dan aplikasi streaming. Ia dapat diskalakan secara horizontal, toleran terhadap kesalahan, sangat cepat, dan berjalan dalam produksi di ribuan perusahaan.... Kafka menyediakan bus pesan layanan streaming penerbit-pelanggan (pub/sub) di dalam cloud, yang memungkinkan beberapa pelanggan data (misalnya, aplikasi, basis data) untuk menggunakan data dari sumber data yang sama.

Layanan streaming berbasis Kafka memerlukan topik yang mengidentifikasi aliran, penerbit untuk menghasilkan data ke aliran, dan satu atau lebih pelanggan untuk menggunakan data dari aliran. Terdapat beberapa topik dengan produsen dan konsumen berbeda yang beroperasi pada bus Kafka yang sama. Kafka memanfaatkan Zookeeper untuk menyimpan metadatanya. Serupa dengan etcd yang dijelaskan di bagian berikut, Zookeeper menawarkan koordinasi terpusat untuk data konfigurasi layanan terdistribusi, dan dalam hal ini, untuk kluster Kafka.

Etcd

Etcd adalah penyimpanan kunci-nilai terdistribusi yang digunakan untuk menyimpan dan mendistribusikan data di seluruh kluster mesin. Aplikasi dapat membaca dan menulis data ke dalam penyimpanan data etcd menggunakan REST API. Misalnya, penyimpanan data etcd dapat menyimpan informasi Kubernetes tentang detail pod dan layanan, namespace, dan pekerjaan terjadwal. Nilai-nilai ini dapat dipantau, memungkinkan aplikasi untuk mengonfigurasi ulang dirinya sendiri ketika nilai-nilai ini diperbarui. Etcd menggunakan algoritma konsensus Raft sehingga semua instans etcd yang terdistribusi di seluruh kluster pod dapat memperoleh tampilan yang sama. Etcd banyak digunakan di komunitas cloud dan di hampir semua penerapan Kubernetes.



Analisis Data Besar

Komunitas sumber terbuka telah lama menjadi bagian dari platform analisis data besar. Apache Spark, Hadoop, dan Kafka adalah komponen perangkat lunak umum yang ditemukan dalam penerapan data besar. Baru-baru ini, ONAP telah mendefinisikan DCAE dalam arsitekturnya. DCAE dan PNDA (dijelaskan di bagian berikut) adalah platform analisis data besar. Komponen PNDA dimungkinkan untuk beroperasi di dalam platform data DCAE.

Platform untuk Analisis Data Jaringan

Jaringan otonom diharapkan menghasilkan data dalam jumlah besar untuk diserap oleh platform cloud. Beberapa node yang ada di jaringan akan memancarkan lalu lintas secara real-time untuk diproses segera di cloud. Contohnya adalah telemetri streaming dari kendaraan otonom yang penting untuk keselamatan dan keamanan. Node lain akan mengunggah data secara batch untuk analisis offline atau pelatihan ML. Contohnya adalah unggahan video rekaman oleh drone cerdas yang digunakan untuk perencanaan kota. Platform cloud harus mengakomodasi beragam sumber data dan aplikasi. Platform harus berskala dan selalu tersedia serta tangguh terhadap kegagalan aplikasi atau manusia.

Untuk memenuhi persyaratan tersebut, Nathan Marz merancang Arsitektur Lambda (LA). Arsitektur ini dirancang sebagai "sistem yang tangguh dan toleran terhadap kesalahan, baik terhadap kegagalan perangkat keras maupun kesalahan manusia, mampu melayani berbagai beban kerja dan kasus penggunaan, serta memerlukan pembacaan dan pembaruan latensi rendah. Sistem yang dihasilkan harus berskala linier, dan harus berskala keluar, bukan naik....

Alur sistem LA terdiri dari hal-hal berikut:

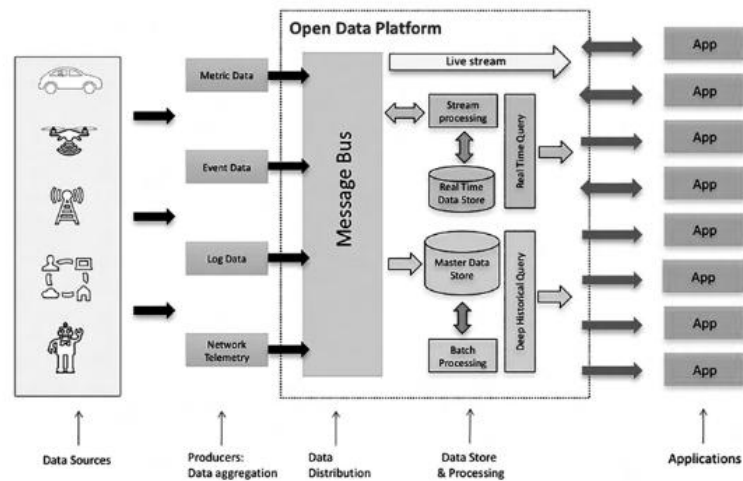
- Aliran data masuk dikirim ke lapisan batch dan lapisan kecepatan.
- Lapisan Batch (1) memelihara basis data yang tidak dapat diubah yang berisi semua data mentah yang diterima oleh sistem dan (2) melakukan prakomputasi tampilan batch yang dapat dikueri dari aplikasi.
- Lapisan kecepatan menangani data terbaru dan menghasilkan tampilan inkremental yang dapat dikueri dari aplikasi.
- Lapisan penyajian dapat mengkueri tampilan batch, tampilan waktu nyata, atau menggabungkan keduanya. Perhatikan bahwa kesalahan apa pun yang muncul dalam tampilan batch lapisan penyajian akan ditimpa karena tampilan batch baru terus-menerus dibangun ulang dari set data master.

LA menghindari status, mutabilitas data, atau kehilangan data. Sebaliknya, semua data mentah disimpan dalam basis data yang tidak dapat diubah dan dihitung sebelumnya atau dihitung ulang sesuai kebutuhan. Kombinasi kekekalan dan komputasi ini mencapai toleransi kesalahan aplikasi dan manusia.

PNDA adalah platform big data sumber terbuka yang diturunkan dari LA. Gambar 6.9 menggambarkan arsitektur PNDA. Dimulai dari kiri, terdapat sejumlah sumber data unik yang mengalirkan informasi unggahan batch ke cloud. Data dikumpulkan, diagregasi, dan dipublikasikan ke bus pesan. Penyimpanan data master berlangganan dan menggunakan



semua data mentah yang dapat diproses dalam mode batch sebagai persiapan untuk kueri historis.



Gambar 6.9 Arsitektur PNDA.

Pemrosesan aliran akan berlangganan dan menerima aliran yang membutuhkan pemrosesan waktu nyata. Selain itu, data langsung dapat dialirkan langsung ke aplikasi yang menggunakannya.

Komponen-komponen PNDA terdiri dari serangkaian paket perangkat lunak yang berbasis pada ekosistem Apache yang telah diterapkan dan matang. Misalnya, Kafka/Zookeeper mendukung bus pesan; Hadoop Distributed File System dapat diatur sebagai penyimpanan data utama; Spark dapat menangani pemrosesan batch dan pemrosesan aliran waktu nyata oleh Spark Streaming. Namun, PNDA adalah sistem terbuka sehingga pengembang dan operator bebas menggunakan komponen apa pun yang mereka rasa dapat melakukan pekerjaan tersebut. Fleksibilitas platform PNDA ditonjolkan oleh keragaman aplikasi yang diimplementasikan. PNDA adalah proyek sumber terbuka yang didukung oleh LF.

Sistem Analisis Jaringan Streaming

Pentingnya Internet yang stabil, berkinerja tinggi, dan aman tidak dapat dilebih-lebihkan. Kecelakaan operator (baik yang disengaja maupun tidak), bug perangkat lunak router, dan jaringan serta kebijakan rute yang direkayasa dengan buruk memengaruhi kinerja dan ketersediaan. Yang lebih mengkhawatirkan, serangan siber yang diinisiasi oleh pelaku kejahatan siber dapat menimbulkan malapetaka. Sumber data penting untuk menganalisis dan mendeteksi potensi masalah adalah data sistem perutean Internet yang mencakup, namun tidak terbatas pada, tabel perutean IPv4 dan IPv6 global, topologi AS, pengumuman/penarikan prefix, detail pengidentifikasi keluarga alamat prefix (IPv4 dan IPv6, antara lain), anomali prefix, kebijakan rute, telemetry router, dan konvergensi jaringan.

Data ini didasarkan pada frekuensi, kecepatan, dan skala yang dihasilkannya. Ada ratusan ribu peristiwa (misalnya, pengumuman prefix) yang terjadi setiap hari dari router BGP pada umumnya. Banyak kemajuan dalam pengumpulan data perutean Internet telah dicapai selama bertahun-tahun. Routeviews adalah umpan data BGP publik yang diperbarui setiap 15



menit . Terdapat paket komersial, skrip/alat operator proprietary (sayangnya, pengikisan layar otomatis sudah umum), dan kerangka kerja sumber terbuka yang melakukan pengumpulan dan menawarkan akses API ke data.

Namun, diperlukan lebih banyak lagi untuk menyediakan pemrosesan, pemantauan, dan visibilitas daring instan terhadap informasi ini. Performa penyerapan data BGP yang dialirkan, penguraian data, dan kecepatan baca/tulis basis data harus mendekati waktu nyata (real-time). Penggunaan sumber daya komputasi, penyimpanan, dan memori perlu dioptimalkan. Distribusi data yang efisien ke banyak konsumen menyederhanakan sistem. Mengekspos API standar (misalnya, REST, Kafka) menawarkan akses yang lebih mudah ke rangkaian teknologi basis data dan analitik yang lebih luas. Pengumpulan telemetri BGP dengan fidelitas yang lebih tinggi memungkinkan presisi yang lebih tinggi dalam analisis. Menerapkan teknik yang telah disempurnakan di domain lain (misalnya, penggunaan bus pesan Kafka dan pemrosesan batch Spark dari PNDA) juga masuk akal untuk memenuhi persyaratan ini.

Sistem Analisis Jaringan Streaming adalah proyek sumber terbuka (sebelumnya dikenal sebagai OpenBMP). SNAS mengumpulkan sejumlah besar data perutean, memeliharanya, dan menyimpannya dalam database untuk menyediakan akses/API berkinerja tinggi untuk pelacakan, visualisasi, dan analitik.

SNAS menyediakan beberapa fitur unik, termasuk:

- Mendukung pengumpul Protokol Pemantauan BGP (BMP). Klien BMP beroperasi pada router BGP dan mengirimkan semua basis informasi perutean BGP ke pengumpul SNAS. Pengumpul juga dapat menggunakan umpan data BGP publik seperti routeview.
- Memproses informasi BMP dan BGP yang diterima dan menghasilkannya ke bus Kafka.
- REST API, konsumen Kafka, dan skema basis data untuk pengembangan aplikasi yang cepat.
- Dukungan untuk BGP, OSPF/ISIS, L3VPN, Segment Routing, IPv4/IPv6, dan MPLS.

Internet harus menyerap peningkatan beban lalu lintas dan miliaran titik akhir dengan beragam persyaratan aplikasi, termasuk latensi rendah, IPv4/IPv6, prediktabilitas, konektivitas ad-hoc, dan keamanan yang diperkenalkan oleh penerapan jaringan otonom. Peralatan daring untuk memproses informasi perutean Internet yang dialirkan dan menyediakan akses API akan dibutuhkan. SNAS didukung oleh LF.

Virtualisasi

Openstack

Dulu ketika penawaran cloud pertama kali muncul, penyediaan sumber daya merupakan tugas yang rumit yang digerakkan oleh skrip, sistem manajemen jaringan (NMS) yang dikustomisasi, dan operator manusia. OpenStack adalah serangkaian komponen untuk mengotomatiskan penerapan dan pengelolaan sumber daya cloud. Proyek-proyek Openstack klasik meliputi komputasi (Nova), penyimpanan (Swift), pencitraan (Glance), dan jaringan (Neutron) dan mungkin yang paling familiar bagi mereka yang telah lama bekerja dengannya. Openstack disebut di beberapa kalangan sebagai Infrastruktur-sebagai-Layanan.



Openstack didukung oleh OpenStack Foundation dan komunitas vendor dan operator yang besar. Openstack telah berkembang menjadi platform sumber terbuka yang canggih dan tangguh. Kini terdapat API untuk orkestrasi kontainer, sehingga memungkinkan penerapan hibrida Openstack dan Kubernetes. Openstack Neutron mendukung plug-in untuk orkestrasi ODL dan komponen FD.io bidang data berkinerja tinggi.

Cloud Native

Cloud Native Computing Foundation menyusun sistem CN sebagai sistem berbasis kontainer, dikelola secara dinamis, dan berorientasi pada layanan mikro. Bab 7 memberikan pembahasan mendetail tentang arsitektur dan teknologi CN. CNF berkinerja tinggi dan orkestrasi Kubernetes menyediakan fondasi dasar untuk solusi aplikasi CN.

Ligato, yang memuluskan pengembangan dan pengelolaan CNF berbasis VPP (dan sebenarnya layanan mikro apa pun), adalah platform sumber terbuka sebuah platform sumber terbuka yang terdiri dari tiga komponen berikut:

- CN-infra. Platform ini mendukung berbagai plugin, termasuk logging, database, penyimpanan data kunci-nilai, pengiriman pesan, API (REST dan *gRPC Remote Procedure Call* [gRPC]), dan telemetri. Plugin khusus aplikasi terpisah dapat ditambahkan, dan bersama dengan Kubernetes mendefinisikan fungsionalitas platform secara keseluruhan.
- Agen VPP. Hal ini telah dibahas di bagian sebelumnya, tetapi secara ringkas, platform ini menyediakan bidang kontrol dan manajemen untuk CNF berbasis VPP.
- Pengontrol rantai fungsi layanan. CNF, kontainer CNF aplikasi, dapat "dijahit" menjadi rantai fungsi layanan yang logis. Kubernetes tidak mendukung penyediaan rantai layanan, jadi ini merupakan tambahan yang sangat baik untuk perkakas penyebaran topologi layanan aplikasi CN.

Untuk tujuan kejelasan, perlu dicatat bahwa Ligato bukanlah sebuah aplikasi. Ini adalah platform perangkat lunak yang terdiri dari komponen dan pustaka yang disebutkan di atas yang dapat digunakan oleh pengembang (beberapa, semua, kustom dalam kombinasi apa pun) sesuai keinginan mereka untuk membuat dan menyebarkan topologi layanan aplikasi berkinerja tinggi. Ini modular, dapat diperluas, dan ditulis dalam Golang bahasa pemrograman yang umum untuk perkakas CN seperti Kubernetes dan etcd.

Produk sampingan dari aplikasi yang dikembangkan CN adalah penempatan layanan mikro penyusunnya (disebut sebagai layanan dalam diskusi ini) di seluruh topologi layanan terdistribusi yang besar. Keuntungannya meliputi devops yang disederhanakan, penyebaran layanan yang cepat, skalabilitas horizontal, dan peningkatan toleransi kesalahan. Namun, dengan keunggulan tersebut muncul serangkaian tantangan terkait kebutuhan komunikasi aplikasi. Intinya adalah bahwa setiap aplikasi (misalnya, layanan web, server konten, atau proses bisnis) bertanggung jawab atas konektivitas layanan, autentikasi, penyeimbangan beban, kontrol lalu lintas, pemantauan, kebijakan keamanan, dan sebagainya, masing-masing. Aplikasi juga harus menangani situasi ketika sejumlah besar instans beban kerja layanan muncul/hilang dalam waktu singkat.



Daripada membebani aplikasi individual dengan semua kerumitan ini, pendekatan yang lebih optimal adalah memindahkan pekerjaan ini ke layanan terpisah yang dapat menangani semua kebutuhan komunikasi layanan-ke-layanan aplikasi. Layanan ini disebut sebagai service mesh. Service mesh adalah lapisan infrastruktur khusus untuk menangani komunikasi antar-layanan aplikasi. Lapisan ini bertanggung jawab atas pengiriman permintaan yang andal melalui topologi layanan kompleks yang membentuk aplikasi CN modern. Singkatnya, service mesh berperan sebagai overlay L7. Beberapa fungsi yang didukung oleh service mesh meliputi:

- Visibilitas yang dimungkinkan oleh telemetri, pelacakan, dan pelaporan statistik. Fungsi observabilitas di mana fungsi pelacakan dan pencatatan L7 diaktifkan di bidang data dengan hasil yang dikirim melalui sesi bidang kontrol ke pengumpul hulu dan aplikasi analitik.
- Ketahanan dan toleransi kesalahan terhadap fluktuasi kinerja dan ketersediaan jaringan/aplikasi. Batas waktu, pemeriksaan kesehatan, dan pemutus sirkuit untuk mendeteksi dan jika perlu mengalihkan rute untuk kegagalan cepat sehingga dapat memulai tindakan pemulihan.
- Kontrol lalu lintas termasuk penyeimbangan beban dan pengalihan rute lalu lintas dinamis. Fungsi bidang data untuk mengoptimalkan pemanfaatan sumber daya dan efisiensi penerapan (melalui pengujian canary).
- Penegakan kebijakan aplikasi yang disediakan. Memusatkan definisi kebijakan yang diikuti dengan penerapan ke bidang data service mesh.
- Keamanan menyeluruh. Sertifikat dan kunci diterapkan dalam service mesh untuk menawarkan komunikasi bidang data yang aman antar layanan mikro melalui penggunaan proksi tepi sidecar.

Service mesh melindungi pengembang dari kekhawatiran tentang membangun telemetri, keamanan, dan fungsionalitas jaringan yang disesuaikan ke dalam aplikasi mereka. Perusahaan dan operator menyukai service mesh karena mereka akhirnya memiliki observabilitas lapisan aplikasi dan sarana untuk mengalihkan kebijakan utama dan jaringan dari jaringan ke lapisan aplikasi. Ketika digabungkan dengan identitas layanan, autentikasi (misalnya, *Secure Production Identity Framework for Everyone*) dan otorisasi berbasis kebijakan (misalnya, *Open Policy Agent*), service mesh menjanjikan untuk memberikan keamanan yang konsisten dan portabel di seluruh platform cloud.

Service mesh juga menyediakan tunneling dan enkripsi untuk komunikasi tepercaya untuk melindungi dari ancaman bahkan dalam *Zero Trust Network* (ZTN). Gagasan ZTN berasal dari proyek internal di Google dengan tujuan menyediakan akses pengguna akhir dan perangkat ke sumber daya intranet tanpa menggunakan tunneling jaringan pribadi virtual. Secara konseptual, ZTN mengasumsikan bahwa semua jaringan bersifat bermusuhan, ancaman internal dan eksternal selalu ada, dan bahwa lokalitas jaringan (yaitu, di balik firewall) tidak cukup untuk menentukan siapa yang dapat dipercaya atau tidak.

Komunikasi yang aman antara dua pihak melalui ZTN mengharuskan masing-masing pihak mengautentikasi identitas pihak lain secara bersamaan. Hal ini dicapai melalui



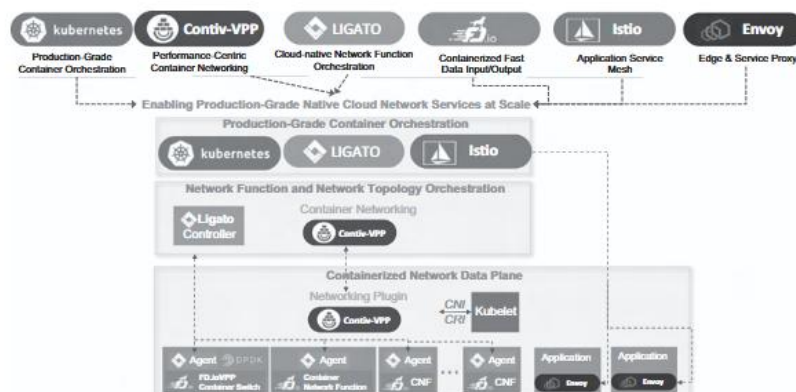
mekanisme autentikasi dua arah dengan *Mutual Transport Layer Security* (mTLS) sebagai pendekatan umum yang digunakan dalam lingkungan service mesh.

Dalam skenario ini, informasi sertifikat/kunci diberikan kepada para pihak sehingga mereka dapat memulai jabat tangan mTLS, mengautentikasi identitas masing-masing, dan membangun komunikasi antar sesi melalui saluran aman. Ini dapat dianggap sebagai zona kepercayaan dua pihak. Tentu saja, dimungkinkan untuk membentuk zona kepercayaan dari beberapa pihak selama mereka diberikan kredensial sertifikasi/kunci yang sesuai. Dengan demikian, zona kepercayaan dapat dibentuk hanya antara pihak-pihak yang memerlukan komunikasi dengan layanan atau layanan-layanan tertentu.

Sejumlah perusahaan skala web seperti Twitter, Google, dan Netflix selama bertahun-tahun telah menggunakan implementasi service mesh internal mereka sendiri dengan beberapa berkontribusi pada proyek sumber terbuka. Istio adalah solusi service mesh yang diciptakan oleh Google yang menyediakan definisi dan penerapan kebijakan, manajemen kunci, dan ekspor telemetri/pelaporan. Pada dasarnya, ini adalah bidang kontrol untuk service mesh. Komponen Envoy dikembangkan oleh Lyft dan berfungsi sebagai proxy tepi yang ditempatkan di bidang data service mesh untuk mencegah semua lalu lintas masuk dan keluar. Hal ini memungkinkan perutean, penyeimbangan beban, kontrol akses, pengumpulan statistik, dan fungsionalitas bidang data berbasis kebijakan.

Istio dengan proksi tepi Envoy diterapkan menggunakan model sidecar. Sidecar dapat dipandang sebagai serangkaian fungsi nilai tambah yang dipasok ke layanan mikro. Sidecar disuntikkan ke dalam layanan mikro oleh Istio. Fungsi-fungsi tersebut dapat beroperasi sebagai proses di dalam kontainer layanan mikro atau dapat ditempatkan di dalam kontainernya sendiri. Karena Envoy berada di bidang data, ia dapat menggabungkan VPP CNF untuk throughput dan kinerja maksimum. Envoy juga bekerja dengan Kubernetes dan komponen Istio-Auth untuk mengelola pertukaran kredensial sertifikat/kunci sehingga komunikasi antarlayanan yang aman dapat berlangsung.

Gambar 6.10 menempatkan semua bagian ini ke dalam Tumpukan Jaringan Cloud Native yang lengkap dengan mesh layanan aplikasi dan fungsi bidang data CNF di bagian bawah, mesh layanan Ligato, dan fungsi pengontrol CNF di bagian tengah, dan orkestrasi siklus hidup CN secara keseluruhan di bagian atas.



Gambar 6.10 Tumpukan jaringan cloud native.

Dengan adanya Cloud Native Network Stack, kita dapat mulai memikirkan bagaimana hal ini dapat membantu dalam memahami dan membangun solusi untuk jaringan otonom berbasis AI. Pertimbangkan hal berikut:

- ZTN. Kita dapat berasumsi bahwa banyak jaringan otonom akan berada dalam kondisi zero trust dan node akan menjadi titik akhir yang bertanggung jawab atas komunikasi dengan pihak tepercaya.
- Service mesh. Polimorfisme ini akan hadir dalam jaringan otonom. Jaringan ini dapat diorkestrasi atau ad-hoc. Kemungkinan akan ada berbagai bentuk identity mesh yang disatukan.
- Portabilitas cloud. Sistem CN dapat beroperasi di public cloud yang besar, enterprise cloud, fog cloud, IoT cloud, atau hibrida dari kombinasi apa pun secara umum, arsitektur yang sama, model yang sama, OSS yang sama. Jaringan otonom pasti memiliki sumber daya komputasi onboard atau di dekatnya. Instans layanan mikro dapat bersifat sementara atau berjangka panjang.

Diharapkan inovasi, pengembangan, dan penerapan aplikasi CN lebih lanjut akan memfasilitasi lebih banyak solusi jaringan berbasis AI/otonom.

OpenFog

Terdapat sejumlah tantangan yang ditimbulkan oleh proliferasi dan pertumbuhan jaringan IoT dan antisipasi kedatangan jaringan otonom yang canggih. Pengurangan ukuran, berat, dan daya tahan baterai membatasi sumber daya komputasi dan penyimpanan pada perangkat IoT kecil yang dalam beberapa kasus diperlukan untuk menjalankan program AI. Batas latensi yang tidak dapat diterima dapat terjadi jika jarak antara perangkat lokal dan entitas jarak jauh (misalnya, cloud terpusat atau mesin-ke-mesin) terlalu jauh. Jaringan mungkin mengalami tekanan jika aplikasi yang haus bandwidth muncul dalam jumlah besar meskipun ada kemajuan dalam infrastruktur jaringan 5G baru. Orang bertanya-tanya apakah sifat latensi rendah, aman, dan andal dari Internet taktik akan menjadi kenyataan.

Salah satu solusi menjanjikan yang akan membantu adalah menempatkan sumber daya komputasi, penyimpanan, dan jaringan sedekat mungkin dengan pengguna. Paradigma komputasi tepi ini telah ada sejak lama khususnya yang berkaitan dengan operator seluler yang telah menerapkan platform komputasi tepi seluler di dekat menara seluler. Dengan kepemilikan jaringan akses radio dan backhaul, operator seluler dapat menawarkan kepada pelanggan mereka serangkaian fungsi baru dan yang ditingkatkan, termasuk waktu respons yang lebih cepat, pembongkaran lalu lintas, dan kedekatan perangkat virtual.

Komputasi fog dibangun berdasarkan gagasan komputasi edge terdistribusi dan menanamkannya di dalam arsitektur cloud. Pada dasarnya, semua orkestrasi, fungsi, dan layanan cloud dapat diterapkan di lokasi mana pun antara cloud dan jaringan atau perangkat IoT. Tidak seperti komputasi edge, fog bersifat hierarkis dan dapat mendukung semua/setiap fungsi cloud, termasuk komputasi, jaringan, penyimpanan, dan CN. Komputasi fog bersifat agnostik terhadap jenis jaringan dan perangkat.



Untuk memfasilitasi pemahaman, penerimaan, pengembangan produk, interoperabilitas, dan penerapan, konsorsium OpenFog dibentuk. Komputasi kabut adalah "arsitektur horizontal tingkat sistem yang mendistribusikan fungsi komputasi, penyimpanan, kontrol, dan jaringan lebih dekat kepada pengguna di sepanjang kontinum cloud-to-thing..." Misinya adalah mengembangkan dan menghadirkan arsitektur komputasi kabut yang dapat digunakan oleh pihak-pihak yang berkepentingan, termasuk peneliti, industri, penyedia layanan/infrastruktur, dan banyak lainnya, untuk berinovasi dan membangun layanan baru seperti realitas tertambah CV.

Contoh kasus penggunaan dan aplikasi yang dapat memanfaatkan sumber daya yang lebih dekat yang disediakan oleh komputasi kabut adalah sebagai berikut:

- Realitas tertambah
- Jaringan Vehicle to Anything, infrastruktur pendukung, caching unduhan *over-the-air* (OTA)
- Keamanan dan pengawasan visual lokal
- Otomatisasi pabrik
- Jaringan pintar
- Aplikasi AI
- Hosting dan operasi aplikasi multi-penyewa proksimal
- Pembongkaran bandwidth jaringan
- Aplikasi keselamatan dan kesehatan latensi rendah

Konsorsium OpenFog telah mengembangkan dan menghadirkan dokumen arsitektur.

Integrasi Sistem

Integrasi Berkelanjutan/Pengiriman Berkelanjutan

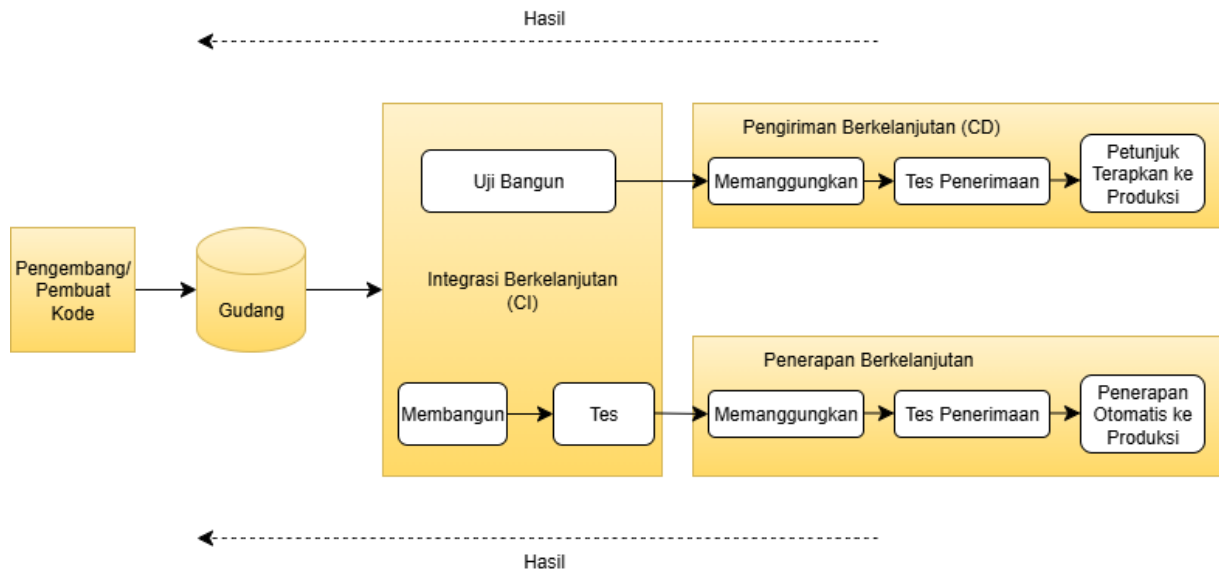
Devops mempercepat waktu antara pengembangan dan penerapan perangkat lunak melalui penggunaan alat dan otomatisasi untuk menyederhanakan proses pengodean, integrasi, pengujian, dan penerapan. Integrasi Berkelanjutan/Pengiriman Berkelanjutan (CI/CD) merupakan hal mendasar bagi misi dan keberhasilan proses DevOps.

Integrasi Berkelanjutan/Pengiriman Berkelanjutan dapat digambarkan sebagai sebuah alur kerja seperti yang ditunjukkan pada Gambar 6.11. Proses ini dimulai dengan pengembang yang mengomitmen perubahan kode mereka kembali ke cabang utama aplikasi yang terletak di repositori bersama. Integrasi berkelanjutan dipanggil ketika komitmen ke repositori bersama terdeteksi. Server CI khusus secara otomatis membuat citra build baru dan melakukan serangkaian pengujian otomatis terhadap build tersebut. Jika pengujian gagal, maka komitmen baru tersebut merusak sesuatu dan pengembang dapat memperbaikinya. Jika pengujian berhasil, maka pengembang dapat melanjutkan pekerjaan mereka.

Pengiriman berkelanjutan memperluas proses CI dengan mempersiapkan dan menguji build baru untuk penerapan produksi. Hal ini melibatkan pemindahan build yang telah diuji CI ke lingkungan staging tempat pengujian tambahan (misalnya, API, beban, dan keandalan) sebelum penerapan dapat dilakukan. Jika semuanya berjalan lancar, pengembang akan menandatangani secara manual dan kode siap untuk diterapkan ke lingkungan produksi



langsung. Penerapan berkelanjutan menghilangkan persetujuan pengembang secara manual, dan seluruh proses rilis, mulai dari kode hingga penerapan produksi, akan diotomatisasi.



Gambar 6.11 Pipeline CI/CD.

Virtualisasi Fungsi Jaringan Terbuka (VNF)

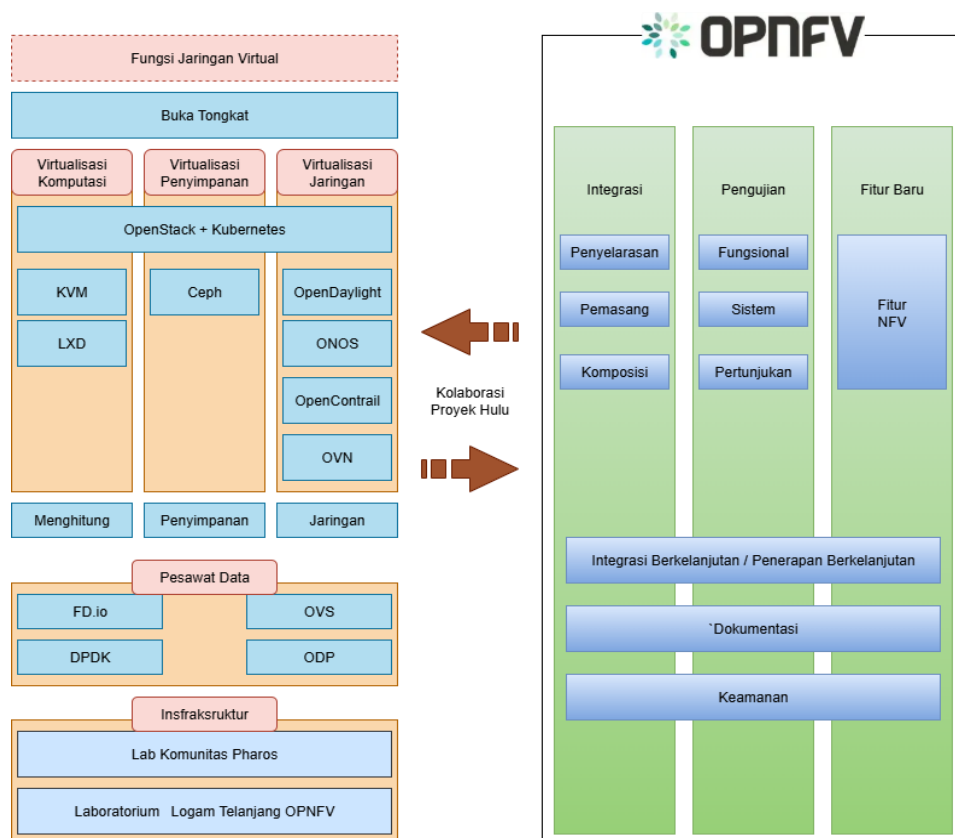
Menyusun, menguji, dan memvalidasi sistem cloud dalam persiapan untuk penerapan dapat menjadi proses yang rumit dan mahal. Ada banyak pertimbangan yang perlu dipertimbangkan, yang semuanya harus diperhatikan untuk memastikan penerapan yang sukses. Salah satu contoh berikut saja dapat menjadi tantangan untuk diselesaikan:

- Persyaratan yang dijabarkan di awal proyek dengan tujuan akhir yang telah ditentukan.
- Siklus hidup dan alur kerja sistem yang telah ditentukan.
- Integrasi platform dan komponen sumber terbuka.
- Instalasi, penerapan, pemutakhiran, dan pemecahan masalah VNF. Perlu diingat bahwa VNF adalah instans virtual dari fungsi jaringan atau layanan (misalnya, firewall, penyeimbangan beban), sehingga cara mengelola dan mengoperasikan komponen-komponen ini di lingkungan cloud menjadi penting.
- Menetapkan rantai layanan berbasis VNF dan konfigurasi overlay jaringan virtual dengan underlay jaringan fisik.
- Dokumentasi sistem yang lengkap.
- Penggunaan alat standar dan praktik terbaik untuk integrasi berkelanjutan/penyebaran berkelanjutan.
- Keamanan sistem.

Open NFV (OPNFV) adalah proyek sumber terbuka untuk mengatasi tantangan ini. Meskipun namanya mungkin menyiratkan platform sumber terbuka untuk penyebaran NFV oleh operator di cloud, kenyataannya tidak. OPNFV adalah platform referensi sumber terbuka yang tersedia bagi operator untuk merancang, mengintegrasikan, menguji, dan memvalidasi sistem cloud yang terdiri dari komponen NFV untuk mendukung penyebaran aktual (Gambar 6.12).



Pada dasarnya, berbagai komponen komputasi, penyimpanan, dan virtualisasi jaringan sumber terbuka terintegrasi dalam OPNFV. Ide di balik hal ini diilustrasikan pada Gambar 6.12. Rilis OPNFV tertentu akan mengidentifikasi proyek hulu apa saja yang didukung. Hulu berarti bahwa sebuah proyek sumber terbuka telah setuju untuk berkolaborasi dengan komunitas OPNFV dengan kata lain, operator yang menggunakan OPNFV untuk menguji dan memvalidasi sistem dapat memanfaatkan satu atau lebih proyek sumber terbuka (misalnya, Kubernetes dan kontainer) yang diidentifikasi di sebelah kiri sebagai bagian dari upaya pengembangan perangkat lunak mereka.



Gambar 6.12 Arsitektur OPNFV.

Di sebelah kanan adalah pilar-pilar fungsional yang didukung oleh OPNFV, termasuk integrasi dan pengujian beserta fitur-fitur baru. Integrasi berkelanjutan lintas komunitas digunakan untuk menjalankan siklus hidup OPNFV. Fitur-fitur tambahan OPNFV yang bernilai tambah mencakup penggunaan testbed otomatis, kerja sama SDO (misalnya, IETF, ETSI NFV, dan MEF), dan inisiatif keamanan yang didukung industri melalui Inisiatif Infrastruktur Inti yang didukung LF. Open NFV adalah sebuah proyek LF.

Peralatan

Peralatan untuk integrasi sistem perangkat lunak sumber terbuka (yaitu, CI/CD) bergantung pada alur kerja devops yang digunakan dalam lingkungan rekayasa. Github adalah repositori bersama sumber terbuka yang digunakan di banyak lingkungan pengembangan. Server CI Jenkins mendukung ratusan plug-in dan cukup fleksibel untuk beroperasi di semua jalur CI/CD. Netflix Spinnaker mendukung pengiriman berkelanjutan untuk penerapan CN. Dan



terakhir, *Amazon Web Services (AWS)* dan *Google*, di antara operator cloud lainnya, menawarkan berbagai alat dan platform CI/CD.

Pembelajaran Mesin

Banyak pustaka ML sumber terbuka yang tersedia dan digunakan saat ini. Aplikasi transportasi pintar PNDA menggunakan Theano. Aplikasi PNDA lainnya untuk analitik keamanan menggabungkan Apache SPOT. Aplikasi lainnya termasuk pytorch dan Apache Spark MMLlib. Tensorflow yang diciptakan oleh Google mungkin merupakan pustaka ML yang paling terkenal. Para pengembang merasa nyaman menggunakannya, menikmati dukungan komunitas yang luas, dan hadir dalam banyak penawaran layanan ML cloud. Penggunaan pembelajaran mesin akan terus tumbuh seiring jaringan menjadi lebih cerdas. Kita dapat terdorong oleh fakta bahwa vendor sedang membangun dan mengirimkan aplikasi ML; operator akan memperluas penggunaan teknologi ML mereka.

Dan berkat platform otomatisasi seperti ONAP, ML dapat digunakan untuk mengurangi waktu pemrosesan dari input data hingga output dan penerapan kebijakan yang dihasilkan. Kita juga melihat peningkatan dalam penelitian AI dan, secara anekdot, lebih banyak insinyur muda memasuki sektor swasta dengan minat, keterampilan, dan pelatihan ML.

Kecerdasan Umum Buatan

Metode AI yang ada dianggap sebagai "AI sempit" (alias AI lemah). Ini berarti metode tersebut dapat melakukan satu tugas dengan sangat baik, seperti rekomendasi pembelian atau prakiraan cuaca. Kecerdasan umum buatan (AGI) dianggap sebagai "AI kuat". Tanpa membahas detailnya, AGI sangat mirip dengan kecerdasan dan penalaran manusia. AGI dapat mengatasi masalah berekor panjang masalah yang belum pernah ditemukan dan bahkan mungkin tampak seperti bentuk kreativitas atau penalaran. Untuk mengilustrasikan perbedaan antara AI sempit dan kuat, mari kita asumsikan terdapat jaringan otonom bot keamanan lingkungan (NSB).

Tujuan mereka adalah berpatroli di lingkungan tersebut dan mengawasi aktivitas mencurigakan. Jika ada aktivitas yang terdeteksi, mereka akan memberi tahu pihak berwenang. Dengan AI sempit, NSB dilatih untuk mencari aktivitas spesifik yang diklasifikasikan sebagai mencurigakan dan tidak lebih. Jika terjadi pembobolan dan pencuri menggunakan taktik yang tidak dilatih untuk dideteksi oleh NSB, kejahatan tersebut tidak dilaporkan.

Di sisi lain, AI kuat memberdayakan NSB untuk berpikir, bernalar, dan belajar. Ketika menemukan pencuri yang menggunakan taktik baru mereka, NSB, yang belum pernah melihat hal ini sebelumnya, mungkin bertanya-tanya apa yang sebenarnya terjadi. NSB dapat mencari informasi dari sumber data lain (misalnya, catatan kejahatan polisi, CCTV dari lingkungan lain yang menjadi korban pembobolan), membandingkan aktivitas keluar masuk lingkungan yang normal dengan aktivitas baru ini, atau menghubungi tetangga untuk menanyakan dan memastikan bahwa apa yang diamati tidak biasa. Dengan mempertimbangkan semua hal tersebut, NSB dapat mencapai kesimpulan logis bahwa ada sesuatu yang tidak beres di sini dan melaporkannya kepada pihak berwenang. NSB akan belajar dari pengalaman baru ini dan akan proaktif dalam menjangkau dan mencari informasi untuk memperkuat pengetahuan ini.



Sistem Penalaran Non-Aksiomatik

Kecerdasan umum buatan berdasarkan Sistem Penalaran Non-Aksiomatik (NARS) bersifat non-aksiomatik dalam arti bahwa data empiris dapat sewaktu-waktu mengesampingkan keyakinan sebelumnya dalam sistem. Hal ini merupakan persyaratan bagi sistem apa pun yang harus beroperasi di dunia nyata secara waktu nyata dengan sumber daya komputasi terbatas. Setiap data sensor baru yang mengesampingkan keyakinan yang ada akan memperbarui pengetahuan yang ada. Dapat dibayangkan bahwa sensor yang beroperasi jauh di dalam jaringan otonom akan bergantung pada kemampuan ini. Terdapat versi sumber terbuka NARS dan komunitas peneliti serta makalah yang diterbitkan yang terus berkembang.

OpenCog

OpenCog membedakan dirinya dengan memanfaatkan Teori Sinergi Kognitif, yang dapat dianggap menyatukan berbagai jenis memori. Jenis memori yang diintegrasikan OpenCog meliputi deklaratif (misalnya, bola merah di atas meja lebih besar daripada bola biru), prosedural (prosedur menumpuk balok di atas balok lain), sensorik, episodik (misalnya, rangkaian tindakan yang dilakukan Bill ketika ia membangun menara kemarin), atensi (rangkaian objek yang tampaknya penting saat ini dalam konteks permainan yang sedang dimainkan Bob dan Bill), dan intensional (tujuan membangun menara yang tidak runtuh). Masing-masing jenis pengetahuan ini memiliki mekanisme penciptaan pengetahuan yang berbeda dan saling terhubung sedemikian rupa untuk saling membantu mengatasi ledakan kombinatorial spesifik jenis memori.

OpenCog mengatasi salah satu isu inti dalam penskalaan sistem pakar. Jika kita mempertimbangkan potensi Web 2.0 dengan halaman web yang dianotasi secara semantik, satu isu yang jelas adalah kebutuhan manusia untuk menambahkan informasi semantik ini. OpenCog mengimplementasikan berbagai mekanisme penciptaan pengetahuan dan menghubungkannya sedemikian rupa untuk menghindari keterbatasan spesifik jenis memori. OpenCog adalah komunitas lain yang tertarik untuk meneliti dan mengeksplorasi AGI.

Ringkasan Open Source dan Kecerdasan Buatan/Jaringan Otonom

Banyak yang dibahas di bagian ini tentang berbagai jenis perangkat lunak sumber terbuka. Gambar 6.4 di awal bagian ini memaparkan komponen dan posisinya dalam tumpukan. Dari hal tersebut dan diskusi selanjutnya, terdapat beberapa hal menarik yang perlu dipertimbangkan:

- Otomatisasi adalah kunci untuk menghadirkan layanan baru dengan cepat. OPNFV dan CI/CD menyederhanakan penerapan VNF dan alur rilis perangkat lunak.
- Kerangka kerja referensi wajib. Operator cloud skala besar dapat memodelkan arsitektur dan alur desain/penerapan/otomatisasi berdasarkan ONAP, sehingga menghemat waktu dan biaya.
- Analisis big data harus ditingkatkan kapasitas dan pemrosesannya agar dapat mengimbangi. PNDA dan SNAS adalah dua solusi yang dirancang untuk memenuhi tantangan tersebut.
- CN memanfaatkan layanan mikro terkontainerisasi untuk penerapan aplikasi yang skalabel, tangguh, dan aman.



Dan terakhir, solusi AI/ML sumber terbuka tersedia dan sedang digunakan. Kami memperkirakan penggunaannya akan meningkat karena kebutuhan (diperlukan dalam jaringan otonom) dan karena sumber daya serta keahlian menjadi lebih mudah tersedia untuk pelatihan dan pengembangan model. Operator cloud publik yang dibahas di bagian selanjutnya akan membantu dalam hal ini.

6.4 OPERATOR CLOUD, KECERDASAN BUATAN, DAN JARINGAN OTONOM

SP Cloud berperan penting dalam mengaktifkan dan mendukung jaringan otonom berbasis AI. Mereka memiliki kemampuan pengumpulan dan pemrosesan data yang masif untuk mendukung aplikasi AI (berserta sumber daya komputasi), menggunakan, melakukan pelatihan model, dan menghasilkan wawasan, perilaku, serta prediksi. Penyedia cloud melakukan riset dan pengembangan ekstensif terhadap fungsi-fungsi AI baru. Produk mereka mendukung AI-as-a-Service yang tersedia saat ini. Mereka saat ini dan akan terus menyediakan lingkungan pengembangan menggunakan perangkat dan kerangka kerja AI sumber terbuka terbaru. Perangkat dan lingkungan ini dapat mempercepat kemajuan menuju pembangunan sistem loop tertutup otomatis dalam jaringan otonom.

Penyedia cloud dapat mengembangkan kemitraan bagi hasil dengan pihak lain untuk menciptakan pasar baru dan menawarkan layanan baru. Pengetahuan dan pengalaman yang luas di bidang AI dapat ditawarkan sebagai semacam layanan konsultasi untuk membantu pelanggan memulai upaya mereka. Tidak dapat disangkal bahwa otomatisasi, operasi, produk, dan aplikasi yang terinspirasi AI saat ini sedang meletakkan fondasi bagi arsitektur dan layanan cloud baru yang digerakkan oleh AI. Rekrutmen besar-besaran oleh operator cloud untuk insinyur perangkat lunak AI dan ilmuwan data sedang berlangsung dan diperkirakan tidak akan berkurang dalam waktu dekat bahkan mungkin selamanya.

Amazon Web Services

Dengan cadangan sumber dayanya yang besar dan tampaknya tak terbatas, AWS secara agresif mengejar strategi "pembelajaran mesin untuk semua orang" di berbagai pasar, yang banyak di antaranya akan berdampak pada domain, termasuk jaringan otonom.

Bagi perusahaan dan organisasi yang memulai perjalanan AI, AWS menawarkan beragam penawaran unik:

- API untuk layanan yang siap untuk ML, termasuk penerjemahan ucapan, visi komputer, dan interaksi chatbot.
- Beban kerja big data AI dapat diintegrasikan dengan solusi platform lain, seperti penyimpanan objek S3 dan gudang data Redshift untuk menyelidiki dan mengembangkan aplikasi baru.
- Layanan pembelajaran mesin Amazon, yang menyediakan alat visualisasi dan panduan untuk membantu memandu pengguna awam melalui proses pembuatan model ML tanpa harus mempelajari algoritma dan teknologi ML yang rumit.
- Kerangka kerja ML yang terkenal, termasuk TensorFlow dan Apache MXNet, dikemas dalam perangkat virtual khusus pembelajaran mendalam yang disebut Amazon

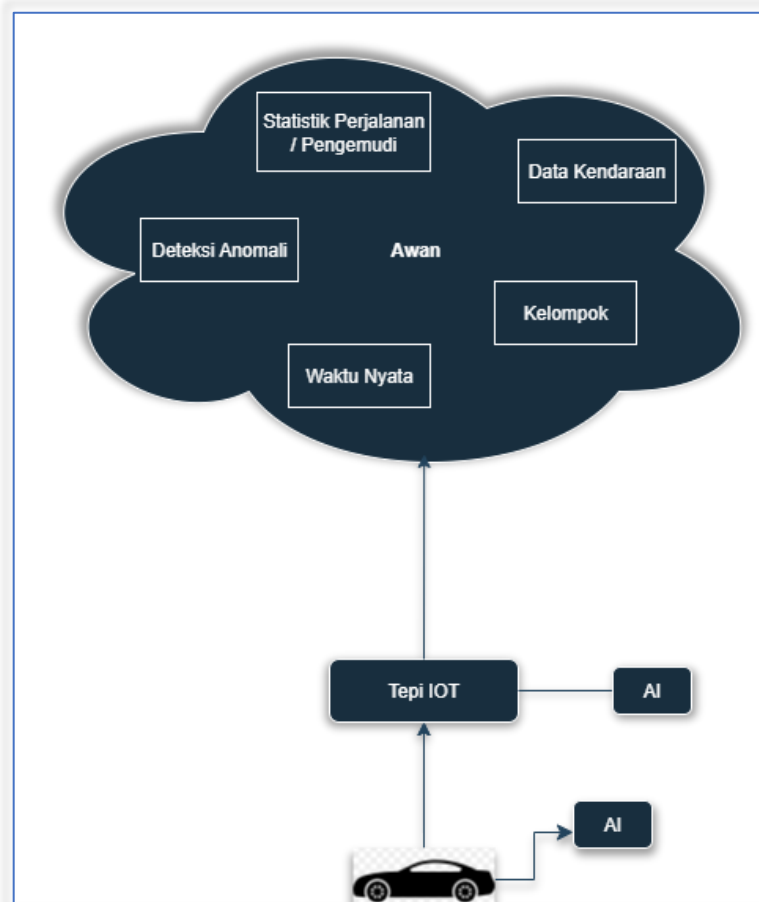


Machine Images. Klaster Amazon Machine Images dapat diterapkan di seluruh spektrum opsi komputasi, mulai dari sumber daya GPU hingga prosesor siap pakai.

- SageMaker memudahkan peneliti dan pengembang untuk membangun, melatih, menguji, dan menyetel model dalam aplikasi dan alur kerja.

Tentu saja, semua ini dan lebih banyak lagi dapat disajikan dalam platform "bayar sesuai pemakaian" berupa layanan dan dukungan AWS yang aman dan fleksibel.

Contoh lain dari teknologi AI AWS yang sedang diterapkan adalah penerapan layanan keamanan mutakhir. Amazon Macie, yang telah diluncurkan pada penyimpanan AWS S3, adalah layanan yang memanfaatkan ML untuk secara otomatis menemukan, mengklasifikasikan, dan melindungi data pengguna yang sensitif. Macie menggunakan pemrosesan bahasa alami (NLP) untuk mengurai data pengguna yang disimpan di S3 guna mengidentifikasi informasi identifikasi pribadi yang sensitif, seperti kunci SSN dan SSH. Layanan ini secara otomatis mengklasifikasikan data ke dalam tingkat risiko tinggi hingga rendah, lalu memulai pemantauan akses data berisiko tinggi. Jika analisis pola akses data menemukan anomali yang mengindikasikan potensi akses tidak sah, Macie akan memperingatkan pengguna melalui layanan CloudWatch-nya. Sistem ini berpotensi menjadi penyelamat, terutama jika kredensial akses tidak sah untuk pengguna dengan izin kerahasiaan tinggi disimpan di lokasi yang tidak dijaga keamanannya.



Gambar 6.13 Arsitektur cloud kendaraan terhubung AWS.



Amazon menjangkau kehidupan kita sehari-hari dan kemampuan AI mereka akan meluas ke komunikasi dan transportasi manusia/mesin. Ada Alexa, asisten rumah yang menggunakan AI untuk meningkatkan pengenalan suara manusia. Drone tanpa awak yang mengirimkan barang dari gudang Amazon tidak akan 100% dikendalikan oleh pilot manusia. IoT dan mobil terhubung yang ditunjukkan pada Gambar 6.13 dapat menggunakan komputasi lokal Greengrass dan solusi pemrosesan ML mereka dan digabungkan dengan aplikasi yang dihosting cloud (pengumpulan data, analitik, deteksi anomali, dan operasi pengguna) dapat memaksimalkan utilitas, menurunkan biaya, dan memberikan pengalaman manusia terbaik.

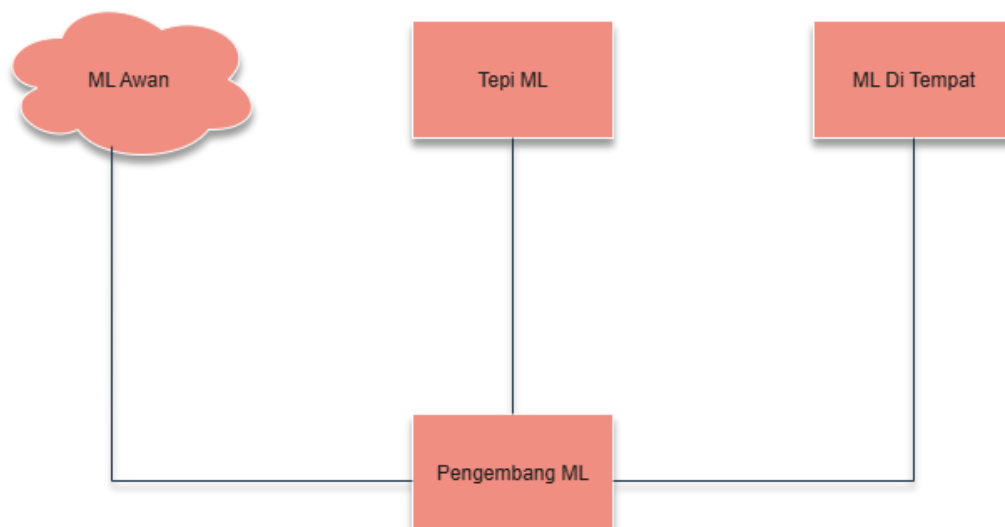
Sebagai operator cloud terbesar dalam hal pengguna dan layanan yang ditawarkan, AWS akan memainkan peran besar seiring berkembangnya jaringan otonom. Mereka saat ini menawarkan dan akan terus mengembangkan dan menghadirkan kemampuan ML baru. Jelas mereka bermaksud untuk memproyeksikan dan mendistribusikan banyak aplikasi dan layanan ini sedekat mungkin dengan pengguna.

Microsoft Azure

Platform ML Microsoft Azure menawarkan seperangkat alat, layanan, dan infrastruktur yang lengkap untuk memungkinkan siklus pengembangan yang lebih cepat dan sederhana bagi pengembang AI. Platform ML Azure mendukung hal-hal berikut:

- Machine Learning Workbench
- Layanan Eksperimen Machine Learning
- Layanan Manajemen Model Machine Learning
- Pustaka Machine Learning untuk Apache Spark
- Alat Visual Studio Code untuk AI

Gambar 6.14 mengilustrasikan posisi layanan-layanan ini dengan layanan ML ditampilkan di sebelah kiri dan opsi pelatihan dan penerapan opsional di sebelah kanan.



Gambar 6.14 Layanan machine learning Microsoft Azure.



Visual Studio Code Tools for AI digunakan untuk pembuatan aplikasi AI. VS Code Tools terintegrasi dengan layanan ML Azure, menyediakan tampilan terpadu untuk pelatihan dan penerapan model. Pengembang dapat memulai proyek Microsoft Cognitive Toolkit baru dari VS Code, mengirimkannya ke layanan Pelatihan AI Azure Batch untuk pelatihan, lalu menerapkan model yang telah dilatih secara lokal, atau ke cloud. Hal ini menjawab kebutuhan pengembang akan Lingkungan Pengembangan Terpadu (IDE) AI generasi berikutnya, yang menyediakan alur kerja desktop ke cloud yang lengkap untuk rekayasa perangkat lunak AI.

Solusi menarik lainnya yang sedang dikembangkan Azure adalah platform IoT Edge, yang bertujuan untuk memperluas kecerdasan cloud ke perangkat edge. Dengan memindahkan analitik data berbasis AI dan pengambilan keputusan ke edge, perusahaan dapat mengurangi biaya bandwidth, sekaligus tetap mendapatkan manfaat dari wawasan yang diperoleh dari data IoT.

Platform Microsoft CV merupakan peluang lain bagi Azure untuk memperluas dukungan cloud mereka ke jaringan otonom berbasis CV. Terdapat banyak ruang untuk inovasi, termasuk menawarkan layanan pemeliharaan prediktif berbasis ML dan fungsi bantuan pengemudi secara real-time. Namun, akan ada banyak tantangan yang harus diatasi seiring peluncuran layanan ini. Pertama, keamanan harus terjamin dan harus menyeluruh, mulai dari komponen di dalam kendaraan hingga ke cloud. Kedua, platform cloud harus mampu menangani komunikasi terencana (misalnya, pembaruan OTA, unggahan telematika normal) dan tidak terencana (misalnya, lonjakan telemetri berbasis peristiwa, patch OTA yang tidak terencana).

Google Cloud

Google adalah pemimpin industri dalam mengembangkan dan memperkenalkan inovasi dan solusi AI. Google Cloud menghosting dan menyediakan banyak kemampuan menarik berikut:

- Cloud AutoML. Ini adalah rangkaian produk berbeda yang menyederhanakan pelatihan model ML dan pengembangan aplikasi. Layanan ini mendapatkan popularitas dengan klaim bahwa autoML adalah kode ML yang dapat menghasilkan kode ML.
- Cloud Machine Learning Engine. Ini adalah layanan terkelola untuk membangun dan melatih model pada data apa pun, jenis apa pun, dan ukurannya berapa pun. Model yang dihasilkan Tensorflow dapat dilatih pada kluster sumber daya berskala besar sebagai persiapan untuk penerapan aplikasi analisis prediktif daring atau batch.
- Layanan ini beroperasi bersama fitur Google Cloud lainnya untuk penerapan produksi. Selain perangkat lunak, Google juga berinvestasi dalam perangkat keras yang dioptimalkan untuk AI. Google Cloud telah menerapkan *Tensor Processing Unit* (TPU) yang dirancang khusus untuk performa dan skalabilitas ML yang maksimal. Google mendukung komunitas riset dengan menyediakan layanan pelatihan gratis untuk program-program tertentu di TensorFlow Research Cloud, yang mengoperasikan kluster yang terdiri dari 1000 Cloud TPU.

Langkah awal Google dalam mobil swakemudi telah mendorong industri dan memberi mereka wawasan berharga tentang teknologi pengoperasian kendaraan otonom. Terdapat siklus berkelanjutan pengumpulan set data, pelatihan model, dan pembaruan model yang



terjadi antara kendaraan dan cloud. Data mentah yang dikumpulkan dari sensor, radar, visi komputer, dan sumber data lainnya harus dikirim ke cloud untuk pelatihan model.

Volume set data tertinggi terjadi selama pengembangan dan pengujian. Namun, siklus pelatihan model berkelanjutan tidak akan hilang, sehingga cloud harus menyediakan sumber daya yang memadai untuk penyerapan set data, baik data terkompresi maupun data mentah. Cukuplah untuk mengatakan bahwa Google akan terus mendorong solusi berbasis AI agar lebih dekat dengan pengguna. Platform Google Home adalah salah satu contohnya investasi besar mereka dalam kendaraan otonom menandakan hal lain.

IBM Cloud

Layanan cloud ML IBM dirancang berdasarkan platform Watson-nya. Watson ML menawarkan banyak fitur yang sama dengan operator cloud lainnya, termasuk pembuatan/pelatihan model, pustaka sumber terbuka, dan alat untuk membangun alur kerja AI guna mengekstrak dan memvisualisasikan wawasan.

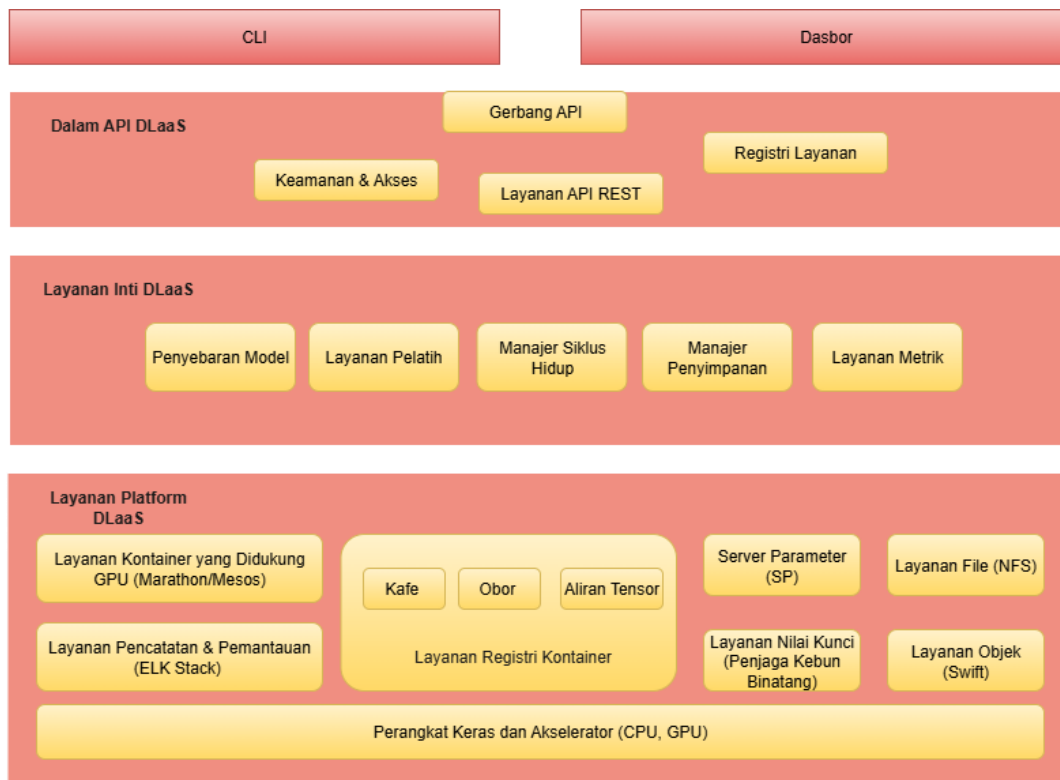
Platform Watson menyediakan API Layanan Kognitif dan Pemrosesan Bahasa Alami yang dibahas di Bab 2. Membangun alur kerja menggunakan API bisa jadi rumit, sehingga IBM menawarkan layanan AppConnect. Ini adalah alat konfigurasi grafis bagi pengguna dengan berbagai tingkat keahlian untuk membangun aplikasi ML.

Misalnya, seorang pemasar dapat mengembangkan aplikasi yang menganalisis nada bicara pelanggan dalam sebuah pesan dan menentukan tindakan selanjutnya. Jika pelanggan menggunakan bahasa positif, permintaan survei yang ramah dapat muncul. Namun, jika Watson menentukan nadanya negatif, aplikasi akan menemukan manajer akun pelanggan dan memberi tahunya melalui IBM Message Hub, menyarankan agar mereka menghubungi pelanggan untuk menyelesaikan masalah tersebut. Jika digunakan dengan benar, aplikasi yang didukung Watson akan meningkatkan efisiensi operasional dan meningkatkan kepuasan pelanggan.

IBM mendukung gagasan bahwa kompleksitas, biaya, dan persyaratan pemrosesan ML dapat dijalankan secara efisien dalam layanan berbasis cloud. *Deep-Learning-as-a-Service* (DLaaS) menyediakan fleksibilitas dan performa yang tersedia dari cloud dengan berbagai alat, kerangka kerja ML, model, dan data pelatihan untuk mengembangkan, menguji, dan mengoperasikan aplikasi ML. Alur kerja DLaaS dasar meliputi persiapan model, pengunggahan model dan data pelatihan ke cloud, pembuatan dan pemantauan pekerjaan pelatihan, lalu pengunduhan model yang telah dilatih.

Arsitektur DLaaS ditunjukkan pada Gambar 6.15. Terdapat API di atasnya untuk akses UI dan aplikasi. Layanan inti DLaaS adalah serangkaian layanan mikro yang mendukung alur kerja pekerjaan pelatihan dari awal hingga akhir. Layanan platform DLaaS merupakan fondasi untuk menjalankan dan mengelola pekerjaan pelatihan. Ketika pekerjaan pelatihan DLaaS dimulai, *Life Cycle Manager* (LCM) akan meminta Kubernetes untuk mengalokasikan sumber daya. Data disalurkan melalui serangkaian layanan mikro pembelajar paralel yang mengimplementasikan kerangka kerja yang diinginkan pengguna. Secara berkala, bobot yang diperoleh dari komputasi ML dimasukkan/ditarik dari server parameter. Zookeeper adalah penyimpanan data kunci-nilai yang digunakan untuk mengatur siklus hidup pelatihan.





Gambar 6.15 Arsitektur IBM DLaaS.

Alibaba

Dipilih sebagai platform AI nasional Tiongkok untuk mengembangkan "City Brain", Alibaba Cloud (Aliyun) adalah salah satu pelopor dalam menyediakan AI sebagai layanan. Pada Oktober 2016, Alibaba memprakarsai proyek Hangzhou City Brain, yang bertujuan untuk menerapkan pembelajaran mesin (ML) guna membantu pemerintah Hangzhou mengatasi masalah lalu lintas yang signifikan. Proyek ini berbasis teknologi pengenalan gambar, memanfaatkan kamera CCTV untuk melacak aktivitas mobil dan manusia. Melalui analisis perilaku dan masukan data, sistem menganalisis arus lalu lintas serta mendeteksi dan memprediksi kemungkinan kejadian yang mengakibatkan kemacetan. Untuk menutup lingkaran setan, pengetahuan yang dikumpulkan digunakan untuk menyesuaikan lampu lalu lintas di kota dengan tujuan meningkatkan efisiensi lalu lintas secara real-time. Ini adalah contoh nyata yang menggambarkan potensi AI dalam mendukung jaringan IoT otonom dari cloud. Selain City Brain, Alibaba Cloud menyediakan layanan AI lainnya, mulai dari analisis tren media sosial, statistik real-time untuk streaming NBA, hingga penilaian risiko keuangan dan deteksi anomali dalam produksi industri.

Tencent

Tencent Cloud telah menjadi pemain kuat dalam teknologi AI, seperti visi komputer dan NLP. Serupa dengan Alibaba, Tencent berfokus pada penyediaan AI sebagai layanan pada platform cloud-nya, dengan sebagian besar layanannya dikhususkan untuk bidang analisis citra medis, AI gim video, dan media sosial. Salah satu contoh menarik dari layanan tersebut adalah API Image to Speech mereka. API ini digunakan oleh aplikasi untuk menganalisis konten



gambar dan mengubahnya menjadi ucapan bahasa alami, sehingga menyediakan aksesibilitas bagi mereka yang memiliki gangguan penglihatan. Layanan menarik lainnya adalah layanan analisis opini publik "Wenzhi", yang didasarkan pada repositori data untuk satu miliar pengguna. Dengan bantuan platform Elastic MapReduce dari Tencent Cloud, layanan ini mampu menganalisis opini publik dan peristiwa di internet secara waktu nyata.

6.5 TANTANGAN KE DEPAN

Pasar dan Lanskap Baru

Memprediksi masa depan memang tidak mudah. Siapa yang bisa meramalkan bagaimana peralihan paket eksperimental di akhir tahun 1960an berevolusi menjadi Internet Global yang mengubah dunia, baik saat ini maupun di masa mendatang? Hal yang sama berlaku untuk ponsel pintar perangkat yang tampaknya kita gunakan terus-menerus 24/7 ini memiliki komputasi, penyimpanan, dan memori yang jauh lebih besar daripada pendahulunya, mainframe seukuran ruangan berpendingin udara. Inovasi dan penemuan lain yang belum terbayangkan sebelumnya antara lain media sosial, streaming musik, unduhan video, dan mata uang kripto virtual tanpa bank.

Prakiraan dan proyeksi pasar menawarkan gambaran pada suatu titik waktu. Berdasarkan apa yang telah dibahas dalam bab ini dan di bagian lain buku ini, bab ini mungkin cocok untuk memberikan komentar tentang bagaimana jaringan otonom berbasis AI dapat memengaruhi pasar dan lanskap teknologi baru yang kita perkirakan di masa mendatang.

- Cloud tetap berada di puncak rantai makanan data. Mereka memiliki sumber daya komputasi, penyimpanan, memori, dan jaringan yang unggul semuanya diperlukan untuk mengonsumsi, menyimpan, dan memproses data dalam jumlah besar. Di sinilah konten berada dan tempat kita sebagai pengguna melalui ponsel pintar atau peralatan rumah tangga mengaksesnya. Aplikasi media sosial berada di cloud, begitu pula aplikasi perusahaan dan infrastruktur TI. Aplikasi AI baru yang canggih beroperasi di cloud. Waktu dan biaya untuk desain dan penerapan layanan cloud ini dan yang akan datang akan berkurang drastis melalui otomatisasi dan virtualisasi berkat platform sumber terbuka baru seperti ONAP dan Kubernetes. Platform, devops, toolchain, dan keterampilan rekayasa perangkat lunak manusia dan otomatis akan mendorong inovasi baru, pasar baru, dan peluang produk.
- Data adalah raja. Jelas ada pertumbuhan data yang sangat besar sebagaimana ditunjukkan dalam setiap proyeksi lalu lintas. Ini hanya akan meningkat seiring dengan semakin banyaknya perangkat IoT, sensor, dan node otonom yang online. Diperkirakan bahwa masing-masing CV akan menghasilkan 0 (gigabyte) data untuk konsumsi internal dan ekspor ke cloud. Di sisi lain, unduhan OTA berbasis cloud, peta 3D, dan infotainment streaming akan menjadi hal yang umum. Operator cloud akan menerapkan sensor pintar, jaringan AI, dan algoritma untuk mengumpulkan dan menganalisis segala jenis lalu lintas aplikasi. Pemrosesan dan monetisasi data ini akan menghadirkan banyak peluang dan pasar baru.



- Menuju Edge. AI cloud terpusat yang serba tahu dan serba melihat yang bekerja pada data jaringan otonom yang diunggah mungkin kurang cocok untuk beberapa aplikasi jika latensi tinggi dan sumber daya bandwidth rendah. Untuk mengatasi masalah ini, komputasi edge dapat menawarkan AI tingkat menengah yang terletak di antara perangkat AI cloud dan IoT. Dengan sumber daya AI lokal atau regional, perangkat IoT dapat memperoleh keluaran dan wawasan yang lebih cepat berdasarkan perangkat dan kondisi lokal. Inilah yang didukung oleh konsorsium OpenFog, dan argumennya sangat meyakinkan.
- CV dilengkapi dengan daya komputasi yang lebih besar. Aplikasi AI kendaraan dapat memanfaatkan data lokal untuk deteksi anomali dan memungkinkan pengendaraan yang lebih aman dan efisien. Jelas bahwa kendaraan otonom telah dan akan terus menggunakan aplikasi AI di dalam kendaraan.

Mari kita terapkan AI pada ponsel pintar. Ada beberapa versi Tensorflow untuk ponsel dan bahkan chip AI, dengan contoh yang paling menonjol adalah perangkat keras Apple iPhone untuk aplikasi pembelajaran mesin (ML) "on-the-fly" seperti pengenalan wajah. Dalam contoh lain, pengguna ponsel pintar Android pertama kali di India sering kali menghabiskan ruang penyimpanan perangkat dengan mengirimkan pesan "selamat pagi" kepada teman dan keluarga mereka setiap pagi. Untuk mengatasi masalah ini, Google menggunakan AI untuk melatih aplikasi ponsel pintar agar dapat mencari dan menghapus data yang tidak relevan, sehingga mengosongkan ruang penyimpanan ponsel pintar.

Industri ini semakin berupaya memperluas lanskap AI. LF mendukung proyek Acumos berbasis ML. OpenAI adalah proyek lain yang mengkaji cara untuk memperkenalkan AGI guna memecahkan masalah nyata. Tak perlu dikatakan lagi, AI menjadi arus utama berkat sumber terbuka. Hal ini akan mempercepat penerapan jaringan otonom.

Rekayasa Perangkat Lunak

Tantangan mendasar bagi rekayasa perangkat lunak AI terletak pada sifat kompleks modelnya. Perangkat lunak deterministik dapat diuji secara offline. Dengan model ML yang terlatih, hampir mustahil untuk memprediksi secara tepat bagaimana aplikasi akan berperilaku sebelum divalidasi dalam lingkungan produksi. Beberapa petunjuk tentang akurasi model mungkin ada selama fase pelatihan; namun demikian, ketidakpastian tetap ada mengenai bagaimana model akan bekerja pada data nyata, yang merupakan tantangan besar.

Metode pengembangan Agile yang populer, yang melibatkan sprint dua minggu dari tim pengembangan ke tim operasi, akan sulit untuk sepenuhnya memeriksa kinerja dan akurasi model dalam jangka waktu tersebut, terutama jika set data pelatihan yang memadai tidak memadai. Meskipun alur kerja devops telah diotomatisasi sampai batas tertentu, banyak proses masih dilakukan oleh manusia, sehingga mengganggu aspek rekayasa perangkat lunak yang lebih kreatif dan memuaskan. Kecerdasan buatan dapat meringankan para insinyur dari beberapa pekerjaan rutin dalam siklus hidup devops ini. Untungnya, mereka yang berinvestasi dalam rekayasa perangkat lunak AI dan manfaatnya yang luar biasa memberikan solusi seperti berikut:



- Alur kerja dan rantai alat otomatis baru yang membentuk jalur "kendaraan dengan okupansi tinggi" untuk mengurangi putaran umpan balik devops sedang dikembangkan dan diterapkan. Meskipun dapat diterapkan untuk semua upaya pengembangan, devops AI juga dapat memperoleh manfaat.
- Layanan berbasis ML, termasuk pengembangan model, pengujian, dan validasi yang semuanya didukung dan ditawarkan oleh operator cloud yang dibahas di bagian sebelumnya, akan membuat perbedaan besar. Layanan ini memiliki alat dan sumber daya yang disesuaikan (komputasi dan penyimpanan) yang diperlukan untuk menguji dan memvalidasi kinerja serta akurasi model. Hal ini akan melindungi pengembang aplikasi dari kompleksitas dan misteri ML dan membantu meringankan pekerjaan padat karya para insinyur perangkat lunak, sehingga memungkinkan inovasi dan pengembangan perangkat lunak AI yang lebih cepat.

Seiring bertambahnya populasi insinyur perangkat lunak dan devops AI, kita dapat mengharapkan lebih banyak inovasi di bidang ini. Tidak diragukan lagi mereka akan bergantung pada otomatisasi berbasis AI dan manajemen siklus hidup.

Identitas dan Privasi

Untuk keperluan diskusi ini, mari kita asumsikan cloud mengacu pada sumber daya komputasi apa pun yang dapat diakses oleh pengguna dengan kredensial yang diperlukan. Ini mencakup cloud klasik, sistem komputasi edge, dan yang ada di dalam node otonom seperti kendaraan.

Bentuk manajemen identitas cloud yang paling umum adalah kriptografi asimetris, di mana kunci privat merupakan satu-satunya sumber identifikasi. Di dunia nyata, pengguna tidak selalu mengamankan kunci dan kata sandi. Oleh karena itu, peretas memanfaatkan fakta tersebut. Setelah diterapkan di bidang identifikasi seperti pengenalan wajah, AI dapat berguna dalam memverifikasi identitas cloud. Kunci penerapan verifikasi berbasis AI adalah mempertimbangkan lebih dari sekadar kata sandi dan kunci. Dengan terus memantau perilaku akses pengguna dan membedakan antara pola perilaku aman dan berbahaya, AI dapat menemukan potensi penipu, meskipun mereka mungkin memiliki kredensial dan hak akses yang benar. Setelah perilaku yang berpotensi berbahaya teridentifikasi, pengguna akan diberi tahu dan tindakan yang tepat akan diambil. Akurasi sistem ini belum diungkapkan saat ini, tetapi sistem ini memberikan lapisan keamanan tambahan bagi mereka yang enggan menerapkan praktik terbaik.

Proses yang dijelaskan sebelumnya menyediakan cloud hosting dengan akses ke data pengguna. Ini menimbulkan pertanyaan: dapatkah kita mempercayai cloud AI dengan data pengguna kita?. Kita melihat iklan bertarget yang berasal dari AI sepanjang waktu, jadi ada manfaat finansial. Privasi telah ditandatangani setelah persyaratan layanan diterima. Ini memberikan hak bagi cloud untuk memasukkan atau memproses data pengguna untuk tujuan apa pun, selama tetap berada di dalam cloud. Namun, tren terbaru dalam AI hanya memperburuk masalah privasi, karena data adalah sumber daya yang berharga, yang dimanfaatkan perusahaan untuk mendorong model AI mereka. Model-model ini kemudian akan dibungkus menjadi produk (misalnya, asisten rumah atau CV), yang dijual di cloud melalui



API. Meskipun produk-produk ini tidak secara langsung mengekspos privasi pengguna, mereka berisi wawasan yang diperoleh dari data pengguna.

Meskipun masih spekulatif, mungkin saja untuk membangun prosedur kontrol akses privasi yang tepat dengan bantuan AI dan blockchain. Pertama, alih-alih menyimpan data pengguna dalam basis data terpusat di cloud pribadi perusahaan, blockchain dapat dibangun antara perusahaan dan pengguna untuk menyimpan data pribadi terenkripsi, yang hanya dapat diakses melalui kontrak pintar yang telah ditetapkan. Kedua, alih-alih perjanjian persyaratan layanan, cloud dapat menggunakan AI untuk mengajukan pertanyaan kepada pengguna dalam bahasa alami. Misalnya, salah satu pertanyaannya adalah, "Siapa nama gadis ibumu?" Setelah pengguna menjawab pertanyaan-pertanyaan ini, persyaratan layanan yang disesuaikan akan dibuat antara pengguna dan perusahaan.

Perjanjian persyaratan layanan baru ini kemudian akan diterjemahkan ke dalam kode dan dikemas dalam kontrak pintar pada blockchain yang disebutkan di atas, yang dapat divalidasi oleh semua orang. Setiap kali perusahaan ingin menggunakan data pribadi, perusahaan harus melalui kontrak pintar, yang hanya akan memberikan akses jika skenarionya memenuhi persyaratan tujuan yang dinyatakan. Prosedur ini akan menjamin bahwa perusahaan dapat mengakses data pribadi pengguna untuk tujuan yang telah disetujui secara eksplisit oleh pengguna, sehingga memberikan perlindungan data yang kuat dan terperinci bagi pengguna.

6.6 MEMBANGUN JARINGAN OTONOM DENGAN OPEN SOURCE

Perangkat Lunak dan komunitas Open Source berinovasi dengan cepat dan pesat. Berdasarkan sebagian besar bab ini, tidak berlebihan untuk mengatakan bahwa open source telah membentuk banyak fondasi untuk jaringan otonom berbasis AI. Banyak kemajuan telah dicapai, termasuk infrastruktur otomatis (ONAP, Kubernetes), dan platform pemrosesan/penyimpanan data canggih (Apache Spark dan Hadoop). Komunitas OSS dan proyek-proyeknya telah muncul sebagai mekanisme industri de facto untuk inovasi perangkat lunak, yang sebagian besar akan berakhir di jaringan otonom berbasis AI.

Namun, perlu disadari bahwa peserta OSS dan penerima manfaatnya dari kalangan akademisi dan industri memang mempertahankan komponen kepemilikan mereka demi potensi keuntungan. Tujuan-tujuan ini telah memunculkan pertanyaan lain: Apa sebenarnya peran OSS dalam membangun jaringan otonom?

Untuk menjawab pertanyaan ini, pertama-tama kita perlu mengingat kembali beberapa keterbatasan kolaborasi OSS.

- Faktor persekutuan terkecil. OSS dibangun oleh berbagai perusahaan dan individu yang membentuk komunitas proyek. Mempertahankan partisipasi ini berarti OSS harus memenuhi serangkaian persyaratan minimum yang umum bagi semua. Akibatnya, OSS yang sukses seringkali cenderung tetap generik agar bermanfaat dalam berbagai skenario. OSS menawarkan fungsi yang "cukup baik" bagi sebagian besar pengguna. Misalnya, Linux sukses karena dapat digunakan di berbagai industri termasuk otomotif,



cloud, dan seluler. Fleksibilitas Linux telah menjadikannya salah satu proyek sumber terbuka yang paling banyak diinvestasikan di dunia.

- Spesialisasi di luar kotak. Filosofi "cukup bagus" mendorong partisipasi yang luas, tetapi kecenderungan ini menimbulkan tantangan spesialisasi sehingga tidak cocok untuk lingkungan tertentu. Misalnya, Linux tidak cocok untuk aplikasi waktu nyata (real-time), sehingga untuk sistem tertanam, para insinyur harus mengembangkan patch RT. Linux tidak cocok untuk laptop perusahaan karena tidak sepenuhnya kompatibel dengan aplikasi Microsoft. Masih jauh dari menggantikan Autosar sebagai OS yang beroperasi di dalam CV meskipun ada Linux kelas otomotif, yang mungkin cocok untuk keperluan infotainment.

Meskipun demikian, OSS tidak boleh dianggap sebagai solusi pamungkas untuk semua, melainkan harus dianggap sebagai fondasi bagi operator cloud yang menyediakan Perangkat Lunak sebagai Layanan mereka sendiri. Hal ini juga dapat menjadi peluang yang baik bagi penyedia untuk menawarkan langganan pemeliharaan dan dukungan OSS, sehingga terhindar dari persaingan perangkat lunak bisnis. Untuk masa mendatang yang dapat diperkirakan, jaringan otonom akan dikembangkan dengan menggabungkan lapisan OSS dan perangkat lunak berpemilik. Kerangka kerja kecerdasan buatan seperti TensorFlow dan Acumos akan menguntungkan partisipasi komunitas, dan tentunya ketika diterapkan, akan memanfaatkan platform otomasi besar seperti ONAP.

Kesimpulannya, perangkat lunak sumber terbuka akan memainkan peran besar dalam evolusi jaringan otonom berbasis AI. Cloud akan mengandalkan teknologi ini untuk meningkatkan skalabilitas, menerapkan layanan baru, dan memperkenalkan solusi AI/ML yang semakin canggih dalam siklus tertutup otomatis. Platform otomasi skala besar beserta aplikasi dan jaringan yang dikembangkan menggunakan prinsip-prinsip CN akan menjadi hal yang umum.



BAB 7

MESIN DAN CLOUD NATIVE DENGAN MIKRO LAYANAN

7.1 PENDAHULUAN

Bab ini membahas beberapa implikasi penerapan kecerdasan buatan (AI) dan khususnya pembelajaran mesin (ML) pada jaringan komunikasi berbasis cloud yang menyediakan layanan komputasi kompleks. Dari satu aspek, penerapan AI, dan khususnya ML, pada jaringan mencakup dua bidang: AI dalam manajemen jaringan, dan layanan jaringan berbasis AI.

Kecerdasan buatan dalam manajemen jaringan didasarkan pada beberapa asumsi yang dianggap fakta, tetapi sebenarnya tidak. Diasumsikan bahwa jaringan berbasis cloud memungkinkan fleksibilitas, kompleksitas, dan dinamika yang melampaui kemampuan manusia untuk mengendalikan jaringan secara waktu nyata. Di sini, fleksibilitas mengacu pada kemampuan untuk menyesuaikan atau mengonfigurasi layanan jaringan sesuai kebutuhan pengguna. Kompleksitas mengacu pada jumlah konfigurasi layanan jaringan yang berjalan secara paralel dengan kemungkinan interaksi langsung dan tidak langsung. Dan dinamika mengacu pada laju perubahan konfigurasi ini, apa pun alasannya. Diasumsikan bahwa kombinasi fleksibilitas, kompleksitas, dan dinamika (serta fitur serupa lainnya) memerlukan otomatisasi, misalnya dalam bentuk kontrol loop tertutup. Diasumsikan juga bahwa otomatisasi prosedural tidak memadai, sehingga diperlukan AI. Lebih lanjut diasumsikan bahwa manusia mungkin terlalu lambat untuk mengodekan aturan AI secara manual, sehingga ML harus digunakan untuk mempelajari perilaku jaringan dan mengotomatiskan pembuatan aturan manajemen jaringan.

Mengembangkan, menguji, menerapkan, dan mengamankan layanan komunikasi berbasis AI untuk jaringan otonom berbasis cloud merupakan tantangan dan kompleks. Teknologi cloud, termasuk kontainer, cloud native, dan layanan mikro, memungkinkan penggunaan kembali aplikasi, mengurangi waktu pemasaran, penggunaan infrastruktur yang optimal, dan adaptasi terperinci terhadap kebutuhan pengguna; namun, kompleksitas administrasi membutuhkan AI.

Memfasilitasi AI dan ML membutuhkan data pelatihan yang sangat besar, dan dalam banyak kasus, label target, terutama untuk pembelajaran terawasi. Data pelatihan harus dikumpulkan, dibersihkan, diatur, dan diprovisi dengan sangat hati-hati. Kualitas dan keunggulan data pelatihan jelas sangat penting bagi kinerja dan ketahanan jaringan.

Di tingkat nasional dan internasional, jaringan telekomunikasi merupakan sistem yang sangat kompleks. Jaringan komunikasi berbasis cloud bahkan lebih kompleks, dan terlebih lagi kompleks jaringan cloud, sebagaimana akan dijelaskan lebih lanjut di bagian selanjutnya. Oleh karena itu, kendali loop tertutup berbasis AI pada jaringan praktis merupakan jaringan yang terdiri dari banyak loop tertutup, di mana tidak ada loop yang benar-benar tertutup.

Kecerdasan buatan dalam layanan jaringan mengacu pada entitas jaringan, seperti layanan mikro berbasis AI, yang dapat berperilaku sangat berbeda untuk pelanggan atau



situasi yang berbeda. Mengaktifkan, memantau, memprediksi (perilaku), dan mengelola layanan mikro berbasis AI merupakan tantangan yang kompleks. Kita tentu dapat membayangkan bidang ketiga yang mencakup manajemen layanan berbasis AI.

Seperti yang kita alami dengan perdagangan algo saham yang diduga memperburuk kejatuhan pasar saham tahun 1987 serta beberapa "flash crash" setelahnya, kegagalan AI yang sederhana dapat dengan cepat berubah menjadi pemadaman jaringan total. Yang mungkin lebih buruk lagi, kegagalan AI yang sederhana dapat berkembang menjadi kegagalan jaringan "berjalan" yang intermiten. Kegagalan berjalan ini mungkin memiliki efek sementara yang sangat berbeda untuk layanan yang berbeda. Misalnya, peningkatan latensi saat bepergian dapat menunda beberapa masukan dari sistem *Internet-of-Things* (IoT) industri yang menyebabkan kesalahan perhitungan sementara, yang berubah saat kegagalan merambat melalui jaringan.

7.2 KEMBANGKITAN KOMPUTASI AWAN DAN KOMUNIKASI

Dari satu aspek, rekayasa adalah seni optimasi, menukar kekurangan yang tidak perlu dengan kekurangan yang diperlukan. Pertimbangan optimasi berkembang dan berubah seiring perubahan teknologi yang mendasarinya, dan kemudian kembali dalam bentuk yang sedikit berbeda.

Tujuan dasar sistem operasi (OS) adalah transparansi perangkat keras (yang sekarang disebut "virtualisasi"), yang juga mendukung multiprosesing dan multiprograming. Dalam teknologi cloud, suatu proses dalam multiprosesing (atau tugas dalam multitasking) disebut server. Ini mengacu pada kemampuan untuk memecah perangkat lunak menjadi proses-proses yang dapat diproses secara paralel. Multiprograming mengacu pada menjalankan beberapa program berbeda secara bersamaan pada komputer yang sama. OS menyediakan penggunaan kembali perangkat keras dan perangkat lunak dengan mengorbankan peralihan konteks dan bahaya thrashing, di mana komputer memproses kode OS lebih banyak daripada memproses kode aplikasi.

Prinsip-prinsip ini diterapkan ke lingkungan cloud. Dengan prosesor tunggal, pemrosesan paralel bersifat virtual, dan prosesor mengeksekusi proses secara bergantian (dengan peralihan konteks di antaranya). Sistem multiprosesor sebenarnya dapat mengeksekusi proses secara paralel jika seluruh sistem dibangun untuk pemrosesan terdistribusi. Selain prosesor multicore, klien-server adalah arsitektur pemrosesan terdistribusi yang paling umum digunakan. Pemrosesan terdistribusi memperkenalkan konektivitas, bandwidth komunikasi, latensi, arsitektur jaringan, dll. Internet memperkenalkan globalisasi ke komputasi, sehingga pemrosesan dapat dilakukan di mana saja, daring. Dengan komputasi seluler, pemrosesan terdistribusi telah mencapai saku kita (dalam bentuk ponsel pintar) dan sebagian besar perangkat lunak klien sekarang diproses saat bepergian, mengambil sekitar setengah dari lalu lintas Internet.

Pelapisan perangkat lunak, atau penumpukan, sangat umum sebagai sarana abstraksi, yang memungkinkan peningkatan lapisan tanpa memengaruhi lapisan lainnya. *Open Systems Interconnection* (OSI) adalah model konseptual jaringan yang dikembangkan di Organisasi



Internasional untuk Standardisasi (ISO) sebagai ISO/IEC 7498-1. Komputasi awan membawa teknologi informasi (TI) ke skala ekonomi yang ekstrem. Diasumsikan bahwa mengurangi biaya akan meningkatkan konsumsi hingga ke titik di mana sumber daya selalu terbatas. Organisasi beralih ke cloud karena tiga keuntungan utama: pengurangan Belanja Modal (CAPEX), skalabilitas, dan kemudahan atau kecepatan penerapan serta akuisisi layanan berbasis perangkat lunak. Tentu saja, ada alasan lebih lanjut, yang sebagian besar didasarkan pada tiga keunggulan cloud ini. Komputasi awan didasarkan pada prinsip berbagi, yang berarti bahwa setiap orang memengaruhi setiap orang. Tujuan jaringan otonom adalah meminimalkan dampak yang dirasakan sementara efisiensi perangkat keras, efisiensi perangkat lunak, dan fleksibilitas layanan terus meningkat.

Komputasi Awan sebagai Bisnis dan Layanan

Dari kejauhan, konsep yang sekarang kita kenal sebagai komputasi awan sama tuanya dengan komputasi digital. Namun, kami lebih tertarik untuk memahami pola revolusi saat ini. Dari satu aspek, ekonomi dapat dibagi menjadi nilai tambah (produk dan layanan) dan nilai intrinsik. Ekonomi nilai tambah didasarkan pada pembagian biaya, dan harga produk atau layanan terutama didasarkan pada biaya penyediaannya. Ekonomi nilai intrinsik (juga dikenal sebagai ekonomi hasil) didasarkan pada pembagian pendapatan, dan harganya didasarkan pada manfaat bagi pelanggan. Kedua jenis ekonomi ini telah ada sejak lama dan jelas akan tetap ada; namun, tampaknya ekonomi nilai intrinsik saat ini sedang berkembang, dan revolusi komputasi awan serta jaringan otonom perlu diselidiki dari aspek ini.

Kita dapat mengamati keinginan operator telekomunikasi untuk beralih ke ekonomi nilai intrinsik. Alih-alih menyediakan infrastruktur dasar yang dapat memindahkan bit, operator ingin menyediakan layanan canggih yang secara khusus disesuaikan dengan kebutuhan pengguna. Istilah bandwidth, latensi, dan jaringan (dan komputasi) serupa akan digantikan oleh istilah bisnis dan gaya hidup yang spesifik bagi pengguna. Ide ini sama sekali bukan hal baru, tetapi tampaknya semakin populer.

Teknologi komputasi telah berevolusi dari beberapa mainframe raksasa di seluruh dunia menjadi superkomputer yang dapat diakses semua orang. Namun, perkembangan yang sesungguhnya adalah migrasi menuju layanan kebutuhan pengguna sesuai permintaan. Bisnis bukanlah infrastruktur komputasi, bukan perangkat lunak, dan bukan pemrosesannya; melainkan hasil, nilai intrinsik, dan solusi untuk masalah pengguna saat ini secara real-time. Layanan ini dapat melibatkan sejumlah perangkat masukan dan berbagai jenis perangkat keluaran.

Kemungkinan konfigurasi komputasi tertentu bersifat satu kali, unik, dan selamanya. Tampaknya, layanan semacam itu tidak dapat disediakan oleh satu entitas bisnis saja, dan oleh karena itu, kemungkinan, konfigurasi bisnis tertentu yang menyediakan layanan satu kali ini juga dibuat sesuai permintaan. Tidak ada cara lain untuk mencapai layanan bernilai intrinsik tersebut selain melalui cloud, atau lebih tepatnya, melalui jaringan berbasis cloud yang otonom. Sebagaimana telah disebutkan sebelumnya, kombinasi fleksibilitas, kompleksitas, dan dinamika membutuhkan otomatisasi penyediaan layanan berbasis AI menggunakan ML dan teknologi serupa.



Konsep dan Komponen Cloud

CAPEX merupakan alasan utama bagi organisasi untuk beralih ke cloud; namun, sudut pandang finansial yang agak sempit ini agak menyesatkan. Pusat data privat membutuhkan perencanaan, pengadaan, instalasi, dan pemeliharaan yang kompleks dan berjangka panjang. Beralih ke cloud dan ke Biaya Operasional (OPEX) diduga menyederhanakan masalah. Secara praktis, sebuah solusi memindahkan fokus masalah ke bagian lain dari keseluruhan kompleksitas, yang mungkin tampak, untuk sementara waktu, lebih nyaman dan lunak terhadap kebutuhan tertentu. Namun, dalam jangka panjang, inti masalahnya tetap ada.

Ketergantungan bisnis pada operasi TI mereka semakin meningkat, membutuhkan sistem yang sangat andal dan perangkat lunak yang teruji secara ketat. Jaringan telekomunikasi adalah salah satu contoh layanan yang sangat andal. Namun, keandalan datang dengan biaya finansial yang tinggi dan biaya operasional yang lebih tinggi lagi. Pusat TI menjadi penyebab utama latensi bisnis, yang merupakan penyebab utama hilangnya pangsa pasar dan profitabilitas. Bisnis berada di bawah tuntutan yang semakin meningkat untuk kelincahan, dan pusat TI berada di bawah tekanan khusus untuk mendukung kelincahan bisnis. Janji utama komputasi awan adalah kelincahan, didukung oleh pengembangan kolaboratif dan komputasi kolaboratif. Namun, kolaborasi yang rumit dan ketergantungan silang memiliki harga, memperkenalkan efek stokastik yang terputus-putus dan tidak stabil. Namun, tidak diragukan lagi, awan menawarkan penskalaan yang lebih cepat, setidaknya di mana penskalaan "horizontal" adalah solusi yang valid.

Teorema toleransi konsistensi, ketersediaan, dan partisi (CAP) (alias teorema Brewer) menyatakan bahwa tidak mungkin bagi penyimpanan data terdistribusi untuk secara bersamaan menyediakan lebih dari dua dari tiga jaminan konsistensi data, ketersediaan data (atau latensi), dan toleransi partisi. Konsistensi data berarti bahwa setiap operasi baca mengambil nilai yang paling diperbarui. Ketersediaan data berarti bahwa setiap operasi dilayani dengan segera. Toleransi berarti bahwa layanan data "toleran" terhadap beban (meskipun partisi). Secara praktis, teorema CAP menyatakan bahwa partisi menurunkan konsistensi data (untuk menyediakan ketersediaan data), atau ketersediaan data (untuk menyediakan konsistensi data), atau keduanya (misalnya, konfigurasi yang dioptimalkan). Setiap penyimpanan data dipartisi; namun, cloud mengagungkan partisi (misalnya, penskalaan horizontal), sehingga menjadikan teorema CAP relevan.

Untuk contoh sederhana tentang apa yang dimaksud teorema CAP, kita dapat membayangkan pemrosesan transaksi, dan khususnya situasi seperti pemesanan ganda kursi pesawat yang sama, atau kamar hotel yang sama, atau slot yang sama pada pakaian operasi rumah sakit, yang mengubah masalah "toleransi" menjadi panggilan bisnis. Sistem kendali misi-kritis (misalnya, koordinasi kereta api atau kontrol lalu lintas udara) jelas tidak toleran terhadap aspek apa pun dari konsistensi dan ketersediaan data.

Konsistensi data dalam arti luas dapat dibagi menjadi konsistensi satu item data mengenai berbagai proses paralel, dan koherensi data (atau konkurensi) yang mempertimbangkan konsistensi sejumlah item data yang berbeda dan saling terkait. Misalnya, suatu basis data dikatakan koheren ketika semua transaksi telah selesai. Toleransi dapat



merujuk pada kesalahan intrinsik dari setiap nilai terhitung akibat inkoherenensi akibat beberapa transaksi yang dieksekusi secara paralel, sehingga menimbulkan ketidakpastian data sebagai sebuah fenomena, alih-alih sebagai kesalahan. Namun demikian, komputasi awan dengan berbagai keunggulannya yang tak tergantikan dan berbagai permasalahannya yang spesifik akan tetap ada dan berkembang pesat. Jaringan komunikasi juga bermigrasi ke awan, dan kompleks jaringan awan akan segera menggunakan AI untuk mengelola kompleksitasnya secara real-time. Di Bagian 7.3, kami akan mencoba menjelaskan beberapa aspek dari perpaduan AI, logika fuzzy, pembelajaran mesin berkelanjutan, ketidakpastian data intrinsik, dan trade-off CAP yang sedang berkembang ini.

Cloud Native

Cloud native adalah sebuah konsep di mana perangkat lunak dirancang, dikembangkan, dan dioptimalkan untuk memanfaatkan teknologi awan. Teknologi awan pada dasarnya adalah arsitektur terdistribusi, dan cloud native dirancang khusus untuk pemrosesan dan penyimpanan data terdistribusi. Cara lain untuk mempertimbangkan cloud native adalah metodologi aplikasi dua belas faktor yang dikembangkan di Heroku (<https://12factor.net/>). Tujuannya adalah membagi sistem menjadi sebanyak mungkin entitas pemrosesan independen yang berkolaborasi melalui infrastruktur cloud. Oleh karena itu, setiap entitas pemrosesan cloud membuat asumsi sesedikit mungkin tentang lingkungannya. Dengan demikian, memungkinkan penyediaan, pemutakhiran, skalabilitas, penghentian yang lancar, dan keusangan yang sederhana dan otomatis. Cloud native juga mengasumsikan status tanpa status, mengandalkan layanan pendukung untuk setiap aspek "memori jangka panjang" seperti status dan data. *Cloud Native Computing Foundation*, yang didirikan pada tahun 2015 dan kini beranggotakan lebih dari 550 orang, mempromosikan konsep cloud native, khususnya penggunaan kontainer, sebagai sarana untuk mencapai batasan yang ketat.

Menurut Cloud Native Computing Foundation, sistem cloud-native harus memiliki karakteristik berikut:

1. Dikemas dalam kontainer. Menjalankan aplikasi dan proses dalam kontainer perangkat lunak sebagai unit penerapan aplikasi yang terisolasi, dan sebagai mekanisme untuk mencapai tingkat isolasi sumber daya yang tinggi. Meningkatkan pengalaman pengembang secara keseluruhan, mendorong penggunaan kembali kode dan komponen, serta menyederhanakan operasi untuk aplikasi cloud-native.
2. Dikelola secara dinamis. Dijadwalkan dan dikelola secara aktif oleh proses orkestrasi terpusat. Meningkatkan efisiensi mesin dan pemanfaatan sumber daya secara radikal sekaligus mengurangi biaya yang terkait dengan pemeliharaan dan operasional.
3. Berorientasi pada layanan mikro. Terhubung secara longgar dengan dependensi yang dijelaskan secara eksplisit (misalnya, melalui titik akhir layanan). Meningkatkan kelincahan dan pemeliharaan aplikasi secara keseluruhan secara signifikan.

Layanan Mikro

Layanan mikro merepresentasikan modularitas cloud dan implementasi khas dari konsep cloud-native. Layanan mikro adalah perangkat lunak yang dapat diterapkan dan dihapus, yang konon tanpa memengaruhi pengoperasian seluruh sistem perangkat lunak. Satu



layanan mikro dapat mengajukan layanan dari layanan mikro lain melalui infrastruktur komunikasi antarproses cloud. Sebuah layanan mikro terlindungi dari lingkungannya, begitu pula sebaliknya.

Membagi sistem perangkat lunak menjadi beberapa layanan mikro independen memungkinkan beberapa keuntungan penting komputasi awan:

- Pengembangan independen setiap layanan mikro, memungkinkan siklus pengembangan yang singkat dan rilis versi baru yang cepat. Teknologi komputasi baru (misalnya, bahasa perangkat lunak, perangkat lunak pendukung manajemen, dan metodologi) dapat diperkenalkan secara bertahap ke dalam sistem, dan sistem dapat dengan mudah dan aman mencakup beberapa teknologi komputasi.
- Skalabilitas, konfigurasi sistem yang fleksibel, dan kemudahan migrasi layanan mikro antara platform perangkat keras dan perangkat lunak.
- Ketahanan seluruh sistem terhadap kegagalan layanan mikro tertentu.

Meskipun lebih mudah dan cepat untuk mengembangkan dan menerapkan satu layanan mikro (daripada sistem monolitik), jauh lebih sulit untuk mengelola seluruh sistem dari banyak layanan mikro dengan berbagai versinya yang berjalan secara paralel. Untuk mencapai keunggulannya, layanan mikro memerlukan batasan yang ketat. Layanan mikro bekerja sama menggunakan protokol Antarmuka Pemrograman Aplikasi (API), sebaiknya berorientasi pada *Representational State Transfer* (REST) (REST adalah serangkaian batasan dan properti berbasis HTTP yang mendukung interoperabilitas antarsistem komputer di internet) dan sebaiknya tanpa status (stateless). Jika tidak sepenuhnya tanpa status, layanan mikro juga menganjurkan manajemen data terdesentralisasi di mana setiap layanan mikro mengelola penyimpanan datanya sendiri. Jika layanan mikro B lain membutuhkan data yang dikelola oleh layanan mikro A, layanan mikro tersebut harus menggunakan API layanan mikro A untuk memanggilnya guna mendapatkan data yang diperlukan.

Meskipun tangguh, panggilan jarak jauh pada dasarnya lambat, dan batasan yang ketat meningkatkan latensi, serta persyaratan bandwidth dan pemrosesan, sehingga meningkatkan topologi dan kompleksitas konfigurasi. Meskipun setiap layanan mikro lebih mudah dipahami, kompleksitasnya dialihkan ke interkoneksi antarlayanan mikro, yang meningkatkan debugging dan kompleksitas operasional seperti yang terkait dengan pemrosesan transaksi, penguncian data, deteksi dan pemulihan deadlock, dan critical race. Beberapa akronim tiga huruf, seperti *Enterprise Application Integration* (EAI), *Service-Oriented Architecture* (SOA), dan *Enterprise Service Bus* (ESB), dimaksudkan untuk memberikan solusi yang memadai untuk masalah komunikasi antarproses.

Solusi tersebut memecahkan kebutuhan untuk mendefinisikan, mengembangkan, dan menguji API antara dua layanan mikro yang berinteraksi. Implementasi EAI/SOA/ESB yang efektif bertindak sebagai jembatan antara dua layanan mikro yang berinteraksi, sehingga meningkatkan peralihan konteks, beban pemrosesan, beban komunikasi, latensi, dan satu titik kegagalan, serta membentuk kemacetan pemrograman. Mengimplementasikan EAI/SOA/ESB sebagai layanan mikro yang benar-benar tanpa status dan membuat layanan mikro sesuai



dengan beban dapat membantu mengatasi masalah beban. Di sisi lain, solusi EAI/SOA/ESB juga dapat digunakan sebagai fasilitas penguncian data, yang sama sekali tidak tanpa status.

Dengan batasan yang ketat, layanan mikro menguntungkan untuk integrasi berkelanjutan dan pengiriman berkelanjutan (CI/CD) yang dibahas di bagian berikut. Rilis cepat menunjukkan pengujian yang sangat terbatas, dan ekspektasi bahwa sistem toleran terhadap kegagalan layanan mikro tunggal. Namun, sistem yang kompleks, dan khususnya basis data yang kompleks, sama sekali tidak toleran. Kembali ke masalah koherensi data, kesalahan perangkat lunak dapat menyebabkan basis data menjadi tidak koheren, sehingga menciptakan proliferasi inkonsistensi data dan kerusakan berulang.

Aspek lain dari layanan mikro adalah keamanan siber, dan dilemanya adalah apakah lebih aman menyimpan aset di satu tempat dengan perlindungan terbaik, atau apakah pelanggaran keamanan pada satu layanan mikro tidak membahayakan seluruh sistem. Di satu sisi, batasan yang ketat, jika ditetapkan dengan tepat, baik untuk melawan serangan siber; namun, topologi yang berubah cepat dan perilaku sistem yang kurang dapat diprediksi membuat keamanan siber jauh lebih sulit. Sekali lagi, AI mungkin menawarkan solusi, meskipun masih meningkatkan biaya pemrosesan dan latensi.

Mesin Virtual dan Kontainer

Mesin virtual adalah teknologi komputasi lama yang digunakan sebagai lapisan antara OS dan perangkat keras, sehingga memvirtualisasikan perangkat keras. Konsep ini telah dikembangkan lebih lanjut sebagai lapisan antara OS dasar (atau hosting) dan OS tamu (atau hosting). Teknologi ini, juga dikenal sebagai hypervisor, memungkinkan beberapa OS hosting untuk berjalan secara bersamaan pada perangkat keras yang sama, melalui OS bersama (dasar, hosting).

Mesin virtual menjadi alat infrastruktur cloud yang sangat berguna, bahkan esensial, yang memungkinkan fleksibilitas di mana aplikasi dapat berjalan di mana saja, terlepas dari perangkat keras dan OS yang mendasarinya. Namun, virtualisasi ini juga berarti bahwa aplikasi tidak dapat dioptimalkan untuk arsitektur perangkat keras atau OS tertentu. Akibatnya, ketika aplikasi dimigrasi ke arsitektur perangkat keras atau OS yang berbeda, kinerjanya dapat menurun secara signifikan.

Menjalankan setiap aplikasi di atas OS tamunya sendiri juga menyediakan isolasi yang efektif antar aplikasi. Namun, seperti halnya sistem perangkat lunak dengan tumpukan tinggi, isolasi ini mahal, karena peralihan konteks sangat membebani. Menjalankan setiap aplikasi di atas OS tamunya sendiri dan mesin virtualnya sendiri membutuhkan administrasi sistem yang ekstensif. Konsep kontainer diciptakan untuk menyederhanakan masalah. Alih-alih menumpuk dan mengelola beberapa lapisan perangkat lunak, aplikasi dikemas bersama dengan semua alat sistem, pustaka, pengaturan, dll., dan hanya yang dibutuhkan aplikasi.

Sebuah kontainer dapat berjalan langsung di atas OS kernel. Oleh karena itu, kontainer dikelola dan dimigrasikan dengan lebih mudah, namun, mengurangi kemampuan untuk mengoptimalkan perangkat lunak ke perangkat keras dan OS tertentu. Keuntungan lain dari kontainer adalah ukurannya. Menggunakan mesin virtual per aplikasi juga membutuhkan memori untuk seluruh mesin virtual dan seluruh OS tamu, yang sebagian besar tidak



digunakan. Kontainer hanya mengambil bagian-bagian OS yang digunakan aplikasi dan karenanya jauh lebih kecil. Akibatnya, lebih banyak kontainer dapat berjalan di server yang sama, sejauh kebutuhan pemrosesan dan peralihan konteks memungkinkan. Namun, perbandingan ini agak menyesatkan, karena mengasumsikan pendekatan cloud-native di mana setiap bagian aplikasi merupakan layanan mikro yang dikontainerisasi secara independen. Jika suatu sistem dirancang sebagai beberapa subsistem (aplikasi, layanan mikro) yang berjalan di atas sistem operasi bersama, maka perbandingan ukuran dapat berubah karena kontainer mengemas kode sistem operasi yang sama berulang kali.

Keuntungan ketiga kontainer berasal dari kesederhanaan dan kecepatan instansiasi, yang memungkinkan respons lebih cepat terhadap perubahan permintaan, seperti beban pemrosesan, latensi, dan bandwidth komunikasi. Konfigurasi ulang topologi sistem yang lebih cepat berarti pemanfaatan infrastruktur perangkat keras yang lebih baik, biaya yang lebih rendah, dan layanan yang lebih baik, serta pemulihan kesalahan yang lebih cepat. Perlu dicatat bahwa pendekatan sistem OS bersama yang disajikan pada paragraf sebelumnya biasanya jauh kurang fleksibel. Beban pemrosesan dan latensi pemrosesan yang terkait dengan peralihan konteks harus dibandingkan antara beberapa kontainer dan mesin virtual yang menjalankan sistem operasi bersama dan beberapa subsistem. Perbandingan semacam itu mungkin menunjukkan keunggulan solusi mesin virtual.

Keamanan umumnya dianggap sebagai kelemahan pendekatan kontainer, karena kernel yang mendasarinya mungkin rentan terhadap pelanggaran keamanan dan dengan demikian menyediakan pintu belakang ke aplikasi yang terkontainerisasi. Namun, cara untuk mengatasi masalah keamanan tersebut sedang dikembangkan, yang menambah kompleksitas dan mengurangi fleksibilitas. Sekali lagi, pendekatan mesin virtual memperoleh keunggulan lebih lanjut ketika sejumlah subsistem dilindungi oleh OS bersama. *Open Container Initiative* (OCI) didirikan pada tahun 2015 di bawah naungan Linux Foundation untuk mengembangkan standar industri bagi format kontainer dan perangkat lunak runtime kontainer untuk semua platform. Sponsor proyek ini antara lain Amazon Web Services, Google, IBM, HP, Microsoft, VMware, Red Hat, Oracle, Twitter, HP, CoreOS, dan Docker yang menyumbangkan sekitar 5 persen basis kodenya untuk proyek tersebut agar dapat berjalan.

Sistem Manajemen dan Orkestrasi

Docker mungkin masih merupakan alat paling populer untuk mengemas kode layanan mikro dalam kontainer, tetapi ada beberapa alat lain yang semakin populer. Di atas alat kontainerisasi, terdapat alat untuk mengelola kontainer, seperti Kubernetes dan Docker Swarm. Terdapat juga alat serupa untuk mengelola mesin virtual seperti VMware vSphere suite. Praktisnya, setiap platform cloud memiliki perangkat manajemen mesin virtualnya sendiri serta beberapa alat pihak ketiga. Demikian pula, seiring dengan semakin populernya teknologi kontainer, semakin banyak alat kontainerisasi dan alat manajemen kontainer yang memasuki pasar, dan kebutuhan akan standar menjadi nyata.

Alat manajemen sistem harus menampilkan dan mengelola semua server host (perangkat keras) dan semua mesin virtual serta kontainer yang beroperasi dalam infrastruktur tertentu. Berbagai sistem manajemen menawarkan layanan otomatis, seperti menghasilkan



peringatan, mengkonsolidasikan mesin virtual ke server host sesedikit mungkin jika memungkinkan, mendeteksi penyimpanan yang tidak terpakai, kode, dll. Fitur-fitur tersebut biasanya bergantung pada beberapa kait perangkat lunak di dalam mesin virtual atau kontainer.

OCI mempromosikan standar dan spesifikasi untuk teknologi kontainer guna mencapai interoperabilitas untuk teknologi kontainer dan alat manajemen. Saat ini, upaya tersebut lebih mengarah pada proyek sumber terbuka daripada standar. Rilis pertama format gambar dan spesifikasi yang dikeluarkan pada Juli 2017 mendefinisikan persyaratan minimal untuk kontainer yang portabel di berbagai OS, perangkat keras, arsitektur CPU, cloud publik, dll., dengan proses sertifikasi formal yang direncanakan. Saat ini, belum ada rencana mengenai jaringan, komunikasi antarkontainer, dan distribusi. *Cloud Native Computing Foundation* juga mempromosikan antarmuka jaringan kontainer (CNI) dan standar antarmuka penyimpanan kontainer (CSI). CNI menstandarisasi jaringan di bawah penerapan kontainer multihost yang memungkinkan kerja sama dan pertukaran antar perangkat jaringan proprietary (misalnya, Weave, Contrive, Docker Network); CSI adalah standar serupa untuk infrastruktur data (misalnya, Rook).

Virtualisasi Fungsi Jaringan

Virtualisasi fungsi jaringan (NFV) yang dipromosikan oleh *European Telecommunications Standards Institute* pada dasarnya adalah arsitektur cloud untuk layanan telekomunikasi yang menciptakan jaringan yang lebih fleksibel dan dinamis dibandingkan jaringan komunikasi konvensional saat ini. Dalam jaringan berbasis NFV, layanan yang dapat diimplementasikan dalam perangkat lunak dieksekusi melalui perangkat keras cloud generik sebagai fungsi jaringan virtual (VNF). Dalam hal ini, VNF seperti layanan mikro yang beroperasi di atas mesin virtual atau kontainer. Seperti layanan mikro, VNF mudah dipasang, dihapus, dan dimigrasikan antar fasilitas perangkat keras, sesuai permintaan, di bawah kendali perangkat lunak orkestrasi NFV (NFV-O).

Dari sudut pandang operator telekomunikasi, NFV memindahkan jaringan ke teknologi cloud. Dari sudut pandang TI, jaringan publik yang menghubungkan situs-situs cloud atau cloud hibrida telah divirtualisasikan. Beragamnya sistem perangkat lunak manajemen cloud kini harus bekerja sama dengan masing-masing NFV-O, dan setiap NFV-O harus mempertimbangkan masing-masing sistem perangkat lunak manajemen cloud yang terhubung. Karena jaringan telekomunikasi saling terhubung, masing-masing NFV-O harus mengoordinasikan fungsionalitas dan aktivitasnya. Oleh karena itu, dinamika NFV menambah kompleksitas dinamika komputasi cloud.

Virtualisasi fungsi jaringan mendorong dekomposisi fungsionalitas node jaringan semaksimal mungkin. Aliran paket didekomposisi secara vertikal menjadi pensinyalan, kontrol, data, streaming, dll., yang masing-masing ditangani di bidang yang berbeda. Pemrosesan elemen-elemen ini di setiap bidang didekomposisi untuk memungkinkan distribusi horizontal ke berbagai unit perangkat keras, sehingga memungkinkan skalabilitas yang optimal, dll. Aplikasi yang berhubungan dengan komunikasi didekomposisi menjadi fungsi-fungsi yang



dapat dipilih untuk memungkinkan instansiasi terpisah dari setiap fungsi guna mengoptimalkan kustomisasi, mengurangi biaya, dll.

Sebagai aturan praktis, dalam hal kinerja per dolar, perangkat keras lebih efisien. Dalam hal fleksibilitas, perangkat lunak lebih efisien. Oleh karena itu, untuk tujuan bisnis, operator harus memanfaatkan fleksibilitas perangkat lunak dengan terus mengkonsolidasikan fungsionalitas perangkat lunak (misalnya, VNF, layanan mikro) melalui perangkat keras bersama, sambil terus mempertimbangkan perjanjian tingkat layanan (SLA), persyaratan kualitas layanan (QoS), ketersediaan, masalah keamanan, dll. Selain itu, NFV harus berkembang menjadi platform yang efektif untuk layanan komunikasi canggih yang baru dikembangkan yang memungkinkan diferensiasi bisnis dan keunggulan kompetitif di pasar. Memahami bahwa teknologi cloud dan NFV mendorong kelincahan bisnis dan pengembangan teknologi yang lebih cepat, diharapkan fungsi dan fitur baru akan dihadirkan secara berkelanjutan dengan laju yang semakin meningkat, praktis melampaui kemampuan untuk menyebarluaskan dan mengasimilasi pengetahuan.

DevOps, Integrasi Berkelanjutan, dan Penerapan Berkelanjutan

Kelincahan bisnis merupakan pendorong utama untuk beralih ke teknologi cloud dan NFV. Kelincahan bisnis berarti mempersingkat waktu pemasaran sekaligus mengurangi biaya penyampaian fungsi baru, mulai dari definisi fungsi hingga penagihan. Membawa suatu fungsi dari definisi hingga penagihan melibatkan banyak disiplin ilmu yang memiliki jaringan motivasi dan pertimbangan yang rumit.

Dua pendekatan umum adalah "dibangun dengan benar sejak awal" dan "prototipe yang diulang dengan cepat". Mengingat keterbatasan tenaga kerja dan anggaran, jelas bahwa hanya sebagian kecil proyek yang menggunakan metode prototipe yang diulang dengan cepat yang akan menghasilkan bisnis yang menguntungkan. Namun, tempat ini memiliki dua keuntungan penting: "biarkan pasar memutuskan" dan "mengubur yang mati", keduanya berkontribusi pada keuntungan bisnis.

DevOps dan CI-CD adalah konsep yang dipromosikan oleh penelitian dan pengembangan (R&D) untuk meningkatkan laju definisi fungsi hingga penerapan. DevOps mendorong integrasi yang erat antara pengembangan perangkat lunak, jaminan kualitas, dan operasi pusat data, yang secara tradisional sengaja dipisahkan. Mengintegrasikan fungsi-fungsi bisnis ini cenderung memungkinkan siklus pengkodean-integrasi-pengujian-pengiriman-penerapan yang lebih cepat dan otomatis dengan tujuan untuk menyelesaikan sebanyak mungkin siklus per hari. Perbedaan mendasar antara integrasi berkelanjutan, pengiriman, dan penerapan adalah sejauh mana otomatisasi menjangkau siklus tersebut. Secara praktis, DevOps mencoba mereorganisasi bisnis agar R&D dapat mendorong fitur ke pelanggan dengan kecepatan lebih cepat tanpa mengorbankan kualitas layanan. Siklus yang sepenuhnya otomatis (seperti dalam CD) berarti bahwa penerapan ke pelanggan dimulai oleh programmer tanpa intervensi atau pertimbangan lebih lanjut, tentu saja, jika pengujian otomatis menyetujuinya.

DevOps dan CI-CD selaras dengan dekomposisi perangkat lunak, layanan mikro, dan konsep cloud-native. Namun, jika NFV dipertimbangkan, belum terlihat bahwa operator telekomunikasi mengizinkan programmer vendor untuk meluncurkan rilis baru layanan mikro



langsung ke produksi klien utama dengan SLA yang ketat dan sangat detail. Salah satu elemen penting DevOps dan CI-CD adalah pemantauan berkelanjutan. Jika sebuah modul, misalnya, VNF atau layanan mikro, dirilis beberapa kali sehari, dengan jumlah modul yang cukup banyak di dalam sistem dan proses penerapan berkelanjutan yang membutuhkan waktu yang cukup lama, versi sistem akan dirilis secara paralel. Analisis log daring, pemantauan, dan pelaporan di hulu proses DevOps CI-CD menjadi krusial.

Sekilas tentang Komputasi Tepi

Di satu sisi, komputasi tepi menentang konsep terpusat komputasi awan sebagai pusat data global. Di sisi lain, komputasi tepi merupakan puncak dari konsep distribusi horizontal dan dinamika migrasi awan. Dalam pengertian yang lebih lanjut ini, komputasi tepi mewujudkan semua permasalahan komputasi awan demi mengurangi latensi, waktu pemrosesan, dan bandwidth. (Juga biaya pemrosesan data lokal tanpa memindahkan data besar ke inti.)

Ada dua aspek utama komputasi tepi:

- Komputasi tepi jaringan seperti C-RAN dan V-RAN, yang merupakan fungsi tepi yang melekat pada operator telekomunikasi/jaringan/situs.
- Komputasi tepi awan, di mana operator telekomunikasi/jaringan/situs tidak memiliki dan tidak bertanggung jawab atas perangkat lunak tepi selain menyediakan infrastruktur pemrosesan, penyimpanan, dan jaringan (misalnya, Infrastruktur sebagai Layanan). Terdapat juga kotak putih lokal, yang merupakan bentuk awan tepi mini.

Satu pertanyaan utama adalah bagaimana komputasi tepi jaringan berbeda dari NFV, selain komputasi tepi jaringan seluler yang bergantung pada mobilitas terminal (misalnya, roaming).

Komputasi tepi awan, yang beroperasi pada infrastruktur komputasi yang jauh lebih kecil dan terbatas, kurang efisien dibandingkan komputasi awan terpusat, dan oleh karena itu lebih mahal. Komputasi tepi awan juga lebih berisiko daripada komputasi awan terpusat, terutama saat roaming. Oleh karena itu, kecuali terjadi perubahan infrastruktur, aplikasi dapat bermigrasi ke tepi hanya jika komputasi awan terpusat tidak dapat menyediakan layanan yang diperlukan karena keterbatasan, seperti bandwidth, latensi, dan turnaround. Perubahan infrastruktur yang dapat menyebabkan komputasi tepi awan menjadi lebih efisien daripada komputasi awan terpusat adalah meningkatnya biaya komunikasi, yang saat ini tampaknya tidak mungkin. Perubahan infrastruktur yang dapat menyebabkan komputasi tepi awan menjadi lebih efisien daripada komputasi terminal telah meningkatkan efisiensi daya (J/bit) jaringan akses radio. Hal ini mungkin menyarankan penggunaan komputasi tepi awan untuk mengurangi konsumsi daya baterai perangkat klien berbasis baterai.

Fasilitas komputasi tepi awan dapat berada di antara fasilitas switching utama jaringan inti dan klien, di mana pun operator jaringan memiliki jenis "real estat", termasuk stasiun pangkalan, lemari di tepi jalan, dan ruang bawah tanah gedung, dengan daya pemrosesan dan penyimpanan terdistribusi yang bervariasi sesuai permintaan. Contoh gedung real estat dapat mencakup kantor pusat dan pusat data regional dan nasional.

Antifragilitas

Konsep antifragilitas diperkenalkan oleh Nassim Taleb dalam bukunya *Antifragil* (Random House/Penguin) sebagai kualitas sistem yang semakin kuat ketika terpapar stresor.



Hal ini serupa dengan sistem kekebalan tubuh manusia, yang semakin kuat ketika terpapar patogen. Secara praktis, antifrabilitas dapat diwujudkan melalui emulasi beban stokastik. Submodul Chaos Monkey dari proyek Simian Army Netflix menyuntikkan kegagalan acak ke dalam komponen produksi untuk mengidentifikasi kelemahan dalam arsitektur perangkat lunak dan kemudian menghilangkannya.

Memasuki Kecerdasan Buatan

Sebagaimana dibahas sebelumnya, tujuan kompleks jaringan cloud adalah kelincahan bisnis melalui kompleksitas sistem. Peningkatan stratifikasi vertikal, peningkatan distribusi horizontal, dan CI-CD yang lebih cepat, semuanya menambah kompleksitas dan dinamika sistem, sehingga menjadikan penggunaan teknologi AI wajib. Namun, mengingat besarnya tantangan yang ada, teknologi AI masih dalam tahap awal. Masalah mendasarnya adalah orkestrasi jaringan cloud berbasis AI, ketika dikombinasikan dengan manajemen pemrosesan cloud berbasis AI, menciptakan jaringan sistem manajemen berbasis AI yang saling terhubung.

7.3 JARINGAN PEMBELAJARAN LAYANAN MIKRO

Dorongan untuk kelincahan bisnis mendorong teknologi yang memungkinkan pemrosesan (cloud) dan dinamika jaringan. Sejumlah besar layanan mikro terhubung secara rumit satu sama lain, dengan penyimpanan data dari berbagai jenis, dan dengan infrastruktur pemrosesan, penyimpanan, dan jaringan bersama, serta terikat pada berbagai batasan, seperti SLA. Layanan mikro ini dapat terus ditingkatkan atau diganti, sementara kebutuhan, persyaratan layanan, dan beban pelanggan juga terus berubah, dengan pemeliharaan kesalahan (serta pemeliharaan prediktif) yang menambah ketidakpastian sistem secara keseluruhan. Di atas masalah yang sangat kompleks ini, terdapat beban yang terus-menerus untuk mengurangi biaya keseluruhan, baik dalam hal CAPEX maupun OPEX.

Dinamika cloud dan jaringan yang terus berkembang membutuhkan sistem manajemen dan orkestrasi yang canggih, dengan pemahaman bahwa sistem-sistem ini praktis saling terhubung. Tidak ada satu pun sistem manajemen atau orkestrasi yang berdiri sendiri. Setiap keputusan rekonfigurasi memiliki implikasi langsung pada proses yang dikelola oleh sistem manajemen dan orkestrasi di sekitarnya, yang akan segera merespons dan saling memengaruhi berulang kali.

Pertanyaannya adalah apakah otomatisasi konvensional yang canggih dapat memecahkan masalah atau apakah AI penting. Untuk tujuan ini, otomatisasi menggunakan algoritma atau aturan buatan manusia, sementara sistem AI mempelajari jaringan cloud dan secara otomatis menghasilkan aturannya sendiri (misalnya, tanpa campur tangan manusia). Salah satu aspek dari perbedaan ini adalah kemungkinan basis aturan dari berbagai sistem manajemen dan orkestrasi untuk konvergen, atau divergen, serta pertanyaan, Apa yang lebih baik untuk meningkatkan kemungkinan masing-masing sistem tersebut untuk mengalami crash? Dengan kata lain, apa cara yang lebih baik untuk anti-fragilitas? Kecerdasan buatan dalam layanan mikro menambah situasi kompleks yang telah dijelaskan sebelumnya. IoT diharapkan dapat melibatkan AI secara intensif dan dapat memberikan gambaran sekilas tentang kompleksitas yang dihasilkan. IoT dapat dipandang sebagai distribusi perangkat yang



sangat besar, seperti sensor dan aktuator, di mana setiap perangkat melapor, secara langsung maupun tidak langsung, ke beberapa prosesor. Setiap prosesor terus-menerus menganalisis data yang dikumpulkan dari sejumlah besar masukan untuk membuat keputusan. Oleh karena itu, sebuah prosesor dapat mengelola penyimpanan datanya sendiri tetapi juga dapat mengakses penyimpanan data lain. Dengan asumsi bahwa beberapa prosesor ini menggunakan AI, akses data bersifat intensif, dan cakupan data yang diakses dapat berubah "secara tak terduga".

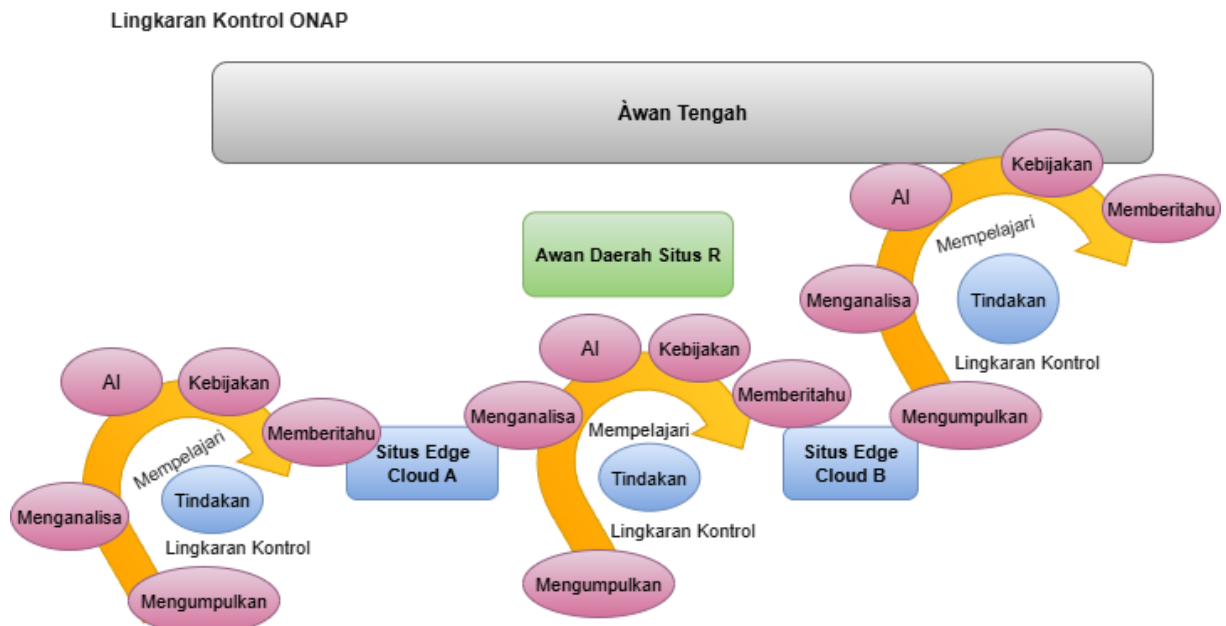
Lingkungan jaringan cloud kompleks yang dikelola, oleh karena itu, secara praktis menggabungkan sejumlah sistem manajemen dan orkestrasi jaringan cloud (begitu pula ekosistem layanan mikro berbasis AI). Maka, jelas, AI dari sistem manajemen dan orkestrasi ini juga digabungkan. Penggabungan ini memengaruhi data yang dianalisis dan basis aturan yang dibuat. Selain pertanyaan "siapa yang memiliki apa?", terdapat kebutuhan untuk beberapa "standarisasi" data, kualitas data, dan keunggulan data, serta fitur-fitur serupa dari basis aturan. Cloud native mengacu pada upaya yang selaras dengan *Cloud Native Computing Foundation*. Misinya adalah untuk mendorong penyelarasan di seluruh kontainer dan layanan mikro yang memanfaatkan salah satu dari beberapa kerangka kerja untuk menyederhanakan desain, pembuatan, dan penerapan aplikasi terdistribusi yang skalabel dengan memanfaatkan cloud. Aspek kuncinya adalah bagaimana aplikasi diatur untuk memungkinkan berbagai bagian aplikasi diterapkan secara optimal agar sesuai dengan persyaratan kinerja, biaya, dan skala. Ketika arsitektur cloud native dipertimbangkan dalam desain, pengembangan, dan penerapan, pemanfaatan sumber daya cloud yang mendasarinya dapat menjadi 70% lebih efisien.

Kasus Penggunaan Jaringan Pembelajaran Layanan Mikro

Kecerdasan buatan (AI) menjadi otak penggerak cloud native. Pengembang perangkat lunak menanamkan layanan mikro dengan kecerdasan berbasis AI. Kemampuan ini penting bagi jaringan otonom untuk memungkinkan layanan mikro utama ditempatkan dan diskalakan secara optimal guna memenuhi persyaratan aplikasi. Layanan mikro kecerdasan buatan dapat dipecah terlebih dahulu menjadi teknologi fungsional yang lebih kecil (misalnya, regresi, klasifikasi, dan prediksi). Blok/layanan mikro fungsional ini kemudian dapat dikontainerisasi dan diterapkan secara dinamis dalam lingkungan cloud native. Interaksi antar layanan mikro ini biasanya difasilitasi menggunakan bus pesan.

Jika kita melihat *Open Network Automation Platform* (ONAP) misalnya, terdapat beberapa tempat seperti bagian Pengumpulan Data, Analisis, dan Peristiwa serta bagian Platform Manajemen Otomasi Loop Kontrol dan Kebijakan yang terdistribusi. Secara spesifik, loop kontrol (CL) adalah tempat kita dapat menemukan aplikasi AI untuk jaringan telekomunikasi otonom. Sebagai bagian dari jaminan layanan, berbagai jenis CL memantau fungsi jaringan untuk kinerja dan manajemen kesalahan. CL ini menjangkau berbagai bagian jaringan untuk menyelesaikan fungsi yang dibutuhkan. Untuk tindakan berkecepatan tinggi, CL seringkali perlu didistribusikan dekat dengan tepi jaringan. Untuk layanan yang kompleks, CL seringkali menjangkau beberapa lokasi dan pusat regional. Untuk pemantauan layanan yang lengkap, CL dapat menjangkau seluruh tepi jaringan dan komponen yang terdistribusi secara regional dan terpusat (Gambar 7.1).





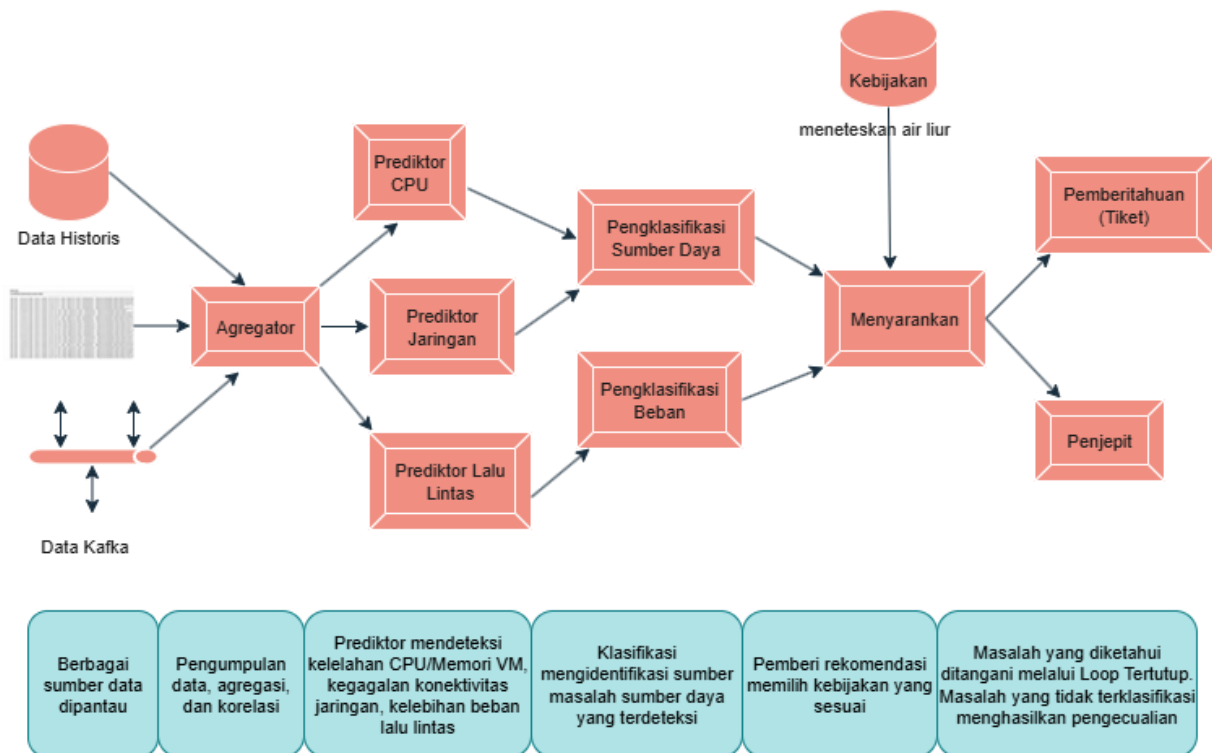
Gambar 7.1 Contoh loop kontrol ONAP yang menunjukkan penerapan cloud tepi, regional, dan pusat.

Alasan untuk memanfaatkan cloud native untuk layanan mikro AI adalah karena banyak CL dan analitik AI yang berbeda memiliki persyaratan yang berbeda pula. CL harus efisien saat diskalakan dan didistribusikan sesuai permintaan layanan atau jaringan. Persyaratan lain untuk AI dan model prediktor meliputi pemrosesan set data besar, pemindahan data atau analitik ke data sesuai kebutuhan. Hal ini seringkali memerlukan pemanfaatan perangkat keras berkinerja khusus (misalnya, GPU/TPU) dan koneksi ke akselerator untuk AI, seperti pembelajaran mendalam. Kerangka kerja berbasis cloud dan runtime memastikan eksekusi dengan cara yang tidak bergantung pada vendor, penyedia, tumpukan, lokasi, dan teknologi, sekaligus memberikan intelijen penting kembali ke ONAP untuk memperkuat perannya dalam “memutuskan”.

Mari kita lihat contoh ONAP *Virtual Machine* (VM) Resource Predictor. Terdapat sejumlah sumber data dan kolektor. Data diproses dan diintegrasikan untuk mempersiapkannya bagi analitik yang dikontainerisasi sebagai layanan mikro. Dalam hal ini, terdapat tiga kelas prediktor (CPU, Jaringan, Lalu Lintas) yang dibangun di pustaka SciKit Learn dan TensorFlow. Masing-masing merupakan model AI spesifik yang dilatih menggunakan data nyata. Pengklasifikasi mengidentifikasi perangkat spesifik untuk memetakan tindakan ke bagian layanan yang tepat. Pemberi rekomendasi bekerja dengan kebijakan untuk menentukan tindakan spesifik yang akan diambil (Gambar 7.2). Mungkin terdapat opsi yang berbeda, tergantung pada lokasi atau keadaan lain yang perlu diperhitungkan selama proses pengambilan keputusan.

Untuk cloud native, ini berarti setiap kolektor dapat ditempatkan dalam kontainer dengan sumber datanya membentuk paket efisien yang selaras dengan kebutuhan sumber daya (overhead rendah).





Gambar 7.2 Alur aliran data.

Prediktor terdiri dari beberapa artefak yang memanfaatkan struktur cloud native untuk elastisitas fungsional, konektivitas, dan abstraksi infrastruktur. Misalnya, prapemrosesan data juga dapat ditempatkan oleh pengumpul data untuk meminimalkan pergerakan data tambahan. Model pembelajaran dapat ditempatkan dan dijalankan dalam kontainer tempat sekelompok inti tertentu ditugaskan untuk memenuhi kebutuhan model dan menjamin persyaratan kinerja. Pelatihan dapat ditempatkan dalam kontainer yang memanfaatkan sumber daya GPU untuk menangani persyaratan pembelajaran mendalam. Kemudian, pelaporan dapat ditempatkan dalam kontainer di cloud pusat untuk menyederhanakan akses, interkoneksi, dan persyaratan pencatatan.

Singkatnya, cloud native akan memungkinkan tingkat efisiensi dan manajemen yang lebih tinggi dari solusi layanan mikro terdistribusi yang akan diterapkan dalam kontainer yang sangat sesuai dengan kebutuhan pemrosesan. Dengan pertumbuhan layanan mikro, efisiensi dan kemudahan pengelolaannya akan sangat penting untuk meningkatkan skala solusi.

Sejumlah Penyimpanan Data. Setiap Penyimpanan Data Tersebut

dikelola oleh satu layanan mikro, yang bertanggung jawab atas semua akses data ke penyimpanan data tersebut. Dalam beberapa kasus, penyimpanan data tersebut merupakan basis data yang digunakan bersama oleh sejumlah proses yang berbagi VM yang sama. Setiap penyimpanan data tersebut dapat menjadi bagian dari beberapa koleksi data besar. Setiap koleksi penyimpanan data dapat dianalisis oleh beberapa layanan mikro AI. Oleh karena itu, jaringan harus membawa sejumlah besar data yang dianalisis oleh sejumlah besar layanan mikro AI.



Setelah menganalisis data besar masing-masing dan menghasilkan hasil analisis, seperti aturan yang dapat ditindaklanjuti, layanan mikro tersebut harus kembali menganalisis data waktu nyata, membuat keputusan, dan mengomunikasikan keputusan tersebut melalui jaringan yang kelebihan beban.

Pencerminan (mirroring) adalah solusi brute force yang mahal. Pendekatan lain yang mungkin konsisten dengan konsep dekomposisi berbasis cloud adalah memisahkan proses analisis pembelajaran AI dari proses analisis waktu nyata AI, dan mungkin juga dari bagian kontrol menjadi dua atau tiga layanan mikro yang "terhubung secara longgar". Demikian pula, penyimpanan data dapat dibagi menjadi penyimpanan data jangka panjang dan penyimpanan data jangka pendek, masing-masing melayani layanan mikro pembelajaran jangka panjang dan layanan mikro analisis waktu nyata, yang ditempatkan bersama dengan penyimpanan datanya masing-masing. Setiap penyimpanan data juga dapat memiliki dua (atau lebih) layanan mikro, satu (atau lebih) bertanggung jawab untuk mengumpulkan data dan yang lainnya (atau lebih) bertanggung jawab untuk menyebarkan data.

Sebagai contoh, pertimbangkan ekosistem kota pintar yang mengelola sejumlah basis data dan repositori, termasuk penyimpanan gambar dan aliran video, serta beberapa aplikasi AI publik (misalnya, kotamadya) dan swasta independen yang menyediakan berbagai kombinasi data dan konten ini. Mungkin, jika tidak masuk akal, penyimpanan data ini (dan layanan mikro pengelolanya) dan layanan mikro AI dapat berjalan pada layanan cloud yang terpisah. Bahkan untuk layanan cloud yang sama, beberapa perangkat lunak manajemen cloud dapat digunakan. Lokasi layanan cloud dapat saling terhubung oleh lebih dari satu operator jaringan. Setiap kota diharapkan memiliki ekosistem yang serupa, meskipun tidak identik, di mana banyak layanan mikro direplikasi. Beberapa hasil pembelajaran mungkin berlaku di kota yang berbeda, sehingga memerlukan pertukaran hasil analisis, yang menjadikan berbagi data mentah antar kota juga bermanfaat. Pertukaran data mentah memerlukan standarisasi kualitas data, keunggulan data, dan ukuran serupa untuk nilai data yang dianalisis.

Oleh karena itu, terdapat jaringan "sistem". Setiap sistem mencakup beberapa subsistem, beberapa mungkin berbasis AI (dan yang lainnya melibatkan proses TI reguler). Subsistem AI dapat mencakup beberapa proses AI, seperti analitik big data, untuk membuat aturan dan analisis jangka pendek menggunakan aturan tersebut secara real-time. Proses-proses ini dapat diimplementasikan sebagai layanan mikro independen yang dapat diinstansiasi secara kolokasi dengan penyimpanan datanya, serta direplikasi untuk pelanggan yang berbeda (yang memiliki penyimpanan data yang berbeda).

Oleh karena itu, layanan mikro dari sistem yang sama dapat diinstansiasi di bawah perangkat lunak manajemen cloud yang berbeda serta di lokasi cloud yang berbeda dan saling terhubung oleh lebih dari satu jaringan telekomunikasi. Ketika sistem perangkat lunak manajemen cloud berbasis AI "mempelajari" perilaku cloud-nya, dan ketika perangkat lunak orkestrasi jaringan menganalisis lalu lintas jaringan, mereka sebenarnya mempelajari kombinasi perilaku layanan mikro AI di wilayah mereka, efek dari layanan mikro AI di luar wilayah mereka, dan efek yang mendasarinya dari sistem manajemen cloud dan orkestrasi jaringan di sekitarnya. Sementara aplikasi berorientasi konsumen baru yang sukses mungkin



membutuhkan waktu setidaknya berbulan-bulan untuk berkembang dan memengaruhi dunia dan jaringan cloud, jaringan layanan mikro berbasis AI dapat memiliki efek yang cukup besar dalam hitungan jam, jika tidak menit. Tantangannya adalah bahwa proses pembelajaran yang saling terkait ini konvergen, alih-alih meledak. Tantangan tersebut dapat dipenuhi, misalnya, melalui "umpan balik negatif" yang dipertukarkan antara sistem manajemen cloud dan orkestrasi jaringan.

Penanganan Alarm Palsu dalam Jaringan Pembelajaran Bertingkat

Sejauh ini, kita telah membahas beberapa permasalahan yang terkait dengan bisnis agile, cloud agile, dan jaringan agile. Kita telah membahas contoh-contoh permasalahan pemrosesan data, penyimpanan data, dan komunikasi, yang terkait dengan dinamika cloud dan jaringan yang dirancang untuk kelincuhan bisnis. Kita juga secara singkat membahas permasalahan struktur cloud manajemen yang rumit dan jaringan perantaranya. Kemudian, kita membahas beberapa aspek AI, sebagaimana AI dapat diimplementasikan dalam aplikasi cloud, aplikasi jaringan, manajemen cloud, dan manajemen jaringan (atau orkestrasi). Salah satu aspek manajemen cloud dan jaringan adalah pembuatan dan penanganan alarm, yang keduanya (secara independen) bergantung pada teknologi AI, analitik big data, pembelajaran mesin (ML), dll. Penanganan alarm merupakan bisnis yang sangat sensitif dan sangat mahal. Alarm palsu, baik positif maupun negatif palsu, dapat merugikan bisnis.

Dari sudut pandang sistem pemantauan, kompleks agile cloud dan agile network pada dasarnya tidak stabil, terutama ketika memantau sebagian dari kompleks tersebut. Kita mungkin memperkirakan sejumlah besar sistem pemantauan (kesalahan) menghasilkan berbagai jenis alarm (misalnya, terkait perangkat keras, perangkat lunak, beban, keamanan, atau konten). Setiap instansi manajemen cloud, dan setiap instansi manajemen jaringan, dikaitkan dengan sejumlah sistem pembangkit alarm yang berbeda, masing-masing menangani jenis alarmnya masing-masing.

Karena sifat tugas manajemen cloud dan jaringan yang tumpang tindih dan saling terhubung, satu penyebab tunggal dapat menghasilkan banyak alarm. Satu penyebab tunggal dapat memengaruhi perilaku modul dari berbagai lapisan cloud dan lapisan jaringan yang berbeda, serta modul yang terurai secara vertikal atau multi-instansiasi dan komponen cloud dan jaringan yang digunakan bersama oleh berbagai sistem manajemen. Mungkin terdapat sejumlah jenis alarm serupa, yang dihasilkan oleh sistem pemantauan dengan jenis yang sama, yang terkait dengan sistem manajemennya masing-masing. Penyebab yang sama juga dapat memicu sejumlah alarm dengan memantau sistem dari berbagai jenis yang melayani sistem manajemen cloud atau jaringan yang sama. Oleh karena itu, diperlukan lapisan pemantauan peringatan yang mengkonsolidasikan berbagai alarm dan memastikan bahwa alarm ditangani secara efisien.

Beberapa jenis alarm dipicu langsung oleh unit yang dipantau (perangkat keras atau perangkat lunak). Namun, sistem deteksi kesalahan cenderung menggunakan AI untuk menganalisis dan mendeteksi inkonsistensi sebagai sarana untuk memprediksi perkembangan kesalahan, mengantisipasi perkembangan beban, mendeteksi pelanggaran keamanan terselubung, dll. Mekanisme tersebut harus memahami dinamika cloud dan jaringan, karena



keduanya diubah oleh sistem manajemennya masing-masing. Selain masalah umum alarm positif palsu dan negatif palsu (kekurangan alarm), terdapat juga masalah alarm yang tidak tepat waktu, yang mungkin prematur atau terlambat. Alarm positif palsu dan tidak tepat waktu jelas mahal, tetapi juga menyebabkan "kelelahan alarm" (atau kelelahan peringatan) pada tim respons, yang menurunkan efisiensi, sehingga menambah biaya keseluruhan. Mengubah ambang batas alarm meningkatkan probabilitas kejadian negatif palsu, yang bahkan lebih mahal.

Tujuan dari semua sistem analisis kesalahan prediktif adalah untuk menilai kejadian yang terkait dengan kesalahan dan menyediakan alarm yang lebih akurat dan tepat waktu berdasarkan sejumlah indikasi. Indikasi tersebut dapat berasal dari serangkaian kejadian dengan jenis yang sama atau dari sejumlah kejadian yang berbeda namun terkait. Tujuan AI dan ML dalam deteksi kesalahan adalah untuk menentukan hubungan antara berbagai kejadian dan kemungkinan kesalahan, lalu menilai mereka berdasarkan hal tersebut. Penilaian menunjukkan parameter statistik terjadinya kesalahan, seperti probabilitas terjadinya kesalahan dalam jangka waktu tertentu (misalnya, waktu rata-rata antar kegagalan).

Hubungan antara kejadian dan kesalahan jelas terkait erat dengan dinamika lingkungan cloud dan jaringan, yang membutuhkan koordinasi yang erat antara sistem manajemen dan sistem deteksi kesalahan. Sistem manajemen melaporkan pemeliharaan umum, seperti kompresi log, pengumpulan sampah, dan rekonstruksi data, serta penerapan dan migrasi layanan mikro, sementara sistem analisis prediksi kesalahan melaporkan pemeringkatan probabilitas kesalahan untuk memungkinkan sistem manajemen membuat perencanaan ke depan. Mempertimbangkan AI dan khususnya ML, jelas bahwa seluruh kompleks jaringan cloud terus belajar, sebagai jaringan mesin pembelajaran independen yang terhubung. Perlu dicatat bahwa jaringan ini tidak dirancang dalam artian tidak ada entitas bisnis atau organisasi tertentu yang bertanggung jawab atas topologi jaringan ini. Oleh karena itu, pengamanan siber terhadap operasi jaringan menjadi tantangan tersendiri.

Mendebug Jaringan Pembelajaran dalam Lingkungan Layanan Mikro Penyebaran Berkelanjutan

Pengujian sistem memiliki tiga aspek. Satu aspek, yang relatif sederhana, adalah menguji apakah sistem melakukan apa yang disyaratkan spesifikasi. Aspek lain yang lebih rumit adalah menguji apakah sistem tidak melakukan apa yang dilarang spesifikasi. Aspek ketiga yang bermasalah adalah menguji "kesalahan yang tidak ditentukan." Kesalahan semacam itu biasanya diakibatkan oleh urutan operasi tertentu, kemungkinan besar dua atau lebih urutan operasi yang bersamaan, serta situasi terkait beban, seperti yang memengaruhi mekanisme penguncian data atau balapan kritis. Berbeda dengan "kesalahan yang ditentukan" yang cenderung stabil dan terlokalisasi, "kesalahan yang tidak ditentukan" cenderung tidak menentu, terputus-putus, dan "berkepanjangan". Kesalahan semacam itu dapat menyebabkan inkonsistensi data yang memburuk menjadi inkoherensi basis data yang menjalar melalui sistem sambil menyebarkan data yang tidak beraturan yang mungkin sulit ditemukan dan dipulihkan.



Pendekatan cloud-native dekomposisi horizontal dan vertikal dari tugas pemrosesan dan penyimpanan data, ditambah dengan CI-CD meningkatkan kemungkinan terjadinya "kesalahan yang tidak ditentukan". Menguji layanan mikro atau bahkan sistem dalam sandbox yang terisolasi tidak akan menghasilkan skenario kompleks yang mengarah ke "kesalahan yang tidak ditentukan". Pada akhirnya, dengan mempertimbangkan kesalahan yang tidak ditentukan, layanan mikro diuji "langsung", sambil menyediakan layanan. Oleh karena itu, sistem manajemen cloud, sistem manajemen jaringan, sistem pemantauan kesalahan, dll., selain mengawasi kompleks jaringan cloud, juga memantau efek dari sistem pengujian yang menguji layanan mikro atau sistem langsung.

Namun, sistem supervisi harus mengidentifikasi dan memisahkan efek pengujian dari perilaku jaringan cloud lainnya. Tentu saja, data pemantauan yang digunakan untuk ML tidak boleh mencakup efek pengujian. Oleh karena itu, catatan log harus menunjukkan apakah sumbernya nyata atau hasil pengujian. Karena kesalahan yang tidak ditentukan dihasilkan dari serangkaian operasi yang kompleks, atau bahkan beberapa rangkaian, sulit untuk mengaitkan kesalahan dengan sumber tertentu. Namun, identifikasi sumber tersebut penting untuk debugging, manajemen runtime, dan proses ML. Sistem pengujian juga diharapkan menggunakan AI untuk mengembangkan rangkaian data pengujian yang kemungkinan menyebabkan kesalahan yang tidak ditentukan. Di satu sisi, sistem pengujian menganalisis kode untuk menemukan kemungkinan kerentanan, dan di sisi lain, sistem pengujian menganalisis berkas log sistem deteksi kesalahan untuk menentukan situasi yang dapat menyebabkan kesalahan yang tidak ditentukan. Sistem pengujian kemudian menyuntikkan data pengujian sambil mensimulasikan masalah terkait perangkat keras, masalah terkait beban, dll.

Dapat diperkirakan bahwa setiap sistem layanan mikro mungkin memiliki setidaknya satu, dan mungkin lebih dari satu, sistem pengujian berbasis AI otomatis yang beroperasi secara runtime. Sistem-sistem ini mengganggu operasi runtime dan dalam fase pembelajaran, sehingga membutuhkan koordinasi yang ketat. Dengan mempertimbangkan stratifikasi kompleks jaringan awan, proses yang dijelaskan sebelumnya untuk lapisan pemrosesan aplikasi berulang di tingkat jaringan. Layanan mikro jaringan serta layanan mikro orkestrasi jaringan terus dirilis, diuji secara runtime, dan diawasi untuk deteksi kesalahan, sekaligus menyediakan layanan jaringan krusial untuk berbagai aplikasi.

7.4 MASA DEPAN JARINGAN AI BERBASIS CLOUD

Dari sudut pandang yang agak sinis, jaringan otonom mewujudkan kekeliruan terbalik dari komputasi terdistribusi:

- Jaringan awan tidak dapat diandalkan. Setidaknya ada satu kesalahan yang sekarang dimulai di suatu tempat.
- Latensi jaringan awan (terpisah dari komunikasi dan pemrosesan) cukup besar, berubah-ubah, lokasinya bervariasi, dan tidak dapat diprediksi.
- Bandwidth jaringan awan (terpisah dari komunikasi dan pemrosesan) terbatas, berubah-ubah, lokasinya bervariasi, dan tidak dapat diprediksi.



- Jaringan cloud tidak aman dan baru saja disusupi.
- Sejumlah administrator yang tidak terkoordinasi dengan baik (kebanyakan otomatis dan agak tidak terduga) kini mengelola jaringan cloud. Topologi jaringan cloud sedang diubah, oleh lebih dari satu administrator.
- Biaya transportasi dalam jaringan cloud sangat besar, berubah-ubah, lokasinya bervariasi, dan tidak dapat diprediksi (mengenai latensi, keamanan, keandalan, dll.). Hal yang sama berlaku untuk migrasi (mengenai waktu perubahan dan bandwidth, latensi, keamanan, keandalan, dll. yang dihasilkan).
- Jaringan tidak homogen dalam aspek apa pun.

Oleh karena itu, jaringan cloud pada dasarnya tidak stabil, dan konsistensi-ketersediaan-toleransi data tidak diketahui (atau diketahui dengan beberapa ketidakpastian terbatas). Tujuan jaringan otonom adalah untuk menyelesaikan masalah ini secara waktu nyata (real-time), sekaligus mengurangi ketidakstabilan jaringan cloud secara keseluruhan (dan ketidakpastian data), serta mengurangi biaya. Meskipun dunia mungkin tampak semakin tidak terprediksi, janji dari jaringan cloud otonom adalah ketika dampak buruk dirasakan, solusinya sudah dalam proses.

7.5 MENATAP KE DEPAN: PERUSAHAAN DI TEPI

Di tepi jaringan telekomunikasi publik terdapat jaringan perusahaan, dan tepi kedua jaringan tersebut cenderung saling mengaburkan. Operator jaringan telekomunikasi publik dapat menyediakan layanan seperti jaringan privat virtual (VPN), pengiris jaringan (network slicing), dan komputasi tepi kepada perusahaan; mengelola entitas perangkat keras dan perangkat lunak di lokasi; dan menyediakan komputasi seluler terobosan lokal.

Batas-batasnya juga kabur dari sudut pandang perusahaan. Dengan aplikasi perusahaan yang berjalan di ponsel pintar dan perangkat yang terhubung, serta berkomunikasi dengan perangkat IoT berbasis rumah, batas antara komputasi perusahaan dan lokasi pelanggan menjadi kabur. Demikian pula, batas antara lingkungan komputasi perusahaan dapat kabur ketika, misalnya, layanan komputasi yang berbeda dapat berkomunikasi dengan IoT publik yang sama (misalnya, kotamadya) dan perangkat IoT industri.

Kecerdasan buatan memasuki transformasi digital perusahaan, dan disiplin internal perencanaan, produksi, promosi, dan penyediaan pun menjadi kabur. Karena semua batas menjadi kabur satu sama lain, AI sendiri pun larut ke dalam batas-batas yang kabur ini, baik dalam arti pemrosesan komputer maupun proses bisnis. Bab selanjutnya membahas AI dalam transformasi digital perusahaan.



BAB 8

PLATFORM PERANGKAT LUNAK DAN KERAS AI

8.1 PENDAHULUAN

Kecerdasan buatan (AI) memainkan peran penting dalam mendorong jaringan otonom dengan kapabilitas seperti pemulihan mandiri dan otomatisasi loop tertutup. Kecerdasan buatan membutuhkan investasi yang signifikan pada platform perangkat keras dan perangkat lunak untuk memungkinkan solusi yang cepat, skalabel, dan hemat biaya. Bab ini akan menjelaskan platform perangkat keras yang sedang berkembang, seperti unit pemrosesan grafis (GPU) dan prosesor Xeon berkinerja tinggi dari Intel, serta kerangka kerja perangkat lunak standar industri, seperti Hadoop, TensorFlow, dan Spark. Selain itu, kami akan membahas peluang untuk akselerasi fungsi tetap beban kerja AI.

Bagian 8.2 menyajikan ikhtisar arsitektur sistem AI, termasuk prinsip-prinsip pembelajaran mesin statistik dan pembelajaran mendalam beserta kasus-kasus penggunaan utama. Bagian 8.3 memperkenalkan seni ilmu data dan alur kerja menyeluruh. Bagian 8.4 menunjukkan kerangka kerja, pustaka, dan primitif perangkat lunak pembelajaran mesin dan pembelajaran mendalam. Bagian 8.5 menjelaskan berbagai kapabilitas perangkat keras dan kesesuaian arsitekturnya untuk kasus-kasus penggunaan utama. Bagian 8.6 mengeksplorasi bagaimana perangkat lunak AI berinteraksi dengan perangkat keras untuk beban kerja utama. Dan Bagian 8.7 merangkum inovasi baru dan penelitian mendatang di bidang ini.

8.2 KECERDASAN BUATAN

Pembelajaran mesin dan AI telah hadir di pusat data. Kemampuan komputasi yang belum pernah ada sebelumnya dan teknik akselerasi perangkat keras menyediakan daya komputasi yang lebih besar untuk melatih sistem pembelajaran mesin, sebuah proses yang melibatkan pemrosesan data dalam jumlah besar.



Gambar 8.1 Taksonomi AI.



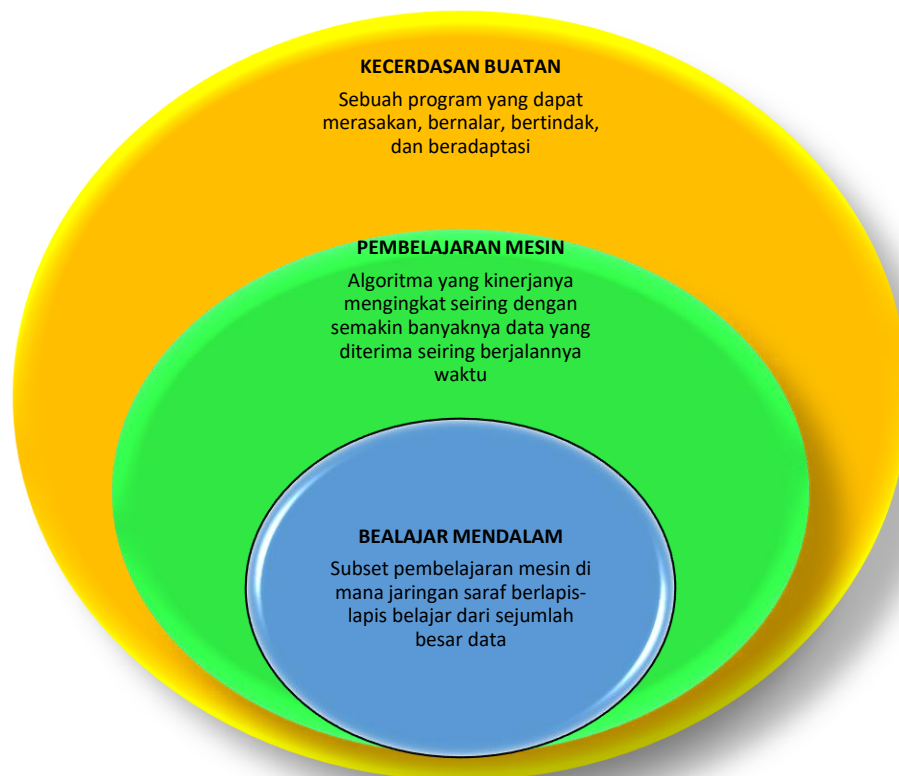
Kecerdasan buatan kini mentransformasi cara pusat data dibangun dan dioperasikan, sekaligus memberdayakan setiap aspek internet. Pembelajaran mesin juga mentransformasi cara pengembang membangun aplikasi cerdas yang bermanfaat bagi pelanggan dan konsumen, yang menyentuh hampir setiap aspek kehidupan kita sehari-hari. Salah satu aspek AI yang paling menarik adalah potensinya untuk merevolusi bukan hanya industri komputasi, atau industri perangkat lunak, tetapi juga setiap industri yang menyentuh kehidupan kita. AI akan mentransformasi masyarakat dengan cara yang sama seperti revolusi industri revolusi teknis dan revolusi digital telah mengubah setiap aspek kehidupan sehari-hari. Bagi pengembang, perluasan bidang AI berarti Anda berpotensi menerapkan minat dan pengetahuan AI Anda ke industri yang juga Anda minati, seperti musik, olahraga, atau perawatan kesehatan. Dalam bab ini, kita akan mulai dengan menjelaskan arsitektur sistem AI. Namun, sebelum itu, mari kita pahami terlebih dahulu taksonomi AI berikut (Gambar 8.1).

Meskipun terdapat banyak cara berbeda untuk memahami AI dan beragam teknik untuk mendekatinya, kunci kecerdasan mesin adalah ia harus mampu merasakan, bernalar, dan bertindak, lalu beradaptasi berdasarkan pengalaman. Setiap tahapan pemrosesan data ini dijelaskan sebagai berikut:

- **Merasakan:** Mengidentifikasi dan mengenali objek atau konsep yang bermakna di tengah data yang sangat besar. Apakah ada lampu lalu lintas dalam video? Apakah itu tumor atau jaringan normal dalam gambar?
- **Bernalar:** Memahami konteks yang lebih luas dan membuat rencana untuk mencapai tujuan. Misalnya, jika tujuannya adalah menghindari tabrakan, mobil harus menghitung kemungkinan kecelakaan berdasarkan perilaku kendaraan, jarak, kecepatan, dan kondisi jalan.
- **Bertindak:** Merekomendasikan atau langsung memulai tindakan terbaik. Misalnya, berdasarkan analisis kendaraan dan lalu lintas, sistem dapat mengerem, mempercepat, atau menyiapkan beberapa mekanisme keselamatan lainnya.
- **Beradaptasi:** Terakhir, kita harus mampu mengadaptasi algoritma di setiap fase berdasarkan pengalaman, melatihnya kembali agar semakin cerdas. Algoritma kendaraan otonom harus dilatih ulang untuk mengenali lebih banyak titik buta, memasukkan variabel baru ke dalam konteks, dan menyesuaikan tindakan berdasarkan insiden sebelumnya.

Dua pendekatan paling populer yang digunakan pengembang untuk memprogram mesin agar dapat melakukan tugas-tugas sebelumnya adalah pembelajaran mesin statistik dan pembelajaran mendalam (Gambar 8.2).





Gambar 8.2 Konsep kecerdasan buatan, pembelajaran mesin, dan pembelajaran mendalam.

Pembelajaran Mesin Statistik

Dalam pembelajaran mesin, algoritma pembelajaran membangun model dari data. Model-model ini dapat terus ditingkatkan seiring dengan semakin banyaknya data yang terpapar seiring waktu. Ada empat jenis utama pembelajaran mesin: terawasi, tak terawasi, penguatan, dan pembelajaran berkelanjutan. Dalam pembelajaran mesin terawasi, algoritma belajar mengidentifikasi data dengan memproses dan mengkategorikan sejumlah besar data berlabel. Dalam pembelajaran mesin tak terawasi, algoritma mengidentifikasi pola dan kategori dalam sejumlah besar data tak berlabel. Pembelajaran penguatan memungkinkan mesin atau agen perangkat lunak untuk mempelajari perilakunya berdasarkan umpan balik dari lingkungan. Dan pembelajaran berkelanjutan (juga disebut pembelajaran seumur hidup) dibangun di atas gagasan belajar secara terus-menerus dan adaptif tentang dunia luar.

Pembelajaran mesin statistik menggabungkan statistik dengan ilmu komputasi. Ini adalah ruang lingkup dalam pembelajaran mesin yang berkaitan dengan pengembangan algoritma dan teknik untuk tujuan pembelajaran dari data yang diamati dengan membangun model stokastik yang dapat digunakan untuk membuat prediksi dan keputusan.

Pembelajaran Mendalam

Pembelajaran mendalam adalah subset lain dari pembelajaran mesin di mana jaringan saraf berlapis-lapis dilatih dengan sejumlah besar data. Pembelajaran mendalam (juga terkadang disebut sebagai pembelajaran terstruktur dalam atau pembelajaran hierarkis) adalah salah satu bagian dari keluarga metode pembelajaran mesin yang lebih luas yang



didasarkan pada pembelajaran representasi data, bukan algoritma khusus tugas. Keempat jenis metode pembelajaran mesin seperti yang disebutkan sebelumnya dapat menyertakan model pembelajaran mendalam dalam implementasinya (Gambar 8.3).



Gambar 8.3 Contoh kasus penggunaan pembelajaran mendalam.

Sistem Penalaran

Berbeda dengan sistem pembelajaran mesin, sistem penalaran adalah sistem perangkat lunak yang menghasilkan kesimpulan dari pengetahuan yang tersedia menggunakan teknik-teknik logis, seperti deduksi dan induksi. Sistem penalaran memainkan peran penting dalam implementasi AI dan sistem berbasis pengetahuan. Terdapat dua jenis sistem penalaran: (1) penalaran berbasis memori dan (2) penalaran berbasis logika. Penalaran berbasis memori bertujuan untuk mengidentifikasi kasus-kasus serupa dari pengalaman atau ingatan masa lalu, dan menerapkan informasi dari pengalaman masa lalu tersebut pada permasalahan yang dihadapi. Penalaran berbasis logika bertujuan untuk mempelajari pengetahuan berbasis aturan, yang disebut hipotesis, dari observasi (positif dan negatif), menggunakan pengetahuan latar belakang dan batasan integritas yang ada.

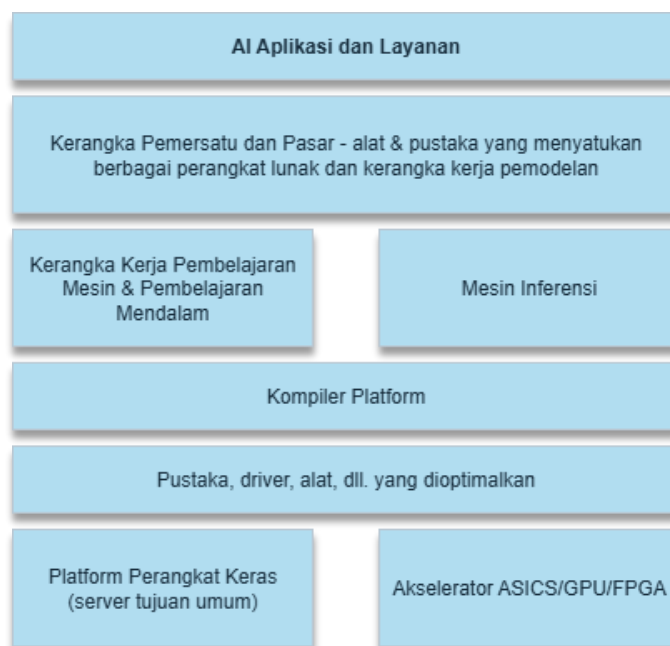
Arsitektur Sistem Kecerdasan Buatan

Beberapa tahun ke depan akan menyaksikan pertumbuhan pesat AI yang didorong oleh munculnya teknologi 5G dan aplikasi serta layanan baru di tepi jaringan, seperti ritel, layanan



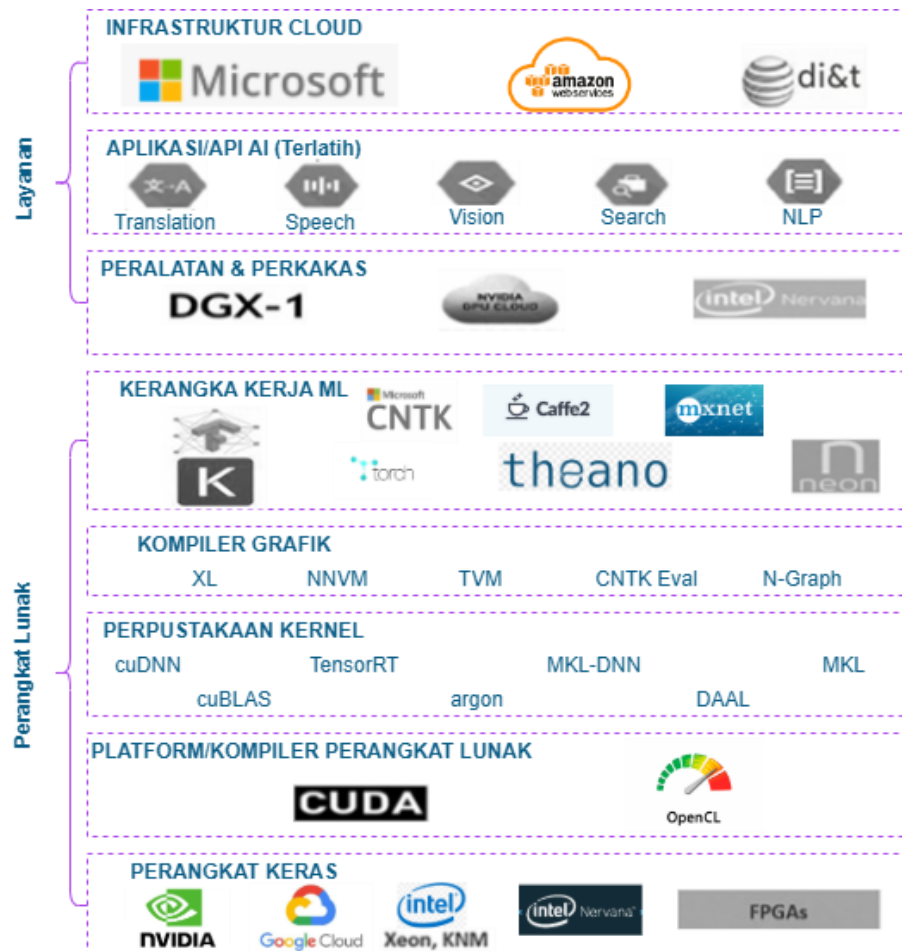
kesehatan, mengemudi otonom, pemrosesan bahasa alami, pengawasan dan keamanan, cloud gaming, *augmented reality* (AR)/*virtual reality* (VR), dan masih banyak lagi. Untuk pembuatan, penerapan, dan adopsi aplikasi AI yang cepat, diperlukan arsitektur modular dan skalabel yang didefinisikan dengan mempertimbangkan beberapa prinsip dasar:

- Perangkat keras serbaguna untuk mengurangi biaya dan mendorong siklus inovasi yang lebih cepat
- Akselerator perangkat lunak dan perangkat keras, seperti sirkuit terpadu khusus aplikasi (ASIC) atau *field-programmable gate array* (FPGA), merupakan solusi yang memberikan manfaat signifikan dari segi kinerja dan biaya
- Kemampuan untuk memanfaatkan model AI yang dibangun dengan berbagai kerangka kerja perangkat lunak, seperti Caffe2, Theano, atau TensorFlow
- Memungkinkan penerapan model pembelajaran mesin terlatih yang optimal pada berbagai platform perangkat keras melalui mesin inferensi
- Menciptakan "komunitas model AI" sebuah pasar tempat para pengembang berkontribusi dan menyempurnakan model dan perangkat lunak AI
- Mendorong inovasi aplikasi dan layanan yang lebih cepat dengan perangkat lunak sumber terbuka Berikut adalah arsitektur sistem AI modular tingkat tinggi (Gambar 8.4).



Gambar 8.4 Arsitektur sistem AI modular.





Gambar 8.5 Solusi perangkat keras dan perangkat lunak untuk pembelajaran mendalam.

Gambar di atas adalah pemetaan arsitektur sistem sebelumnya ke solusi perangkat keras dan perangkat lunak yang digunakan untuk membangun aplikasi pembelajaran mesin saat ini (Gambar 8.5). Bagian selanjutnya akan memberikan detail lebih lanjut tentang arsitektur perangkat keras dan perangkat lunak.

8.3 SENI ILMU DATA

Ilmu data merupakan perpaduan keahlian dalam interpretasi data, pengembangan algoritma, dan pemrograman perangkat lunak untuk memecahkan masalah analitis yang kompleks secara komputasi. Seni ilmu data melibatkan empat langkah utama: (1) Akuisisi data, (2) Persiapan data, (3) Pengembangan/validasi model, dan (4) Penerapan dan inferensi model (Gambar 8.6).

Pengumpulan Data

Miliaran pengguna dan perangkat menghasilkan data dalam jumlah besar setiap hari dalam berbagai format dengan kecepatan yang berbeda-beda. Langkah pertama dalam proses ilmu data adalah memperoleh data yang bermakna dan berdampak dari berbagai sumber tersebut, misalnya, sensor di *Internet of Things* (IoT), cloud, dan Web. Ilmuwan data biasanya menyimpan kumpulan data di penyimpanan cloud privat atau publik, seperti bucket Amazon Web Service S3.





Gambar 8.6 Alur kerja ilmu data.

Transformasi Data

Mengagregasi dan mengkurasi set data merupakan salah satu langkah yang paling memakan waktu dalam alur kerja ilmu data. Untuk pembelajaran mesin terawasi, set data yang terkumpul perlu diberi label dan diformat. Sebagai contoh pelabelan data, untuk mendeteksi tingkat korosi baut di anjungan minyak secara otomatis, ilmuwan data perlu membuat set data sampel pelatihan dengan mengekstrak bingkai dari rekaman video, menggambar kotak pembatas di sekitar setiap baut dengan koordinat, dan menetapkan tingkat intensitas korosi secara manual.

Pembuatan Model

Selanjutnya, ilmuwan data akan mengeksplorasi model yang ada untuk mengekstrak nilai dari set data dengan cepat. Seringkali, mereka menggunakan model dari model zoo (yang merupakan kumpulan model yang sudah terlatih) dan menyesuaikannya agar paling sesuai dengan set data mereka. Setelah model yang tepat diidentifikasi, mereka menggunakan persentase tertentu (berdasarkan konvensi, biasanya sekitar 70%) dari set data masukan mereka untuk melatih model.

Pelatihan Model

Pelatihan model merupakan salah satu langkah komputasi yang paling intensif dalam alur kerja. CPU atau GPU umumnya digunakan untuk mendukung langkah pelatihan, yang dapat berlangsung selama beberapa jam hingga beberapa hari, tergantung pada kompleksitas model dan ukuran set data pelatihan. Setelah tingkat akurasi yang diharapkan tercapai, model diharapkan siap untuk validasi. Sisa set data input (misalnya, 30%) biasanya digunakan untuk validasi model.

Inferensi Model

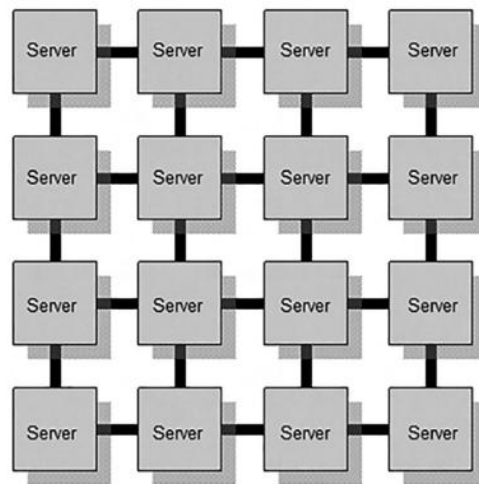
Setelah tingkat akurasi yang diharapkan dari prediksi model diverifikasi dengan set data validasi, langkah selanjutnya dalam alur kerja adalah inferensi model. Tergantung pada solusi yang sedang dikembangkan, model diterapkan di sisi server atau di perangkat edge, seperti ponsel atau kamera pintar. Tergantung pada arsitektur prosesor perangkat edge, teknik seperti kuantisasi dapat diterapkan pada model yang telah dilatih untuk mengecilkan ukurannya agar sesuai dengan ruang memori yang terbatas tanpa kehilangan tingkat akurasi prediksinya. Langkah inferensi memasukkan sampel data baru ke dalam model dan menghasilkan skor prediksi dengan tingkat akurasi dan keyakinan. Semakin banyak data yang diproses, model disempurnakan menjadi lebih akurat. Misalnya, saat mobil otonom melaju, aplikasi menarik informasi waktu nyata melalui sensor, GPS, perekaman video 360 derajat, dan lainnya, yang kemudian dapat digunakan untuk mengoptimalkan prediksi di masa mendatang.



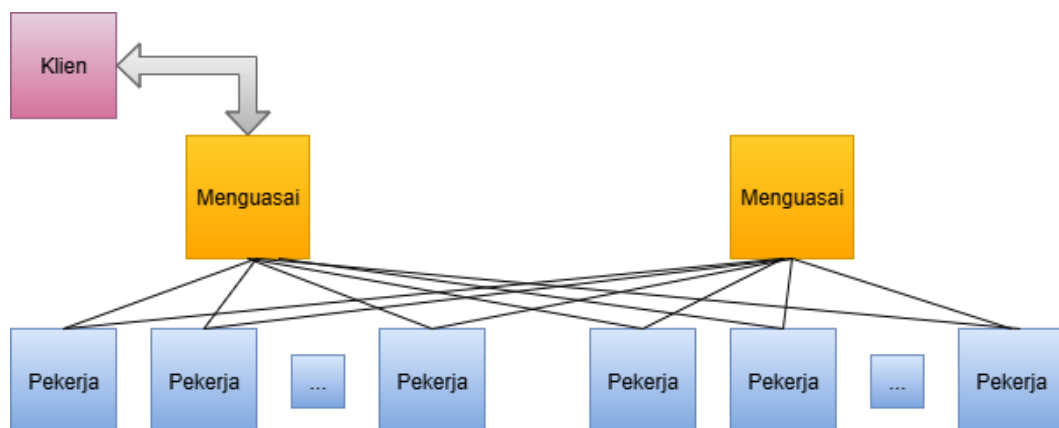
Kecerdasan Buatan Terdistribusi pada Skala Hiper

Banyak penyedia layanan cloud besar menerapkan algoritma AI di beberapa aplikasi mereka, seperti sistem rekomendasi, analisis perilaku/sentimen konsumen, peningkatan efisiensi operasional, dan penempatan iklan. Biasanya, model AI mereka kompleks dan kumpulan data sampel berukuran beberapa terabyte. Melatih model AI dengan kumpulan data besar tersebut pada satu CPU atau GPU hingga mencapai tingkat akurasi tinggi membutuhkan waktu beberapa hari atau minggu. Untuk mengurangi waktu pelatihan, beberapa kerangka kerja perangkat lunak AI telah diaktifkan dengan dukungan multi-CPU atau GPU.

Implementasi multiprosesor ini menggunakan prinsip komputasi terdistribusi seperti antarmuka penyampaian pesan seperti yang ditunjukkan pada Gambar 8.7 atau pemanggilan prosedur jarak jauh seperti yang ditunjukkan pada Gambar 8.8 untuk memungkinkan komunikasi multinode all-to-all dan/atau arsitektur master-slave.



Gambar 8.7 Komunikasi all-to-all, menggunakan antarmuka penyampaian pesan.



Gambar 8.8 Arsitektur master-slave, menggunakan pemanggilan prosedur jarak jauh.

8.4 KERANGKA KERJA PERANGKAT LUNAK KECERDASAN BUATAN

Seni AI telah memikat imajinasi kita sejak pertengahan abad ke-20, menginspirasi para ilmuwan komputer untuk menciptakan teknologi komputasi yang semakin kompleks, sekaligus



menciptakan solusi baru yang berorientasi pelanggan untuk meningkatkan kehidupan sehari-hari masyarakat. Kemajuan dalam tiga bidang berikut telah mengubah lanskap komputasi AI secara fundamental dan mendorong demokratisasi pembelajaran mesin: (1) Biaya unit komputasi telah menurun dari tahun ke tahun sehingga menjadikan AI lebih terjangkau; (2) Demokratisasi data di mana lembaga dan pemerintah bersedia berbagi data (dengan syarat kepatuhan dan privasi data tertentu) demi kebaikan bersama, yang mempercepat adopsi algoritma pembelajaran mesin; dan (3) Konvergensi kerangka kerja perangkat lunak sumber terbuka telah menyederhanakan implementasi algoritma pembelajaran mesin dan pembelajaran mendalam yang sebelumnya kompleks.

Beberapa kerangka kerja perangkat lunak sumber terbuka AI yang populer telah muncul dalam tiga tahun terakhir, terutama dari kalangan akademisi dan penyedia layanan cloud besar, seperti Google, Facebook, dan Microsoft. Selain itu, beberapa set data terbuka berlabel telah tersedia, misalnya, set data ImageNet telah memacu penemuan-penemuan penting pada topologi jaringan saraf tiruan.

Sebagian besar kerangka kerja AI populer mendukung bahasa pemrograman terinterpretasi seperti Python. Ilmuwan data dapat menggunakan kembali banyak modul komputasi numerik Python yang didukung dan juga mengimpor modul perangkat lunak dari ilmuwan data lain dengan sangat mudah. Meskipun kerangka kerja perangkat lunak ini, pada tingkat tinggi, mendukung sebagian besar algoritma dan operasi pembelajaran mesin dan pembelajaran mendalam yang umum, banyak di antaranya juga unggul dalam satu atau lebih implementasi tertentu, seperti analitik gambar dan video, atau pemrosesan ucapan dan bahasa alami.

TensorFlow

TensorFlow (<https://www.tensorflow.org/>) awalnya dikembangkan oleh para peneliti dan teknisi di Tim Google Brain dalam organisasi riset Kecerdasan Mesin Google. Beberapa aplikasi Google seperti layanan Search Engine, YouTube, Home, dan Advertisement mereka menggunakan TensorFlow secara intensif untuk tujuan melakukan riset pembelajaran mesin dan jaringan saraf tiruan dalam serta mendorong pengembangan dan penerapan model. Kasus penggunaan meliputi rekomendasi, pemahaman bahasa alami, pengenalan ucapan, klasifikasi gambar, dll.

TensorFlow diimplementasikan dalam Python dan menggunakan grafik aliran data untuk komputasi numerik. Karena adopsi yang lebih luas di Google dan komunitas eksternal, TensorFlow saat ini merupakan framework sumber terbuka terdokumentasi terbaik yang tersedia. TensorFlow memiliki arsitektur modular dan antarmuka pengguna yang mudah digunakan dengan dukungan untuk jaringan saraf konvolusional (CNN) dan jaringan saraf berulang (RNN). Selain sejumlah algoritma pembelajaran mendalam, TensorFlow juga mendukung algoritma pembelajaran mesin statistik selektif. TensorBoard adalah serangkaian alat visualisasi yang dapat digunakan untuk memahami, men-debug, dan mengoptimalkan program yang berjalan pada kode TensorFlow.



Caffe

Caffe (<http://caffe.berkeleyvision.org/>) dikembangkan oleh Berkeley AI Research dan kontributor komunitas. Caffe ditulis dalam C++ dengan antarmuka Python. Caffe merupakan proyek sumber terbuka yang dihosting di GitHub. Caffe banyak digunakan dalam proyek penelitian akademis dan prototipe rintisan, serta aplikasi industri skala besar di bidang visi, ucapan, dan multimedia. Pada April 2017, Facebook mengumumkan Caffe2 (<https://caffe2.ai/>), yang mencakup fitur-fitur baru, seperti RNN.

Neon

Neon (<https://www.intelnervana.com/neon/>) adalah kerangka kerja pembelajaran mendalam referensi Intel Nervana yang dirancang untuk ekstensibilitas dan kemudahan penggunaan. Neon adalah kerangka kerja sumber terbuka yang mendukung Python dan beberapa implementasi pembelajaran mendalam, seperti CNN, RNN, model memori jangka pendek panjang, dan auto-encoder dalam. Neon akan menjadi salah satu kerangka kerja pertama yang menawarkan dukungan untuk produk perangkat keras ASIC pembelajaran mendalam Intel Nervana yang saat ini sedang dalam pengembangan.

Hadoop

Proyek Apache™ Hadoop® (<http://hadoop.apache.org/>) mengembangkan perangkat lunak sumber terbuka untuk komputasi terdistribusi yang andal dan skalabel. Pustaka perangkat lunak Apache Hadoop adalah kerangka kerja yang memungkinkan pemrosesan terdistribusi kumpulan data besar di seluruh kluster komputer menggunakan model pemrograman sederhana. Kerangka kerja Hadoop mendukung beberapa algoritma pembelajaran mesin statistik melalui pustaka Perangkat Lunak Mahout dan dirancang untuk ditingkatkan skalanya dari satu server menjadi ribuan mesin, yang masing-masing menawarkan komputasi dan penyimpanan lokal. Hadoop dirancang untuk menangani ketidakstabilan perangkat keras dan dengan menghadirkan redundansi melalui teknik perangkat lunak seperti replikasi data rangkap tiga, memberikan layanan dengan ketersediaan tinggi di atas kluster komputer, yang masing-masing mungkin rentan terhadap kegagalan.

Proyek Hadoop mencakup beberapa modul (Gambar 8.9). Hadoop Umum: Utilitas umum yang mendukung modul Hadoop lainnya. *Hadoop Distributed File System* (HDFS™): Sistem berkas terdistribusi yang menyediakan akses throughput tinggi ke data aplikasi. Hadoop YARN: Kerangka kerja untuk penjadwalan pekerjaan dan manajemen sumber daya kluster. Hadoop MapReduce: Sistem berbasis YARN untuk pemrosesan paralel kumpulan data besar.





Gambar 8.9 Modul kerangka kerja Hadoop.

Spark

Apache Spark™ adalah mesin yang cepat dan umum untuk pemrosesan data skala besar. Mesin ini dibangun oleh komunitas pengembang yang besar dari lebih dari 200 perusahaan (dengan komiter dari 19 organisasi). Sejak 2009, lebih dari 1000 pengembang telah berkontribusi pada Spark.

Spark menawarkan lebih dari 80 operator tingkat tinggi yang memudahkan pembuatan aplikasi paralel. Dan Anda dapat menggunakannya secara interaktif dari shell Scala, Python, dan R. Spark mendukung setumpuk pustaka termasuk SQL dan Data Frames, MLlib untuk pembelajaran mesin, BigDL untuk pembelajaran mendalam, GraphX, dan Spark Streaming. Pustaka-pustaka ini dapat digabungkan dengan mulus dalam aplikasi yang sama. Spark berjalan di Hadoop, Mesos, mandiri, atau di cloud. Spark dapat mengakses beragam sumber data, termasuk HDFS, Cassandra, HBase, dan S3, yang menawarkan kemampuan untuk membawa komputasi ke node yang memiliki data, sehingga meminimalkan atau menghilangkan kebutuhan untuk memindahkan data.

Scikit-Learn

Scikit-Learn (<http://scikit-learn.org/>) adalah kerangka kerja pembelajaran mesin berbasis Python, yang dibangun di atas pustaka seperti NumPy, SciPy, dan matplotlib. Kerangka kerja ini menyediakan alat yang sederhana dan efisien untuk tugas-tugas analisis data, seperti klasifikasi, regresi, dan pengelompokan, serta prapemrosesan data, reduksi dimensionalitas, dan pemilihan model. Kerangka kerja ini bersifat sumber terbuka dan dapat digunakan secara komersial.

MXNet

MXNet (<https://mxnet.incubator.apache.org/>) adalah kerangka kerja pembelajaran mendalam yang dikembangkan oleh kolaborator dari berbagai perusahaan dan universitas, termasuk Microsoft, Nvidia, Baidu, Intel, Carnegie Mellon University, University of Alberta, dan



University of Washington, yang saat ini merupakan proyek inkubator Apache. MXNet mendukung berbagai bahasa pemrograman, termasuk R, Python, Julia, dan Scala. MXNet memiliki dukungan GPU yang lebih canggih dibandingkan kerangka kerja lain dan juga relatif cepat dalam hal runtime algoritma pembelajaran mendalam. MXNet memiliki kemampuan pemodelan untuk CNN dan RNN.

H2O

H2O (<https://www.h2o.ai/>) adalah perangkat lunak sumber terbuka untuk analitik data besar. Perangkat lunak ini ditulis dalam Java, Python, dan R. Proyek H2O mengklaim mengembangkan antarmuka analitis untuk komputasi awan, dan antarmuka pengguna grafisnya kompatibel dengan sebagian besar peramban internet yang umum digunakan. H2O memungkinkan pengguna untuk menganalisis dan memvisualisasikan seluruh rangkaian data dan memiliki algoritma statistik, seperti pengelompokan K-means, model linier umum, hutan acak terdistribusi, peningkatan gradien, Bayes naif, dan analisis komponen utama.

Pustaka (Intel MKL/MKL-DNN, Nvidia CUDA)

Kerangka kerja perangkat lunak kecerdasan buatan bergantung pada pustaka perangkat lunak dari vendor silikon, seperti Intel dan Nvidia, untuk mengakses arsitektur set instruksi tanpa harus memprogram pada tingkat perakitan. Sebagai praktik umum, vendor silikon membuat pustaka tingkat rendah yang mencakup sejumlah kernel matematika dan numerik (kode assembly) dan mengeksposnya ke aplikasi tingkat tinggi melalui antarmuka pemrograman aplikasi umum.

Intel *Math Kernel Library* (MKL) memiliki fungsi yang sangat optimal, berulir, dan tervektorisasi untuk aljabar linear padat dan jarang (BLAS, LAPACK, dan PARDISO), Transformasi Fourier Cepat (FFT), matematika vektor, statistik ringkasan, primitif jaringan saraf dalam (DNN), dan banyak lagi. MKL ditawarkan sebagai format biner yang dapat dihubungkan dengan bahasa pemrograman tingkat tinggi saat mengompilasi untuk memanfaatkan sepenuhnya kemampuan silikon Intel.

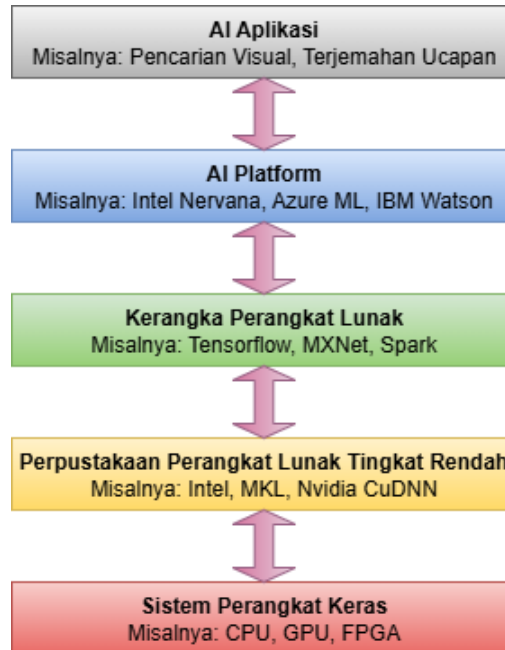
Intel MKL-DNN adalah pustaka kinerja sumber terbuka untuk akselerasi kerangka kerja pembelajaran mendalam pada arsitektur Intel®. Ini mencakup primitif yang sangat tervektorisasi dan terulir untuk implementasi DNN dengan antarmuka C dan C++. Nvidia menawarkan koleksi pustaka GPU (misalnya, cuDNN, cuFFT, cuSPARSE, cuBLAS, cuRAND) yang ditujukan untuk aplikasi pembelajaran mesin dan pembelajaran mendalam. Semua kerangka kerja pembelajaran mesin populer terhubung ke pustaka CUDA ini untuk memanfaatkan sepenuhnya kapabilitas perangkat keras GPU.

Platform

Saat ini terdapat cukup banyak platform perangkat lunak AI yang tersedia, masing-masing memiliki fitur dan domain aplikasi yang unik. Di antaranya, Amazon Machine Learning, Microsoft Azure Learning Studio, dan IBM Watson adalah beberapa yang paling terkenal. Platform Acumos yang baru-baru ini diumumkan (<https://www.acumos.org/>) bertujuan untuk menciptakan standar industri agar aplikasi AI dapat digunakan kembali dan mudah diakses oleh pengembang mana pun.



Gambar 8.10 mengilustrasikan tumpukan solusi AI lengkap, yang mencakup (dari tingkat rendah hingga tingkat tinggi) sistem perangkat keras, pustaka tingkat rendah, kerangka kerja, platform, dan aplikasi.



Gambar 8.10 Tumpukan solusi platform kecerdasan buatan.

8.5 PILIHAN PERANGKAT KERAS KECERDASAN BUATAN

Aplikasi kecerdasan buatan membutuhkan perangkat keras komputer yang canggih, yang berarti pasar chip AI berpotensi sangat besar. Itulah sebabnya banyak perusahaan, mulai dari raksasa semikonduktor hingga perusahaan rintisan, ikut serta. Berbagai pilihan perangkat keras dapat dikategorikan ke dalam empat jenis perangkat dasar: CPUs, GPUs, FPGAs, dan ASICs (Gambar 8.11).

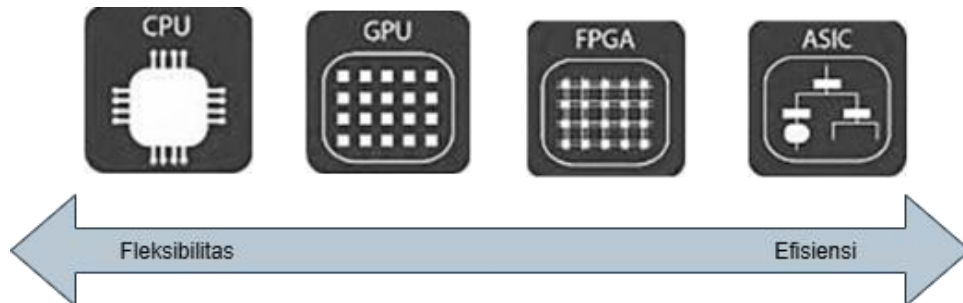
Secara heuristik, langkah pengumpulan data, transformasi, dan pembuatan model dalam alur kerja AI dijalankan pada CPU tujuan umum bersama aplikasi web dan basis data. Untuk langkah pelatihan intensif komputasi, ilmuwan data dapat memilih dari empat jenis sebelumnya. Mayoritas aplikasi pembelajaran mesin statistik dapat berjalan pada CPU tujuan umum. Kerangka kerja pelatihan pembelajaran mendalam memerlukan komputasi intensif dalam memori (khususnya, perkalian matriks), dan arsitektur GPU seringkali merupakan solusi yang lebih sesuai.

CPU (Arsitektur x86)

CPU adalah platform yang paling fleksibel dan mudah digunakan, yang berfungsi sebagai "otak" dari berbagai sistem komputasi, mulai dari seluler, tablet, konsumen (laptop/desktop), hingga server perusahaan. Saat ini, CPU adalah platform yang paling umum digunakan untuk pembelajaran mesin klasik (baik pelatihan maupun inferensi), serta inferensi pembelajaran mendalam (yaitu, penerapan run-time). Dua tahap pelatihan dan inferensi dalam pembelajaran mendalam sangat berbeda, dan masing-masing memerlukan gaya



pemrosesan yang berbeda. Secara khusus, inferensi pada dasarnya merupakan kalkulasi sekuensial (meskipun beberapa paralelisme dapat dicapai selama satu operasi inferensi, hal ini dibatasi oleh dependensi data yang ditentukan oleh arsitektur jaringan dan tingkat paralelisme umumnya rendah).



Gambar 8.11 Pilihan perangkat keras untuk aplikasi AI.

Sebagian besar ilmuwan data tidak memerlukan perangkat yang dioptimalkan untuk inferensi kecuali mereka berencana untuk melakukan pemrosesan data volumetrik di pusat data. Aplikasi seperti mengemudi otonom, pengawasan waktu nyata, dan yang berjalan pada perangkat IoT edge akan melakukan inferensi sekuensial, alih-alih paralel secara masif. Inferensi item data individual akan didominasi oleh kinerja sekuensial perangkat. Dalam kasus seperti itu, CPU merupakan pilihan yang jauh lebih baik daripada perangkat paralel masif, seperti akselerator.

Di sisi pelatihan, CPU dapat memberikan kinerja yang lebih tinggi daripada GPU pada banyak masalah pelatihan, terutama ketika ukuran data pelatihan relatif kecil. Akselerator mencapai kinerja tinggi ketika mereka memiliki sejumlah besar utas bersamaan untuk dieksekusi. Ketika pelatihan dengan set data yang hanya berisi ratusan hingga puluhan ribu contoh, hanya sebagian kecil dari paralelisme akselerator yang dapat digunakan. Dalam situasi seperti itu, kinerja yang lebih baik dapat dicapai pada prosesor banyak-inti dengan cache cepat dan subsistem memori bertumpuk seperti prosesor Intel Xeon atau Xeon Phi. Oleh karena itu, penting untuk mempertimbangkan berapa banyak data yang akan tersedia untuk pelatihan ketika memilih perangkat keras Anda. Faktor penting lainnya yang perlu dipertimbangkan adalah bahwa seringkali ternyata kinerja pelatihan dibatasi oleh kinerja cache dan memori daripada kemampuan floating-point, karena setelah sistem memori dan cache jenuh, kemampuan floating-point tambahan terbuang sia-sia. Mengingat ketergantungan perhitungan gradien pada memori, pengguna perlu berhati-hati tentang apa kinerja floating-point yang sebenarnya (bukan puncak) ketika perhitungan didominasi oleh kinerja bandwidth memori dan cache. Dengan pertimbangan ini, CPU dengan memori yang lebih besar mungkin menjadi pilihan yang lebih disukai daripada GPU untuk kasus penggunaan tertentu.

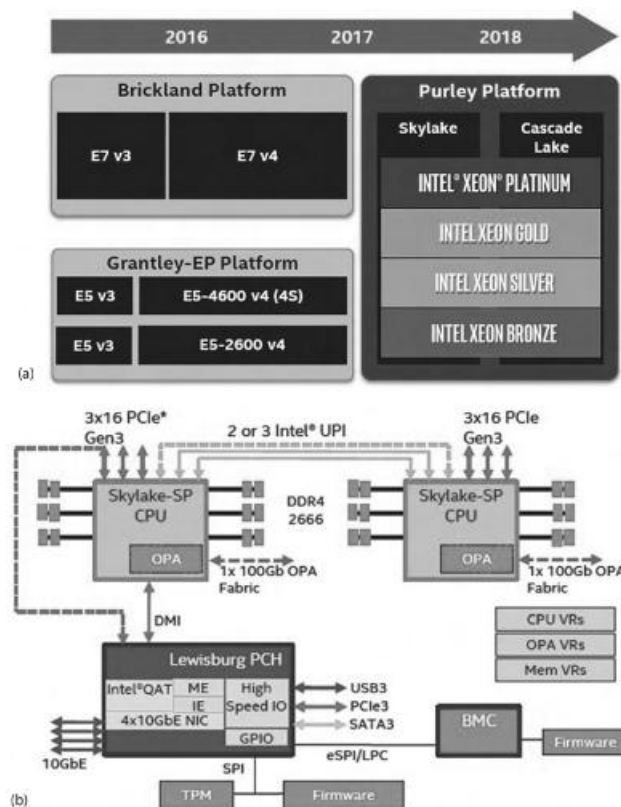
CPU Intel generasi baru telah berevolusi untuk mengakomodasi beban kerja AI. Bahkan tanpa perangkat keras khusus, telah terbukti bahwa penyertaan instruksi vektor AVX-512 yang lebih luas dan saluran memori tambahan dapat meningkatkan kinerja pelatihan dan inferensi per inti lebih dari dua kali lipat pada prosesor Intel Skylake tanpa memerlukan peningkatan



ukuran set data untuk memanfaatkan paralelisme. Menggunakan lebih banyak inti akan memberikan peningkatan kinerja tambahan (terdapat hingga 28 inti pada setiap CPU). Baik SKU produk Intel Xeon Phi maupun Intel Xeon (Skylake) menawarkan antarmuka Intel Omni-Path dalam paket, yang seharusnya mengurangi biaya sistem dan latensi jaringan sekaligus meningkatkan bandwidth jaringan. Ini merupakan kabar baik bagi mereka yang perlu melatih (atau melakukan inferensi volume) di seluruh jaringan atau di dalam cloud. Gambar 8.12a menunjukkan peta jalan server Intel saat ini, dan Gambar 8.12b menunjukkan platform Purley terbaru.

Unit Pemrosesan Grafis

Unit pemrosesan grafis telah menjadi pilihan paling populer untuk pelatihan pembelajaran mendalam (*deep learning*) dengan tujuan mempercepat pelatihan untuk mencapai "waktu-ke-model" tersingkat. Semua perangkat keras di pasaran menggunakan paralelisme untuk mempercepat pelatihan. Akselerator mencapai kinerja floating-point yang tinggi ketika memiliki sejumlah besar thread konkuren untuk dieksekusi. GPU telah membawa tren multicore ke tingkat ekstrem dengan menambahkan ribuan inti kontrol yang lebih kecil ke dalam satu perangkat. Selain itu, GPU memiliki sumber daya unit logika aritmatika (ALU) yang lebih besar dibandingkan dengan CPU. GPU Nvidia pada umumnya memiliki ribuan inti, yang memungkinkan eksekusi cepat dari operasi yang sama di beberapa inti. Inti dari pembelajaran mendalam adalah sejumlah besar perkalian matriks (perkalian matriks umum) dan konvolusi yang dapat dilakukan secara paralel, yang menjadikan GPU pilihan yang sangat baik.



Gambar 8.12 (a) Peta jalan server Intel. (b) Arsitektur platform server Intel Purley.



Secara umum, untuk beban kerja paralel inheren dengan operasi matriks berat yang memerlukan bandwidth tinggi ke memori, GPU dapat menawarkan kinerja yang lebih baik daripada CPU jika beban kerja dapat masuk dalam memori GPU, biasanya 16GB. Namun, jika beban kerja melibatkan ukuran set data yang lebih besar dari 16GB, perangkat GPU perlu mengambil data dari memori melalui CPU, yang secara substansial dapat memperlambat kinerja pekerjaan. Misalnya, produsen peralatan medis menangani gambar 3D beresolusi tinggi dalam gigapiksel, dan perusahaan farmasi terkemuka memiliki gambar 1024×1024 dengan lebih dari 10 saluran untuk implementasi CNN terkait penemuan obat. Waktu pelatihan pembelajaran mendalam berbanding lurus dengan ukuran batch gambar yang dapat ditampung setiap node pemrosesan dalam memori lokalnya. Untuk gambar beresolusi tinggi, seiring bertambahnya ukuran batch, GPU saat ini tidak memiliki memori yang cukup besar untuk menyimpan data, sehingga menghasilkan waktu pelatihan yang lebih lama.

Salah satu tantangan dalam penggunaan GPU adalah jarak yang lebih jauh dari memori utama server, sehingga menyebabkan penundaan pergerakan data (sering kali dilakukan melalui protokol PCI Express atau PCIe). Untuk mempercepatnya, perusahaan seperti Nvidia telah mengembangkan interkoneksi yang lebih cepat yang disebut NVLink. Tantangan lain dengan GPU meliputi kurangnya skalabilitas antar kartu dan antar server, harga yang biasanya lebih tinggi, dan kebutuhan konsumsi daya yang lebih tinggi. Misalnya, Nvidia menilai kartu grafis Titan X mereka berdaya 250W, dan merekomendasikan catu daya sistem sebesar 600W

Field-Programmable Gate Array (FPGA)

Field-programmable gate array (FPGA) adalah sirkuit terpadu yang blok logikanya dapat diprogram dan dikonfigurasi ulang menggunakan bahasa deskripsi perangkat keras. Keuntungan paling signifikan yang dimiliki FPGA dibandingkan CPU dan GPU adalah efisiensi komputasi, yang berarti operasi yang lebih cepat dan mengonsumsi lebih sedikit energi dalam banyak aplikasi. Diamati bahwa FPGA menawarkan kinerja/watt yang jauh lebih tinggi daripada GPU, karena meskipun tidak dapat bersaing dalam kinerja murni, mereka memiliki penggunaan daya yang jauh lebih sedikit (seringkali puluhan watt). Selain itu, FPGA dapat menunjukkan beberapa latensi inferensi terendah, ditambah lagi mereka dapat ditingkatkan di lapangan.

Fleksibilitas struktur FPGA memungkinkan koneksi langsung berbagai masukan, seperti kamera, tanpa memerlukan perantara. Ini menghilangkan pergerakan data yang tidak perlu, memungkinkan latensi sistem yang lebih rendah sehingga sistem dapat memproses data dan membuat keputusan secara real-time. Sementara itu, FPGA juga memiliki efisiensi daya yang luar biasa, untuk aplikasi seperti pusat data skala besar di mana total biaya kepemilikan merupakan faktor kunci. Memindahkan algoritma pembelajaran mendalam, seperti AlexNet (model CNN), dapat mengurangi daya yang dibutuhkan hingga 80%. Untuk perangkat tepi yang sensitif terhadap daya, FPGA dapat diprogram sebelumnya untuk mendukung berbagai pembelajaran mesin statistik dan algoritma pembelajaran mendalam.

Selain itu, struktur fleksibel dan kemampuan konfigurasi ulang FPGA memungkinkan pengguna untuk membuktikan perangkat keras mereka di masa mendatang. Dalam bidang yang serba cepat dan berkembang seperti AI, hal ini memaksimalkan laba atas investasi



pelanggan, memungkinkan kinerja luar biasa untuk kebutuhan saat ini dan masa depan. Misalnya, seseorang dapat menjalankan topologi jaringan saraf saat ini dan apa pun di masa mendatang pada perangkat keras yang sama. FPGA juga fleksibel untuk semua jenis presisi, memungkinkan seseorang untuk mengoptimalkan akurasi, kinerja, dan kecepatan sesuai kebutuhannya. Terakhir, FPGA dapat digunakan untuk pemrosesan inline dan off-road.

Seiring dengan terus berkembangnya implementasi jaringan saraf, mendefinisikan ulang arsitektur silikon untuk setiap evolusi dapat menjadi sangat mahal. FPGA, karena sifat kemampuan pemrograman ulangnya, merupakan platform alternatif yang lebih baik, terutama untuk pemrosesan inline deterministik, real-time, dari data streaming tanpa buffering. Namun, satu kelemahan utama FPGA adalah kesulitan dalam pemrograman (memerlukan keahlian pengembangan tingkat transfer register), yang telah mencegah FPGA menjadi solusi komputasi umum (Gambar 8.13).

Sirkuit Terpadu Khusus Aplikasi

Sirkuit terpadu khusus aplikasi merupakan pilihan perangkat keras yang paling tidak fleksibel, tetapi berkinerja paling tinggi. Sirkuit ini juga paling efisien dalam hal kinerja/dolar dan kinerja/watt, tetapi membutuhkan investasi besar dan biaya rekayasa non-berulang yang membuatnya hemat biaya hanya dalam jumlah besar. ASIC dapat dirancang dan dikodekan secara keras untuk pelatihan atau inferensi. Meskipun bertujuan tunggal, ASIC memiliki efisiensi komputasi tertinggi. Jaringan saraf tiruan dapat berjalan lebih cepat dan mengonsumsi daya lebih sedikit ketika dipasangkan dengan chip yang dirancang khusus untuk menangani serangkaian besar perhitungan matematika yang dibutuhkan sistem AI.

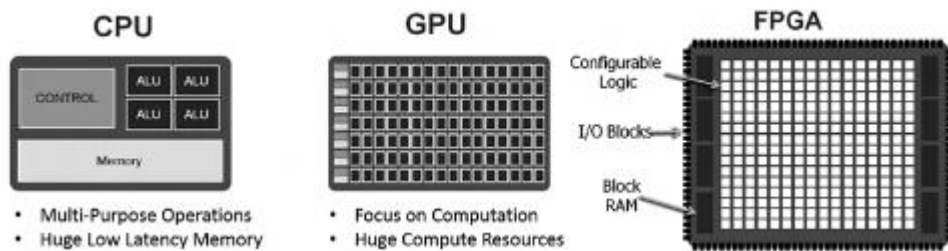
Pengumuman terbaru tentang perangkat keras AI khusus (misalnya, unit pemrosesan tensor Google [TPU] dan prosesor jaringan saraf tiruan Intel [NNP]) telah meningkatkan kesadaran bahwa solusi khusus dapat melampaui kinerja CPU dan GPU untuk beberapa aplikasi jaringan saraf tiruan. Namun, pengguna mungkin ingin mempertimbangkan jenis model jaringan yang akan dilatih ketika mereka membuat pilihan perangkat keras. Misalnya, manfaat perangkat keras khusus yang menjalankan operasi tensor dengan presisi rendah hanya berlaku untuk beberapa jenis arsitektur neural seperti CNN. Penting untuk memahami apakah pekerjaan Anda memerlukan penggunaan jenis arsitektur neural tersebut.

TPU1 Google hanya menargetkan inferensi, sementara TPU2 atau "TPU awan" mereka mendukung pelatihan dan inferensi. Setiap chip TPU2 mendukung 45 *Tera Floating Point Operations per Second* (TFLOP) dan hadir dalam papan berisi empat chip, yang memiliki total 180 TFLOP (untuk floating point setengah presisi 16-bit). TPU tersebut tidak akan dijual secara komersial, tetapi dapat diakses melalui platform komputasi awan Google.

Chip NNP Intel, yang dirancang khusus untuk AI, adalah yang pertama dari jenisnya (yaitu, dirancang khusus untuk pembelajaran mendalam) yang tersedia secara komersial. Chip ini menghilangkan hierarki cache standar dan menggunakan perangkat lunak untuk mengelola memori on-chip guna mencapai waktu pelatihan yang lebih cepat untuk model pembelajaran mendalam. Perangkat keras ini berspesialisasi dalam arsitektur titik fleksibel, yang akan diekspos ke berbagai kerangka kerja perangkat lunak pembelajaran mendalam seperti Neon, TensorFlow, dan MXNet melalui lapisan penyusun grafik yang disebut Nervana Graph.



Beberapa generasi produk Intel Nervana NNP yang sedang dalam pengembangan baru-baru ini diumumkan, yang akan memberikan kinerja yang lebih tinggi dan memungkinkan tingkat skalabilitas baru untuk model AI. Pada saat yang sama, vendor juga sedang membangun chip untuk membantu mengeksekusi jaringan saraf pada ponsel pintar dan perangkat lainnya. Misalnya, IBM sedang membangun chip semacam itu. Dan Intel baru-baru ini mengakuisisi Movidius, sebuah perusahaan yang telah memasukkan chip ke dalam perangkat.



Gambar 8.13 Karakteristik CPU, GPU, dan FPGA.

8.6 KARAKTERISTIK BEBAN KERJA

Secara umum, beban kerja AI dapat dibagi menjadi dua kelas: pembelajaran mesin klasik dan pembelajaran mendalam. Pembelajaran mesin klasik merupakan mayoritas, yang mencakup algoritma paling umum yang digunakan oleh ilmuwan data, seperti model linear, pengelompokan k-means, dan pohon keputusan. Metode-metode tersebut banyak digunakan dalam deteksi penipuan, personalisasi pemasaran, prediksi indikator kinerja utama, keamanan siber, dll. Aplikasi pembelajaran mendalam yang umum meliputi pengenalan gambar, pengenalan suara, permainan otomatis, dan kendaraan otonom.

Contoh Performa dan Latensi

Pada tahap inferensi AI, tujuan performa yang terpenting adalah latensi. Untuk meminimalkan waktu respons ujung ke ujung jaringan, inferensi biasanya mengelompokkan input dalam jumlah yang jauh lebih kecil daripada pelatihan untuk mendukung pemrosesan mendekati waktu nyata. Misalnya, persyaratan latensi untuk seluruh alur Google Autosuggest kurang dari 200 milidetik. Alur kerja mencakup frontend, penyeimbangan beban, pemahaman kueri, dan saran pelengkapan otomatis yang didukung oleh pembelajaran mendalam, serta penelusuran tumpukan pencarian lengkap untuk menampilkan hasil yang akan diperoleh jika pengguna benar-benar mencari salah satu kueri yang disarankan otomatis. Aplikasi penting seperti kendaraan otonom menerapkan persyaratan latensi yang lebih ketat pada alur pemrosesan, di mana pembelajaran mendalam digunakan sebagai salah satu komponennya. Waktu pemrosesan yang cepat untuk aplikasi ini memengaruhi keselamatan dan nyawa manusia. Sebagai contoh lain, antarmuka ucapan memerlukan model pengenalan dan sintesis ucapan untuk mengembalikan hasil tanpa penundaan yang terasa oleh pengguna.

Untuk membantu pengembang memanfaatkan perangkat keras mereka dengan lebih baik, vendor telah menyediakan serangkaian optimasi inferensi untuk GPU atau CPU (misalnya, pustaka cuDNN Nvidia dan pustaka MKL-DNN Intel). Sebagai contoh, versi terbaru cuDNN mencakup penyempurnaan pada inferensi batch kecil dengan membagi dimensi tambahan.



Hal ini mengurangi jumlah komputasi per blok utas dan memungkinkan peluncuran blok yang jauh lebih banyak, sehingga meningkatkan okupansi dan kinerja GPU.

Peningkatan signifikan lainnya dicapai melalui operasi floating point presisi rendah. Dalam beberapa tahun terakhir, para peneliti menemukan bahwa penggunaan representasi floating point presisi rendah atau bahkan integer point (FP16 atau INT8) untuk penyimpanan aktivasi lapisan dan representasi yang lebih tinggi (FP32) untuk komputasi tidak mengorbankan akurasi klasifikasi. Penurunan presisi ini secara signifikan meningkatkan kinerja dalam situasi keterbatasan bandwidth dan mengurangi jejak memori keseluruhan yang diperlukan untuk menjalankan jaringan saraf tiruan dalam.

Secara algoritmik, salah satu pendekatan yang menjanjikan untuk mengurangi latensi inferensi dan kebutuhan memori (dan karenanya konsumsi daya) adalah kompresi model, yang telah menjadi topik penelitian yang hangat akhir-akhir ini. Tren penting yang perlu disebutkan adalah seiring dengan terus berkembangnya AI di cloud, industri akan mengalami pergeseran dalam tahap inferensi dari penggunaan server khusus kelas atas ke prosesor umum di tepi cloud jaringan. Tren ini didorong oleh kebutuhan untuk mengurangi latensi pada jaringan yang mengalami hambatan bandwidth dan mengatasi isu privasi dan kepatuhan yang memanas (yaitu, memproses data secara lokal alih-alih mengirimkannya ke cloud untuk mengontrol akses ke data).

Waktu Pelatihan

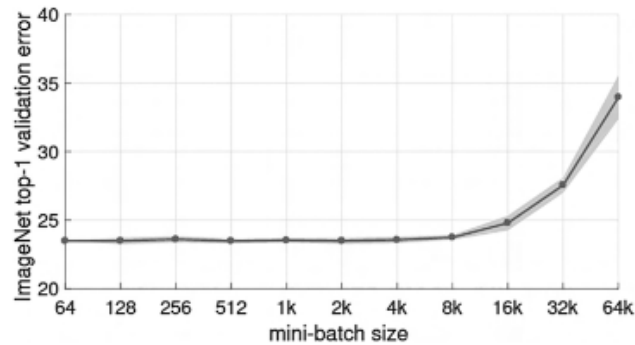
Untuk tahap pelatihan algoritma pembelajaran mesin dan pembelajaran mendalam, "waktu pelatihan" merupakan metrik penting dalam banyak skenario penggunaan, selain akurasi. Metrik ini ditentukan oleh sejumlah faktor, misalnya, model jaringan, pengaturan hiperparameter, kerangka kerja, optimasi pustaka, dan platform perangkat keras. Terdapat sejumlah trik algoritmik yang dapat mempercepat proses pelatihan. Dari perspektif perangkat keras, kapabilitas floating point puncak (yaitu, TFLOP, metrik untuk perangkat keras pembelajaran mesin, pada dasarnya jumlah operasi floating point yang dapat didukung perangkat keras per detik) sering dianggap sebagai spesifikasi terpenting. Namun, kinerja memori dan cache, serta bandwidth I/O, dapat sama pentingnya dalam sebagian besar keadaan.

Skalabilitas di beberapa server dan beberapa akselerator merupakan faktor kunci lain dalam mencapai terobosan dalam waktu pelatihan. Telah dipublikasikan bahwa tim Facebook mengurangi waktu pelatihan model pembelajaran mendalam ResNet-50 pada ImageNet dari 29 jam menjadi satu jam yang mereka lakukan dengan mendistribusikan pelatihan dalam mini-batch yang lebih besar di lebih banyak GPU. Sebelumnya, batch berisi 256 gambar tersebar di delapan GPU Tesla P100, tetapi penelitian saat ini menunjukkan tingkat akurasi yang sama ketika pelatihan dengan ukuran mini-batch besar hingga 8.192 gambar yang didistribusikan di 256 GPU (Gambar 8.14).

Pada tanggal 7 November 2017, para peneliti dari UC Berkeley, U-Texas, dan UC Davis mempublikasikan hasil pelatihan mereka yang juga melatih ResNet-50 dalam waktu rekor 31 menit dan AlexNet dalam waktu rekor 11 menit pada CPU dengan akurasi tercanggih. Hasil ini diperoleh pada prosesor Intel® Xeon® Scalable (nama kode Skylake-SP). Pendorong utama



kecepatan performa ini adalah: (1) Peningkatan kapasitas komputasi dan memori prosesor Intel Xeon Scalable dibandingkan generasi sebelumnya; (2) Optimalisasi perangkat lunak dalam Intel® MKL-DNN dan dalam kerangka kerja pembelajaran mendalam yang populer; dan (3) Kemajuan terkini (misalnya, Intel® Machine Learning Scaling Library) dalam algoritma pelatihan terdistribusi untuk beban kerja pembelajaran mendalam yang diawasi.



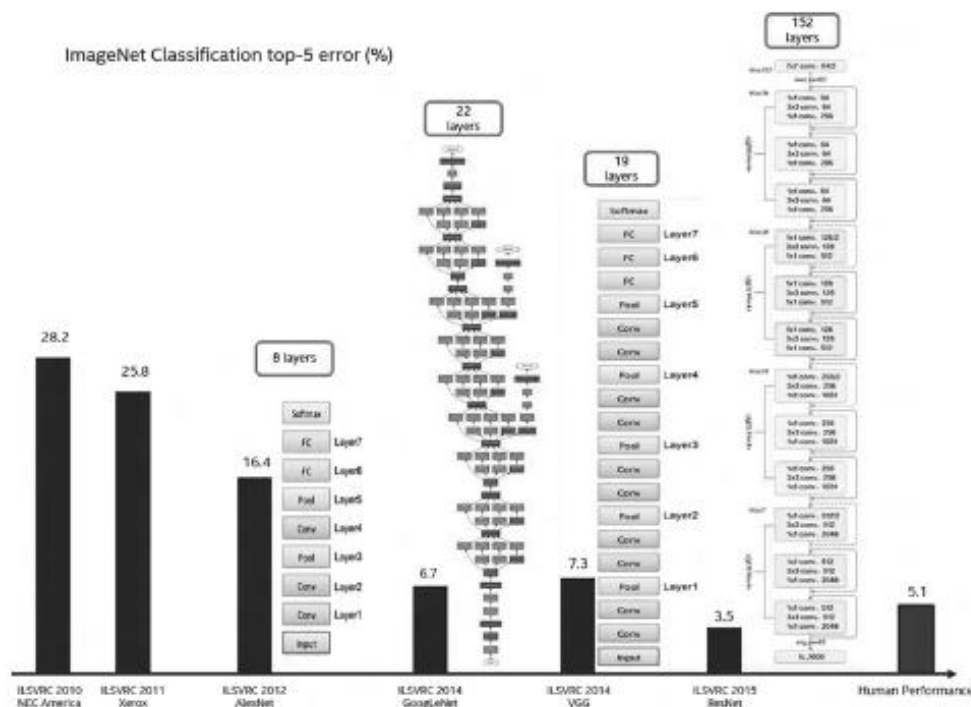
Gambar 8.14 Kesalahan validasi ImageNet top-1 versus ukuran mini-batch.

Akurasi Model Mutakhir

Dengan peningkatan yang luar biasa dalam pengembangan algoritma pembelajaran mendalam selama lima hingga delapan tahun terakhir, akurasi model mutakhir telah meningkat secara dramatis, bahkan melampaui akurasi manusia dalam beberapa aplikasi. Sebagai contoh, gambar berikut menunjukkan perkembangan model jaringan saraf tiruan konvolusional untuk tugas pengenalan gambar dan akurasinya yang ditunjukkan oleh kesalahan 1 teratas dan 5 teratas.

Akurasi 1 teratas adalah akurasi konvensional di mana jawaban model (yang memiliki probabilitas tertinggi) harus sama persis dengan jawaban yang diharapkan dan 5 teratas mewakili lima jawaban dengan probabilitas tertinggi model Anda harus sama persis dengan jawaban yang diharapkan. Seperti yang ditunjukkan pada Gambar 8.15, akurasi yang dicapai dengan model ResNet-152 adalah kesalahan 5 teratas sebesar 3,5%, yang secara signifikan lebih baik daripada kesalahan manusia sebesar 5,1%. Gambar 8.16 mengilustrasikan model CNN yang populer akurasinya (sumbu y), jumlah operasi (sumbu x), dan ukuran modelnya (ditunjukkan oleh ukuran gelembung).





Gambar 8.15 Model jaringan saraf konvolusional untuk pengenalan gambar peningkatan akurasi.

Hambatan Performa dan Cara Mengatasinya

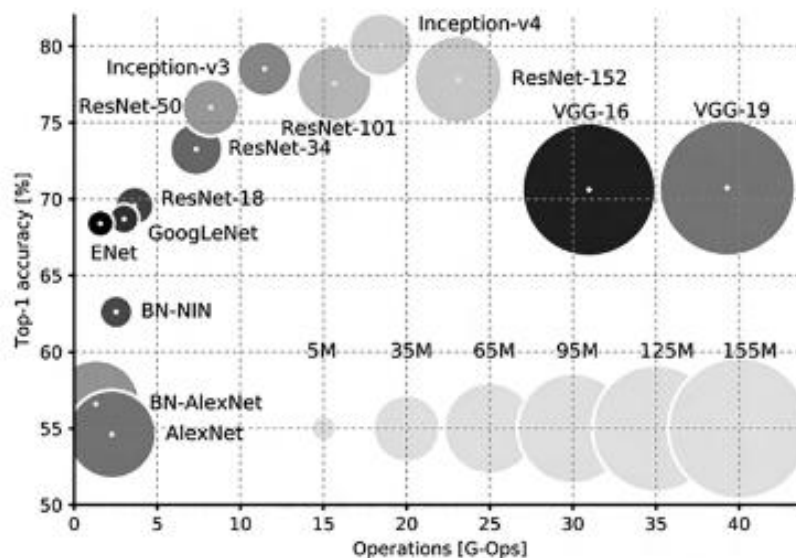
Hambatan performa pembelajaran mendalam terletak pada area berikut:

- **Komputasi dan memori:** Hal ini tetap menjadi tantangan meskipun kita sedang mengembangkan chip yang disesuaikan, karena permintaan akan chip yang lebih banyak terus meningkat. Memori merupakan salah satu tantangan terbesar dalam jaringan saraf tiruan dalam saat ini. Para peneliti kesulitan dengan keterbatasan bandwidth memori dari perangkat *Dynamic Random-Access Memory* (DRAM) yang harus digunakan oleh sistem saat ini untuk menyimpan bobot dan aktivasi dalam jumlah besar di DNN. Arsitektur komputer telah dikembangkan dengan chip prosesor yang مخصوص untuk pemrosesan serial dan DRAM yang dioptimalkan untuk memori berdensitas tinggi. Antarmuka antara kedua perangkat ini merupakan hambatan utama yang menimbulkan keterbatasan latensi dan bandwidth serta menambah beban konsumsi daya yang cukup besar. Para peneliti telah menerapkan berbagai teknik penghematan memori untuk mengurangi kebutuhan memori. Misalnya, pendekatan penggunaan kembali memori telah dikembangkan oleh para peneliti di Google DeepMind dengan RNN. RNN adalah jenis DNN khusus yang memungkinkan siklus dalam strukturnya untuk mengodekan perilaku pada urutan masukan. Untuk RNN, perhitungan ulang telah terbukti mengurangi memori hingga 20 kali lipat untuk urutan dengan panjang 1000 dengan overhead kinerja hanya 30% .
- **Set data:** Dengan algoritma saat ini yang sebagian besar diawasi, terdapat kebutuhan akan set data yang lebih besar untuk terus mendorong perkembangan penelitian baru. Perusahaan seperti Google telah berupaya untuk menghasilkan dan menerbitkan lebih



banyak set data (misalnya, mereka baru-baru ini merilis satu set data video dan dua set data untuk robotika). Di sisi lain, pembelajaran tanpa pengawasan, yang tidak memerlukan data pelatihan berlabel, telah menjadi topik penelitian yang aktif.

- **Algoritma:** Kemajuan algoritma yang ada, seperti gagasan ReLU, Dropout, Sequence to Sequence, dan GAN telah membawa perubahan besar. Namun, peningkatan komputasi 1000 kali lipat dengan peningkatan perangkat keras tradisional murni tidak mungkin tercapai. Diperlukan perancangan bersama algoritma dan komputasi. Misalnya, akan sangat efektif jika membuat model dengan parameter 1000 kali lipat lebih banyak, tetapi hanya menggunakan komputasi 10 kali lipat lebih banyak.
- **Keahlian:** Mengeksperimen konfigurasi yang digunakan untuk meningkatkan kinerja seperti yang ditunjukkan di bagian sebelumnya dianggap sebagai trik. Konfigurasi ini mungkin tidak cukup umum untuk diterapkan pada set data dan topologi jaringan lain.



Gambar 8.16 Akurasi Top-1 versus jumlah operasi yang diperlukan untuk satu kali penerusan model pengenalan gambar CNN populer.

8.7 RINGKASAN

Kita hidup dalam ekonomi data. Ada sejumlah besar data di mana-mana, perangkat pribadi, seperti ponsel pintar dan PC, server perusahaan dan pusat data, infrastruktur jaringan, serta cloud privat dan publik. Munculnya komunikasi mesin-ke-mesin dan perangkat IoT semakin menambah banjir data. Dengan demikian, memperoleh kecerdasan dari semua data terstruktur dan tidak terstruktur dapat memungkinkan aplikasi dan model bisnis baru dalam ekonomi data ini. Dalam bab ini, kami telah menjelaskan pembelajaran mesin dan pembelajaran mendalam sebagai elemen dasar yang memungkinkan AI. Kami telah menjelaskan bagaimana model dapat dibuat untuk meniru otak manusia, dan kami dapat menggunakan berbagai jenis data historis untuk "melatih" model-model ini.

Setelah dilatih, model-model ini mirip dengan otak manusia dan dapat menarik inferensi untuk membuat keputusan yang cerdas. Kami juga membahas arsitektur perangkat



keras yang memungkinkan solusi pelatihan dan inferensi tersebut platform x86 untuk keperluan umum, GPU, FPGA, dan ASIC yang dibuat khusus, semuanya memiliki tempat dalam kontinum solusi. Kami juga membahas kerangka kerja perangkat lunak sumber terbuka dan populer, seperti Tensorflow, Caffe, H2O, Scikit Learn, Neon, Hadoop, Spark, dan MXNet, serta platform Acumos, yang memungkinkan pengembang generasi berikutnya membangun aplikasi AI. Kami juga secara singkat membahas pertimbangan kinerja dan latensi utama untuk berbagai beban kerja. Seiring dengan peluncuran teknologi nirkabel 5G dan penerapan edge cloud, AI akan terus berkembang selama dekade berikutnya. Pasti akan ada inovasi baru baik inovasi teknologi maupun model bisnis.

Bab selanjutnya menjelaskan bagaimana AI dapat diterapkan untuk menjadikan infrastruktur jaringan lebih cerdas. Beberapa tahun terakhir telah menyaksikan virtualisasi infrastruktur jaringan dengan teknologi *Network Function Virtualization* (NFV) dan *Software Defined Networking* (SDN) pada platform server standar bervolume tinggi. Seiring dengan semakin homogennya infrastruktur jaringan dan terdefinisi perangkat lunak, terdapat peluang untuk menerapkan pembelajaran mesin, membuat keputusan cerdas dan real-time, serta mendorong otomatisasi jaringan loop tertutup. Hal ini akan menghasilkan efisiensi operasional yang signifikan pada penerapan jaringan di masa mendatang. Penerapan teknik AI untuk mendorong efisiensi ini menjadi fokus utama bab selanjutnya.



BAB 9

KECERDASAN BUATAN UNTUK OPERASI JARINGAN

9.1 PENDAHULUAN

Dalam bab ini, kami mengeksplorasi peran kecerdasan buatan (AI) untuk operasi jaringan. Kami mendefinisikan dan membahas jaringan zero-touch dan mengidentifikasi bagaimana jaringan yang ditentukan perangkat lunak (SDN) dan AI bersama-sama menyediakan teknologi inti dan kecerdasan untuk mewujudkan jaringan tersebut.

Operasi jaringan adalah peran yang biasanya tersembunyi dalam memastikan jaringan dan layanan beroperasi untuk memberikan pengalaman layanan yang lancar kepada pelanggan. Operator jaringan seringkali merupakan ahli yang secara diam-diam dan lancar mengambil desain dan teknologi jaringan tingkat tinggi dan mewujudkannya mengidentifikasi di mana kapasitas harus diterapkan dan diperluas, menerapkan teknologi jaringan, dan memastikan semuanya beroperasi dengan lancar dan lancar bahkan dalam menghadapi pertumbuhan lalu lintas yang luar biasa, perubahan teknologi yang berkelanjutan, beragam teknologi, kegagalan perangkat lunak, perangkat keras, dan perangkat lunak yang kompleks, bahkan bencana alam dan bencana yang disebabkan oleh manusia.

Saat ini, terdapat otomatisasi operasi jaringan yang sangat besar dalam jaringan skala besar, meskipun masih terdapat cukup banyak intervensi manusia. Jaringan itu sendiri memiliki sejumlah besar otomatisasi dan kecerdasan, seperti protokol perutean dan kecerdasan dalam fungsi jaringan untuk secara otomatis memulihkan layanan dengan merutekan ulang lalu lintas jika terjadi kegagalan. Namun, saat ini manusia terlibat dalam tugas-tugas kompleks yang terkait dengan merancang dan merencanakan jaringan, memperbaiki jaringan (menambahkan fungsi jaringan baru, memperbaiki perangkat lunak, dan memperbaiki cara jaringan dikonfigurasi), dan memperbaiki fungsi jaringan yang terganggu. Karena jaringan terus berkembang, akan semakin mustahil untuk meningkatkan keterlibatan manusia dalam fungsi-fungsi ini. Sebaliknya, otomatisasi jaringan tingkat lanjut dan AI akan semakin meresap ke dalam tugas-tugas operasi jaringan, membawa kita semakin dekat ke visi jaringan otonom atau tanpa sentuhan. Dalam jaringan tanpa sentuhan, semua tugas sehari-hari diotomatisasi, dengan AI dan pembelajaran mesin yang memungkinkan pengambilan keputusan. Operator jaringan memberikan panduan tingkat tinggi untuk otomatisasi menggambarkan maksud dan kebijakan operasi mereka. Otomatisasi dan AI melanjutkannya dari sana menerjemahkan maksud dan kebijakan ke dalam keputusan dan tindakan yang memperluas jaringan, mengaktifkan layanan dan fitur baru, memulihkan jaringan dan layanan saat menghadapi kegagalan dan gangguan, serta melindungi jaringan dari potensi serangan dan pengaruh eksternal.

Bab ini menjelaskan bagaimana AI dan SDN akan digunakan bersama untuk mendorong tercapainya tujuan jaringan zero-touch. Bagian 9.2 menyajikan ikhtisar latar belakang tugas-tugas yang terlibat dalam pengoperasian jaringan. Di Bagian 9.3, kami menyajikan visi "jaringan zero-touch" bagaimana AI dan SDN merevolusi cara pengoperasian jaringan. Bagian 9.4



membahas teknologi yang dibutuhkan dan perjalanan industri untuk mewujudkan visi ini. Terakhir, Bagian 9.5 menjelaskan dampaknya terhadap tim operasi, dan bagaimana peran mereka berkembang seiring terwujudnya jaringan zero-touch.

9.2 PENGERTIAN OPERASI JARINGAN

Kita mulai dengan mendefinisikan tugas-tugas operasi jaringan. Operasi jaringan adalah proses di mana jaringan, aplikasi, dan layanan disebarkan dan dipelihara. Tim operasi jaringan bertanggung jawab untuk memastikan bahwa jaringan memiliki kapasitas terpasang dan operasional yang memadai, untuk menjalankan pembaruan dalam jaringan dan untuk mengelola fungsi dan layanan jaringan, sehingga tingkat kinerja layanan yang diinginkan tercapai.

Kami mengkategorikan fungsi-fungsi ini sebagai penyediaan jaringan (menyediakan jumlah elemen jaringan yang dikonfigurasi dengan tepat untuk memberikan layanan yang ekonomis dan berkualitas), pemeliharaan jaringan (melaksanakan aktivitas yang direncanakan dan mencegah serta memperbaiki kesalahan dan gangguan), dan administrasi jaringan (memastikan bahwa jaringan memberikan tujuan layanan yang telah ditetapkan, bahkan ketika menghadapi perubahan pengaruh eksternal).

Penyediaan jaringan mencakup perencanaan jaringan, rekayasa, dan implementasi fungsi dan layanan jaringan. Tujuan utamanya adalah untuk menentukan bagaimana jaringan akan berkembang, menentukan di mana kapasitas harus dikerahkan, dan kemudian mewujudkan rencana ini dalam jaringan. Pemeliharaan jaringan mengacu pada tugas-tugas yang terlibat dalam melayani fungsi jaringan dan interkoneksinya, baik secara reaktif (ketika masalah muncul) maupun secara preventif (untuk mencegah masalah). Terakhir, administrasi jaringan berkaitan dengan pengelolaan bagaimana topologi jaringan dan perutean lalu lintas diadaptasi untuk memberikan tujuan layanan yang telah ditetapkan, bahkan ketika menghadapi perubahan pengaruh eksternal (misalnya, perubahan lalu lintas masuk), aktivitas jaringan (gangguan dan aktivitas pemeliharaan), dan bencana.

Penyediaan Jaringan

Bagi penyedia layanan yang menyediakan beragam layanan internet kabel, jaringan privat virtual (VPN), suara dan data seluler, suara melalui protokol internet (IP), dan layanan TV terdapat ratusan jenis fungsi jaringan (yang secara tradisional dikenal sebagai elemen jaringan). Fungsi jaringan ini biasanya berupa peralatan dan/atau perangkat lunak yang disediakan oleh beberapa vendor, bahkan untuk peran fungsi jaringan yang sama. Mungkin terdapat puluhan jenis fungsi jaringan yang mendukung layanan tertentu misalnya, jalur data pengguna seluler dapat melintasi hingga 50 jenis fungsi jaringan yang unik antara lokasi seluler dan internet. Seiring kemajuan industri dalam memvirtualisasikan semakin banyak fungsi jaringan, jaringan juga akan terdiri dari fungsi jaringan fisik dan virtual (VNF). Seiring waktu, VNF diperkirakan akan semakin umum. Namun, fungsi jaringan fisik akan selalu ada, karena mustahil untuk memvirtualisasikan semua fungsi (misalnya, serat optik dan radio). Semua jenis fungsi jaringan yang berbeda ini di sepanjang jalur layanan harus disediakan, dengan kapasitas yang memadai pada masing-masing fungsi agar layanan dapat didukung.



Kami berfokus di sini pada penyediaan jaringan dan berasumsi bahwa penyediaan konektivitas dan layanan pelanggan berada di luar cakupan bab ini. Penyediaan jaringan adalah proses yang digunakan operator jaringan untuk memutuskan di mana dan kapan akan menerapkan perangkat lunak dan perangkat keras baru ke dalam jaringan (perencanaan kapasitas jaringan), lalu mengeksekusi penerapan ini meningkatkan kapasitas dan mengonfigurasinya agar menjadi bagian dari jaringan. Proses penyediaan jaringan mengasumsikan pilihan teknologi jaringan dan desain jaringan dasar sebagai input; hal ini biasanya diputuskan oleh teknisi jaringan di luar tim operasi jaringan. Perencanaan kapasitas jaringan harus menentukan berapa banyak kapasitas setiap fungsi jaringan yang akan diterapkan di setiap lokasi dan bagaimana menghubungkan fungsi-fungsi jaringan di dalam dan antar lokasi. Jika interval waktu yang lama diperlukan untuk menerapkan kapasitas, perencana harus memperkirakan kebutuhan kapasitas di masa mendatang sebelum melakukan perencanaan kapasitas. Setelah kapasitas direncanakan, kapasitas tersebut harus diterapkan. Fungsi jaringan fisik tradisional memerlukan penerapan peralatan secara fisik di lokasi, yang mengharuskan pengiriman teknisi.

Penerapan semacam itu biasanya memakan waktu hingga berbulan-bulan, sehingga perencanaan kapasitas harus memperkirakan permintaan dan merencanakannya berbulan-bulan atau lebih sebelumnya. Namun, VNF dapat dijalankan dari jarak jauh sesuai permintaan dengan syarat tersedia kapasitas cloud yang memadai (server, penyimpanan, dan interkoneksi jaringan). Dengan demikian, perencanaan kapasitas juga dapat dilakukan "sesuai permintaan", dengan kapasitas yang dipindahkan seiring perubahan permintaan dari waktu ke waktu dan bahkan sepanjang hari. Baik fisik maupun virtual, fungsi jaringan harus dikonfigurasi secara logis dan kemudian divalidasi sebelum dapat digunakan untuk membawa lalu lintas dan memberikan layanan. Konfigurasi mencakup hal-hal seperti menentukan identitas fungsi jaringan (misalnya, alamat IP), menentukan protokol perutean yang akan digunakan untuk berkomunikasi dengan fungsi jaringan lain (di mana perutean dilakukan di dalam fungsi jaringan), dan mengonfigurasi kebijakan seperti lalu lintas apa yang akan dibawa atau diblokir. Setelah konfigurasi selesai, "uji dan aktifkan" memvalidasi bahwa fungsi jaringan siap untuk mendukung lalu lintas dan kemudian memperbarui konfigurasi jaringan untuk mulai merutekan lalu lintas melalui fungsi jaringan.

Pemeliharaan Jaringan

Pemeliharaan jaringan sangat mirip dengan pemeliharaan mobil pemeliharaan preventif perlu dilakukan secara berkelanjutan dan proaktif untuk mengurangi risiko masalah, dan perlu dilakukan ketika masalah muncul. Pemilik mobil yang bertanggung jawab membawa mobil mereka untuk diservis secara berkala, mengganti oli, memperbarui perangkat lunak mobil, dan memeriksa kondisi mobil dengan mengganti komponen sebelum kerusakan fatal dapat memengaruhi kemampuan mobil untuk beroperasi. Mereka juga menanggapi pemberitahuan penarikan kembali merawat kendaraan sebagai respons terhadap masalah yang telah diidentifikasi oleh produsen. Namun, terlepas dari upaya terbaik bahkan dari pemilik mobil yang paling bertanggung jawab sekalipun, kegagalan komponen penting atau pengaruh eksternal dapat menyebabkan mobil berkinerja buruk atau bahkan rusak. Pemilik



mobil kemudian perlu melakukan pemeliharaan reaktif memecahkan masalah dan kemudian memperbaiki atau mengganti komponen yang rusak.

Jaringan secara tradisional dikelola dengan cara yang kurang lebih sama dengan operator melakukan pemeliharaan preventif dan reaktif. Pemeliharaan rutin preventif atau terencana dilakukan untuk mengurangi risiko kegagalan, misalnya, generator cadangan dijalankan untuk memastikan generator tersebut dapat menjalankan fungsinya saat diperlukan dan pembaruan perangkat lunak diterapkan untuk memperbaiki bug perangkat lunak. Pemeliharaan jaringan reaktif, yang sering disebut sebagai manajemen kesalahan, berkaitan dengan pemulihan dari gangguan dalam jaringan. Ini mencakup kegagalan keras dan lunak kegagalan keras mengacu pada saat fungsi jaringan gagal dan sepenuhnya tidak berfungsi, sedangkan kegagalan lunak mengacu pada saat fungsi jaringan menurun, yaitu, lalu lintas mungkin terputus atau rusak, tetapi tidak sepenuhnya gagal. Tujuan utama pemeliharaan preventif dan reaktif adalah manajemen kualitas layanan memastikan dampak minimal pada layanan yang dialami pelanggan.

Pemeliharaan Terencana

Untuk jaringan berskala besar, aktivitas pemeliharaan merupakan realitas yang berkelanjutan, dengan ribuan aktivitas yang biasanya terjadi setiap bulan. Aktivitas pemeliharaan terencana diperlukan untuk meluncurkan perangkat lunak baru guna memperkenalkan fitur jaringan dan layanan baru serta menyediakan perbaikan untuk bug perangkat lunak. Konfigurasi perangkat diubah, seiring dengan perubahan topologi jaringan, untuk mengatasi masalah perangkat lunak dan memperbaiki kesalahan konfigurasi serta mengaktifkan layanan baru. Pemeliharaan rutin juga diperlukan untuk memastikan perangkat keras yang tidak aktif, seperti generator dan baterai, berfungsi kita tidak ingin mengambil risiko mengetahui ada yang tidak beres ketika fungsi jaringan gagal dan generator diaktifkan.

Aktivitas terencana perlu dirancang, dijadwalkan, dan kemudian dijalankan di seluruh fungsi jaringan yang berpotensi sangat besar. Aktivitas perlu dikelola di seluruh fungsi jaringan untuk meminimalkan dampak terhadap pelanggan misalnya, aktivitas pemeliharaan perlu dikoordinasikan di antara mereka sendiri dan mempertimbangkan kondisi jaringan saat ini untuk meminimalkan risiko gangguan layanan jaringan. Aktivitas pemeliharaan juga perlu dipantau secara ketat, dengan tindakan pemulihan seperti rollback (kembali ke kondisi sebelumnya) yang akan dilakukan jika aktivitas tersebut tidak berjalan semestinya.

Pemeliharaan Reaktif

Kegagalan dan gangguan fungsi jaringan terjadi hampir terus-menerus dalam jaringan skala besar bug perangkat lunak, kerusakan perangkat keras, dan cuaca hanyalah beberapa penyebab masalah tersebut. Operasi jaringan perlu terlebih dahulu mendeteksi dan mengidentifikasi kegagalan dan gangguan tersebut mencari tanda-tanda yang menunjukkan masalah dalam sejumlah besar alarm, log, dan data pengukuran yang diterima dari jaringan dan perangkat pemantau layanan. Setelah masalah teridentifikasi, operasi jaringan memiliki dua tujuan utama: (1) memitigasi dampak layanan bagi pelanggan dan (2) memperbaiki masalah secara permanen dan mengembalikan kapasitas yang dipulihkan ke jaringan.



Di sebagian besar jaringan, terdapat kapasitas redundan bawaan yang dapat mengalirkan lalu lintas saat terjadi kegagalan dan gangguan. Jika terjadi kegagalan, atau gangguan signifikan pada fungsi jaringan atau konektivitas antara dua fungsi jaringan, jaringan biasanya akan secara otomatis memulihkan lalu lintas pelanggan dengan mengalihkannya. Dengan demikian, layanan mungkin mengalami dampak minimal karena jaringan pulih secara otomatis. Peran operasional kemudian adalah memperbaiki fungsi jaringan yang gagal sehingga kapasitas dikembalikan ke jaringan. Dan yang terpenting, dari sudut pandang operator, ini berarti ada lebih sedikit tekanan untuk melakukan perbaikan cepat selama fungsi jaringan yang gagal diperbaiki sebelum terjadi kesalahan lain di wilayah yang sama yang mungkin berdampak pada pelanggan jika kapasitas yang tersedia tidak mencukupi setelah kegagalan awal.

Namun, tidak semua kegagalan atau gangguan dapat dipulihkan secara otomatis saat ini. Dalam beberapa kasus yang jarang terjadi, fungsi jaringan mungkin gagal mengenali kegagalan dan mungkin diam-diam menjatuhkan lalu lintas tanpa mekanisme failover diaktifkan. Dalam skenario lain, mungkin tidak ada kapasitas pemulihan yang cukup tersedia untuk memulihkan layanan baik tidak ada mekanisme pemulihan (misalnya, router tepi atau router pelanggan mungkin tidak dikonfigurasi dengan redundansi karena biaya dan kompleksitas penyediaan mekanisme pemulihan kegagalan tersebut), pemadaman mungkin terlalu besar, atau di tepi jaringan mungkin tidak ada kapasitas yang dapat digunakan (misalnya, pelanggan di tepi jaringan mobilitas mungkin tidak dapat gagal ke menara alternatif jika tidak ada kekuatan sinyal yang cukup di menara tetangga atau jika kegagalan terlalu besar secara geografis). Interaksi yang kompleks antara fungsi jaringan juga dapat memicu mode kegagalan tak terduga yang dapat sangat rumit untuk dipahami, atau kesalahan operator atau bahkan perilaku jahat dalam mengonfigurasi fungsi jaringan dapat mengakibatkan mereka gagal membawa lalu lintas seperti yang diharapkan. Dalam skenario seperti itu, operator jaringan harus mendeteksi, melokalisasi, dan memecahkan masalah dengan cepat, lalu mengambil tindakan untuk memulihkan layanan. Sementara itu, pelanggan berpotensi mengalami penurunan layanan. Tergantung pada kompleksitas masalah yang dihadapi, melokalisasi dan memecahkan masalah dapat menjadi tugas yang sangat memakan waktu dan seringkali sangat menantang, dilakukan di bawah tekanan waktu yang sangat besar.

Untuk fungsi jaringan fisik, memulihkan fungsi jaringan setelah kegagalan perangkat keras biasanya berarti memperbaiki perangkat keras fisik jika rusak atau terganggu. Tindakan sementara mungkin diperlukan untuk memulihkan layanan jika perbaikan permanen mungkin membutuhkan waktu lebih lama. Misalnya, lalu lintas dapat dialihkan di sekitar fungsi jaringan yang rusak, atau generator tetap atau portabel dapat digunakan dalam situasi di mana listrik terputus. Banyak dari perbaikan perangkat keras ini mengharuskan teknisi di lokasi dengan peralatan yang rusak yang biasanya mahal. Teknisi harus bertugas di kantor 24/7, atau teknisi harus dikerahkan ke kantor ketika perangkat keras perlu diperbaiki, yang menyebabkan penundaan terkait. Jika respons harus cepat (misalnya, jika pelanggan akan terdampak hingga perbaikan selesai), maka hal ini bisa sangat mahal.



Namun, di dunia virtual, VNF dapat dipulihkan bahkan saat terjadi kegagalan perangkat keras (server, penyimpanan, dan kegagalan jaringan yang mendasarinya) dengan menjalankan VNF baru untuk menggantikan VNF yang rusak atau terdegradasi. Ini menjadi fungsi yang dikelola sepenuhnya dari jarak jauh mendeteksi VNF yang rusak lalu menginstansiasinya kembali pada perangkat keras baru. Meskipun perangkat keras yang rusak pada akhirnya harus diperbaiki, hal ini dapat dijadwalkan secara semi-reguler, dengan banyak perbaikan/penggantian dijadwalkan di lokasi yang sama, misalnya, sebulan sekali. Hal ini secara signifikan mengurangi overhead operasional dan biaya terkait penanganan kegagalan perangkat keras. Bahkan, dalam banyak skenario kegagalan, ini dapat menjadi fungsi yang sepenuhnya otomatis (berbasis perangkat lunak) tanpa keterlibatan operator manusia, berbeda dengan fungsi jaringan fisik di mana kegagalan perangkat keras perlu dilokalisasi dan diisolasi secara akurat ke akar penyebab spesifik, dan teknisi perlu dikerahkan (jika belum berada di kantor) untuk segera mengganti perangkat keras yang rusak.

Kegagalan perangkat lunak biasanya dapat dipulihkan dari jarak jauh baik untuk elemen jaringan fisik maupun virtual. Untuk elemen jaringan fisik, perangkat lunak secara inheren terikat dengan perangkat keras, sehingga operator hanya dapat melakukan boot ulang perangkat atau kartu jaringan, atau mengubah konfigurasi perangkat. Tindakan perbaikan semacam itu dan lainnya juga tersedia untuk VNF.

Administrasi Jaringan

Administrasi jaringan memastikan bahwa jaringan memenuhi tujuan layanan yang telah ditetapkan, bahkan ketika menghadapi perubahan pengaruh eksternal dan internal. Secara khusus, kami berfokus pada bagaimana topologi jaringan dan perutean lalu lintas diadaptasi untuk mempertahankan tingkat layanan jika memungkinkan. Kami membahas tiga fungsi inti manajemen lalu lintas, penanganan kejadian khusus, dan pemulihan bencana.

Bagian inti dari administrasi jaringan adalah manajemen lalu lintas, yang memastikan bahwa jaringan tetap berfungsi dengan baik (memenuhi tujuan layanan) bahkan ketika menghadapi abnormalitas lalu lintas (misalnya, lonjakan lalu lintas yang tiba-tiba) atau kegagalan dan gangguan jaringan. Tugas operator jaringan dapat mencakup pengalihan rute lalu lintas sebagai respons terhadap pemadaman jaringan, aktivitas pemeliharaan, kejadian khusus, dan bencana. Respons semacam itu dapat dilakukan secara reaktif atau proaktif, tergantung pada apakah pengetahuan sebelumnya tersedia atau tidak.

Dalam kasus ekstrem, operator jaringan perlu merespons bencana saat bencana tersebut melanda suatu wilayah. Dalam situasi seperti itu, dapat terjadi lonjakan lalu lintas secara tiba-tiba yang menghantam jaringan, berpotensi ketika kapasitas juga berkurang akibat bencana. Kontrol manajemen lalu lintas digunakan untuk memastikan bahwa kemacetan tidak menyebar ke area lain dalam jaringan, dan bahwa kapasitas jaringan dimanfaatkan secara efektif untuk mengelola komunikasi di dalam dan di dalam/keluar area bencana. Misalnya, kontrol biasanya diterapkan untuk memastikan bahwa panggilan yang berasal dari area bencana mendapatkan perlakuan istimewa dibandingkan panggilan yang berasal dari luar area tersebut tetapi ingin berakhir di sana.



Banyak kontrol manajemen lalu lintas yang saat ini diotomatisasi terutama ketika respons cepat sangat penting untuk meminimalkan dampak terhadap pelanggan. Namun, beberapa kontrol masih memerlukan intervensi manusia, terutama ketika penilaian manusia diperlukan (misalnya, pemulihan bencana). Dalam beberapa situasi, sekadar mengalihkan lalu lintas di dalam jaringan tidak cukup untuk menangani peningkatan sementara dalam permintaan lalu lintas. Acara sosial besar, seperti konser, acara olahraga (misalnya, maraton), dan acara politik (misalnya, pelantikan presiden atau protes/pawai) dapat menarik populasi besar ke suatu area dalam waktu singkat. Mengingat arus masuk orang yang tiba-tiba dan perangkat jaringan seluler terkait jaringan seluler berisiko mengalami kelebihan beban di tepi jaringan. Untuk menghindari hal ini, operator jaringan seluler berupaya keras untuk mengidentifikasi acara mendatang dan kemudian memperkenalkan kapasitas jaringan sementara, menggunakan teknologi seperti menara seluler beroda atau COW (Gambar 9.1). Perencanaan yang signifikan terlibat di sini mengidentifikasi dan memprediksi di mana acara tersebut akan terjadi, menentukan berapa banyak kapasitas yang dibutuhkan, dan kemudian menjadwalkan sumber daya yang diperlukan.



Gambar 9.1 Menara seluler beroda.

Untuk setiap kejadian, COW perlu diangkut ke lokasi yang relevan dan dihubungkan ke jaringan, kemudian dipantau dan dikelola dengan tepat. Pengalihan kapasitas sesuai permintaan yang serupa dapat diperlukan dalam menanggapi bencana. Layanan komunikasi sangat penting bagi fungsi masyarakat dan tidak dapat hilang begitu saja bahkan dalam menghadapi bencana alam atau bencana buatan manusia yang besar. Oleh karena itu, penyedia layanan seperti AT&T berupaya keras untuk memastikan bahwa mereka dapat memulihkan layanan dengan cepat bahkan dalam keadaan yang paling buruk, seperti setelah tornado, badai, dan serangan teroris. Ini dikenal sebagai pemulihan bencana.

Pemulihan bencana membutuhkan perencanaan yang signifikan dan pelaksanaan yang kompleks. Bencana tentu saja sulit diprediksi, tetapi tim operasi jaringan bertanggung jawab untuk melakukan perencanaan skenario yang cermat dan untuk membuat rencana pemulihan



bencana yang komprehensif. Sebagai bagian dari perencanaan pemulihan bencana, AT&T telah menciptakan kantor portabel dibangun di atas truk yang dapat diangkut di sekitar jaringan dan dihubungkan sesuai kebutuhan. Truk pemulihan bencana ini membawa aset jaringan fisik *Reconfigurable Optical Add-Drop Multiplexer (ROADM)*, router, sakelar suara, dan banyak lagi yang perlu dihubungkan jika terjadi bencana untuk memulihkan layanan jaringan. Ini dilengkapi dengan generator dan COW sehingga jaringan dapat diciptakan kembali dalam menghadapi bencana (Gambar 9.2).



Gambar 9.2 Truk pemulihan bencana yang tiba di Puerto Riko setelah Badai Maria pada tahun 2017.

Tim operasi bertanggung jawab untuk merancang skenario bencana dan membuat rencana untuk menanggapi, bersama dengan mengeksekusi rencana tersebut jika terjadi bencana. Tim melakukan latihan rutin untuk memastikan bahwa mereka siap ketika bencana terjadi. Dan tentu saja, tim operasi bertanggung jawab atas pemulihan jaringan ketika rencana tersebut dijalankan. Sebagai contoh, lebih dari seribu teknisi dikerahkan selama badai AS tahun 2017 untuk memulihkan layanan dengan cepat bersama dengan sel pada truk ringan dengan backhaul satelit, COW, dan sejumlah besar generator. AT&T juga mengerahkan helikopter "sel di sayap" pertamanya selama pemulihan pasca Badai Maria di Puerto Riko, yang digambarkan pada Gambar 9.3.



Gambar 9.3 Sel bersayap (juga dikenal sebagai flying COW) beraksi di Puerto Riko setelah Badai Maria (2017).



9.3 JARINGAN ZERO-TOUCH: OTOMATISASI & KECERDASAN

Seiring beban lalu lintas jaringan dan kapasitas jaringan yang terkait terus tumbuh dengan pesat, otomatisasi fungsi operasi jaringan menjadi sangat penting. Apa yang diperlukan untuk mencapai jaringan zero-touch? Dan bagaimana AI berperan dalam hal ini?

Mari kita bayangkan sebuah dunia di mana jaringan sepenuhnya otonom, di mana perancang dan operator jaringan memberikan masukan kunci (maksud/kebijakan) untuk memandu bagaimana jaringan beroperasi dan kemudian otomatisasi mengambil alih. Jaringan secara otomatis dan otonom memutuskan di mana dan kapan kapasitas dibutuhkan dan menjalankan dan/atau secara fisik menyebarkan kapasitas jaringan baru, lalu membuat perubahan yang diinginkan di seluruh jaringan untuk mewujudkan tujuan bisnis. Kapasitas yang tidak diperlukan secara otomatis dihentikan layanannya, dan perangkat keras dimatikan atau "ditidurkan" untuk menghemat energi saat tidak diperlukan. Jaringan memantau kondisi jaringan dan layanan, dan secara otomatis dan cepat memperbaiki masalah jaringan dan layanan dengan dampak minimal terhadap pelanggan. Jika memungkinkan, jaringan juga memprediksi dan memulihkan gangguan jaringan yang akan datang sebelum terjadi. Teknologi yang muncul dalam bentuk mobil otonom dan drone memberikan fungsi jaringan fisik di tempat yang dibutuhkan dan tanpa campur tangan manusia misalnya, sel di atas roda yang digunakan saat ini untuk menambah kapasitas untuk acara khusus atau untuk mengganti kapasitas jaringan yang gagal. Robot bahkan menyebarkan dan memperbaiki peralatan fisik, memasang dan memperbaiki server, memasang serat dan menghubungkan koneksi fisik atau memperbaiki potongan serat. Bayangkan saja apa yang telah dicapai manusia dengan robot kita telah mampu menyebarkan teknologi robot di planet-planet terpencil jutaan mil jauhnya! Robot yang memperbaiki serat dan menyebarkan peralatan bukanlah hal yang tak terbayangkan.

Seiring dengan semakin terwujudnya visi jaringan zero-touch, efisiensi operasional akan meningkat. Fungsi-fungsi jaringan akan diterapkan sesuai kebutuhan, sehingga pengeluaran modal dapat digunakan secara efisien karena kapasitas tepat waktu (*just-in-time*) disesuaikan dengan kebutuhan real-time. Pelanggan akan mendapatkan manfaat dari kualitas layanan superior yang dimungkinkan oleh respons cepat dan otomatis terhadap gangguan jaringan dan peningkatan kapasitas yang cepat. Konsumsi daya akan ditingkatkan dengan implikasi lingkungan selanjutnya. Dan terakhir, inovasi layanan berkelanjutan dimungkinkan, melalui peluncuran perangkat lunak, teknologi, dan layanan baru yang cepat.

Teknologi jaringan yang ditentukan perangkat lunak akan menjadi inti untuk mewujudkan visi zero-touch. Mengingat keragaman dan kompleksitas jaringan dan layanan penyedia layanan saat ini, sangat penting bahwa teknologi kontrol jaringan yang secara cerdas mengotomatiskan fungsi operasi memanfaatkan kecerdasan yang dapat digunakan kembali di berbagai fungsi dan layanan jaringan. Operator juga harus dapat menentukan bagaimana aktivitas operasi jaringan tersebut dilakukan untuk setiap fungsi dan layanan jaringan tertentu dengan menentukan maksudnya melalui kebijakan. Misalnya, kebijakan dapat digunakan untuk menentukan bagaimana VNF ditempatkan termasuk aturan afinitas. Kebijakan lain



menentukan masalah apa yang harus dideteksi dan bagaimana cara mengatasinya. Sistem operasi jaringan, seperti *Open Network Automation Platform* (ONAP) dari Linux Foundation, memungkinkan hal ini.

Tingkat "berpikir" dan "belajar" yang canggih perlu melengkapi otomatisasi jaringan berbasis kebijakan untuk mewujudkan visi jaringan tanpa sentuhan. Kecerdasan buatan akan memprediksi permintaan pelanggan di masa mendatang, menerjemahkannya ke dalam permintaan jaringan, dan secara cerdas memutuskan di mana dan kapan akan meningkatkan dan menerapkan kapasitas baru. Algoritme akan memantau sejumlah besar data jaringan secara real-time, memprediksi kemungkinan kondisi gangguan di masa mendatang, dan memperbaikinya sebelum terjadi. Algoritma pembelajaran mesin juga akan sangat penting untuk mempelajari kondisi jaringan dan layanan normal serta mengidentifikasi kondisi anomali yang memerlukan perbaikan. Dan pada akhirnya, kecerdasan jaringan harus secara otomatis mempelajari cara menangani situasi anomali, termasuk kegagalan dan dampak pemeliharaan terencana, untuk mendorong perbaikan mandiri. Bagian berikut akan membahas lebih lanjut visi ini.

Visi Penyediaan Jaringan Otonom

Dalam jaringan zero-touch, penyediaan jaringan merupakan proses yang sepenuhnya otomatis menentukan di mana dan kapan kapasitas jaringan harus dikerahkan, menyebarkan fungsi-fungsi jaringan, lalu menjalankan "uji coba dan aktifkan" sebelum menyatakan fungsi-fungsi jaringan siap untuk membawa lalu lintas produksi.

Sebagian dari proses ini saat ini sudah otomatis operator manusia memperkirakan permintaan lalu lintas di masa mendatang, mengumpulkan permintaan saat ini, dan mengekstrapolasi pertumbuhan di berbagai layanan yang dibawa melalui jaringan. Operator menjalankan berbagai perangkat perencanaan kapasitas untuk menentukan di mana akan menyebarkan kapasitas baru. Fungsi jaringan fisik dikerahkan oleh teknisi secara manual memasang perangkat keras ke dalam ruang-ruang di kantor pusat. Konfigurasi perangkat ini dapat diotomatisasi atau setidaknya semi-otomatis terutama jika skalanya signifikan. Dan terakhir, operator biasanya akan menjalankan serangkaian skrip untuk menguji fungsi jaringan sebelum memperkenalkan fungsi dan kapasitas jaringan baru ke dalam jaringan melalui pembaruan konfigurasi perutean.

Perencanaan Kapasitas

Pertama-tama, kami mempertimbangkan fungsi jaringan yang dapat dibangun atau dikonfigurasi secara dinamis mendekati waktu nyata, seperti VNF dan konektivitas antar fungsi jaringan yang ditetapkan sebagai "bandwidth on demand". Virtualisasi secara radikal mengubah proses perencanaan kapasitas dibandingkan dengan elemen jaringan fisik historis; dengan asumsi terdapat kapasitas cloud yang memadai di dalam pusat data, fungsi jaringan baru dapat dijalankan dalam hitungan menit atau kurang. Jika kapasitas jaringan antar kantor juga dapat diprovisi dengan cepat untuk menghubungkan VNF, maka kebutuhan historis untuk prakiraan lalu lintas berbulan-bulan atau bahkan bertahun-tahun dapat dihilangkan untuk VNF. Dengan demikian, kami berasumsi bahwa terdapat kapasitas cadangan yang cukup sehingga ketika fungsi jaringan baru ini diperlukan, fungsi tersebut dapat diinstansiasi dengan



memanfaatkan server atau bandwidth cadangan yang tidak digunakan atau sedang digunakan oleh layanan yang dapat diprioritaskan.

Dengan demikian, kebutuhan kapasitas VNF di sini dapat diidentifikasi sesuai permintaan. Beban lalu lintas dipantau dan di-tren dari waktu ke waktu untuk memprediksi permintaan lalu lintas jangka pendek di masa mendatang, dengan keputusan yang dibuat secara otomatis mengenai kapan kapasitas baru harus secara otomatis diaktifkan (diinstansiasi) pada pusat data atau kapasitas jaringan yang telah diterapkan sebelumnya. Pembelajaran mesin sangat penting untuk memutuskan kapan dan di mana akan mengaktifkan VNF baru, dengan mempertimbangkan kapasitas server dan jaringan interkoneksi yang tersedia, persyaratan pelanggan (misalnya, persyaratan latensi), dan banyak lagi.

Kita semua tahu bahwa memprediksi masa depan bisa menjadi tantangan, dan ini juga berlaku untuk peramalan lalu lintas. Dengan demikian, menghilangkan proses peramalan yang panjang ini untuk fungsi-fungsi jaringan ini mengurangi atau bahkan menghilangkan kesalahan peramalan. Ini berarti bahwa kita dapat mengurangi atau bahkan menghilangkan sejumlah besar kapasitas terlantar yang berpotensi disebabkan oleh kesalahan peramalan di lokasi tertentu, sementara berebut kapasitas di lokasi lain. Demikian pula, lalu lintas dapat dirutekan pada rute optimal jika kapasitas jaringan tersedia di lokasi yang diinginkan, berbeda dengan kesalahan historis yang memaksa lalu lintas dirutekan pada jalur suboptimal, yang mengakibatkan latensi yang lebih lama bagi lalu lintas pelanggan dan pemanfaatan kapasitas yang tidak efisien bagi penyedia layanan.

Virtualisasi juga membuka peluang untuk menghapus fungsi jaringan ketika tidak diperlukan untuk jangka waktu tertentu atau untuk memigrasikan fungsi dari waktu ke waktu bahkan berpotensi berdasarkan waktu untuk fungsi yang mendukung layanan yang tidak sensitif terhadap penundaan/lokasi (dengan asumsi kapasitas interkoneksi yang memadai tersedia). Semua ini membutuhkan pembelajaran mesin algoritma canggih untuk mempelajari pola lalu lintas dan membuat keputusan yang cerdas. Jaringan virtual juga dapat dioptimalkan ulang secara berkala, dengan menempatkan kembali VNF secara dinamis untuk mempertahankan pemanfaatan sumber daya yang efisien.

Fungsi jaringan virtual dan interkonektivitas dinamis dibangun dengan menggunakan kapasitas fisik cadangan yang tersedia di jaringan sesuai kebutuhan. Aset fisik ini perlu di-pre-deploy agar tersedia saat dibutuhkan. Menyebarkan aset fisik termasuk operator menara seluler (kapasitas), server, penyimpanan, dan ROADM merupakan proses yang jauh lebih lama daripada membangun VNF dan dapat memakan waktu berminggu-minggu hingga berbulan-bulan (atau bahkan lebih lama). Gedung, serat optik, dan lokasi menara seluler harus dibeli dan disebarkan dalam skala waktu yang lebih lama lagi bahkan bisa memakan waktu bertahun-tahun untuk mendapatkan atau menyewa properti dan menegosiasikan hak jalan dan lokasi menara seluler. Peramalan permintaan pengguna menyediakan titik awal untuk proses penyediaan jaringan aset fisik. Permintaan tersebut dibuat dengan memprediksi beban di masa mendatang berdasarkan pertumbuhan permintaan historis dan memasukkan aberasi yang diketahui atau diprediksi dari tingkat pertumbuhan ini. Kecerdasan buatan dan pembelajaran mesin akan berperan penting dalam hal memprediksi permintaan lalu lintas di



masa mendatang termasuk melihat faktor-faktor eksternal yang dapat menyebabkan variasi tinggi pada permintaan dasar yang sedang tren. Contoh faktor eksternal dapat mencakup pelanggan tertentu yang dapat membawa fungsi langkah besar dalam permintaan di lokasi tertentu atau pergeseran populasi besar di kota tertentu karena perubahan industri besar yang dapat diprediksi dengan memantau tren industri. Pembelajaran semacam itu membutuhkan analisis data dari jauh di luar jaringan termasuk data sosial, populasi, dan ekonomi.

Perencanaan kapasitas di jaringan masa depan akan menjadi fungsi yang sepenuhnya otomatis dan sangat optimal yang secara bersamaan memperhitungkan berbagai lapisan jaringan. Kebutuhan kapasitas server dan penyimpanan, kapasitas jaringan lokal, fungsi jaringan fisik (misalnya, menara seluler), dan konektivitas antar kantor (misalnya, permintaan fiber/*Dense Wavelength Division Multiplexing*) diprediksi berdasarkan perkiraan permintaan pengguna di berbagai layanan yang didukung oleh penyedia layanan; hal ini kemudian diterjemahkan menjadi kebutuhan kapasitas untuk berbagai teknologi (peralatan) di lokasi tertentu dalam jaringan.

Menyebarkan, Mengonfigurasi, dan Mengaktifkan Fungsi Jaringan

Otomatisasi jaringan yang diimplementasikan melalui platform sistem operasi jaringan akan menyediakan mekanisme untuk secara otomatis meningkatkan kapasitas jaringan virtual, tanpa campur tangan manusia. Otomatisasi jaringan juga akan memulai penyediaan fungsi jaringan fisik. Identitas dan detail konfigurasi lainnya dari fungsi jaringan akan ditentukan secara otomatis, menerjemahkan deskripsi tingkat tinggi tentang maksud operator jaringan, yang kemudian diterjemahkan ke konfigurasi khusus perangkat. Otomatisasi akan menginstal konfigurasi pada fungsi jaringan dan kemudian memvalidasi instansiasi fungsi jaringan yang berhasil melalui serangkaian pengujian otomatis, sebelum membuat kapasitas yang baru disebarkan tersedia untuk membawa lalu lintas produksi.

Pembelajaran mesin dapat melengkapi pengujian yang ditentukan operator untuk mengidentifikasi fungsi jaringan yang dikonfigurasi secara anomali atau berperilaku anomali. Dengan menganalisis konfigurasi dan perilaku fungsi jaringan yang telah disebarkan, kita dapat mempelajari apa yang normal untuk jenis fungsi jaringan tertentu; Membandingkan fungsi jaringan normal ini dengan fungsi jaringan yang baru diinstansiasi dapat mengidentifikasi situasi di mana fungsi jaringan salah konfigurasi atau tidak berfungsi atau berperilaku konsisten dengan fungsi jaringan yang ada. Faktanya, perbandingan jenis ini dapat dilakukan secara berkelanjutan tidak hanya saat fungsi jaringan baru diinstansiasi.

Visi Pemeliharaan Jaringan Otonom

Tugas operasi jaringan yang paling memakan waktu dan kompleks yang dilakukan saat ini adalah pemeliharaan jaringan terutama pemeliharaan reaktif. Kami percaya bahwa di sinilah peluang paling signifikan untuk otomatisasi dan AI berada.

Pemeliharaan Terencana

Pemeliharaan terencana merupakan bagian penting dari siklus hidup jaringan, aplikasi, dan layanan diperlukan untuk memperkenalkan layanan dan fitur baru serta mencegah kegagalan tertentu. Banyak kegagalan umum dapat dihindari jika pemantauan/diagnosis dan pemeliharaan preventif/terencana yang tepat dilakukan.



Secara historis, aktivitas pemeliharaan terencana sering kali direncanakan oleh teknisi jaringan yang menyediakan spesifikasi tertulis tingkat tinggi tentang apa yang dibutuhkan dari setiap aktivitas pemeliharaan dan daftar elemen jaringan tempat aktivitas pemeliharaan tertentu akan dijalankan. Dari sana, operator jaringan membuat jadwal lokal untuk mengeksekusi perubahan pada fungsi jaringan yang relevan, berdasarkan sumber daya manusia yang tersedia dan batasan spesifik fungsi jaringan (misalnya, interval waktu di mana perjanjian pelanggan mengizinkan pemeliharaan fungsi jaringan dilakukan). Operator kemudian harus bekerja lintas tim operasi untuk "mengurai konflik" jadwal memastikan bahwa aktivitas pada elemen jaringan yang berbeda tidak saling bertentangan. Langkah terakhir ini biasanya merupakan proses yang sangat manual dan memakan waktu. Akhirnya, eksekusi biasanya sepenuhnya atau sebagian diotomatisasi terutama ketika terdapat banyak elemen jaringan yang akan dieksekusi pemeliharaannya, seperti di tepi jaringan.

Aktivitas pemeliharaan terencana di masa mendatang akan berjalan sendiri operator jaringan menentukan maksud dan cakupan pemeliharaan, lalu otomatisasi mengambil alih. Otomatisasi berkoordinasi di seluruh aktivitas secara cerdas menjadwalkan (dan menjadwalkan ulang) berbagai aktivitas pada fungsi jaringan yang berbeda, dengan mempertimbangkan berbagai batasan pelanggan, topologi jaringan, dan layanan. Pada waktu yang dijadwalkan, otomatisasi menjalankan aktivitas pemeliharaan secara otomatis menerapkan rilis perangkat lunak baru, mengubah konfigurasi jaringan, memvalidasi efektivitas perubahan, dan menangani dampak negatifnya.

Aktivitas pemeliharaan terencana, setidaknya dalam waktu dekat, kemungkinan akan dimulai oleh operator jaringan. Operator ini akan memutuskan pemeliharaan apa yang perlu dilakukan, menjelaskannya secara rinci kepada sistem otomatisasi yang kemudian mengambil alih untuk menjadwalkan dan menjalankan aktivitas tersebut. Namun, bahkan mungkin suatu hari nanti jaringan memiliki kecerdasan untuk memutuskan sendiri pemeliharaan apa yang perlu dilakukan lalu melanjutkan, menjadwalkan, dan menjalankan aktivitas tersebut.

Setelah keputusan tentang pemeliharaan apa yang perlu dilakukan telah dibuat, otomatisasi perlu menjadwalkan aktivitas dan menjalankannya. Tujuan penting dari fungsi-fungsi ini adalah meminimalkan dampak terhadap pelanggan selama pemeliharaan dengan demikian, aktivitas yang akan menyebabkan fungsi jaringan tidak tersedia dijadwalkan dengan cermat untuk memastikan bahwa aktivitas tersebut dilakukan pada waktu yang meminimalkan dampak terhadap pelanggan, dan aktivitas tersebut dikoordinasikan sedemikian rupa untuk meminimalkan dampak terhadap pelanggan. Misalnya, satu menara seluler dapat dihentikan layanannya untuk pemeliharaan tanpa dampak negatif terhadap pelanggan pelanggan di menara seluler tersebut secara otomatis terhubung ke menara di sekitarnya ketika menara yang sedang dipelihara tidak tersedia. Namun, jika terlalu banyak menara yang dihentikan layanannya di lokasi yang sama, maka mungkin tidak ada menara seluler yang dapat memenuhi permintaan pelanggan, sehingga pelanggan mungkin tidak dapat melakukan panggilan telepon, menonton video, atau terhubung ke Internet. Dengan demikian, meminimalkan dampak terhadap pelanggan memerlukan perencanaan yang cermat tentang fungsi jaringan mana yang dapat secara bersamaan melakukan aktivitas pemeliharaan.



Perhatikan bahwa jika pemeliharaan dapat dilakukan tanpa waktu henti pada fungsi jaringan, seperti yang dijanjikan dengan virtualisasi jaringan dan migrasi VNF tanpa hitless, penjadwalan aktivitas yang canggih akan menjadi kurang penting.

Koordinasi berbagai aktivitas pemeliharaan dicapai melalui penjadwalan aktivitas di berbagai jenis fungsi jaringan dan di jenis fungsi yang sama (misalnya, menara seluler yang berbeda). Algoritma penjadwalan memperhitungkan periode waktu yang tersedia, kendala pelanggan, topologi dan jalur layanan, dampak pelanggan yang diprediksi, dan durasi aktivitas yang diharapkan untuk mengidentifikasi kapan pemeliharaan harus dilakukan pada berbagai fungsi jaringan. Namun, gangguan dan masalah tak terduga selama pemeliharaan dapat terjadi, sehingga penjadwalan ulang aktivitas pemeliharaan mungkin diperlukan. Algoritma untuk mencapai semua ini bisa sangat rumit, terutama ketika bekerja di beberapa lapisan jaringan dan dengan topologi jaringan serta jalur layanan yang kompleks.

Pembelajaran mesin dapat digunakan untuk menentukan apakah aktivitas yang diluncurkan di seluruh jaringan menyebabkan potensi dampak terhadap pelanggan selama pelaksanaan perubahan informasi yang dapat digunakan untuk memutuskan kapan aktivitas tersebut harus dilakukan (misalnya, apakah aktivitas tersebut perlu dijalankan selama periode lalu lintas pelanggan rendah, biasanya pada malam hari?). Pembelajaran mesin dan analitik tingkat lanjut juga dapat digunakan untuk memprediksi dampak terhadap pelanggan ketika fungsi jaringan tidak dapat beroperasi selama aktivitas pemeliharaan misalnya, memprediksi apakah menara di sekitarnya dapat melayani lalu lintas yang terdampak oleh serangkaian menara seluler yang sedang menjalani pemeliharaan.

Pembelajaran mesin juga dapat digunakan untuk menentukan periode waktu terbaik per lokasi untuk melakukan aktivitas pemeliharaan tersebut biasanya saat lalu lintas berada pada titik terendah, seringkali larut malam, meskipun faktor lain seperti aktivitas penting pelanggan tertentu juga dapat diperhitungkan. Prediksi tersebut perlu memperhitungkan aktivitas normal di lokasi tersebut, yang dapat diperoleh dari data jaringan, beserta informasi yang diperoleh dari sumber non-jaringan tentang aktivitas yang mungkin terjadi di dalam atau di sekitar wilayah terdampak (misalnya, acara sosial yang dapat menimbulkan permintaan yang luar biasa tinggi). Demikian pula, pembelajaran mesin dapat menggunakan data eksternal non-jaringan untuk memprediksi kapan faktor eksternal dapat memengaruhi pemeliharaan misalnya, jika diprediksi akan terjadi cuaca buruk.

Eksekusi aktivitas yang direncanakan akan diotomatisasi termasuk memeriksa kondisi jaringan saat ini untuk menentukan apakah "aman" untuk melanjutkan aktivitas yang direncanakan, dan juga menangani ketika terjadi kesalahan (yang kami sebut sebagai fallout). Yang kedua adalah yang paling rumit untuk diotomatisasi solusi sederhananya adalah dengan mengembalikan perubahan, tetapi ini tidak selalu memungkinkan dan juga bukan solusi. Pembelajaran mesin dapat digunakan untuk mempelajari tindakan yang paling tepat untuk menangani fallout yang tidak terduga.

Fallout seringkali sangat mudah diidentifikasi jika kita tidak dapat mencapai fungsi jaringan, atau menjalankan perubahan yang diinginkan, atau jika fungsi jaringan gagal akibat aktivitas pemeliharaan, segera terlihat jelas bahwa terdapat fallout. Namun, terkadang



perubahan tersebut dapat dijalankan dan menimbulkan dampak negatif yang lebih halus. Dengan demikian, pembelajaran mesin digunakan untuk menentukan apakah perubahan jaringan memiliki dampak negatif jangka panjang yang tidak terduga, yang mungkin merupakan indikasi bug perangkat lunak atau kesalahan konfigurasi. Hal ini dapat diidentifikasi dengan membandingkan kinerja sebelum dan sesudah perubahan baik pada satu fungsi jaringan maupun di berbagai fungsi jaringan yang semuanya telah mengalami perubahan serupa. Perbandingan tersebut perlu memperhitungkan fluktuasi kinerja yang normal, seperti perubahan waktu atau musim.

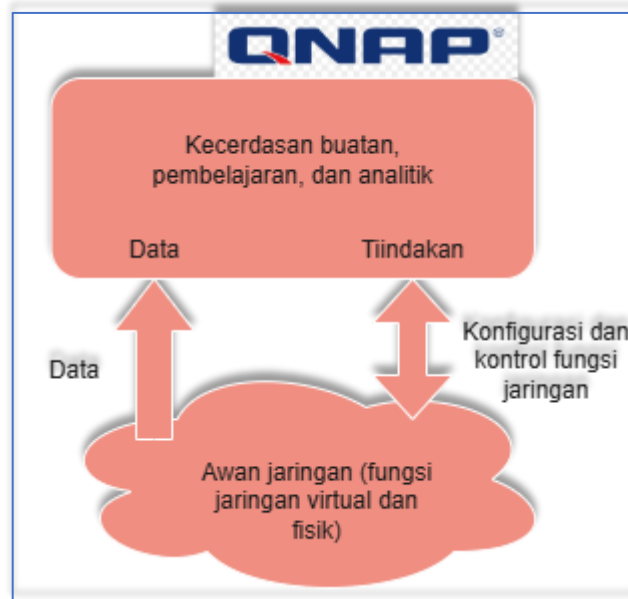
Pemeliharaan Reaktif

Sumber daya operator yang signifikan biasanya terlibat dalam pemeliharaan reaktif saat ini. Operator biasanya merespons tiket, yang biasanya dibuat oleh sistem otomatis yang mendeteksi dan (sebagian besar) melokalisasi kondisi jaringan. Sistem otomatisasi proses mengisi informasi tambahan ke dalam tiket, menyederhanakan pengumpulan dan analisis data bagi operator, sehingga mempercepat investigasi pemecahan masalah. Operator mendalami data dan informasi yang terkumpul dan menyelidiki lebih lanjut masalah tersebut untuk mengidentifikasi apa yang terjadi dan menentukan tindakan pemulihan potensial. Operator mengambil tindakan yang tepat terkadang harus mencoba berbagai alternatif untuk melaksanakan perbaikan.

Sebaliknya, kami membayangkan dunia di mana mesin secara otomatis mendeteksi, mengisolasi, dan mengambil tindakan pemulihan sebagai respons terhadap dan sebelum gangguan jaringan dan layanan yang diprediksi. Loop tertutup semacam itu diimplementasikan menggunakan platform otomatisasi jaringan, di mana otomatisasi memantau kondisi jaringan dan layanan, mendeteksi dan melokalisasi masalah, membuat keputusan cerdas mengenai tindakan apa yang harus diambil, dan kemudian melaksanakan tindakan tersebut.

Dalam jaringan zero-touch, semua masalah diselesaikan secara otomatis melalui tindakan loop tertutup tanpa campur tangan manusia. Namun, kami memperkirakan tidak semua masalah akan diselesaikan secara otomatis dalam jangka pendek hingga menengah, yang mengarah pada loop kontrol yang lebih umum, yang memungkinkan kontrol loop terbuka dan tertutup. Loop terbuka mengacu pada saat otomatisasi gagal dan operator manusia terlibat. Loop terbuka dapat terjadi karena sistem tidak dapat menentukan tindakan loop tertutup yang tepat untuk dicoba, atau karena tindakan loop tertutup yang dicoba tidak berhasil menyelesaikan masalah. Loop terbuka juga digunakan saat tindakan loop tertutup (otomatis) sedang dipertimbangkan sistem otomatisasi merekomendasikan tindakan yang kemudian divalidasi (atau tidak) oleh operator jaringan sebelum tindakan diambil; memungkinkan operator jaringan untuk mendapatkan keyakinan dalam tindakan loop tertutup baru sebelum otomatisasi dimanfaatkan sepenuhnya.





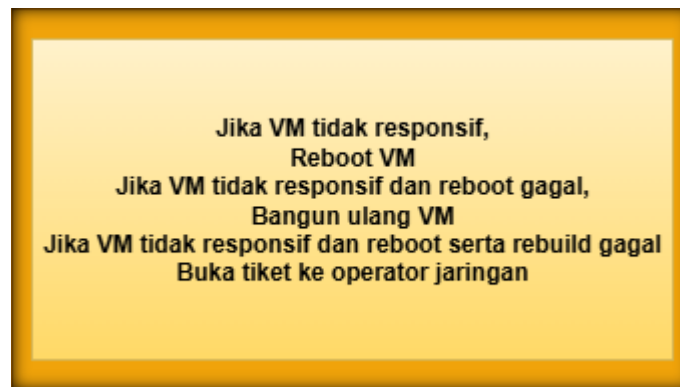
Gambar 9.4 Loop tertutup berbasis pembelajaran mesin.

Kecerdasan buatan dan pembelajaran mesin akan menjadi bagian integral dari loop kontrol di masa mendatang. Seperti yang diilustrasikan pada Gambar 9.4, pembelajaran mesin dan analitik lainnya akan terus memantau sejumlah besar data terkait kondisi jaringan dan layanan untuk mendeteksi kondisi anomali dan memprediksi gangguan di masa mendatang. Kecerdasan buatan akan digunakan untuk mendiagnosis akar penyebab kondisi yang terdeteksi secara otomatis (sesuai kebutuhan), dan analitik berbasis pembelajaran mesin akan mengukur dampak layanan (pelanggan) dari gangguan yang terdeteksi. Terakhir, kami berharap pembelajaran mesin juga akan menjadi kunci untuk mempelajari secara cerdas tanda-tanda apa yang mengindikasikan masalah dan tindakan apa yang harus dilakukan sistem sebagai respons terhadap masalah yang terdeteksi. Pembelajaran otomatis perlu beradaptasi dengan kondisi yang berubah, seperti ketika perangkat lunak fungsi jaringan diperbarui yang akan menimbulkan bug perangkat lunak baru dengan tindakan pemulihan yang berbeda.

Dalam visi masa depan yang sesungguhnya, sistem akan melengkapi pembelajaran berdasarkan pengalaman historis dengan eksperimen aktif otomatis dari tindakan pemulihan alternatif, seperti yang akan dilakukan operator manusia ketika menghadapi gangguan tanpa resolusi yang jelas dan nyata. Sistem juga akan belajar kapan harus mengambil tindakan pemulihan yang berbeda mempelajari dari waktu ke waktu dampak yang ditimbulkan oleh tindakan pemulihan terhadap pelanggan dan membandingkannya selama kejadian tertentu dengan dampak yang dialami pelanggan saat ini dan yang diprediksi akan dialami sebagai akibat dari gangguan yang akan diperbaiki. Misalnya, jika tindakan pemulihan direncanakan sebagai respons terhadap kondisi yang berdampak minimal atau bahkan tidak ada sama sekali terhadap pelanggan, maka tindakan pemulihan yang diperkirakan akan mengganggu arus lalu lintas (misalnya, memulai ulang fungsi jaringan) dapat ditunda hingga lalu lintas rendah (misalnya, pada malam hari). Namun, pemulihan segera dapat dilakukan sebagai respons terhadap gangguan serius yang berdampak pada pelanggan.



Namun, dalam jangka pendek hingga menengah, pembelajaran mesin akan dilengkapi dan bahkan diserap oleh pengetahuan domain manusia, karena operator jaringan memberikan informasi yang digambarkan sebagai kebijakan (atau pada akhirnya sebagai maksud yang lebih abstrak) terkait tanda tangan yang menentukan peristiwa yang diinginkan dan tindakan pemulihan yang harus diambil sebagai respons terhadap kondisi jaringan dan layanan yang diketahui. Misalnya, jika operator mengetahui adanya bug perangkat lunak tertentu yang menyebabkan mesin virtual (VM) menjadi tidak responsif dan dapat diatasi hanya dengan me-reboot VM terkait, maka hal ini dapat dinyatakan oleh operator sebagai kebijakan seperti yang diilustrasikan pada Gambar 9.5.



Gambar 9.5 Contoh kebijakan loop kontrol khusus operator.

Seiring kecerdasan berbasis pembelajaran mesin menjadi semakin komprehensif, kebijakan berbasis pengetahuan domain manusia akan berkurang menjadi salah satu sumber data yang darinya kecerdasan mesin mengambil keputusan tindakan. Hal ini analog dengan seorang anak yang tumbuh dewasa dan "aturan" yang diberlakukan oleh orang tuanya mereda seiring waktu, digantikan oleh pengalaman dan penilaiannya sendiri.

Setiap kali tindakan pemulihan dicoba, sistem harus menentukan apakah tindakan tersebut efektif dalam menyelesaikan masalah yang dihadapi. Hal ini biasanya sama dengan menentukan apakah gejala yang memicu loop kontrol telah dihilangkan dan kinerja layanan baik setelah tindakan penyelesaian. Dalam beberapa layanan, seperti mobilitas, kinerja layanan bervariasi seiring waktu dan di berbagai lokasi. Dengan demikian, pembelajaran mesin digunakan untuk menentukan kinerja normal di lokasi dan waktu tertentu dan dapat dibandingkan dengan kondisi setelah tindakan pemulihan dilakukan untuk memvalidasi perbaikan yang berhasil.

Banyak tindakan perbaikan dapat dijalankan sepenuhnya tanpa keterlibatan manusia misalnya, fungsi jaringan (fisik atau virtual) dapat di-boot ulang dan konfigurasinya diperbarui. Namun, VNF menghadirkan peluang yang sama sekali baru, di mana VNF baru dapat dijalankan dengan cepat dan lalu lintas dimigrasikan untuk segera mengatasi kondisi kegagalan. Mencontoh penyedia cloud, ketika gangguan atau kegagalan terdeteksi pada VNF tertentu, VNF pengganti dapat dijalankan dengan cepat dan lalu lintas dimigrasikan ke sana. Misalnya, pertimbangkan router tepi penyedia virtual (vPE) VNF yang berada di tepi jaringan penyedia layanan dan terhubung dengan pelanggan internet. Jika vPE gagal, atau gangguan terdeteksi,



kita dapat menjalankan vPE baru di server yang berbeda, lalu memigrasikan pelanggan ke vPE baru ini. Untuk mencapai hal ini, konfigurasi vPE baru harus sesuai dengan pengaturan fungsional vPE yang gagal, seperti Jaringan Area Lokal Virtual (VLAN), daftar kontrol akses (ACL), dan konfigurasi kualitas layanan (QoS). Hal ini juga mengharuskan lalu lintas dialihkan ke vPE baru. Demikian pula, jika fungsi jaringan mengalami kemacetan meskipun kemacetan ini tidak terduga dan mungkin disebabkan oleh bug perangkat lunak yang menyebabkan kondisi kelebihan beban yang keliru meningkatkan kapasitas VNF tambahan secara cepat mungkin merupakan cara yang sederhana dan sangat cepat untuk memitigasi dampak terhadap pelanggan, meskipun akar penyebab sebenarnya dari masalah tersebut tidak diketahui.

Dengan demikian, fungsi jaringan virtual membuka peluang baru untuk memikirkan kembali pendekatan tradisional, intensif secara manual, dan seringkali menyakitkan dalam memecahkan masalah jaringan. Kemampuan untuk memunculkan contoh baru dari fungsi jaringan yang gagal atau terganggu menjanjikan untuk menyederhanakan atau bahkan menghilangkan proses pemecahan masalah sebelum perbaikan jaringan dan layanan. Daripada menghabiskan banyak waktu untuk mencoba melokalisasi masalah dan menentukan akar penyebabnya, dengan pelanggan yang berpotensi mengalami dampak layanan, kita dapat dengan mudah memunculkan contoh baru atau tambahan dari fungsi jaringan dalam upaya memulihkan layanan. Hal ini dapat menghemat waktu berjam-jam dari tim besar dan dampak terkait terhadap pelanggan, serta sejalan dengan cara penyedia cloud tradisional mengelola layanan. Namun, ini biasanya merupakan perubahan pola pikir yang cukup besar bagi penyedia layanan jaringan!

Demikian pula, kemampuan untuk menjalankan VNF baru sebagai respons terhadap kegagalan perangkat keras menghilangkan kebutuhan akan pengiriman cepat untuk mengganti perangkat keras yang rusak, sehingga sangat menyederhanakan dan merampingkan operasi lapangan. Sebaliknya, kapasitas cloud yang rusak dapat diganti secara berkala misalnya, bulanan tanpa tekanan waktu dan pengiriman teknisi di luar jam kerja yang mahal.

Namun, tidak semua gangguan jaringan dapat ditangani tanpa intervensi waktu nyata di lokasi fisik. Pemutusan serat optik dan kegagalan fungsi jaringan fisik kemungkinan besar memerlukan perhatian di lokasi jika perangkat keras dan kapasitas yang tersedia terbatas, pelanggan mungkin terdampak hingga perbaikan di lokasi selesai. Hal ini memberikan tekanan waktu pada pengiriman untuk pemulihan. Di masa mendatang, pengiriman tidak akan selalu membutuhkan manusia. Dengan munculnya drone, mobil otonom, dan bahkan robot, kami membayangkan dunia di mana sumber daya dapat dikirim dan bahkan diintegrasikan ke dalam jaringan dengan otomatisasi penuh. Misalnya, lokasi seluler di atas roda atau lokasi seluler di atas sayap (drone) dapat dikirim secara otonom untuk menangani pemadaman listrik, skenario bencana, dan untuk menyediakan peningkatan kapasitas untuk acara-acara khusus. Jika tidak memungkinkan, teknisi manusia akan melengkapi pengiriman otonom.

Terlepas dari apakah pengiriman diotomatisasi atau tidak, otomatisasi dan pembelajaran mesin tetap penting dalam mengelola pengiriman; analitik dalam platform otomatisasi menganalisis data jaringan untuk melokalisasi masalah dan mengidentifikasi



tindakan pemulihan yang tepat. Prioritas tindakan harus dilakukan berdasarkan dampak pelanggan masalah dengan dampak pelanggan yang tinggi harus segera ditangani, sementara masalah yang tidak berdampak pelanggan dapat ditangani dengan urgensi yang lebih rendah. Hal ini memerlukan kuantifikasi dampak pelanggan dari pemadaman dan masalah lainnya yang kompleksitasnya bervariasi dengan teknologi jaringan yang berbeda. Analitik pembelajaran mesin tingkat lanjut akan mengkuantifikasi dampak pelanggan di berbagai teknologi, sehingga penyedia layanan dapat menentukan kapan dan bagaimana memprioritaskan distribusi (dan penyelesaian masalah secara umum), dan agar penyedia layanan secara umum dapat memahami dampak pelanggan dari pemadaman jaringan.

Kebijakan yang terkait dengan otomatisasi memanfaatkan informasi tentang dampak terhadap pelanggan dan faktor-faktor lain untuk memutuskan tindakan apa yang perlu diambil, jika ada. Selain memicu pengiriman secara otomatis, AI yang terintegrasi ke dalam sistem akan mengoptimalkan semua kunjungan lapangan berdasarkan lokasi fisik, prioritas (berdasarkan dampak terhadap pelanggan), dan sumber daya yang tersedia. Otomatisasi kemudian akan bertanggung jawab untuk memantau efektivitas pengiriman tersebut pada skala waktu yang berbeda.

Pembahasan sebelumnya berfokus pada respons waktu nyata atau mendekati waktu nyata terhadap kondisi jaringan, dengan fokus pada perbaikan layanan bagi pelanggan. Namun, penting juga bagi kita untuk mengidentifikasi masalah yang terus berulang, dan, jika memungkinkan, menemukan cara untuk menghilangkannya secara permanen atau setidaknya mengurangi frekuensi kemunculannya. Misalnya, jika bug perangkat lunak mengharuskan fungsi jaringan di-boot ulang secara berkala, maka kita perlu memastikan bahwa vendor perangkat lunak memperbaiki bug terkait dan perangkat lunak diluncurkan di seluruh jaringan untuk menghilangkan masalah tersebut secara permanen. Masalah-masalah yang sangat serius tersebut diidentifikasi melalui tren jangka panjang dan analisis gangguan jaringan dan layanan. Pembelajaran mesin selanjutnya dapat menyusun informasi mengenai kondisi yang memicu masalah tersebut dan mengidentifikasi pola untuk diagnosis masalah dan perbaikan bug.

Visi Administrasi Jaringan Otonom

Manajemen lalu lintas, manajemen acara khusus, dan pemulihan bencana merupakan area yang secara historis cukup bergantung pada kecerdasan dan penilaian manusia. Operator manusia melaksanakan perencanaan terperinci untuk acara khusus dan respons bencana serta menentukan cara mengelola lalu lintas sebelum acara khusus dan dalam menanggapi situasi yang tidak biasa. Teknisi bertanggung jawab untuk mengarahkan peralatan dalam menanggapi bencana. Gambar 9.2 menggambarkan beberapa kompleksitas di mana peralatan harus diangkut ke lokasi yang terkadang sulit dijangkau (misalnya, pulau-pulau terpencil seperti Puerto Riko pada tahun 2017). Kecerdasan buatan memberikan potensi untuk mentransformasi beberapa fungsi ini.

Seiring kemajuan pembelajaran mesin, teknologi ini diharapkan tersedia untuk memungkinkan pengambilan keputusan dan pelaksanaan aktivitas yang mendukung manajemen lalu lintas secara otomatis. Kecerdasan canggih akan mampu mendeteksi dan



memprediksi lalu lintas anomali berdasarkan data jaringan dan sosial (misalnya, berita, media sosial) dan menentukan cara mengalihkan lalu lintas untuk mengelola beban. Otomatisasi jaringan akan membuat perubahan yang diperlukan pada fungsi jaringan.

Dalam visi jaringan yang sepenuhnya otonom, kalender komunitas, pusat konvensi, dan arena, serta media sosial akan ditambah secara otomatis untuk mengidentifikasi dan memprediksi acara sosial besar yang akan menyebabkan peningkatan beban lalu lintas yang signifikan di jaringan, dan untuk memprediksi besarnya peningkatan tersebut. Dengan menggunakan prediksi ini, intelijen jaringan akan secara otomatis menentukan kebutuhan kapasitas jaringan tambahan dan lokasi penempatan kapasitas. Intelijen jaringan kemudian akan secara otomatis mengirimkan kapasitas pada waktu yang relevan dan memantau serta bereaksi terhadap kondisi jaringan selama acara berlangsung, hingga akhirnya menonaktifkan kapasitas setelah acara selesai. Mekanisme umpan balik loop tertutup akan memungkinkan intelijen untuk terus mempelajari dan menyempurnakan prediksi acara dan kapasitas, berdasarkan pengukuran yang dilakukan selama acara tersebut.

Terakhir, SDN, virtualisasi, dan AI berpotensi mengubah proses pemulihan bencana. Pemulihan bencana akan jauh lebih mudah di dunia virtual dibandingkan dengan memulihkan aset jaringan fisik. Pusat data dapat dipulihkan dengan menempatkan kembali VNF secara dinamis ke satu atau beberapa pusat data yang berbeda hal ini memerlukan kemampuan untuk mencadangkan citra VNF dan menjalankan VNF baru secara dinamis di lokasi berdasarkan citra cadangan, atau kemampuan untuk membuat ulang citra VNF secara dinamis. Hal ini juga memerlukan kapasitas jaringan cadangan yang cukup untuk merutekan ulang lalu lintas dengan tepat. Jika kapasitas cadangan tidak tersedia di lokasi lain, pusat data di atas roda dapat digunakan secara lokal ke bencana dan VNF dimigrasikan sebagaimana mestinya. Kecerdasan buatan akan membuat keputusan sesuai dengan bagaimana dan di mana memigrasikan kapasitas. Pembelajaran mesin dan AI juga akan sangat penting untuk mengelola sumber daya fisik yang digunakan selama bencana, seperti COW.

Peralatan dapat digunakan terlebih dahulu sebelum bencana alam seperti badai, ditempatkan di lokasi yang dioptimalkan untuk menyederhanakan dan mempercepat penerapan fungsi jaringan selama dan setelah kejadian. Prakiraan cuaca, dampak pelanggan yang diprediksi dari pemadaman peralatan, dan informasi tentang masa pakai baterai dapat bersama-sama dimanfaatkan untuk mengoptimalkan di mana peralatan digunakan sebelum bencana alam, sementara dampak pelanggan yang terukur dan informasi peralatan dapat bersama-sama digunakan untuk memprioritaskan, merancang, dan mengoordinasikan penerapan setelah bencana terjadi.

9.4 MEWUJUDKAN VISI: PERJALANAN MENUJU MASA DEPAN

Saat ini, terdapat otomatisasi yang sangat besar yang tersedia di jaringan vital untuk memungkinkan tim operasi yang relatif kecil mengelola jaringan dan layanan internasional yang masif. Namun, terdapat pula intervensi manusia yang cukup besar yang diperlukan dalam semua aspek operasi, terutama dalam hal penjadwalan aktivitas, pemecahan masalah dan remediasi masalah jaringan, serta pengelolaan kejadian khusus dan bencana. Otomatisasi yang



tersedia saat ini memanfaatkan kombinasi beberapa sistem, skrip ad hoc, dan sangat bergantung pada aktivitas pengembangan. Seiring dengan terus berkembangnya skala dan kompleksitas layanan jaringan, serta seiring dengan tuntutan inovasi yang cepat dari pelanggan dan teknologi, sistem dan otomatisasi tradisional tidak akan mencukupi.

Dengan demikian, industri telekomunikasi sedang memulai perjalanan menuju sistem operasi jaringan berbasis SDN yang berbasis kebijakan. Hal ini akan menjadi fondasi bagi jaringan zero-touch. ONAP dari Linux Foundation adalah "platform komprehensif untuk orkestrasi dan otomatisasi fungsi jaringan fisik dan virtual berbasis kebijakan dan waktu nyata yang akan memungkinkan penyedia dan pengembang perangkat lunak, jaringan, TI, dan cloud untuk mengotomatiskan layanan baru dengan cepat dan mendukung manajemen siklus hidup yang lengkap. Dalam konteks ini, ONAP beserta resep, kebijakan, dan analitik terkait akan menyediakan fondasi bagi kita untuk mengotomatiskan penyediaan, pemeliharaan, dan administrasi jaringan.

Jenis-jenis VNF baru diintegrasikan ke dalam platform ONAP oleh perancang layanan yang bekerja sama erat dengan operator jaringan. Proses integrasi menentukan peran VNF, dan bagaimana instans baru VNF harus dikonfigurasi, dihubungkan ke fungsi jaringan lain (virtual atau fisik), dan siklus hidupnya dikelola. Ini mencakup penentuan bagaimana platform ONAP harus menginstansiasi instans baru VNF dan bagaimana platform ONAP harus merespons berbagai kondisi jaringan yang terkait dengan VNF. Spesifikasi awal ini dimodulasi oleh tim operasi setelah VNF yang baru dirancang diserahkan kepada mereka, berdasarkan pengalaman lapangan. Sekarang kita akan meninjau teknologi yang dibutuhkan dan perjalanan yang sedang kita tempuh untuk menciptakan jaringan berbasis ONAP yang memanfaatkan AI dalam perjalanan menuju jaringan zero-touch.

Mewujudkan Penyediaan Jaringan Otonom

Sebagaimana dibahas sebelumnya, terdapat banyak aset fisik yang dibutuhkan dalam jaringan masa depan server, fiber optik, dan menara seluler sebagai contoh. Bahkan router backbone yang besar mungkin tidak dapat divirtualisasikan dalam waktu dekat, hingga server dapat mencapai throughput yang dibutuhkan. Penyedia layanan juga dapat memilih, karena alasan bisnis lain, untuk hanya memvirtualisasikan sebagian, tetapi tidak semua, elemen jaringan mereka. Aset fisik semacam itu membutuhkan waktu tunggu yang lama dan dengan demikian perencanaan yang berwawasan ke depan; perangkat perencanaan kapasitas diperlukan untuk mendukung hal ini.

Perencanaan kapasitas menggunakan desain jaringan, prakiraan lalu lintas, dan kendala lainnya sebagai masukan untuk menentukan di mana akan menerapkan kapasitas di masa mendatang. Dengan asumsi bahwa desain jaringan dasar sudah ada, maka proses menuju perencanaan kapasitas jaringan yang akurat terbagi menjadi beberapa komponen utama estimasi atau pengukuran matriks permintaan lalu lintas dan pemodelan serta perkiraan pertumbuhan digunakan bersama-sama untuk membuat perkiraan lalu lintas, sementara algoritma desain dan pengoptimalan perutean jaringan membentuk dasar perkakas perencanaan kapasitas. Kecerdasan buatan yang menyederhanakan seluruh proses



dapat dengan mudah mengambil data pengukuran dari jaringan operasi dan menghasilkan rencana peningkatan kapasitas yang akan dieksekusi.

Mari kita mulai dengan memperkirakan permintaan lalu lintas, dengan fokus di sini pada jaringan IP. Kemampuan kita untuk memperkirakan matriks permintaan lalu lintas secara akurat telah berkembang secara signifikan seiring dengan semakin matangnya teknologi pengukuran jaringan dan infrastruktur data. Pengukuran volume lalu lintas berbasis *Simple Network Management Protocol* (SNMP) tidak memberikan informasi sumber dan tujuan arus jaringan. Oleh karena itu, untuk menghitung matriks lalu lintas diperlukan penggabungan pengukuran SNMP dan matriks perutean jaringan, yang dapat diturunkan dari pengukuran bidang kontrol (alias perutean) dan penyelesaian persamaan linear yang kurang dibatasi. Hal ini cenderung menyebabkan kesalahan estimasi yang tinggi.

Jika pengukuran berbasis aliran seperti NetFlow tersedia, matriks lalu lintas dapat dihitung secara langsung hanya dengan kesalahan statistik yang kecil (karena proses pengambilan sampel aliran). Disempurnakan dengan pengukuran inspeksi paket mendalam, matriks lalu lintas sumber-tujuan tradisional dapat didekomposisi lebih lanjut menjadi matriks lalu lintas komponen yang berbeda berdasarkan kelas aplikasi (seperti streaming video, web, dan *Internet-of-Things*) dan perangkat pengguna (seperti ponsel pintar, tablet, dan PC), yang akan memungkinkan pemodelan dan perkiraan pertumbuhan yang lebih rinci. Pemodelan pertumbuhan lalu lintas melibatkan tren permintaan lalu lintas dari waktu ke waktu dan menyesuaikannya dengan model data, misalnya, memperkirakan tingkat pertumbuhan tahunan gabungan. Karena berbagai kelas lalu lintas tumbuh pada tingkat yang berbeda, pemodelan pertumbuhan yang lebih rinci mencapai akurasi perkiraan yang lebih baik.

Memetakan permintaan lalu lintas ke rencana kapasitas jaringan memerlukan penyelesaian simulasi perutean dan masalah optimasi. Untuk jaringan skala besar, kompleksitas untuk melakukan hal ini jauh melebihi kapasitas pemrosesan kecerdasan manusia, bahkan untuk perencana jaringan yang paling berpengalaman sekalipun. Meskipun menemukan rute jalur terpendek atau, ekuivalennya, desain jaringan berbiaya terendah di antara kumpulan kandidat node dan jalur merupakan masalah yang dapat diatasi, kendala praktis seperti peningkatan kapasitas yang berada dalam unit bandwidth diskrit dengan cepat mengubah masalah tersebut menjadi optimasi kombinatorial, yang sangat meningkatkan kompleksitas komputasinya. Pertimbangan tambahan semakin mempersulit analisis perencanaan kapasitas. Salah satu contohnya adalah merancang ketahanan memperhitungkan kegagalan tautan dan node. Model probabilistik yang canggih dapat memperhitungkan tingkat kegagalan komponen tingkat kegagalan tersebut dapat diukur dari jaringan produksi dan dimodelkan berdasarkan data historis. Contoh lain adalah merancang jaringan yang tangguh dan tahan terhadap kesalahan peramalan permintaan lalu lintas yang inheren. Alat perencanaan kapasitas telah menjadi semakin cerdas selama bertahun-tahun, menghasilkan rencana kapasitas yang lebih efisien dengan lebih sedikit masukan manusia.

Namun, alat perencanaan kapasitas tradisional cenderung memperhitungkan satu lapisan pada satu waktu, yang menyebabkan inefisiensi di seluruh lapisan jaringan. Optimasi lintas lapisan, yang menggabungkan beberapa lapisan jaringan (misalnya, mengoptimalkan



lapisan jaringan optik, IP, dan mobilitas secara bersamaan) meningkatkan efisiensi pemanfaatan sumber daya, sehingga meningkatkan latensi jaringan dan belanja modal. Optimasi lintas lapisan bisa sangat menantang. Untuk itu, algoritma baru dan implementasinya yang cermat diperlukan.

Selain peningkatan kapasitas jaringan fisik, perencanaan dan optimasi kapasitas untuk sumber daya fisik lainnya, termasuk komputasi, penyimpanan, memori, daya, dan pendinginan, menghadirkan tantangan dan peluang serupa bagi kecerdasan berbasis pembelajaran mesin. Prinsip yang sama berlaku melakukan pengukuran yang sangat teliti (misalnya, membagi pengukuran CPU berdasarkan fungsi jaringan), membangun model peramalan multivariat dengan memperhitungkan tren, musim, dan faktor bisnis eksternal (misalnya, memodelkan beban kerja untuk streaming video selama liburan), dan melakukan analisis dan optimasi what-if (misalnya, optimasi rak dan optimasi ketahanan).

Berbeda dengan aset fisik yang membutuhkan waktu penerapan yang lama dan dengan demikian peramalan serta perencanaan yang berwawasan ke depan, VNF dapat digulirkan sesuai permintaan, sehingga kapasitas baru dapat diterapkan kapan pun dan di mana pun dibutuhkan dengan asumsi terdapat kapasitas cloud yang mendasarinya sebagaimana dibahas pada paragraf sebelumnya. Hal ini merupakan terobosan dalam hal mewujudkan jaringan otonom tanpa sentuhan mengurangi biaya overhead operator dan mempercepat respons terhadap perubahan permintaan jaringan.

Perancangan dan perencanaan VNF kemudian dapat menjadi proses pengambilan keputusan secara real-time dan tidak memerlukan waktu tunggu yang panjang serta peramalan permintaan lalu lintas jangka panjang yang diperlukan untuk rekan fisiknya. Namun, kompleksitas yang terkait dengan optimasi sumber daya tetap ada. VNF terdiri dari satu atau lebih grup sumber daya dan VM, sebagaimana dijelaskan dalam templat OpenStack Heat. Penempatan VNF yang efisien pada sumber daya fisik memerlukan penyelesaian masalah bin-packing multidimensi yang kompleks. Sebagai solusinya, terdapat beberapa algoritma kecerdasan mesin untuk mengidentifikasi solusi efisien untuk kelas masalah ini, dengan atau tanpa batasan afinitas penempatan tambahan dan/atau opsi untuk memigrasikan VNF yang sudah diterapkan. ONAP secara otomatis memasukkan optimasi penempatan VNF dalam keputusan instansiasi VNF-nya.

Fungsi jaringan harus saling terhubung untuk memungkinkan layanan menyeluruh. Hal ini mengharuskan kapasitas antar kantor dibangun secara otomatis jika kita ingin mewujudkan visi jaringan tanpa sentuhan; teknologi seperti Ethernet (untuk konektivitas kecepatan rendah antar lokasi) dan ROADM menyediakan kapasitas "sesuai permintaan". Setelah kapasitas jaringan diterapkan, kita perlu mengonfigurasi, menguji, dan kemudian mengaktifkannya ke dalam jaringan. Sebagian besar konfigurasi, pengujian, dan pengaktifan elemen jaringan lama secara historis telah diotomatisasi hingga tingkat yang berbeda-beda pada penyedia layanan skala besar. Namun, ONAP menyediakan platform untuk menghadirkan mekanisme yang konsisten dan terintegrasi untuk mencapai fungsi-fungsi ini dan dapat memungkinkan pergerakan ke proses yang sepenuhnya otomatis.



Baik fisik maupun virtual, fungsi jaringan harus dikonfigurasi secara logis dan kemudian divalidasi sebelum dapat digunakan untuk membawa lalu lintas dan memberikan layanan. Konfigurasi mencakup hal-hal seperti menentukan identitas fungsi jaringan (misalnya, alamat IP), menentukan protokol perutean yang akan digunakan untuk berkomunikasi dengan fungsi jaringan lain, dan mengonfigurasi kebijakan fungsi jaringan, seperti lalu lintas apa yang akan dibawa atau diblokir. Salah satu tantangan terbesar yang dihadapi oleh operator jaringan adalah keragaman berbagai jenis fungsi jaringan dan perbedaan yang sangat besar dalam cara konfigurasi dan pengoperasiannya. Salah satu tujuan utama upaya ONAP dan seterusnya adalah untuk menyediakan antarmuka yang konsisten di berbagai fungsi jaringan dari sudut pandang operator jaringan. Standar seperti NetConf dan RestConf semakin banyak diadopsi di seluruh vendor fungsi jaringan, yang mempercepat konvergensi ke antarmuka konfigurasi yang konsisten.

Model data melalui antarmuka konfigurasi dan manajemen dijelaskan dalam bahasa pemodelan, seperti *Yet Another Next Generation* (YANG). Terdapat dua set model yang banyak digunakan dan didukung. *Internet Engineering Task Force* (IETF) menstandardisasi model data untuk setiap protokol, untuk fungsionalitas perutean dasar, dan untuk persyaratan manajemen peralatan umum. Grup OpenConfig mengelola set model data lain, yang sebagian besar tumpang tindih, dan sering kali terkoordinasi dengan model data IETF. Selain model-model ini, setiap vendor juga mendukung set model khusus vendor dan peralatan yang sering kali dapat diunduh dari situs dukungan mereka. Penggunaan model YANG ini memungkinkan pengontrol manajemen berbasis model seperti yang ada di ONAP.

Setelah konfigurasi selesai, "uji dan aktifkan" memvalidasi bahwa fungsi jaringan siap mendukung lalu lintas dan kemudian memperbarui konfigurasi jaringan untuk mulai merutekan lalu lintas melalui fungsi jaringan baru. "Uji dan aktifkan" dicapai melalui serangkaian pengujian yang dirancang untuk memvalidasi konfigurasi dan kesehatan fungsi jaringan serta layanan yang melewatinya. Setiap fungsi jaringan akan memerlukan pengujian yang berbeda, tergantung lokasinya di jaringan, peran spesifik yang dijalankan oleh fungsi jaringan, dan layanan yang dibawanya. Awalnya, serangkaian pengujian yang akan dilakukan akan ditentukan oleh operator jaringan ditetapkan sebagai maksud atau kebijakan dalam sistem ONAP. Seiring dengan semakin matangnya industri, pembelajaran mesin dapat mengambil alih peran ini secara otomatis mempelajari pengujian apa yang harus dilakukan berdasarkan informasi yang diperoleh dari jaringan, desain layanan, dan pengalaman sebelumnya terkait pemantauan yang efektif.

Pembahasan sebelumnya berfokus pada pengenalan kapasitas baru ke jaringan. Namun, seiring pertumbuhan jaringan secara bertahap, jaringan biasanya menjadi suboptimal. Secara historis, dengan aset jaringan fisik, menata ulang jaringan untuk mengoptimalkan ulang penggunaan kapasitas sangatlah mahal dan memakan waktu. Namun, mengingat VNF dirancang agar mudah dimigrasi, kini kita memiliki peluang untuk mengoptimalkan ulang kapasitas secara berkala. Selain kemampuan untuk memigrasikan VNF dan lalu lintas terkaitnya secara mulus yang dikelola oleh ONAP, algoritma reoptimasi yang canggih akan diperlukan untuk memutuskan kapan dan di mana kapasitas akan dimigrasikan.



Demikian pula, kondisi dapat terjadi ketika lalu lintas menurun baik sementara maupun permanen di suatu lokasi. Dalam situasi seperti itu, VNF dapat diturunkan atau dikurangi ukurannya, dan sumber daya fisik dapat diturunkan atau dikurangi kapasitasnya. Wilayah dengan permintaan musiman seperti lokasi liburan musim panas yang populer mungkin menunjukkan permintaan lalu lintas yang sangat tinggi selama satu musim (misalnya, musim panas), tetapi permintaan rendah untuk sisa tahun tersebut. Dalam situasi seperti itu, kapasitas jaringan dapat dikurangi, dimatikan sementara, atau beberapa sumber daya jaringan (misalnya, operator di lokasi seluler) ditiadakan. Pembelajaran mesin berperan penting dalam hal ini untuk mempelajari kondisi lalu lintas normal, sehingga keputusan dapat diambil terkait kapasitas yang dapat dihilangkan sementara dari jaringan untuk menghemat biaya operasional dan/atau energi (dengan manfaat lingkungan terkait).

Mewujudkan Pemeliharaan Jaringan

Setelah fungsi jaringan diterapkan, fungsi tersebut perlu dikelola. Selanjutnya, kami akan mengkaji pemeliharaan jaringan dan bagaimana perkembangannya menuju visi jaringan otonom.

Pemeliharaan Terencana

ONAP menyediakan cara yang konsisten untuk mengotomatiskan proses manajemen perubahan. Perancang layanan dan/atau operator jaringan merancang aktivitas perubahan dengan membangun alur kerja perubahan untuk aktivitas perubahan tertentu. Menjalankan aktivitas perubahan untuk fungsi jaringan tertentu biasanya melibatkan penentuan terlebih dahulu apakah status jaringan saat ini "aman" untuk melanjutkan pemeliharaan, kemudian menjalankan serangkaian pengujian (pra-pemeriksaan) untuk menangkap konfigurasi dan status fungsi jaringan pra-pemeliharaan, serta kinerja layanan yang akan digunakan sebagai dasar perbandingan setelah aktivitas pemeliharaan. Aktivitas perubahan kemudian dijalankan, dan pasca-pemeriksaan dilakukan untuk memverifikasi bahwa perubahan telah berhasil dijalankan tanpa konsekuensi negatif yang tidak diharapkan. Pemeriksaan pasca-operasi semacam itu perlu dilakukan di seluruh layanan misalnya, jika pemeliharaan dilakukan pada router, maka semua layanan yang dijalankan melalui router tersebut (misalnya, akses internet, layanan mobilitas, *Voice over LTE* [VoLTE], dan layanan TV) harus dipantau dan divalidasi. Jika muncul masalah, seperti fungsi jaringan yang gagal muncul setelah di-boot ulang atau indikator kinerja utama (KPI) menurun, tindakan yang tepat harus dilakukan. Rangkaian tindakan yang harus dilakukan untuk menjalankan perubahan, pengujian apa yang harus dilakukan sebelum dan sesudah aktivitas untuk mendeteksi masalah, dan bagaimana masalah harus diperbaiki jika terjadi, semuanya tertuang dalam resep, yang "direkatkan" sesuai keinginan menggunakan antarmuka grafis.

Setelah aktivitas pemeliharaan dirancang, aktivitas tersebut perlu dijadwalkan. Perancang dan/atau operator layanan menentukan serangkaian fungsi jaringan tempat perubahan tertentu akan dijalankan, dan kebijakan lain yang mengatur bagaimana fungsi-fungsi tersebut harus dijadwalkan. Penjadwalan perubahan kemudian menjadi proses otomatis, yang memanfaatkan algoritma optimasi dan heuristik yang diimplementasikan dalam kerangka kerja optimasi ONAP.



Akhirnya, eksekusi perubahan diotomatisasi melalui ONAP, mengeksekusi alur kerja perubahan yang telah ditentukan oleh perancang layanan pada rangkaian fungsi jaringan sesuai dengan jadwal yang telah dihitung. Eksekusi ini biasanya juga akan menangani dampak otomatisasi, ketika aktivitas perubahan gagal. Teknik pemulihan yang umum adalah dengan mengembalikan perubahan (misalnya, kembali ke perangkat lunak lama selama prosedur pembaruan perangkat lunak).

Kecerdasan buatan meresap ke dalam proses manajemen perubahan mulai dari algoritma optimasi yang digunakan untuk menghitung dan menghitung ulang jadwal hingga eksekusi dan validasi efektivitas aktivitas perubahan. Salah satu area yang sangat menantang yang ditangani oleh AI adalah penjadwalan perubahan. Aktivitas pemeliharaan yang berpotensi berdampak pada pelanggan biasanya dilakukan selama periode pemeliharaan biasanya pada malam hari ketika lalu lintas rendah sehingga dampaknya terhadap pelanggan rendah. Hal ini memberikan periode terbatas di mana aktivitas dapat dijadwalkan; dengan topologi jaringan yang kompleks, beberapa aktivitas pemeliharaan yang berbeda membuat perubahan pada fungsi jaringan, dan dengan skala jaringan yang besar, penjadwalan semacam itu dapat menjadi rumit.

Dampak terhadap pelanggan harus diminimalkan oleh karena itu, koordinasi lintas fungsi jaringan yang berbeda jenis dan lintas lapisan jaringan sangat penting untuk memastikan jalur layanan yang redundan tidak terdampak secara bersamaan. Ketika aktivitas dengan prioritas lebih tinggi muncul, atau ketika aktivitas pemeliharaan berjalan tidak semestinya atau gangguan jaringan mengganggu aktivitas pemeliharaan, jadwal perlu dikerjakan ulang, yang menambah kompleksitas. Penjadwalan semacam itu merupakan masalah koordinasi yang kompleks untuk menentukan waktu yang tersedia bagi aktivitas pada fungsi jaringan tertentu di seluruh aktivitas perubahan. Algoritme untuk menjadwalkan aktivitas perubahan diimplementasikan dalam ONAP di atas kerangka kerja optimasi algoritme ini memperhitungkan jalur layanan dan kendala lainnya untuk memastikan bahwa aktivitas perubahan yang dijalankan secara bersamaan pada fungsi jaringan yang berbeda tidak menyebabkan dampak yang tidak perlu terhadap pelanggan.

Efektivitas aktivitas perubahan idealnya dipantau dalam berbagai skala waktu; pengujian segera sebelum dan sesudah aktivitas pemeliharaan memeriksa masalah signifikan yang mudah terdeteksi. Namun, beberapa masalah tidak muncul hingga beban lalu lintas meningkat biasanya beberapa jam atau lebih setelah aktivitas perubahan. Dengan demikian, dampak perubahan pada metrik layanan dan jaringan harus dipantau selama periode waktu setelah pemeliharaan, dan dibandingkan dengan kinerja sebelum aktivitas. Analisis berbasis pembelajaran mesin tingkat lanjut seperti Mercury memperhitungkan kinerja spesifik waktu dan lokasi untuk mengukur peningkatan kinerja layanan dan fungsi jaringan yang diinginkan serta menangkap konsekuensi yang tidak diinginkan dari peningkatan tersebut. Keluaran dari validasi pasca-perubahan langsung dan analisis Mercury yang lebih canggih mendorong keputusan apakah tindakan perbaikan atau pemulihan (misalnya, pengembalian) diperlukan dan kapan tindakan tersebut harus dilakukan.



Sebelum meluncurkan perubahan umum di seluruh jaringan, seperti menerapkan peningkatan perangkat lunak atau patch ke semua lokasi seluler di seluruh jaringan, operator jaringan biasanya menjalankan fase aplikasi lapangan pertama (FFA), di mana sebagian perangkat dipilih untuk menerapkan perubahan. Mendeteksi degradasi tak terduga yang sangat halus sekalipun selama FFA sangat penting untuk mencegah dampak layanan dalam skala yang lebih luas. Algoritma pembelajaran mesin canggih seperti yang digunakan dalam Mercury jangka panjang menyediakan kecerdasan untuk melihat berbagai contoh aktivitas perubahan umum guna mengidentifikasi dampak persisten di tengah variasi normal kinerja layanan.

Kecerdasan yang lebih canggih dapat membedakan dampak kinerja layanan akibat faktor eksternal yang independen dari perubahan yang diterapkan versus dampak yang intrinsik terhadap perubahan tersebut. Hal ini dicapai dengan mengungkap struktur regresi di antara contoh-contoh yang berada di dalam dan di luar penerapan FFA. Struktur regresi kemudian digunakan sebagai dasar untuk "memprediksi" indikator kinerja layanan selama interval waktu pasca-perubahan dan di lingkungan pasca-perubahan jika perubahan yang direncanakan tidak terjadi. Akhirnya, dengan membandingkan kinerja yang diprediksi dan diamati, analitik mengkuantifikasi dampak layanan yang dihasilkan dari perubahan yang diterapkan.

Jika degradasi layanan tak terduga terdeteksi selama FFA, penerapan perubahan biasanya akan berhenti atau kembali. Sementara itu, semua data selama FFA akan dianalisis untuk investigasi akar penyebabnya. Pembelajaran mesin berguna dalam mengidentifikasi karakteristik umum dari kejadian-kejadian di mana degradasi layanan diamati, seperti pengaturan konfigurasi umum atau pola lingkungan atau lalu lintas umum. Dimensionalitas ruang karakteristik semacam itu biasanya terlalu besar untuk dieksplorasi dan dinalar secara efektif oleh kecerdasan manusia. Dalam proses sebelumnya, penerapan FFA dan investigasi masalah dilakukan dalam langkah-langkah iteratif dan terpisah. Namun, sistem AI seperti yang dijelaskan dalam dapat secara bersama-sama mengoptimalkan dan mengorkestrasi perencanaan FFA dan analisis masalah yang diidentifikasi selama FFA. Hal ini dicapai dengan meluncurkan batch penerapan FFA berikutnya secara adaptif di lokasi-lokasi yang dianggap lebih mungkin untuk mendeteksi degradasi layanan yang tidak terduga.

Lokasi-lokasi tersebut ditentukan berdasarkan karakteristik yang dipelajari dari penerapan FFA sebelumnya yang memvalidasi perubahan yang sama. Misalnya, jika peluncuran awal perangkat lunak baru ke stasiun pangkalan seluler mengamati sedikit peningkatan panggilan terputus di lokasi-lokasi dengan lalu lintas rendah dan mobilitas tinggi (misalnya, di sepanjang jalan raya), sistem AI dapat memilih untuk menerapkan perangkat lunak baru di lokasi-lokasi dengan lalu lintas rendah dan mobilitas tinggi tambahan untuk mengonfirmasi atau menyangkal observasi tersebut. Sistem AI semacam itu juga akan mengatur peluncuran perubahan apa pun di seluruh jaringan dan secara cerdas membuat keputusan tentang kapan dan di mana akan menerapkan lebih banyak perubahan dan apakah akan menghentikan dan mengembalikan penerapan sebelumnya.



Pemeliharaan Reaktif

Fungsi jaringan menghasilkan sejumlah besar data alarm, syslog, log alur kerja, pengukuran kinerja, dan banyak lagi. Pengukuran tingkat layanan juga dicatat mulai dari catatan panggilan dan transaksi individual, hingga metrik yang dikumpulkan di berbagai tingkat, termasuk berdasarkan status panggilan, lokasi, jenis layanan, model perangkat pelanggan, dan faktor lainnya. Terdapat kekayaan data yang sangat besar sangat banyak jika tidak dikelola dengan cermat. Sebagian besar tantangan dalam pemeliharaan reaktif adalah mendeteksi masalah penting sekaligus meminimalkan alarm palsu, yang secara historis membuang-buang sumber daya operator dan di masa mendatang dapat menyebabkan tindakan yang tidak perlu dilakukan. Memproses data jaringan yang melimpah untuk menemukan jarum di tumpukan jerami masalah yang perlu diselidiki dan akar penyebabnya adalah peran sistem deteksi kejadian dan korelasi alarm.

Deteksi kejadian mungkin sesederhana mendengarkan peringatan dari fungsi jaringan misalnya, "tautan terputus" dari fungsi jaringan merupakan indikator yang jelas dan sederhana. Gangguan kinerja dapat dideteksi dengan mendengarkan KPI yang dihasilkan oleh fungsi jaringan dan menentukan kapan KPI berada di atas/di bawah ambang batas yang telah ditentukan atau mengalami penurunan yang tidak normal. Penurunan layanan diidentifikasi dengan melihat pengukuran layanan teragregasi dalam berbagai dimensi (berdasarkan jenis layanan, lokasi, dll. yang berbeda) dan dengan cara yang sama mengidentifikasi kondisi penurunan yang tidak normal. Misalnya, kita mungkin ingin membunyikan alarm ketika latensi jaringan IP telah melampaui nilai normalnya atau ketika throughput dalam lingkungan seluler telah turun dibandingkan dengan normal untuk lokasi dan waktu tertentu. Latensi dalam jaringan IP bervariasi berdasarkan lokasi, tergantung pada seberapa jauh jaraknya.

Demikian pula, throughput dalam jaringan seluler bervariasi berdasarkan lokasi dan juga waktu. Dengan demikian, kita tidak dapat menggunakan nilai ambang batas sederhana untuk menentukan kapan sesuatu tidak beres dan sebaliknya harus menggunakan mekanisme deteksi anomali yang lebih canggih untuk menentukan kapan sesuatu telah berubah secara negatif dan dengan demikian harus dianalisis dan diperbaiki. Argus adalah kemampuan deteksi anomali deret waktu yang sangat canggih yang digunakan untuk mendeteksi kondisi jaringan dan layanan yang anomali dan dapat ditindaklanjuti.

Analisis deteksi anomali tingkat lanjut sering kali memanfaatkan modul pembelajaran bawaan untuk mengekstrak profil normal secara otomatis dari data historis sehingga meringankan operator jaringan manusia dari kerumitan menentukan dan menetapkan kondisi anomali yang perlu diperingatkan. Profil normal tersebut dapat berupa deret waktu numerik univariat dari volume lalu lintas atau pengukuran kinerja dengan musim yang bervariasi berdasarkan lokasi, seperti profil yang dipelajari oleh Argus. Sebagai alternatif, profil normal dapat berupa pola struktural yang menangkap hubungan antara lokasi entitas yang berbeda dan di berbagai metrik, seperti yang ditangkap oleh matriks kovariansi dalam. Teknologi pembelajaran mesin adalah pendorong utama untuk analisis ini. Selain itu, mengintegrasikan pemahaman domain dalam analisis berbasis pembelajaran mesin ini biasanya dapat membantu meningkatkan efektivitasnya. Argus, misalnya, menyesuaikan pembelajarannya



untuk data kinerja jaringan dan layanan dengan secara eksplisit mengabaikan setiap perubahan pola jangka pendek yang signifikan (anomali) sambil beradaptasi dengan cepat terhadap perubahan pola yang berkelanjutan.

Desain algoritma ini didasarkan pada pemahaman bahwa data kinerja jaringan dan layanan biasanya mengandung anomali akibat kegagalan sementara, aktivitas pemeliharaan, atau permintaan lalu lintas flash crowd, dan status jaringan akan kembali normal setelahnya seolah-olah masalah sementara tidak terjadi. Sementara itu, peningkatan permanen atau perubahan rekayasa lalu lintas akan membawa jaringan ke status baru di mana riwayat sebelum perubahan pola tidak lagi bernilai di masa mendatang. Hal yang sama tidak berlaku untuk data keuangan, seperti volume perdagangan atau harga di pasar saham. Akibatnya, Argus telah terbukti menghasilkan hasil berkualitas tinggi (misalnya, tingkat alarm palsu yang rendah) dalam aplikasi operasi jaringan dibandingkan dengan teknik deteksi anomali deret waktu univariat tujuan umum lainnya.

Bisa dibilang, tugas paling kompleks yang terlibat dalam pengoperasian jaringan saat ini berkaitan dengan pemecahan masalah jaringan dan layanan yang kompleks untuk menentukan asal masalah, dan apa penyebabnya. Setiap masalah jaringan biasanya akan menghasilkan beberapa gejala yang perlu diwaspadai; tantangannya di sini adalah mengidentifikasi akar penyebab yang mendasarinya di antara serangkaian gejala yang diamati. Dengan beragamnya mode kegagalan, sejumlah besar jenis fungsi jaringan yang berbeda yang diterapkan di berbagai lokasi geografis, berbagai lapisan jaringan, dan perangkat lunak yang kompleks dengan interaksi yang lebih kompleks di seluruh fungsi jaringan, mengisolasi di mana masalah terjadi dan apa penyebabnya dua langkah yang sering kali terkait dengan tindakan pemulihan bisa sangat menantang.

Mari kita ilustrasikan hal ini dengan contoh sederhana di mana satu tautan antara dua router mengalami tingkat kesalahan bit yang tinggi; kita juga berasumsi bahwa tautan tersebut tidak dikonfigurasi untuk gagal pada tingkat kesalahan tersebut (seperti yang mungkin terjadi ketika redundansinya rendah). Mari kita asumsikan juga bahwa tautan ini menyediakan konektivitas transport antara situs sel mobilitas dan jaringan inti paket mobilitas, seperti yang digambarkan pada Gambar 9.6. Dalam skenario ini, router di dua ujung tautan fisik yang terganggu akan mendeteksi kesalahan dan membunyikan alarm yang sesuai.

Transmisi uji paket ujung ke ujung yang dieksekusi dalam jaringan IP juga akan membunyikan alarm, seperti halnya pengukuran aktif yang dilakukan pada segmen yang relevan dari jalur antara situs sel dan inti mobilitas. Dan akhirnya, pengukuran pelanggan juga dapat menangkap gangguan, dengan pengalaman pelanggan yang menurun dilaporkan di wilayah tersebut. Tanpa pemijatan (atau korelasi) yang tepat dari alarm ini, tim operasi akan dibanjiri oleh alarm bahkan untuk masalah yang paling sederhana, dengan banyak tim dan individu melihat ke dalam aspek yang berbeda dari apa yang secara inheren merupakan satu peristiwa yang mendasari (dalam hal ini, tautan yang terganggu antara dua router). Ini jelas tidak efisien, mengakibatkan pemborosan sumber daya operator, dan dapat memperpanjang dampak pelanggan.





Gambar 9.6 Contoh korelasi kejadian.

AI khususnya mesin korelasi, bertanggung jawab untuk menyusun alarm terkait menjadi satu peristiwa umum, sehingga dapat diproses sebagaimana mestinya. Mesin korelasi membentuk inti dari proses manajemen kesalahan dan kinerja yang diterapkan saat ini, mengurangi volume alarm berkali-kali lipat untuk menciptakan volume tiket yang dapat dikelola yang memerlukan perhatian operator. Selain mendeteksi dan mengisolasi masalah, kita perlu memahami dampak masalah terhadap pelanggan agar dapat memprioritaskan investigasi dan tindakan perbaikan. Analisis pembelajaran mesin tingkat lanjut dapat mengukur dampak masalah jaringan saat ini dan yang diprediksi terhadap pelanggan.

Kuantifikasi dampak pelanggan tersebut digunakan untuk memprioritaskan aktivitas perbaikan dan bahkan untuk memutuskan apakah pengiriman fungsi jaringan fisik diperlukan untuk jenis kondisi tertentu. TONA AT&T, misalnya, menggunakan analitik berbasis pembelajaran mesin tingkat lanjut dan data tingkat layanan (misalnya, pemblokiran panggilan, putusan, dan throughput) untuk menilai dampak kegagalan jaringan terhadap pelanggan dengan membandingkan kinerja yang diamati setelah kegagalan dengan kinerja yang diharapkan atau dipelajari sebelum kegagalan untuk menentukan tingkat keparahan masalah. AI dalam bentuk loop kontrol menyediakan mekanisme inti untuk memungkinkan pemeliharaan reaktif dalam ONAP. Gambar 9.7 menggambarkan loop kontrol sederhana pengumpul yang digunakan pada Mesin Pengumpulan dan Analisis Data (DCAE) ONAP mengumpulkan sejumlah besar data yang beragam dari fungsi jaringan dan layanan ujung ke ujung.

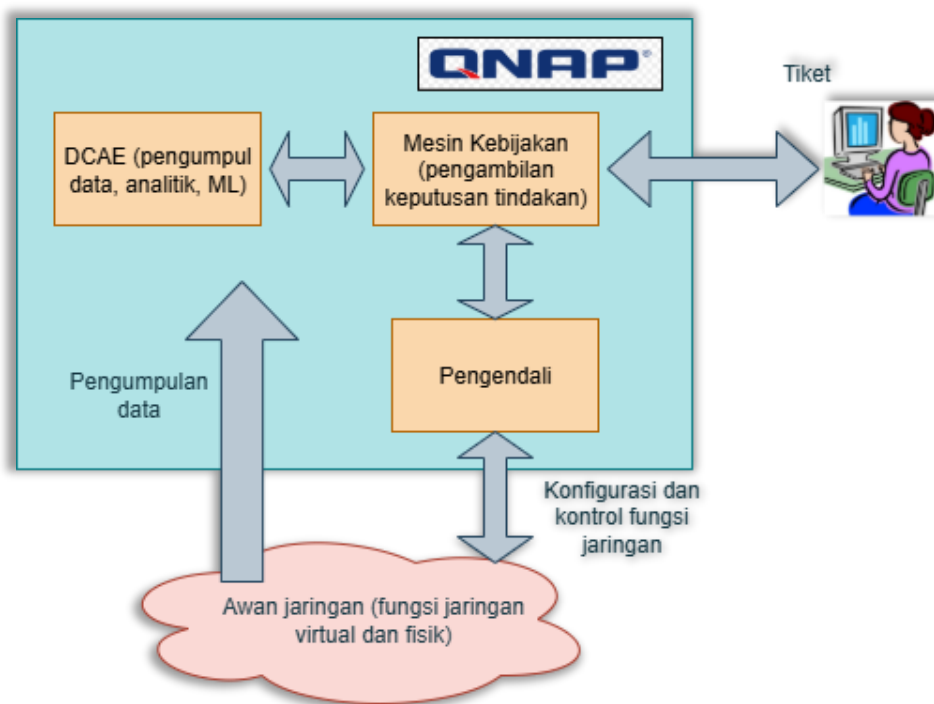
Analisis pada DCAE kemudian memproses data yang masuk untuk mendeteksi dan mengorelasikan peristiwa untuk melokalisasi masalah, dan untuk mengukur dampak masalah terhadap pelanggan. Analisis tersebut mencakup ambang batas sederhana, detektor anomali (misalnya), mesin korelasi, dan TONA. Peristiwa yang dilokalkan diteruskan ke kebijakan untuk memutuskan apa yang harus dilakukan dengannya; baik satu tindakan atau serangkaian tindakan dicoba melalui pengontrol aplikasi SDN atau masalah tersebut ditiketten untuk analisis dan pemulihan oleh manusia.

Operator manusia secara historis bertanggung jawab untuk menjalankan tindakan perbaikan peristiwa yang dikeluarkan dari mesin korelasi ditiketten dan operator jaringan manusia menyelidiki hal ini untuk mengidentifikasi akar penyebab masalah yang paling mungkin dan cara terbaik untuk memperbaiki masalah. Bahkan setelah suatu masalah diidentifikasi dan dilokalisasi, langkah-langkah pemulihan mungkin kurang terlihat operator mungkin perlu mencoba beberapa opsi berbeda sebelum akhirnya menyelesaikan masalah.



Sistem otomasi proses mengotomatiskan sejumlah besar pengumpulan data dan analisis peristiwa, mengisi tiket untuk mempercepat investigasi manusia secara signifikan. Teknologi pembelajaran mesin, seperti Mercury milik AT&T yang awalnya dirancang untuk mengevaluasi efektivitas pemeliharaan terencana juga dapat digunakan untuk menentukan apakah tindakan pemulihan efektif.

Namun, otomatisasi kini telah berkembang hingga mampu memperbaiki kondisi jaringan secara otomatis yang sebelumnya ditangani oleh operator manusia mengambil tindakan secara otomatis, seperti memigrasikan VNF yang gagal atau terganggu ke server baru, me-reboot VM, dan me-reboot fungsi jaringan fisik serta kartu jalur. Respons otomatis semacam itu merupakan bagian integral dari visi ONAP; operator menentukan kebijakan dalam ONAP yang menentukan tindakan otomatis apa yang harus dicoba dan kapan harus mengirimkan tiket untuk memulai investigasi manusia.



Gambar 9.7 Alur peristiwa yang disederhanakan pada ONAP.

Mesin kebijakan ONAP, seperti yang diilustrasikan pada Gambar 9.7, juga dapat memanfaatkan data tambahan yang dikumpulkan sesuai permintaan terkait kondisi yang diamati, membantu dalam memutuskan tindakan apa yang harus diambil. Dalam jaringan zero-touch yang sesungguhnya, semua tindakan perbaikan diotomatisasi dan tidak memerlukan intervensi manusia. Namun, visi semacam itu masih jauh saat ini dan mungkin tidak akan pernah sepenuhnya terwujud sehingga mesin kebijakan akan secara otomatis mengisi tiket untuk memulai analisis dan pemulihan manusia. Perancang layanan dan operator jaringan diharapkan bertanggung jawab untuk menentukan kebijakan yang mendefinisikan loop kontrol ONAP peristiwa yang akan dideteksi, korelasi yang akan dilakukan, tindakan pemulihan yang akan dicoba, dan kapan harus mengajukan tiket untuk investigasi manusia.



Namun, pengetahuan domain yang mendefinisikan loop kontrol ini tidak selalu tersedia; untuk fungsi jaringan yang matang, pengetahuan domain ini mungkin terdistribusi ke beberapa operator jaringan, dan untuk fungsi jaringan baru, pengetahuan ini mungkin tidak ada.

Ketika operator jaringan dengan cepat meluncurkan sejumlah besar jenis VNF yang berbeda dalam waktu singkat, mungkin hampir mustahil untuk menskalakan pengujian lab manual dan analisis terperinci yang secara historis dijalankan dan didokumentasikan untuk mengidentifikasi serangkaian kebijakan awal yang lengkap sebelum penerapan jaringan. Bahkan untuk fungsi jaringan yang matang, rilis perangkat lunak baru dan perubahan kondisi eksternal mungkin memerlukan pengenalan loop kontrol baru.

Pembelajaran mesin dapat membantu identifikasi kebijakan loop kontrol, menciptakan serangkaian kebijakan awal yang kuat untuk fungsi jaringan baru sebelum penerapan lapangan dan terus menyempurnakan dan mematangkan kebijakan untuk fungsi jaringan yang diterapkan di lapangan. Dengan demikian, pembelajaran kebijakan dapat dijalankan di lingkungan lab dan/atau berdasarkan data lapangan. Pembelajaran kebijakan untuk loop terbuka (tanpa tindakan perbaikan otomatis) berfokus pada peristiwa apa yang akan dideteksi, bagaimana peristiwa-peristiwa berbeda ini saling terkait (berkorelasi), dan mengidentifikasi akar penyebabnya (jika memungkinkan) sehingga mendefinisikan kebijakan untuk deteksi peristiwa dan aturan korelasi. Pembelajaran kebijakan untuk loop tertutup juga menggabungkan kebijakan pembelajaran untuk tindakan yang diambil secara otomatis.

Pembelajaran kebijakan untuk loop terbuka bertujuan untuk mengidentifikasi peristiwa yang secara konsisten terjadi bersamaan dan dengan demikian kemungkinan merupakan gejala dari peristiwa mendasar yang sama. Pembelajaran tanpa pengawasan berlaku di sini; label yang digunakan dalam pembelajaran terawasi akan setara dengan aturan korelasi yang ingin kita identifikasi dan dengan demikian akan mengalahkan tujuan penggunaan pembelajaran terawasi. Proses pembelajaran tanpa pengawasan bertujuan untuk mengurangi volume besar alarm, log, status jaringan, dan pengukuran kinerja menjadi peristiwa yang dapat ditindaklanjuti dengan menemukan pola umum dalam data.

Ada banyak algoritma dan teknologi untuk pengelompokan umum dan pengenalan pola, mulai dari Penambangan Aturan Asosiasi prosedural hingga jaringan saraf dalam AutoEncoder. Namun, topologi jaringan dan jalur layanan merupakan kunci korelasi peristiwa; tantangan unik dalam pembelajaran kebijakan loop terbuka adalah mengintegrasikan jalur jaringan dan layanan ke dalam pembelajaran dengan membangun masalah pengenalan pola deret waktu berbasis grafik. Model pengetahuan domain dan representasi data semacam itu juga penting untuk pembelajaran kebijakan loop tertutup. Pola deret waktu berbasis grafik dapat diperiksa oleh operator dan secara otomatis diterjemahkan ke dalam aturan korelasi ONAP (kebijakan). Teknik inferensi kausal, seperti uji kausalitas Granger, dapat diterapkan lebih lanjut untuk mendapatkan pemahaman tentang lokalisasi dan pemecahan masalah peristiwa.

Efektivitas pembelajaran tanpa pengawasan meningkat seiring dengan skala set data. Untuk fungsi jaringan yang telah diterapkan di lapangan, data pengujian yang digerakkan secara manual dari lingkungan lab dapat dibatasi, baik dalam volume maupun keragaman. Hal ini dapat diatasi dengan sistem otomasi yang secara sistematis atau acak menyuntikkan



kesalahan atau mensimulasikan kegagalan dan kondisi yang merugikan, seperti peningkatan permintaan lalu lintas yang tiba-tiba, di lingkungan tersebut. Eksperimen terkontrol yang berulang kemudian dapat menyediakan data yang cukup untuk mem-bootstrap pembelajaran mesin. Tindakan pengendalian injeksi kegagalan harus ditangkap ini memberi kita akar penyebab kegagalan yang diketahui, yang juga dapat digunakan oleh mesin pembelajaran untuk menyempurnakan tanda korelasi dan bahkan tindakan perbaikan yang tepat untuk loop tertutup.

Kebijakan loop tertutup tanda tangan dan tindakan yang diambil untuk mendeteksi dan secara otomatis memperbaiki suatu masalah dapat dipelajari baik di lingkungan laboratorium, maupun dengan "memantau" tindakan yang diambil oleh operator manusia di lapangan. Teknik pembelajaran mesin terawasi sangat cocok untuk pembelajaran dari operator jaringan guna memperoleh kecerdasan yang mengatur loop tertutup. Selagi operator menyelidiki setiap skenario masalah individual dan menerapkan pemahaman serta pengalaman domain mereka dalam menentukan tindakan pemulihan yang sesuai, mereka juga menyediakan "label pelatihan" untuk pembelajaran mesin terawasi. Dengan menggunakan label tersebut, pembelajaran mesin mengevaluasi peristiwa, status jaringan, dan kondisi kinerja, bersama dengan model topologi, untuk membangun aturan dalam menerapkan tindakan pemulihan tersebut.

Memperkenalkan pembelajaran mesin ke dalam loop tertutup akan membutuhkan pembangunan pengalaman dan kepercayaan terhadap teknologi terutama ketika pembelajaran mesin terkait dengan tindakan yang akan diambil secara otomatis dalam jaringan. Membangun pengalaman dan kepercayaan ini bagi operator jaringan akan lebih mudah jika mereka dapat memahami apa yang direkomendasikan dan dilakukan sistem misalnya, jika mereka dapat menafsirkan kebijakan yang direkomendasikan oleh pembelajaran mesin loop tertutup. Pemahaman tentang kebijakan loop tertutup tersebut juga memberikan wawasan berharga kepada operator mengenai akar penyebab masalah yang mendasarinya, seperti bug perangkat lunak yang dipicu oleh beberapa kondisi tertentu, yang dapat mereka gunakan untuk akhirnya menghilangkan akar penyebab (bug) tersebut dari sistem. Berbagai teknik pembelajaran terawasi menghasilkan aturan yang sangat bervariasi dalam hal interpretabilitas untuk pemahaman manusia.

Teknik berbasis pohon keputusan seperti Hutan Acak memiliki logika aturan yang mudah diinterpretasikan, sementara teknik berbasis jaringan saraf seperti memori jangka pendek panjang sangat tidak jelas dalam logika penalaran, dan teknik berbasis grafik seperti Model Markov Tersembunyi dan Jaringan Kepercayaan Bayesian berada di antara keduanya. Teknologi pembelajaran mesin seperti yang ada dalam program AI yang Dapat Dijelaskan DARPA sesuai dengan misi dengan baik karena mereka bertujuan untuk menghasilkan model yang lebih dapat dijelaskan dan memungkinkan pengguna manusia untuk memahami, mempercayai dengan tepat, dan secara efektif mengelola generasi mitra kecerdasan buatan yang muncul, sambil mempertahankan tingkat kinerja pembelajaran yang tinggi (akurasi).

Tergantung pada intrusi tindakan restorasi, persyaratan pada kinerja pembelajaran mesin bervariasi. Beberapa tindakan restorasi memiliki gangguan layanan minimal, seperti



membersihkan konten cache di server proxy, sementara yang lain mungkin harus dilakukan dengan sangat hati-hati, seperti me-reboot perangkat edge di mana pelanggan single-homed. Ketika menyetel titik operasi yang diinginkan untuk pembelajaran kebijakan loop tertutup, operator perlu menentukan keseimbangan antara potensi negatif palsu (kondisi yang hilang di mana tindakan restorasi harus diterapkan) dan potensi positif palsu (tindakan restorasi yang tidak perlu yang menyebabkan dampak layanan yang tidak diinginkan). Meskipun penyetelan tersebut awalnya dapat disediakan oleh operator manusia berdasarkan pengetahuan domain, dapat dibayangkan juga bahwa seiring waktu, kecerdasan pembelajaran mesin yang terpisah dapat mengevaluasi dampak tindakan pemulihan terhadap kinerja layanan dan memasukkan hasilnya ke dalam pembelajaran kebijakan loop tertutup.

Memperlakukan tindakan pemulihan operator manusia sebagai label pelatihan atau kebenaran dasar dapat mengalami masalah kontaminasi label, karena operator manusia tidak selalu mengambil tindakan yang "tepat". Dalam banyak kasus, operator harus mengeksplorasi berbagai tindakan alternatif menggunakan coba-coba untuk menyelesaikan masalah. Kecerdasan berbasis pembelajaran mesin harus mengevaluasi efektivitas tindakan pemulihan untuk menghilangkan derau pada label pelatihan. Untuk loop tertutup reaktif, evaluasi otomatis terhadap efektivitas tindakan pemulihan dapat menggunakan asumsi "reaktif" bahwa tindakan yang dilakukan ditujukan untuk membersihkan peristiwa pemicu atau memulihkan penurunan kualitas layanan, yang ditangkap oleh KPI. Dengan memantau peristiwa dan rangkaian waktu KPI sebelum dan sesudah tindakan, kecerdasan pembelajaran mesin dapat menyimpulkan efektivitas tindakan, yang membantu meningkatkan kinerja pembelajarannya lebih lanjut. Kebijakan loop tertutup juga dapat bersifat proaktif, di mana tanda prediktif terdeteksi dan tindakan loop tertutup diambil sebelum dampak layanan terjadi. Kebocoran memori dapat diatasi menggunakan kebijakan loop tertutup proaktif, di mana tindakan memulai ulang proses atau migrasi akan mencegah potensi kegagalan sistem ketika penggunaan memori mulai mendekati batas sistem.

Pembelajaran mesin untuk kebijakan loop tertutup proaktif dapat memanfaatkan kelas teknik pembelajaran terawasi yang sama seperti dalam pembelajaran kebijakan reaktif, dengan waktu tunggu yang diinginkan untuk memisahkan kondisi fitur dan tindakan atau peristiwa dampak. Namun, pola peringatan dini seperti itu tidak selalu ada di lingkungan operasional, dan masih menjadi tantangan untuk mengukur efektivitas tindakan, terutama mengukur positif palsu tindakan yang tidak perlu yang diambil ketika tidak akan terjadi dampak sistem.

Pembahasan sebelumnya mengasumsikan bahwa pembelajaran kebijakan loop tertutup ONAP mengamati secara pasif tindakan yang dilakukan operator manusia dalam jaringan. Namun, dalam banyak situasi, operator manusia perlu mencoba beberapa tindakan sebelum menentukan tindakan yang dapat memperbaiki masalah tertentu. Kita dapat melangkah lebih jauh dengan pembelajaran mesin dan membiarkan mesin melakukan eksplorasi tindakan pemulihan yang tepat. Hal ini membawa kita ke model pembelajaran mesin daring, di mana sistem pembelajaran itu sendiri terus memperbarui pengetahuannya berdasarkan pengalaman aktif dalam mengeksplorasi berbagai tindakan dalam jaringan.



Mengadaptasi kebijakan loop tertutup daring biasanya dimodelkan sebagai proses pembelajaran penguatan, yang melibatkan pencarian keseimbangan antara eksplorasi dalam menilai efektivitas berbagai tindakan pemulihan dan pemanfaatan kebijakan yang telah diketahui dan diperoleh dari waktu ke waktu. Meskipun tingkat kenyamanan manusia dalam membiarkan pembelajaran mesin mengidentifikasi keseimbangan yang tepat dapat meningkat seiring waktu seiring dengan kematangan teknologi, perjalanannya mungkin dimulai dengan pembelajaran mesin yang hanya melakukan sedikit eksplorasi atau bahkan tidak sama sekali. Seiring dengan meningkatnya tingkat kenyamanan, pembelajaran mesin dapat secara otomatis mempelajari dan mengadopsi tindakan perbaikan jika terdapat keyakinan tinggi bahwa tindakan tersebut tepat. Namun, ketika tingkat keyakinan untuk melakukan tindakan pemulihan loop tertutup rendah dan biaya yang terkait dengan tindakan tersebut berdampak pada pelanggan, pembelajaran mesin akan (setidaknya pada awalnya) melibatkan operator manusia melalui sistem tiket. Setelah operator manusia menindaklanjuti tiket tersebut, pembelajaran mesin menganalisis tindakan dan hasilnya, lalu mengembangkan kebijakan loop tertutupnya. Dengan demikian, loop kontrol ONAP sebagaimana didefinisikan saat ini memanfaatkan kebijakan yang mendefinisikan tanda tangan (peristiwa yang akan dideteksi dan korelasinya) dan tindakan. Namun, pembelajaran mesin dapat ditingkatkan ke tingkat yang sama sekali baru, di mana pembelajaran mesin cukup memantau aliran data yang masuk dan kemudian memutuskan tindakan apa yang akan dicoba tanpa definisi tanda tangan dan tindakan yang eksplisit.

Masalah kronis adalah masalah yang terus berulang masalah ini biasanya perlu diidentifikasi dan dihilangkan secara permanen, jika memungkinkan. Tren jangka panjang pada berbagai agregat temporal, spasial, fungsi jaringan, versi perangkat lunak, dan konfigurasi metrik kinerja jaringan dan layanan, serta tindakan pemulihan yang dilakukan diperlukan untuk mengidentifikasi dan memecahkan masalah kronis tersebut. Analisis tersebut dapat mengungkapkan bug perangkat lunak dalam fungsi jaringan, dan bahkan masalah dengan perangkat pelanggan yang memengaruhi metrik tingkat jaringan dan layanan. Pertimbangan, misalnya, skenario kita sebelumnya di mana ONAP melakukan boot ulang atau migrasi VNF untuk menghindari kegagalan yang diprediksi akibat kebocoran memori. Meskipun otomatisasi mungkin dapat meminimalkan atau bahkan menyembunyikan dampak kebocoran memori ini, jika ini merupakan masalah yang sering terjadi, perangkat lunak harus diperbaiki. Saat masalah ini teridentifikasi, vendor fungsi jaringan dan perangkat pelanggan akan dilibatkan sebagaimana mestinya untuk memperbaiki perangkat lunak tersebut guna menyediakan perbaikan permanen.

Mewujudkan Administrasi Jaringan Otonom

Sebagaimana dibahas sebelumnya, banyak kontrol manajemen lalu lintas yang diotomatisasi saat ini misalnya, jaringan secara otomatis merutekan ulang lalu lintas sebagai respons terhadap kegagalan. Namun, seiring dengan semakin lazimnya protokol jaringan adaptif mandiri, fungsi jaringan elastis yang dapat tumbuh dan menyusut secara dinamis sebagai respons terhadap perubahan beban lalu lintas dan otomatisasi loop tertutup berbasis kebijakan untuk mengelola lalu lintas secara otomatis, pembelajaran mesin akan dimanfaatkan



untuk meningkatkan pengaturan parameter dan mempelajari kebijakan (aturan) secara otomatis dengan cara yang serupa seperti yang dijelaskan sebelumnya untuk pemeliharaan jaringan. Teknik pembelajaran tanpa pengawasan dan terawasi offline (yaitu, adaptasi model batch) dan online (yaitu, adaptasi model berkelanjutan) serta teknik pembelajaran penguatan online yang menggabungkan eksplorasi aktif dapat diterapkan.

Pembelajaran mesin juga merupakan kunci untuk memungkinkan manajemen lalu lintas proaktif memungkinkan otomatisasi jaringan untuk merespons kondisi lalu lintas yang akan datang sebelum kondisi tersebut muncul. Kunci dari manajemen lalu lintas proaktif adalah kemampuan untuk memprediksi volume dan pola lalu lintas. Teknologi pembelajaran mesin seperti Argus dapat digunakan untuk mengidentifikasi pola dan tren musiman, seperti bagaimana permintaan lalu lintas puncak bervariasi di berbagai hari dalam seminggu dan bagaimana pertumbuhannya dari minggu ke minggu. Untuk peningkatan lalu lintas sementara jangka pendek, yang dapat diakibatkan oleh acara di dunia fisik, seperti konser, atau dari acara di dunia virtual, seperti peluncuran game populer baru, operator dapat memanfaatkan pembelajaran mesin untuk mendapatkan parameter model peningkatan lalu lintas di awal acara. Kecerdasan jaringan kemudian dapat memanfaatkan model untuk memproyeksikan permintaan lalu lintas puncak dalam beberapa jam/hari mendatang. Berdasarkan prediksi beban ini, operator atau otomatisasi jaringan dapat menyesuaikan kapasitas jangka pendek.

Jika lonjakan jangka pendek terjadi berulang kali dari waktu ke waktu seperti yang mungkin terjadi di tempat konser atau arena olahraga pembelajaran mesin akan digunakan untuk mempelajari besarnya dan distribusi lalu lintas untuk acara tersebut, sehingga memungkinkan prediksi akurat tentang kapasitas yang diperlukan untuk acara ini. Kapasitas ini dapat dikerahkan dan dipantau secara proaktif memanfaatkan kalender (yang ditambah secara otomatis) yang tersedia untuk lokasi tersebut untuk mengidentifikasi kapan peristiwa yang dapat mengakibatkan beban lalu lintas tinggi akan terjadi, dan prediksi beban lalu lintas untuk menentukan berapa banyak kapasitas tambahan yang diperlukan untuk melayani peningkatan permintaan lalu lintas.

Tingkat terjadinya bencana alam di planet kita meningkat, dan ini memengaruhi frekuensi operator jaringan perlu merespons bencana. Otomatisasi jaringan dan AI membantu menyederhanakan proses manajemen bencana, meningkatkan respons, dan mengurangi biaya terkait.

Sebagian besar perencanaan dan respons bencana secara historis ditangani secara manual. Namun, AI dan otomatisasi semakin banyak digunakan untuk merencanakan respons terhadap bencana secara lebih efektif misalnya, untuk menentukan di mana harus menyebarkan peralatan terlebih dahulu sebelum bencana terjadi untuk memungkinkan pemulihan yang lebih cepat. Laporan cuaca yang berkaitan dengan badai dan topan yang akan datang dapat secara cerdas digabungkan dengan data jaringan dan lalu lintas, data kegagalan historis, dan data sosial (misalnya, data peta yang menunjukkan lokasi rumah sakit dan pusat tanggap darurat) untuk memprediksi lokasi paling efektif untuk menempatkan peralatan (misalnya, generator untuk lokasi seluler) sebelum suatu kejadian. Penempatan peralatan sebelumnya tersebut dapat secara signifikan mengurangi waktu yang dibutuhkan untuk



menyebarkan peralatan setelah bencana, meminimalkan waktu yang dibutuhkan untuk memulihkan layanan penting. Kecerdasan canggih juga dapat digunakan untuk memprioritaskan peristiwa pemulihan setelah bencana terjadi sekali lagi dengan mempertimbangkan data jaringan dan lalu lintas, beserta data sosial (misalnya, lokasi rumah sakit dan pusat tanggap darurat).

Pembelajaran mesin dan otomatisasi cerdas juga dapat digunakan untuk memperpanjang ketersediaan peralatan selama pemadaman misalnya, secara cerdas menidurkan operator untuk menghemat energi selama bencana guna memperpanjang masa pakai baterai dan dengan demikian menunda kebutuhan pengiriman generator. Seiring kita bergerak menuju jaringan zero-touch, ONAP dapat dimanfaatkan untuk mengelola kapasitas dan pengiriman, serta untuk menambah kapasitas jaringan dalam skenario bencana. Alur kerja ONAP akan bertanggung jawab untuk menentukan kapan dan di mana akan menambah kapasitas baru dan secara otomatis akan memantau dampak tindakan pemulihan terhadap pelanggan menggunakan analitik pembelajaran mesin yang berjalan pada DCAE ONAP.

Kualitas Data

Semua tugas manajemen jaringan sebelumnya pada dasarnya bergantung pada data yang tepat waktu dan berkualitas baik. Jika data hilang, tidak lengkap, atau salah, maka keputusan yang salah, termasuk tindakan jaringan yang intrusif, dapat dipicu oleh ONAP, yang berisiko menimbulkan dampak yang tidak perlu pada pelanggan atau bahkan lebih buruk. Oleh karena itu, sangat penting bagi kita untuk memperhatikan kualitas data dengan cermat mendeteksi umpan data yang gagal dengan cepat, mendeteksi data yang buruk dan hilang, dan merancang logika analitik dengan cermat agar dapat menangani data yang hilang, tertunda, dan salah secara efektif.

Sebagian besar data operasional yang dimanfaatkan untuk tugas-tugas yang dibahas dalam bab ini memerlukan data mendekati waktu nyata (hampir real-time) yang menjadi dasar pengambilan keputusan. Khususnya untuk aktivitas seperti pemeliharaan reaktif, membangun umpan data dengan latensi rendah adalah kunci untuk mencapai respons cepat terhadap gangguan jaringan dan layanan.

Menangani kualitas data menuntut tindakan pencegahan dari berbagai sisi. Mari kita mulai dengan mempertimbangkan umpan data ini adalah data yang mengalir masuk, baik sebagai aliran data atau sebagai berkas ke ONAP atau yang setara. Kita perlu mampu mendeteksi dengan cepat ketika satu atau lebih umpan gagal, ketika umpan telah rusak atau kehilangan data, atau ketika mereka tertunda secara berlebihan. Ini memerlukan sistem manajemen umpan yang memantau umpan data yang masuk, mendeteksi perilaku anomali (misalnya, penurunan di bawah volume data normal yang terkait dengan umpan), dan kemudian menjalankan tindakan perbaikan. Tindakan perbaikan idealnya harus diotomatisasi, seperti halnya dengan loop kontrol jaringan; operator manusia akan diperlukan jika otomatisasi gagal. Namun, sumber data yang berbeda secara inheren tiba dengan penundaan yang berbeda dan memiliki tingkat agregasi waktu yang berbeda.

Beberapa sumber data dikaitkan dengan titik waktu (misalnya, tautan turun), sementara yang lain mewakili agregat selama interval waktu (misalnya, jumlah paket yang



dibawa dalam interval 15 menit). Aturan korelasi (kebijakan) yang menggabungkan berbagai jenis data ini secara inheren menantang untuk dikelola, karena kita perlu dapat menentukan cara menangani pengaturan waktu yang berbeda ini. Mengidentifikasi kapan data tertunda hingga tidak dapat digunakan, hilang, atau salah, bahkan lebih menantang.

Selain memantau volume umpan data, kita juga dapat memeriksa isi umpan data. Perubahan format data dapat diidentifikasi, ditandai, dan dibagikan secara otomatis dengan aplikasi yang memanfaatkan data tersebut sehingga mereka dapat beradaptasi dengan tepat dalam memproses data. Demikian pula, data dapat divalidasi secara cepat misalnya, mendeteksi nilai yang berada di luar rentang yang dapat diterima (misalnya, CPU yang melebihi 100% atau kurang dari 0%, throughput yang melebihi kapasitas atau kurang dari nol). Masalah dapat ditandai dan dilaporkan kembali ke aplikasi yang menghasilkan data untuk pemulihan jangka panjang dan ke aplikasi yang menggunakan data untuk menunjukkan bahwa nilai tersebut tidak boleh digunakan atau digunakan dengan hati-hati.

Data yang hilang dapat paling mudah dideteksi melalui nomor urut nomor urut perlu disuntikkan oleh lapisan aplikasi sumber. Namun, jika nomor urut tidak tersedia seperti halnya sebagian besar sumber data saat ini inferensi kompleks terkadang dapat digunakan untuk mendeteksi data yang hilang. Misalnya, ketika pengukuran kinerja dilaporkan pada interval waktu yang kurang lebih teratur, data yang hilang dapat disimpulkan dengan mencari interval tanpa pengukuran yang dilaporkan.

Pada titik agregasi data di mana beberapa aliran data diproses untuk menghasilkan aliran data keluaran, seperti korelasi peristiwa dan komputasi metrik kinerja, analitik harus dirancang untuk mengantisipasi data yang tertunda dan hilang. Karena aliran data masuk yang berbeda mungkin mengalami penundaan yang berbeda, analitik kemungkinan besar perlu menunggu beberapa waktu hingga data yang berbeda tiba sebelum menghitung hasilnya. Biasanya, pilihan desain ini dibuat untuk mencapai keseimbangan antara ketepatan waktu data keluaran (seperti untuk aplikasi yang sensitif terhadap waktu seperti deteksi peristiwa) dan kelengkapan data keluaran (seperti untuk penggunaan yang sensitif terhadap akurasi seperti penagihan).

Di tempat-tempat yang memungkinkan sumber daya komputasi dan penyimpanan, analitik agregator dapat menghasilkan hasil yang mendekati waktu nyata, yaitu dengan waktu tunggu yang singkat berdasarkan data yang dimilikinya, dan hasil yang lebih akurat dan tertunda seiring dengan semakin banyaknya data yang tersedia. Dengan demikian, aplikasi yang perlu bereaksi sangat cepat memanfaatkan hasil agregator yang kurang akurat dan lebih tepat waktu, sementara aplikasi yang membutuhkan hasil yang lebih akurat harus menunggu data tersedia. Dalam semua kasus, merupakan praktik yang baik untuk menyertakan tidak hanya hasil agregasi tetapi juga indikator kualitas data yang terkait dengan hasil tersebut dan menyebarkannya ke analitik hilir (misalnya, berapa banyak aliran data pengumpul yang menjadi dasar keluaran).

Meskipun setiap aplikasi individual harus menentukan trade-off yang diinginkan antara latensi data dan kelengkapan serta akurasi data, desain aliran analitik dan pergerakan data secara keseluruhan membatasi latensi minimum yang dapat dicapai aplikasi. Beralih dari



pemrosesan data terpusat menuju analitik data terdistribusi dan intelijen di tepi (lokal tempat data jaringan atau layanan dikumpulkan dan kontrol diaktifkan) akan mengurangi latensi batas bawah ketika kontrol ultra-waktu nyata diperlukan. Dalam beberapa kasus, data pengukuran "terlalu besar untuk disalin", artinya, tidak diinginkan untuk mengirimkan data melalui jaringan area luas baik karena pemborosan bandwidth atau latensi data yang berkepanjangan. Pendekatan hibrida harus dipertimbangkan di mana pemrosesan tepi melalui agregasi, deteksi anomali atau peristiwa, atau pengambilan sampel cerdas secara signifikan mengurangi volume aliran data dan intelijen terpusat membuat analitik dan kontrol lebih lanjut berdasarkan aliran data yang berkurang.

Terakhir, di sisi analitik aplikasi, penting untuk secara eksplisit merepresentasikan dan menangani data yang hilang dan salah. Kesalahan umum adalah memperlakukan data yang hilang secara sembarangan sebagai nilai data nol, yang mencemari hasil dan berpotensi mendatangkan malapetaka pada jaringan. Demikian pula, data yang jelas-jelas keliru, seperti nilai CPU yang melebihi utilisasi 100% atau utilisasi tautan yang melebihi kapasitas, harus diidentifikasi dan ditangani dengan tepat (dibulatkan atau dihapus seluruhnya). Mendeteksi kesalahan data ini sejak dini meminimalkan apa yang disebut "sampah masuk, sampah keluar". Lebih lanjut, tergantung pada aplikasinya, memanfaatkan analitik yang robust dapat bermanfaat jika memungkinkan: penggunaan statistik median kurang subjektif terhadap data yang terkadang keliru dibandingkan statistik rata-rata; penggunaan regresi robust merupakan alternatif yang lebih baik daripada regresi kuadrat terkecil ketika data terkontaminasi dengan outlier atau observasi yang berpengaruh.

Kebijakan yang Aman

ONAP memanfaatkan kebijakan untuk memungkinkan pengguna menentukan perilaku sistem. Kebijakan, misalnya, digunakan untuk mengonfigurasi cara sistem menginstansiasi VNF (misalnya, kebijakan menentukan lokasi VNF), cara menjadwalkan acara pemeliharaan terencana, serta tanda tangan dan tindakan loop kontrol. Dengan demikian, kebijakan ONAP memberikan kekuatan yang luar biasa kepada pengguna menghilangkan pengembangan kustom, sehingga mempercepat respons terhadap perubahan kondisi jaringan dan mengurangi biaya pengembangan.

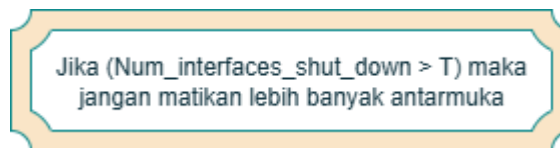
Namun, kekuatan tersebut disertai risiko. Kesalahan yang dimasukkan ke dalam kebijakan oleh operator jaringan atau kebijakan berbahaya dapat menyebabkan kerusakan yang sangat besar, kecuali jika perlindungan diterapkan untuk mengatasinya. Bahkan data buruk yang dikombinasikan dengan kebijakan yang dibuat dengan buruk berisiko menyebabkan kerusakan. Dan karena pembelajaran mesin menggantikan kebijakan eksplisit yang ditentukan manusia dalam loop kontrol dan seterusnya, kita perlu memastikan bahwa hal ini tidak membahayakan jaringan. ONAP memanfaatkan kebijakan yang aman untuk meminimalkan risiko yang terkait dengan kebijakan yang ditentukan manusia dan yang dipelajari mesin.

Seperti halnya teknologi baru lainnya, kebijakan dan pembelajaran mesin untuk loop kontrol perlu dimasukkan dengan hati-hati ke dalam jaringan. Pengujian dan evaluasi ekstensif di lingkungan lab dilakukan. Namun, lingkungan lab dan simulasi offline tidak dapat

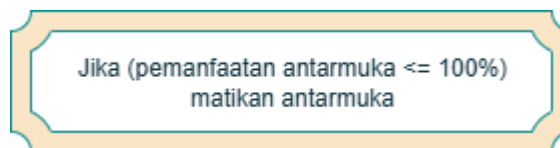


memperkenalkan keragaman, kebisingan, dan skala jaringan produksi. Teknologi baru dengan demikian biasanya diperkenalkan secara bertahap ke dalam jaringan produksi misalnya, sebagai FFA dengan cakupan terbatas (misalnya, terkait dengan sejumlah kecil fungsi jaringan). Teknologi tersebut juga dapat diperkenalkan ke bagian-bagian jaringan di mana mereka dapat memiliki dampak yang lebih kecil jika terjadi kesalahan (misalnya, di tepi jaringan).

Salah satu contoh teknologi yang dapat digunakan untuk melindungi dari kebijakan yang buruk adalah pelindung kebijakan. Sebagai analogi, pagar pembatas di jalan raya pegunungan ditempatkan di tepi jalan raya untuk memberikan perlindungan terhadap mobil yang terlempar dari tepi tebing. Dalam kondisi normal, mobil seharusnya tidak menabrak pagar pembatas. Namun, jika mobil berputar di luar kendali dan menabrak pagar pembatas, pagar pembatas dirancang untuk mencegah hasil yang dahsyat. Pelindung kebijakan serupa; Kebijakan-kebijakan tersebut pada dasarnya merupakan kebijakan pengganti, yang, ketika mendeteksi pelanggaran kondisi invarian (kondisi yang seharusnya selalu benar), mencegah tindakan relevan lebih lanjut untuk diambil. Sebagai contoh, pertimbangkan sebuah pelindung kebijakan yang membatasi jumlah antarmuka jaringan yang dapat dimatikan oleh sistem pada suatu titik waktu tertentu. Kebijakan pelindung ini dapat ditentukan sesuai Gambar 9.8, yang berarti jika jumlah antarmuka yang dimatikan pada suatu titik waktu tertentu melebihi ambang batas yang telah ditentukan, T , maka tidak ada lagi antarmuka yang boleh dimatikan hingga kondisi tersebut teratasi. Dengan demikian, kebijakan pelindung ini mengesampingkan semua kebijakan sistem lain yang mencoba mematikan antarmuka.



Gambar 9.8 Contoh kebijakan pelindung.



Gambar 9.9 Contoh kebijakan "buruk".

Pelindung kebijakan dapat digunakan untuk melindungi dari kesalahan yang disebabkan oleh manusia dan AI atau dari niat jahat. Misalnya, jika seorang operator, baik secara sengaja maupun keliru, memasukkan kebijakan ke dalam sistem, seperti yang didefinisikan pada Gambar 9.9, yang secara efektif mencoba mematikan semua antarmuka dalam jaringan, maka kebijakan pengaman pada Gambar 9.8 akan menggantikannya setelah jumlah antarmuka yang dimatikan melebihi T , sehingga melindungi infrastruktur jaringan.

Mempertimbangkan loop kontrol secara khusus, pengaman kebijakan juga dapat memberikan tingkat perlindungan seiring evolusi kita menuju loop kontrol otonom di mana sistem mempelajari tanda tangan dan tindakan secara mandiri tanpa validasi operator perantara. Pengaman kebijakan dapat membatasi apa yang dapat dilakukan pembelajaran



otonom, khususnya membatasi laju tindakan dan sejauh mana tindakan tersebut beroperasi di jaringan. Hal ini dapat memberikan tingkat kepercayaan diri kepada operator jaringan saat teknologi mutakhir tersebut diterapkan.

9.5 DAMPAK PADA PERAN DAN TANGGUNG JAWAB OPERASI

Kita sekarang melihat melampaui teknologi dan mempertimbangkan dampak otomatisasi dan AI terhadap tim operasi jaringan.

Struktur dan Peran Operasi Jaringan

Peran dan struktur tim operasi jaringan terus berkembang seiring waktu, seiring kemajuan teknologi dan semakin otomatisnya tugas-tugas manajemen jaringan.

Dalam penyedia layanan besar seperti AT&T, dengan beragam jaringan dan teknologi, tim operasi secara historis telah diorganisasikan berdasarkan peran dan teknologi jaringan. Terdapat tim yang merencanakan jaringan, pusat keandalan yang berfokus pada teknologi yang mengelola berbagai teknologi jaringan dan segmen jaringan, pusat manajemen layanan yang mengelola layanan menyeluruh (misalnya, layanan Internet dan VPN, layanan suara, dan distribusi video/TV), dan tim lapangan yang bekerja di kantor pusat, lokasi pelanggan, dan di masyarakat luas yang melaksanakan pekerjaan yang tidak dapat diselesaikan dari jarak jauh.

Tim pemeliharaan jaringan memiliki tingkatan, dengan tingkat keterampilan dan tanggung jawab yang meningkat seiring dengan bertambahnya tingkatan. Tanggung jawab Tier 1 sebagian besar telah diotomatisasi di penyedia layanan saat ini dan mencakup pelaksanaan tugas-tugas seperti meninjau dan menyimpan berkas log, memverifikasi konfigurasi port, dan memverifikasi konektivitas sirkuit. Personel Tier 1 memiliki tingkat keterampilan teknis yang rendah dan hanya membutuhkan pengetahuan domain yang minimal. Organisasi Tier 2 melakukan sebagian besar pekerjaan pemeliharaan dan administrasi jaringan yang belum diotomatisasi. Teknisi Tier 2 lebih terampil daripada Tier 1; mereka adalah tim utama yang bertanggung jawab atas pemeliharaan rutin, terencana, dan reaktif, serta menyediakan garis pertahanan pertama untuk masalah jaringan.

Organisasi Tier 3 menyediakan dukungan operasional serta desain dan perencanaan operasi jaringan dan layanan. Personel Tier 3 ditugaskan untuk membantu menyelesaikan masalah jaringan yang kompleks masalah yang tidak dapat diselesaikan sendiri oleh tim Tier 2 secara tepat waktu. Mereka melakukan forensik teknis mendalam yang diperlukan untuk menentukan akar penyebab gangguan jaringan yang kompleks, dan mereka melakukan analisis pasca-masalah yang terperinci untuk membuat rencana tindakan yang mengatasi setiap celah yang terungkap oleh peristiwa besar. Tim ini juga bertanggung jawab untuk menyelesaikan masalah kronis atau sistemik, bermitra erat dengan vendor fungsi jaringan sesuai kebutuhan. Tim Tier 3 juga bekerja sama erat dengan para insinyur yang merancang layanan jaringan dan jaringan itu sendiri, terutama terkait rencana manajemen jaringan yang mendefinisikan bagaimana jaringan dan layanan seharusnya dioperasikan. Secara khusus, personel Tier 3 merancang dan mensertifikasi prosedur kompleks yang diperlukan untuk menerapkan teknologi jaringan baru, mengganti fungsi jaringan, dan memutakhirkan perangkat lunak jaringan.



Pusat Operasi Teknologi Global (GTOC) adalah pusat operasi jaringan tunggal dengan redundansi lokasi yang menyediakan fungsi pengawasan dan koordinasi menyeluruh. Gambar 9.10 menggambarkan GTOC AT&T. GTOC mengelola eskalasi pemadaman dan menyediakan pengawasan serta koordinasi operasional di berbagai pusat keandalan yang berfokus pada teknologi, layanan pelanggan, dan badan eksternal seperti FCC. GTOC bertanggung jawab atas manajemen lalu lintas, termasuk respons terhadap bencana dan peristiwa khusus. GTOC juga berfungsi sebagai saluran komunikasi tunggal kepada pimpinan senior internal sehingga informasi yang konsisten dan akurat tentang peristiwa jaringan menjangkau semua pemangku kepentingan. Selain itu, GTOC mengembangkan dan memelihara proses-proses penting yang digunakan di seluruh organisasi operasi.



Gambar 9.10 Pusat operasi teknologi global (GTOC).

Perubahan Peran Operasi Jaringan

Meskipun volume lalu lintas terus tumbuh secara eksponensial, pertumbuhan pendapatan penyedia layanan tidak mengikuti kurva yang sama. Dengan demikian, tim operasi biasanya tidak dapat diperluas secara signifikan, bahkan ketika dihadapkan pada layanan baru, teknologi jaringan dan kontrol baru (misalnya, pengenalan VNF), dan perluasan berkelanjutan dalam jumlah fungsi jaringan yang dikelola.

Pengenalan VNF ke dalam jaringan mengubah peran yang dijalankan tim operasi. Peran operasi menjadi lebih erat kaitannya dengan perusahaan perangkat lunak tradisional, seperti Google atau Facebook. Namun, peran keahlian jaringan masih sangat kuat karena aplikasi yang berjalan di "cloud" merupakan fungsi jaringan yang menyediakan kapabilitas jaringan dan memanfaatkan protokol jaringan. Seiring dengan terus berkembangnya otomatisasi melalui ONAP dan AI terkait, tingkat operasional yang lebih rendah (Tingkat 1, Tingkat 2) akan terus berkurang. Visi utama jaringan zero-touch menghilangkan peran-peran ini sepenuhnya, tugas yang sebelumnya dilakukan oleh manusia akan ditangani oleh ONAP dan otomatisasi lainnya. Pengenalan virtualisasi fungsi jaringan juga mengurangi kebutuhan tenaga kerja di tempat dan kunjungan cepat sesuai permintaan ke kantor pusat. Namun, meningkatnya kompleksitas dan ketergantungan pada perangkat lunak mendorong peningkatan keterampilan Tingkat 3, yang mengharuskan personel untuk melakukan penelusuran kesalahan canggih terhadap masalah



kompleks berbasis perangkat lunak. Bergantung pada struktur organisasi, peran, dan tanggung jawab, tim operasi mungkin juga memiliki kebutuhan yang lebih besar akan keterampilan perangkat lunak tingkat lanjut dan pengembang kebijakan yang dapat menggunakan antarmuka kebijakan ONAP untuk memungkinkan otomatisasi baru.

Sebagaimana dibahas sebelumnya dalam bab ini, pembelajaran mesin diperkirakan akan semakin banyak digunakan untuk mempelajari kebijakan yang dimanfaatkan dalam platform otomatisasi jaringan berbasis kebijakan seperti ONAP. Operator perlu melakukan perjalanan untuk bekerja dengan AI seiring perkembangannya dimulai dari peran orang tua di mana operator bertanggung jawab atas penulisan kebijakan dan manajemen alur kerja, bergerak menuju peran mentor, di mana AI meniru dan belajar dari aktivitas dan keputusan manusia, dan akhirnya ke peran konselor di mana operator manusia memberikan panduan korektif dan umpan balik kepada kontrol AI. Mesin dan manusia perlu bertemu di tengah jalan untuk melakukan perjalanan ini operator jaringan manusia perlu semakin nyaman dengan pembelajaran mesin dan semakin memercayai mesin, sementara mesin perlu memberi manusia pemahaman tentang apa yang dihasilkan pembelajaran mesin dan otomatisasi untuk membantu mencapai tingkat kenyamanan yang semakin meningkat tersebut.

Perkembangan alami bagi operator manusia untuk mendapatkan kepercayaan pada kecerdasan berbasis pembelajaran mesin tercermin dalam tiga kelas skenario penggunaan teknologi pembelajaran mesin: (1) pembelajaran mesin menghasilkan model dan data untuk meningkatkan pemahaman manusia tentang data yang kompleks, (2) pembelajaran mesin menyajikan rekomendasi untuk pengambilan keputusan manusia, dan (3) pembelajaran mesin mewujudkan otomatisasi lengkap, melewati operator manusia. Contoh aplikasi kelas pertama adalah penambahan tiket operasi, yang mencakup penguraian dan pemahaman data terstruktur dan tidak terstruktur (seperti catatan operator dalam tiket), menghasilkan statistik pola kegagalan umum atau elemen jaringan, dan menentukan efektivitas dan efisiensi berbagai prosedur dan tindakan respons masalah. Seiring operator mendapatkan wawasan dari penambahan tiket berbasis pembelajaran mesin, kepercayaan diri mereka dalam "berkonsultasi" dengan pembelajaran mesin untuk mendapatkan rekomendasi meningkat. Di sini, kecerdasan mesin membandingkan tiket jaringan yang sedang berlangsung/diberikan dan mengidentifikasi tindakan dan prosedur respons yang kemungkinan berhasil menyelesaikan masalah. Ini diberikan kepada operator jaringan sebagai rekomendasi; terserah kepada operator manusia untuk memutuskan apakah akan mengikuti rekomendasi (atau tidak). Sistem rekomendasi seperti itu disebut sistem pakar.

Sementara teknologi yang mendasari sistem pakar dapat bervariasi, mulai dari logika keputusan berbasis aturan hingga kecerdasan berbasis jaringan saraf dalam, selalu membantu bagi sistem pakar untuk mengekspos faktor-faktor yang mengarah pada rekomendasi dan skor "keyakinan" dari rekomendasi tersebut kepada operator manusia. Membantu operator manusia untuk melewati proses pemeriksaan mereka membantu kecerdasan pembelajaran mesin untuk berkembang ke skenario aplikasi ketiga di mana keputusan berbasis pembelajaran mesin terhubung langsung dengan platform kontrol untuk menjalankan aktivitas kontrol; operator manusia hanya perlu sesekali menyuntikkan beberapa umpan balik korektif ke



kecerdasan pembelajaran mesin. Seiring dengan diperkenalkannya AI dan meningkatnya otomatisasi jaringan ke dalam jaringan, tim operasi jaringan bertanggung jawab untuk memastikan otomatisasi ini bekerja secara efektif dan sesuai rancangan. Hal ini biasanya dicapai dengan meluncurkan otomatisasi secara hati-hati dan bertahap serta memantau otomatisasi dan dampaknya terhadap jaringan secara cermat. Teknologi canggih juga dapat dimanfaatkan pertama kali di tepi jaringan, di mana masalah apa pun dengan otomatisasi kemungkinan besar akan berdampak jauh lebih kecil dan di mana otomatisasi seringkali memberikan dampak yang lebih besar terhadap peningkatan kinerja operasional dan layanan.

9.6 KESIMPULAN

Peran dan tanggung jawab operasi jaringan akan terus berkembang seiring kemajuan kita menuju visi jaringan otonom. Transformasi teknologi yang pesat saat ini sedang berlangsung, dengan virtualisasi, pembelajaran mesin, dan teknologi kontrol SDN berbasis kebijakan seperti ONAP yang semakin matang dan merambah ke dalam penerapan jaringan berskala. Komunitas jaringan dan komunitas AI perlu bekerja sama secara erat untuk secara efektif mencapai visi jaringan otonom.



BAB 10

KECERDASAN BUATAN UNTUK KEAMANAN SIBER

10.1 PENDAHULUAN

Dalam bab ini, kami akan membahas beberapa pertimbangan untuk penggunaan otomatisasi, pembelajaran mesin (ML), dan kecerdasan buatan (AI) misalnya, jaringan saraf tiruan (NN) dalam konteks keamanan jaringan. Kami akan menjelaskan tantangan dan teknologi dalam memerangi serangan siber dan menjelaskan beberapa contoh di mana penggunaan teknologi otomatisasi yang agresif dapat digunakan untuk meningkatkan kemampuan keamanan dalam lingkungan jaringan yang ditentukan perangkat lunak (SDN).

Peluang tersedia di banyak bidang, seperti pembuktian identitas, autentikasi, manajemen kebijakan keamanan, penilaian postur keamanan, dan analitik ancaman. Hal ini dapat diterapkan untuk meningkatkan skala operasi melampaui kapasitas manusia, meningkatkan kemutakhiran konfigurasi keamanan, mengurangi masalah akibat kesalahan, dan mengurangi waktu deteksi dan respons terhadap peristiwa. Kami akan membahas lebih lanjut tentang topik-topik ini dan topik lainnya di bagian selanjutnya.

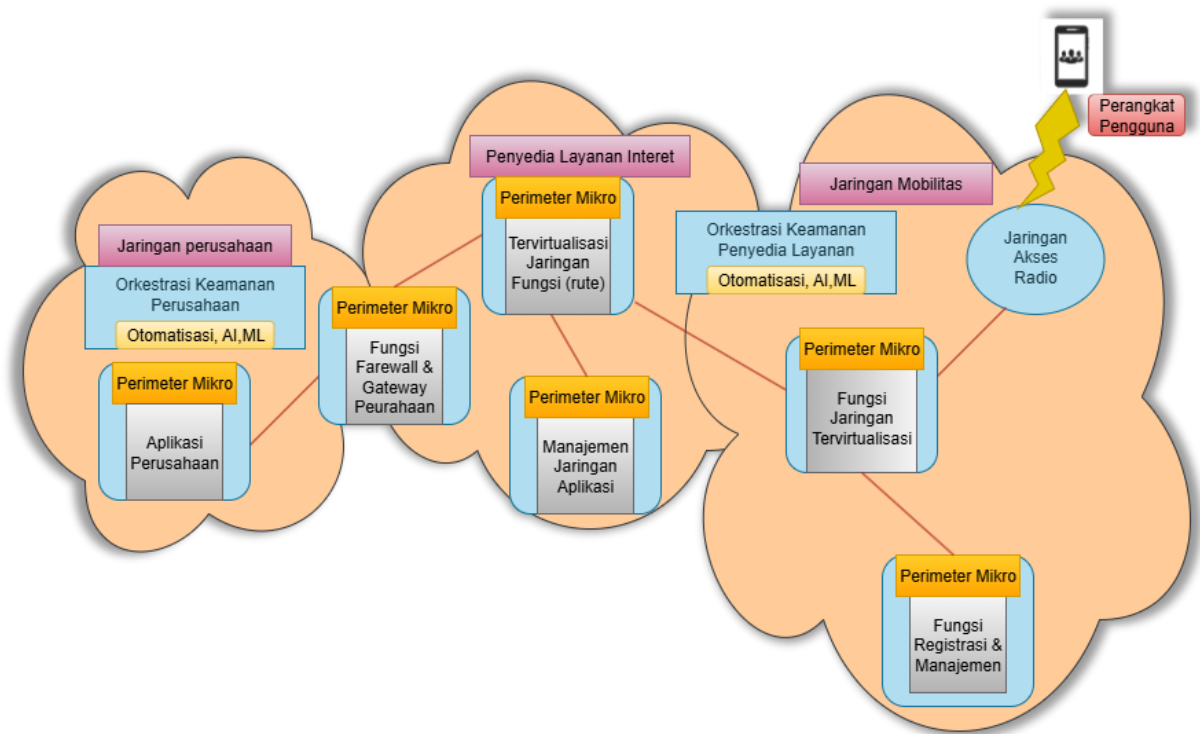
Apa Itu Keamanan Jaringan?

Dalam pembahasan ini, "keamanan jaringan" terutama difokuskan pada memastikan jaringan itu sendiri tahan terhadap serangan siber. Hingga taraf tertentu, keamanan jaringan juga dapat memastikan bahwa layanan jaringan tangguh dan relatif kebal terhadap serangan perutean atau upaya pemantauan lalu lintas jaringan. Keamanan jaringan dapat diterapkan pada jaringan yang disediakan oleh penyedia layanan, misalnya, Penyedia Layanan Internet (ISP). Atau, dapat diterapkan pada jaringan perusahaan yang disediakan oleh bisnis di gedung, kampus, atau di beberapa kampus.

Gambar 10.1 menunjukkan model referensi jaringan umum yang membantu mengidentifikasi beberapa infrastruktur yang diperlukan untuk menyediakan fungsi jaringan. Ilustrasi ini juga memperkenalkan gagasan fungsi jaringan tervirtualisasi (vNF). Virtualisasi atau SDN menyediakan beberapa peluang keamanan baru yang akan dibahas nanti. Terdapat berbagai jenis fungsi jaringan dalam jaringan perusahaan, yang berinteraksi dengan ISP, di dalam ISP, dan di dalam layanan jaringan mobilitas. Masing-masing fungsi ini perlu dilindungi dari serangan. Banyak dari fungsi ini dapat divirtualisasi.

Ilustrasi tersebut juga menunjukkan "mikro-perimeter" sebagai sarana untuk membantu menyediakan kontrol keamanan vNF. Mikro-perimeter adalah serangkaian fungsi yang digerakkan oleh perangkat lunak yang membantu melindungi fungsi tersebut. Kontrol dapat mencakup kontrol kebijakan jaringan, audit kepatuhan keamanan, enkripsi, pengumpulan data aktivitas, dan fungsi keamanan lainnya. Terakhir, ada fungsi orkestrasi keamanan di perusahaan serta penyedia layanan.





Gambar 10.1 Model infrastruktur jaringan umum menunjukkan kontrol keamanan semakin mendekati fungsi jaringan tervirtualisasi (vNF) serta aplikasi perusahaan.

Fungsi orkestrasi keamanan berfungsi untuk mengelola serangkaian kemampuan kolektif di seluruh entitas masing-masing secara kohesif dan holistik. Karena fungsi orkestrasi keamanan adalah sarana utama untuk mengelola keamanan, ada baiknya untuk secara logis menganggap fungsi otomatisasi, AI, dan ML sebagai bagian dari fungsi ini. Mungkin ada fungsi pendukung lain yang memiliki logika kompleks juga, tetapi pada akhirnya, mereka mungkin harus dianggap sebagai bawahan orkestrasi keamanan.

Keamanan jaringan tidak berarti bahwa pengguna jaringan, sistem yang menggunakan layanan jaringan, atau aplikasi yang mengirimkan lalu lintas melalui jaringan akan kebal terhadap serangan siber apa pun. Hal-hal yang terhubung ke jaringan tetap dapat diserang, dan hanya jenis serangan tertentu yang dapat diidentifikasi melalui analisis aktivitas jaringan. Misalnya, jika pengguna menghubungkan perangkat USB yang terinfeksi ke komputer dan menginfeksi komputer tersebut, jaringan tidak dapat berbuat banyak untuk melindungi komputer tersebut dari serangan. Namun, jika komputer yang terinfeksi tersebut terhubung melalui jaringan untuk menyerang komputer lain, ada beberapa langkah yang berpotensi dapat dilakukan jaringan untuk meminimalkan serangan atau mencegah keberhasilan serangan. Untuk melakukannya, jaringan perlu mengetahui lalu lintas mana yang merupakan lalu lintas serangan, dan perlu memiliki kemampuan untuk menyaring lalu lintas yang buruk secara selektif tanpa mengganggu lalu lintas yang baik.

Berikut adalah beberapa langkah yang menjadi peluang penerapan AI untuk keamanan jaringan. Peluang tersebut terbagi dalam kategori berikut:



1. Memantau konfigurasi banyak sistem dan jaringan untuk mengidentifikasi titik lemah dan menerapkan kontrol yang tepat, yang dapat meminimalkan profil serangan.
2. Memantau dampak dan manfaat kontrol keamanan untuk membantu meminimalkan beban kontrol keamanan pada jaringan dan aplikasi.
3. Mempelajari secara kumulatif berbagai skenario serangan dan memantau indikasi ancaman pada jaringan serta meminimalkan waktu respons.

Hal ini cocok untuk AI dan ML, terutama karena keduanya kompleks dan membutuhkan skala serta dinamika yang signifikan. Seiring dengan perubahan jaringan yang besar dan kompleks, kebutuhan konfigurasi keamanan juga perlu berubah secara terus-menerus dan dinamis. Kita akan membahas hal ini lebih lanjut nanti.

Tren Ancaman Siber

Tujuan bagian ini adalah untuk membantu menetapkan kebutuhan akan metode yang lebih canggih untuk mengelola keamanan siber. Secara umum, motif ancaman siber adalah sebagai berikut:

- **Keuntungan finansial:** Penjahat siber telah mengembangkan metode teknologi untuk mencuri dan menjual hampir semua hal berharga di dunia siber, termasuk dana elektronik, pembelian ritel, konten media, kartu kredit, informasi identitas, dan layanan daring. Seluruh ekosistem telah berkembang dengan spesialisasi dalam pengembangan malware, pencurian data, penjualan kembali, pencucian uang, dan pengiriman. Serangan penolakan layanan terdistribusi (DDoS) (yaitu, serangan banjir lalu lintas), pencurian data, dan penghancuran sistem adalah metode populer untuk memeras uang dari korban. Layanan serangan penolakan layanan terdistribusi telah menjadi industri komersial yang memfasilitasi ribuan serangan di seluruh dunia setiap hari dengan biaya tertentu.
- **Pengaruh politik:** Negara-negara bangsa maupun kelompok teroris (atau negara) dapat memperoleh keuntungan melawan musuh dengan memanipulasi media mereka, memanipulasi pemilu, dan menguasai infrastruktur penting. Tentu saja, salah satu infrastruktur penting adalah telekomunikasi. Dalam banyak kasus, ini menjadi ancaman asimetris karena biaya untuk mencegah serangan yang berhasil bisa jauh lebih besar daripada investasi yang dibutuhkan untuk berhasil melakukan serangan.
- **Gangguan:** Beberapa orang meretas untuk bersenang-senang. Meskipun ini bukan ancaman yang paling tangguh, ancaman ini dapat menimbulkan konsekuensi yang tidak terduga. Beberapa metode serangan baru dikembangkan melalui eksperimen independen.

Lanskap ancaman menjadi semakin kompleks. Seiring berkembangnya teknologi dan metode serangan baru, teknik serangan lama tidak akan hilang. Selama manusia terlibat, akan selalu ada kesalahan. Selama solusi dibatasi oleh anggaran, sistem lama mungkin tetap beroperasi. Akibatnya, kerentanan yang diketahui akan tetap ada dan teknik serangan lama akan terus berguna bagi penyerang. Hal ini mungkin tidak terjadi dalam sistem atau operasi jaringan arus utama. Namun, mungkin ada aktivitas di pinggiran perusahaan dan jaringan yang juga perlu dipertimbangkan.



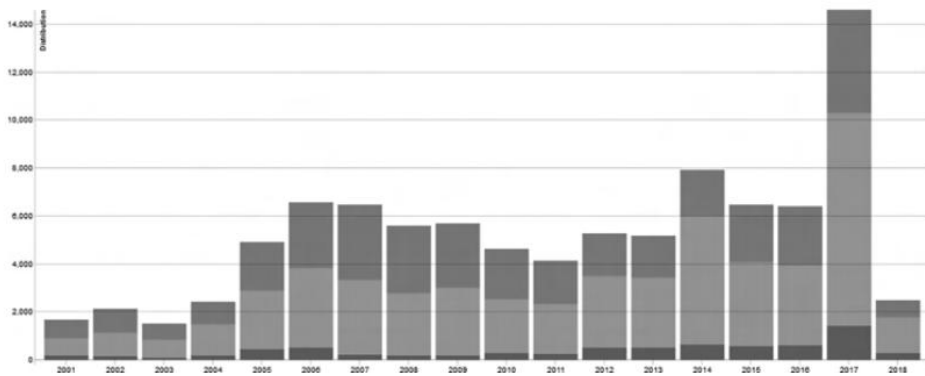
Bagi bisnis (termasuk penyedia layanan jaringan), pelanggaran atau pemadaman semakin merugikan seiring dengan semakin banyaknya otomatisasi yang diterapkan ke dalam proses bisnis. Implikasinya meliputi:

- Dampak operasional dan waktu henti
- Kerusakan reputasi merek
- Hilangnya pendapatan
- Hilangnya pelanggan dan keunggulan kompetitif

Bagi konsumen layanan ini, hal ini dapat berdampak berkelanjutan jika informasi pribadi pelanggan terekspos. Seiring dengan semakin kompleksnya sistem dan perangkat lunak, frekuensi kerentanan juga terus meningkat seperti yang diilustrasikan pada Gambar 10.2. Pertumbuhan *Internet of Things* (IoT) (yaitu, perangkat yang terhubung ke jaringan otonom) memiliki serangkaian konsekuensi tersendiri terhadap keamanan:

- **IoT diterapkan pada semua jenis aplikasi.** Saat buku ini ditulis, hampir mustahil untuk menentukan bagaimana beragamnya aplikasi IoT dapat memengaruhi situasi hidup atau mati, tetapi banyak skenario yang tidak terlihat saat ini pasti akan berdampak pada hidup atau mati.
- **Tidak ada lagi manusia di monitor dan keyboard untuk membuat penilaian yang baik tentang apa yang mungkin terjadi.** Benda yang dikendalikan dari jarak jauh akan bertindak sesuai perintah (benar atau salah). Informasi akan diberikan oleh perangkat (benar atau salah). Kita bergantung pada keamanan perangkat (serta keputusan desain yang baik) untuk memastikan perangkat dapat membuat keputusan yang tepat, setidaknya di sebagian besar waktu. Jika perangkat disusupi, perangkat akan melakukan apa yang diinginkan penyerang.
- **IoT mengubah parameter penskalaan jaringan.** Meskipun jaringan umumnya dirancang untuk mengakomodasi kapasitas byte, kini jaringan harus dirancang untuk mengakomodasi jumlah perangkat yang terhubung. Dunia masih dalam tahap awal pemahaman tentang bagaimana tindakan kecil yang dilakukan oleh banyak perangkat dapat memengaruhi keamanan jaringan dan pengguna. Apa yang terjadi jika seseorang mencuri sedikit informasi dari jutaan tempat?
- **Perangkat IoT dirancang agar murah (terkadang sangat murah).** Hal ini tidak dapat disamakan dengan keamanan yang lebih lemah. Oleh karena itu, diperlukan cara yang paling efisien dan efektif untuk memastikan keamanan perangkat. Ini merupakan tantangan yang signifikan, dan perlu menjadi tujuan dengan prioritas yang sangat tinggi.
- **Sejumlah besar perangkat pasti akan ditemukan memiliki masalah keamanan.** Harus ada metode dan alat untuk memperbaiki kerentanan, memblokir serangan, dan memulihkan perangkat yang disusupi. Memulihkan (atau menghilangkan) masalah keamanan atau perangkat secara manual saja tidak lagi cukup. Perlu ada cara untuk memperbaiki masalah secara otomatis, berpotensi masif, dan sangat cepat.





Gambar 10.2 Kerentanan *Common Vulnerability Scoring System* (CVSS) berdasarkan tahun menunjukkan peningkatan kerentanan yang dilaporkan dari tahun ke tahun.

Secara kolektif, kami telah mencatat beragamnya motivasi yang terus berkembang untuk melakukan serangan siber, lanskap ancaman yang semakin kompleks seiring waktu, implikasi yang semakin besar bagi bisnis yang menjadi korban pelanggaran, dan serangkaian kerentanan perangkat lunak yang semakin meningkat serta implikasi IoT yang belum dapat dipastikan. Tuntutan keamanan ini tidak dapat diatasi dengan metode konvensional. Solusi yang sangat otomatis perlu diciptakan untuk membantu meningkatkan dan menyederhanakan keamanan.

10.2 PERTIMBANGAN AI DAN OTOMATISASI KEAMANAN

Kita telah membahas masalah keamanan dasar yang ada dan menetapkan bahwa masalah tersebut membutuhkan solusi yang lebih otomatis. Di bagian ini, kita akan berfokus pada topik dasar yang perlu dipertimbangkan ketika menerapkan teknik otomatisasi, seperti pembelajaran mesin (ML) dan kecerdasan buatan (AI), untuk memecahkan masalah keamanan. Akan ada banyak tantangan yang harus dipecahkan. Dalam kondisi terkini, penggunaan teknik canggih untuk memecahkan masalah memerlukan beberapa keterampilan khusus dari para pelaksananya. Keterampilan ini relatif jarang, dan ada banyak masalah yang harus dipecahkan. Akibatnya, penting untuk menargetkan peluang yang akan memberikan manfaat paling besar dengan risiko teknis paling kecil. Penting juga untuk mencoba menggeneralisasi solusi semaksimal mungkin, sehingga dapat diterapkan pada berbagai kebutuhan dengan perubahan minimal.

Misalnya, terdapat banyak jenis peristiwa terkait keamanan yang dapat diidentifikasi melalui beberapa variasi deteksi anomali volumetrik, sehingga mungkin bentuk deteksi yang relatif generik yang memanfaatkan ML dan NN dapat diterapkan pada berbagai situasi dengan hasil yang baik. Pencarian bentuk analisis generik adalah tujuan utama dalam keamanan otomatis. Sementara itu, pencarian variasi dan metode serangan tak terduga adalah tujuan utama bagi para penyerang. Seperti yang dapat Anda duga dari hal ini, hal ini akan terus menjadi tantangan yang sulit.

Peran Jaringan Berbasis Perangkat Lunak

Jaringan berbasis perangkat lunak adalah penggunaan teknologi komputasi awan untuk meminimalkan ketergantungan pada perangkat keras khusus. Dengan mengurangi



ketergantungan pada perangkat keras, fungsi jaringan dapat "divirtualisasikan" (yaitu, dijalankan sebagai fungsi perangkat lunak pada perangkat keras komputer komoditas). Ada beberapa keuntungan dari virtualisasi ini yang berkaitan dengan keamanan:

1. **Masalah keamanan dapat diperbaiki dengan cepat.** Karena infrastruktur yang umum digunakan, alih-alih perangkat keras dan perangkat lunak khusus, berbagai jenis masalah keamanan dapat diperbaiki dengan perubahan perangkat lunak di seluruh infrastruktur secara relatif konsisten. Patch perangkat lunak tidak lagi bersifat khusus untuk model peralatan jaringan yang unik. Teknologi manajemen cloud otomatis dapat digunakan.
2. **Akuisisi data keamanan dapat dimodifikasi.** Umumnya, sulit untuk memprediksi jenis data yang dibutuhkan untuk mendeteksi, mengkuantifikasi, mengkarakterisasi, menyelidiki, dan memulihkan serangan. Hal ini terutama berlaku jika serangan belum ditemukan. Terdapat tipe data generik yang berguna, seperti data aliran dan log akses. Namun, selalu ada kebutuhan yang tidak dapat kita antisipasi sebelumnya. Jaringan yang ditentukan perangkat lunak (software-defined networking) dapat memungkinkan modifikasi VNF secara lebih dinamis daripada infrastruktur berbasis perangkat keras. Akuisisi data dapat dilakukan melalui modifikasi VNF yang ada. Sebagai alternatif, fungsi akuisisi data dapat dilakukan melalui service-chaining dengan fungsi yang dirancang khusus untuk menghasilkan, mengumpulkan, dan menganalisis data aktivitas jaringan yang sesuai. Sebagai contoh, kami memperkenalkan mikroperimeter pada Gambar 10.1. Mikroperimeter merupakan contoh solusi rantai layanan yang dapat membantu pengumpulan data dan menyediakan kontrol keamanan untuk fungsi-fungsi tervirtualisasi. Karena fungsi-fungsi tersebut tervirtualisasi, fungsi-fungsi tersebut dapat dimasukkan ke dalam jalur aktivitas jaringan yang sesuai saat dibutuhkan dan dihapus saat tidak dibutuhkan melalui metode jarak jauh. Jika belum jelas, mengumpulkan, menyimpan, dan menganalisis semua aktivitas jaringan tidaklah praktis dan tidak diinginkan. Jaringan yang ditentukan perangkat lunak (software-defined networking) memberikan peluang untuk menjadi seefisien dan seefektif mungkin. Hal ini memungkinkan kemampuan akuisisi dan analisis data SDN tertentu menjadi pembeda antara penyedia layanan jaringan. Dan fleksibilitasnya memberikan banyak peluang yang akan berguna untuk kontrol loop tertutup (yaitu, otomatis).
3. **Kontrol keamanan dapat diubah.** Karena kontrol keamanan baru diperlukan untuk menangani jenis aktivitas serangan jaringan tertentu, kontrol baru dapat diimplementasikan sebagai bagian dari vNF atau (sekali lagi) melalui rantai layanan fungsi-fungsi jaringan khusus. Misalnya, mungkin praktis untuk mengelompokkan protokol tertentu ke dalam fungsi yang dapat menyaring tanda-tanda spesifik dari lalu lintas berbahaya. Jaringan yang ditentukan perangkat lunak memberikan peluang untuk menjadi seefisien dan seefektif mungkin dengan menyesuaikan fungsi pemrosesan dengan kebutuhan spesifik pada saat itu. Hal ini memungkinkan kemampuan kontrol SDN yang spesifik menjadi pembeda antara penyedia layanan jaringan.



Jaringan yang ditentukan perangkat lunak menyediakan kerangka kerja penting untuk otomatisasi yang lebih besar dalam kontrol keamanan. SDN dapat dikonfigurasi untuk menghasilkan data yang dibutuhkan untuk menggunakan teknik AI dan ML sebagai bagian dari solusi keamanan. SDN juga menyediakan sarana untuk mengontrol dan mengadaptasi kapabilitas jaringan dan sistem secara terprogram sebagai respons terhadap arahan AI dan ML. Ketika kita menggunakan metode otomatis untuk mengamati kondisi dan aktivitas, menghitung tindakan korektif atau penyesuaian, dan menerapkan perubahan tanpa campur tangan manusia, hal ini terkadang disebut sebagai kontrol loop tertutup.

Ada pula pertimbangan praktis. Mengubah vNF dapat menyebabkan perubahan perilaku kinerja fungsi jaringan. Menambahkan fungsi baru kemungkinan akan menghabiskan lebih banyak sumber daya CPU atau memori. Fungsi rantai layanan kemungkinan akan menimbulkan lebih banyak latensi dan kompleksitas pada alur jaringan. Karena alasan ini, penting untuk menyeimbangkan nilai dari penerapan kontrol keamanan dan mempertimbangkan apakah ada metode lain yang lebih efisien untuk mencapai tujuan keamanan yang dibutuhkan. Dengan kata lain, jaringan merupakan bagian penting dari keamanan yang baik, tetapi jaringan mungkin bukan jawaban terbaik untuk semua masalah keamanan.

Evolusi Strategi Keamanan Perusahaan

Kita telah membahas manfaat SDN sebagai bagian dari solusi keamanan jaringan. Namun, perusahaan juga sedang bertransformasi. Pada perusahaan lama, sebagian besar aplikasi dihosting di pusat data perusahaan. Kontrol di sekitar perimeter jaringan perusahaan digunakan untuk melindungi aplikasi dari internet. Komunikasi dengan aplikasi tersebut ke dan dari klien perusahaan (misalnya, karyawan) dibatasi aksesnya di dalam jaringan perusahaan. Perusahaan modern telah menjadi jauh lebih kompleks.

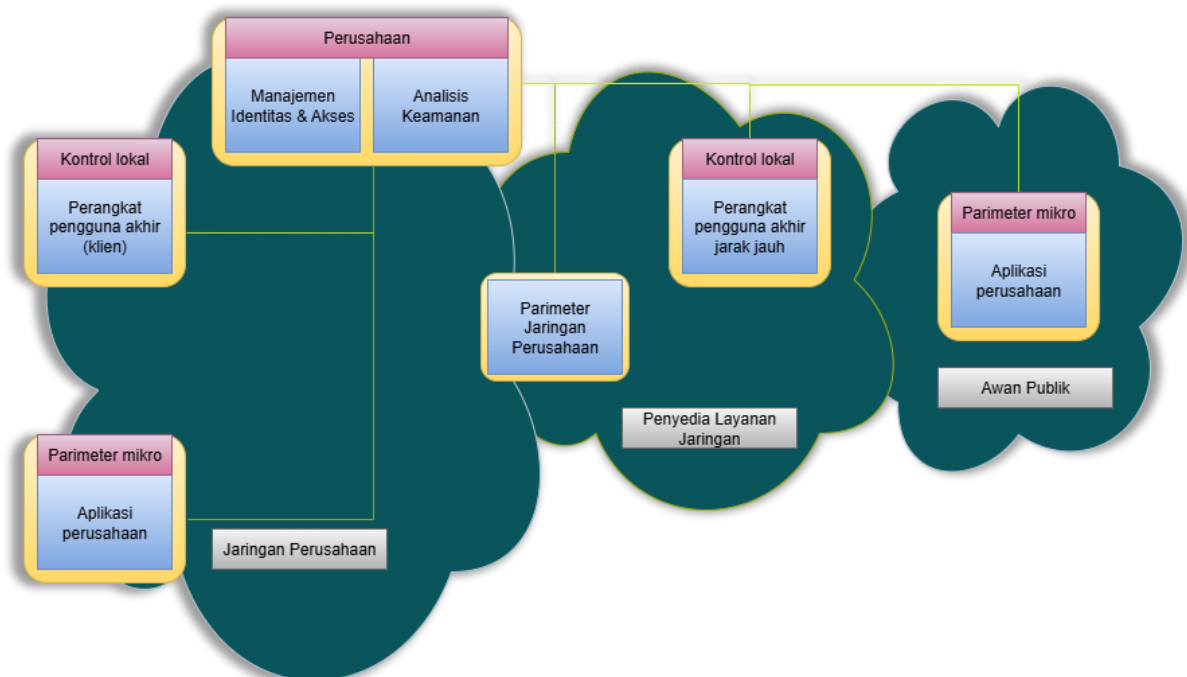
Aplikasi dihosting di jaringan mitra. Aplikasi disediakan sebagai layanan cloud publik yang dihosting di internet. Karyawan perusahaan bekerja dari jarak jauh dan menggunakan perangkat seluler untuk mengakses aplikasi perusahaan, yang berarti membuat aplikasi perusahaan dapat diakses dari internet. Struktur dan pendekatan keamanan untuk perusahaan terpaksa berubah untuk mengakomodasi perubahan struktur perusahaan ini. Meskipun konsep jaringan perusahaan kemungkinan akan tetap ada untuk beberapa waktu, peran dan efektivitas perimeter jaringan perusahaan terus menurun.

Area umum yang perlu ditekankan untuk mengakomodasi perubahan ini diilustrasikan pada Gambar 10.3, yang meliputi:

- ❖ **Virtualisasi Keamanan:** Kontrol keamanan harus diintegrasikan lebih dekat ke aplikasi (ala mikro-perimeter), ke infrastruktur cloud (ala mikro-perimeter), ke klien yang terhubung (misalnya, perangkat anti-malware), dan ke jaringan (ala Jaringan Area Luas Terdefinisi Perangkat Lunak [SD-WAN]).
- ❖ **Autentikasi dan manajemen akses yang kuat** diperlukan untuk mengontrol izin akses aplikasi, terlepas dari di mana atau bagaimana aplikasi dihosting. Kontrol akses berbasis kata sandi terlalu rentan untuk melindungi aplikasi sensitif.



- ❖ **Analisis Keamanan:** Analisis keamanan diperlukan untuk memfasilitasi deteksi status dan perilaku keamanan yang berisiko secara holistik (ala memverifikasi kontrol keamanan sesuai harapan dan efektif). Analisis ini juga berfungsi untuk mendeteksi, menyelidiki, dan merespons ancaman di seluruh infrastruktur, aplikasi, dan aset data perusahaan. Secara tradisional, kami menyebutnya "analisis ancaman", tetapi ada banyak peluang untuk menelusuri data aktivitas jaringan guna menemukan aktivitas yang relevan dengan keamanan yang tidak terkait dengan ancaman. Misalnya, mengidentifikasi dan menghapus peralatan di jaringan yang tidak diketahui oleh sistem kepatuhan sangat membantu untuk memastikan tidak ada kerentanan yang tidak perlu.
- ❖ **Orkestrasi Keamanan:** Seiring dengan semakin beragamnya perusahaan, kontrol keamanan yang diterapkan juga semakin beragam. Untuk mengelola keamanan perusahaan, diperlukan serangkaian kontrol keamanan terpusat untuk mengelola postur keamanan perusahaan. Hal ini sering disebut sebagai "orkestrasi keamanan". Misalnya, seiring diperkenalkannya firewall mikro-perimeter untuk melindungi aplikasi, kontrol ini perlu dikonfigurasi bersama dengan firewall perusahaan lama dan kontrol keamanan di cloud. Hal ini perlu dilakukan secara konsisten dan dinamis untuk mendukung kebutuhan bisnis yang dipadukan dengan postur keamanan yang baik.



Gambar 10.3 Orkestrasi keamanan menyediakan manajemen keamanan holistik di seluruh aset perusahaan.



Motivasi untuk Otomatisasi Kecerdasan Buatan dalam Keamanan

Tujuan utama kami adalah otomatisasi. Penggunaan AI dan ML adalah cara untuk mencapai tujuan ini. Mari kita bahas beberapa hal yang perlu kita perhatikan dalam peluang otomatisasi keamanan AI.

1. **Skalabilitas:** Beberapa pekerjaan dapat dilakukan secara manual secara individual, tetapi tidak dapat diskalakan secara efektif jika terdapat ribuan atau jutaan skenario atau contoh yang perlu dipertimbangkan. Misalnya, sebelumnya kita membahas perlunya kemampuan untuk mengikuti perkembangan serangan baru sekaligus melacak metode serangan lama. Ini adalah tujuan skalabilitas. Kami juga mencatat peningkatan jumlah perangkat IoT yang akan terhubung ke jaringan, yang akan meningkatkan skala kejadian dan masalah keamanan yang mungkin terjadi. Ada juga jenis skalabilitas lain yang akan mengarahkan kita pada solusi AI. Kita akan membahas hal ini dalam konteks langkah-langkah keamanan preventif nanti.
2. **Amplifikasi Talenta:** Sudah menjadi rahasia umum di industri bahwa talenta keamanan siber sangat berharga. Fakta ini akan tetap berlaku untuk beberapa waktu. Dan dalam kondisi terkini, penerapan teknologi ML dan AI yang efektif masih merupakan disiplin ilmu yang relatif terfokus. Dengan kata lain, kita belum memiliki kemampuan untuk menggunakan AI sebagai sarana untuk menginterpretasikan tantangan keamanan siber baru dan secara otomatis menentukan teknologi ML yang tepat untuk diterapkan pada kebutuhan spesifik. Namun, pemanfaatan talenta keamanan siber yang tersedia untuk membantu mengarahkan bagaimana dan di mana menerapkan teknik ML tertentu dapat dicapai. Jika diterapkan secara efektif, hal ini memungkinkan talenta tersebut untuk berupaya mengatasi masalah baru tanpa harus sering mempertahankan kemampuan yang ada. Kita dapat memperkuat kemampuan beberapa orang untuk memecahkan dan beradaptasi dengan lebih banyak tantangan keamanan.
3. **Kompleksitas Praktis:** Beberapa masalah terlalu rumit untuk dipahami atau dipecahkan oleh manusia biasa. Sebagai contoh sederhana, pertimbangkan dimensi matriks. Orang biasanya mampu memahami matriks dua dimensi dan banyak yang dapat memvisualisasikan tiga dimensi dengan relatif baik. Kebanyakan manusia tidak mampu mempertimbangkan bagaimana puluhan atau ratusan dimensi diterjemahkan menjadi solusi spesifik. Namun, penggunaan ML terarah dapat membantu mengembangkan hubungan yang andal dengan skenario dan membantu menentukan atribut data atau dimensi apa yang relevan tanpa memerlukan pemahaman manusia. Ketika mempertimbangkan masalah keamanan seperti analisis malware, manusia mungkin kesulitan memahami berbagai cabang aktivitas dan mengenali teknik pengkodean yang digunakan oleh pengembang malware. Kompleksitas ini mendorong penggunaan metode berbasis aturan dan mungkin analisis kognitif.
4. **Kemudahan:** Ada batasan untuk apa yang dapat dilakukan dengan manusia karena mereka hanya dapat bekerja dengan kecepatan tertentu. Dalam beberapa skenario keamanan siber, kemudahan dalam merespons serangan dapat menjadi pembeda antara gangguan dan pelanggaran besar. Penting untuk bertindak cepat ketika



merespons insiden. Sistem yang merancang respons (setidaknya rekomendasi) berpotensi jauh lebih cepat dan komprehensif daripada bergantung pada manusia untuk mengumpulkan, menafsirkan, dan menganalisis data dan kemudian menyusun bagian-bagian respons. Metode otomatis untuk deteksi dan respons sering digunakan untuk memitigasi serangan DDoS. Dalam sebagian besar keadaan, serangan dapat dideteksi dengan mengukur aktivitas relatif terhadap kapasitas normal dan maksimum sumber daya yang dilindungi (misalnya, situs web). Mitigasi umumnya dapat dilakukan dengan mengenali pola dalam jenis serangan standar dan membuang paket yang sesuai dengan pola tersebut. Dalam kasus seperti ini, bahkan kemampuan mitigasi sederhana yang dilakukan secara otomatis dan cepat pun memiliki nilai. Jika serangan berlanjut, manusia dapat meninjau mitigasi dan menyesuaikannya dengan nuansa apa pun dalam aktivitas serangan untuk pemulihan layanan yang optimal.

5. **Akurasi:** Dengan asumsi model ML dikalibrasi dengan tepat dan dapat mempelajari respons yang tepat untuk skenario tertentu, setidaknya model tersebut akan berperilaku konsisten. Di sisi lain, manusia rentan terhadap kesalahan. Hal ini terutama berlaku dalam kasus di mana manusia hanya memiliki kesempatan terbatas untuk berlatih. Kemampuan manusia bergantung pada pengetahuan, pengalaman, dan keterampilan setiap individu. Dan hanya ada kemampuan terbatas (melalui pelatihan) untuk mentransfer pengetahuan dari satu manusia ke manusia lainnya. Hal ini sangat penting untuk keamanan, karena serangan tertarget yang sebenarnya mungkin sangat jarang. Seseorang mungkin hanya menghadapi jenis ancaman tertentu beberapa kali. Jika mesin diajari suatu respons, ia akan mengingat respons tersebut dan cenderung mengeksekusi respons tersebut secara konsisten meskipun respons tersebut salah.

Seperti yang Anda lihat, ada banyak motivasi untuk menggunakan metode otomatisasi canggih guna memecahkan masalah keamanan. Saat kita memilih aktivitas kandidat untuk menerapkan otomatisasi, akan bermanfaat untuk memahami manfaat apa yang akan diperoleh dari upaya tersebut.

Pertimbangan untuk Mendorong Peluang Otomatisasi Kecerdasan Buatan

Pertimbangan tambahan dan agak acak berikut ini mungkin berguna untuk menetapkan aktivitas otomatisasi keamanan praktis:

1. **Rutinitas tidak boleh menjadi rutinitas:** Setiap kali praktisi keamanan melakukan aktivitas rutin, ini merupakan peluang yang jelas untuk mempertimbangkan bagaimana otomatisasi dapat diterapkan. Perhatikan aktivitas yang paling sering dilakukan tim Anda dan pertimbangkan aspek apa dari aktivitas tersebut yang serupa setiap harinya. Dengan semakin matangnya teknologi AI, cakupan aktivitas kandidat yang dapat dipertimbangkan untuk diotomatisasi dapat meluas. Tidak perlu lagi mendefinisikan aturan atau prosedur khusus untuk melakukan otomatisasi. Misalnya, setiap kali peringatan dihasilkan oleh sistem yang memerlukan validasi atau investigasi lebih lanjut, terdapat langkah-langkah yang relatif standar yang akan dilakukan oleh seorang analis. Tujuan analis adalah untuk menentukan apakah kejadian tersebut positif-benar (yaitu, memastikan benar-benar ada kejadian keamanan yang memerlukan respons),



mengembangkan interpretasi tentang apa yang terjadi, menentukan apakah investigasi lebih lanjut diperlukan, dan menentukan respons yang tepat. Analisis akan dimulai dengan mengumpulkan informasi tentang target dan sumber potensial serangan sebelum dan sesudah peristiwa terdeteksi. Informasi kontekstual lebih lanjut akan dikumpulkan. Informasi tersebut akan diinterpretasikan untuk mendapatkan petunjuk tambahan. Langkah-langkah ini (hingga batas tertentu) relatif standar dan dapat dilakukan oleh sistem secara otomatis dan segera setelah mendeteksi peringatan. Proses ini dapat terus memantau dan melaporkan target dan sumber. Analisis berbasis aturan yang dipadukan dengan pembelajaran mesin kognitif dapat digunakan untuk membantu mengembangkan interpretasi. Kemampuan ini juga dapat berkembang untuk merekomendasikan dan mungkin memulai respons. Semua hal ini dapat menghemat tenaga kerja yang signifikan, meningkatkan akurasi respons, dan mungkin yang terpenting, meningkatkan kecepatan respons. Sebagian besar proses deteksi dan respons standar dapat diotomatisasi. Hal ini memungkinkan praktisi keamanan untuk berfokus pada aktivitas yang lebih kompleks serta pengembangan jenis kemampuan baru.

2. **Penyerang kreatif, DAN mereka ahli dalam otomatisasi:** Seperti yang telah kita bahas sebelumnya, penyerang terus mengembangkan teknik serangan baru. Penyerang juga terbukti terus mengembangkan kemampuan alat serangan mereka. Misalnya, malware baru sering kali menggunakan satu metode eksploitasi untuk mendapatkan akses guna meningkatkan hak istimewa pada suatu sistem. Seiring perkembangan malware, pengembang malware terus mengembangkan malware untuk menggabungkan, memprioritaskan, dan menggunakan metode eksploitasi baru. Sebagian besar eksploitasi saat ini bersifat otomatis, di mana jaringan dipindai untuk mencari sistem yang berpotensi rentan, target potensial diprofilkan berdasarkan kelemahannya, dan eksploitasi yang tepat dilakukan. Setelah itu, pijakan dibangun dengan memuat malware perintah dan kontrol secara otomatis. Sistem target dapat diprofilkan berdasarkan kapabilitasnya (misalnya, memori, penyimpanan, perangkat lunak yang menarik, data), yang kemudian dilaporkan kembali kepada penyerang. Semua ini terjadi dalam hitungan detik dan dapat merekrut ribuan perangkat ke dalam botnet dalam hitungan jam. Hanya ada situasi langka di mana penyerang merevolusi strategi serangan. Umumnya, mereka membuat perubahan dan peningkatan kecil secara bertahap dari waktu ke waktu memperbaiki apa yang mereka ketahui sudah berfungsi. Organisasi yang mengotomatiskan fungsi perlindungan keamanan dapat memperoleh manfaat dengan mengikuti arahan para penyerang. Teruslah mencari peluang untuk mengotomatisasi. Pertimbangkan alat baru untuk memfasilitasi otomatisasi. Mulailah dengan memecahkan masalah kecil dan sederhana, lalu tingkatkan kompleksitasnya seiring waktu. Terus tingkatkan tingkat otomatisasi dalam solusi. Jangan pernah berhenti mengotomatisasi.
3. **Pekerja cenderung memiliki pengetahuan domain yang terbatas:** Umumnya, praktisi keamanan akan memiliki pengetahuan yang luas tentang topik keamanan siber dan



akan mempraktikkan keamanan siber. Hal ini tidak berarti mereka harus memiliki pengetahuan yang memadai tentang ML, AI, atau bahkan keterampilan pemrograman. Demikian pula, personel yang memiliki pengetahuan yang luas tentang pemrograman, ML, dan/atau AI cenderung bukan pakar keamanan. Masing-masing bidang teknologi ini sangat kompleks dan membutuhkan upaya yang signifikan untuk dipelajari secara memadai. Lebih lanjut, bukan hal yang aneh bagi organisasi untuk bersikap bias dalam praktik perekrutan mereka dan mengabaikan kebutuhan akan keberagaman keterampilan tenaga kerja. Misalnya, seorang manajer perekrutan di organisasi keamanan mungkin memiliki pengalaman dan pengetahuan yang signifikan di bidang keamanan. Mereka cenderung mengidentifikasi diri dengan orang lain yang memiliki latar belakang dan keterampilan yang serupa. Upaya yang disengaja diperlukan untuk memastikan organisasi mengembangkan keberagaman keterampilan yang sesuai dan mengikutsertakan kandidat ilmu data dan rekayasa perangkat lunak untuk bekerja bersama praktisi keamanan. Ingat kembali pertimbangan sebelumnya di mana kita tidak perlu berhenti mengotomatisasi. Peran dan aktivitas baru ini bersifat permanen dan tidak boleh diperlakukan sebagai aktivitas konsultasi.

4. **Mesin tidak mengetahui segalanya:** Setidaknya belum. Penting untuk mempelajari detail tentang bagaimana keputusan dibuat ketika mencoba mengotomatisasi aktivitas. Hal ini terutama penting ketika mencoba menggunakan otomatisasi tingkat lanjut seperti AI kognitif untuk membuat keputusan. Dalam beberapa keadaan, AI dapat membantu mengidentifikasi karakteristik yang tidak langsung terlihat oleh manusia. Misalnya, AI mungkin dapat dengan mudah mengenali karakteristik paket jaringan yang tidak valid dan dapat dipertimbangkan untuk diblokir. Namun, situasi sebaliknya juga sering terjadi. Menentukan bahwa paket yang tidak valid merupakan bagian dari suatu serangan bisa jauh lebih sulit ditentukan melalui analisis data. Seringkali diperlukan penggunaan kumpulan data dan informasi lain untuk menyimpulkan penyebab situasi dan tingkat ancaman yang ditimbulkannya. Karena beberapa langkah analisis konvensional bersifat manual, hal ini memungkinkan manusia untuk mengakses sistem lain untuk mendapatkan informasi, melakukan panggilan telepon, dan mewawancarai orang lain. Hal ini menyebabkan kebutuhan akan sistem analisis keamanan yang menerapkan AI. Sistem ini mungkin perlu mengumpulkan serangkaian informasi yang jauh lebih luas untuk membantu diagnosis peristiwa secara otomatis. Mungkin diperlukan ketersediaan data konfigurasi sistem, informasi pemilik sistem, data kerentanan, dan informasi eksploitasi untuk membantu pengambilan keputusan yang lebih konklusif. Hal ini tidak berarti bahwa sistem analisis keamanan perlu mengumpulkan semua data setiap saat, tetapi sistem tersebut mungkin memerlukan serangkaian informasi yang sangat luas agar tersedia sesuai kebutuhan.
5. **Penyerang ahli tidak menciptakan anomali besar:** Kita perlu menyadari bahwa terdapat beragam jenis serangan yang ada karena berbagai motivasi. Beberapa penyerang sengaja menarik perhatian. Beberapa penyerang berusaha keras untuk



menghindari ketahuan. Kita harus menghindari godaan untuk menggeneralisasi serangan secara berlebihan dan metode yang diperlukan untuk mendeteksi berbagai serangan yang ada secara efektif. Lebih lanjut, kita perlu mengantisipasi bahwa serangan tingkat lanjut yang sangat canggih sengaja dirancang agar terlihat seperti aktivitas normal dan mungkin secara praktis mustahil untuk dideteksi melalui analisis aktivitas. Hal ini menunjukkan bahwa metode terbaik mungkin adalah mencegah serangan agar tidak berhasil daripada mencoba mendeteksi dan merespons serangan. Fokuslah pada pencegahan terlebih dahulu. Dan pertimbangkan berbagai pemeriksaan integritas dalam infrastruktur sebagai sesuatu yang kooperatif dan pelengkap untuk analisis aktivitas. Misalnya, tujuan utama penyerang adalah mendapatkan akses ke kredensial pengguna (nama pengguna dan kata sandi) dari pengguna yang sah di suatu sistem, dan mendapatkan akses ke sistem yang terekspos ke Internet melalui protokol akses jarak jauh seperti *secure shell* (SSH) atau *remote desktop protocol* (RDP). Dengan kemampuan ini, sangat mudah bagi penyerang untuk terlihat seperti pengguna yang sah. Tidak ada sistem deteksi intrusi yang akan mendeteksi eksploitasi. Dengan sedikit perawatan, tidak ada anomali aktivitas jaringan yang akan dipicu.

Bagaimana seseorang melindungi diri dari ancaman seperti ini? Alat seperti autentikasi dua faktor, yang membatasi akses dari Internet misalnya, membatasi akses ke ruang alamat protokol Internet (IP) yang diperlukan dapat mencegah penyerang mencuri dan menggunakan kredensial yang sah. Pemeriksaan integritas dalam sistem dapat membantu mengidentifikasi perubahan atau aktivitas dalam sistem yang tidak sesuai dengan penggunaan yang diharapkan. Pembelajaran mesin dapat digunakan untuk membuat profil titik akses normal (alamat IP sumber) bagi pengguna, dan ini dapat digunakan untuk membantu membuat kebijakan akses yang lebih terbatas dan lebih aman, sehingga menolak akses penyerang. Pembelajaran mesin dapat digunakan untuk membuat profil konfigurasi normal sistem, yang dapat membantu menandai perubahan yang berada di luar normal. Kedua jenis pembuatan profil ini tidak akan praktis untuk ratusan ribu sistem tanpa otomatisasi yang signifikan.

6. **Penyerang memiliki tujuan untuk menjadi subversif:** Karena kita berupaya mengotomatiskan keamanan dan menerapkan kontrol loop tertutup (yaitu, implementasi otomatis kontrol keamanan), tidaklah diinginkan bagi penyerang untuk dapat menggunakan kontrol keamanan loop tertutup untuk mencapai tujuan mereka sendiri. Penyerang dapat mencoba secara sengaja menciptakan aktivitas yang tampak seperti aktivitas serangan untuk mengaktifkan tindakan pencegahan loop tertutup dan dengan demikian menyebabkan dampak pada aktivitas yang sah. Sebaliknya, penyerang dapat mencoba menggunakan aktivitas seperti serangan yang halus dan berjangka panjang dengan tujuan melatih alat deteksi serangan untuk menerima aktivitas serangan sebagai aktivitas normal. Penggunaan teknik AI dan ML perlu mempertimbangkan hal ini. Mungkinkah musuh melatih sistem untuk menerima perilaku jahat sebagai hal yang normal? Apa tindakan pencegahan untuk hal ini? Misalnya, jika suatu solusi ditujukan untuk memblokir aktivitas berbahaya setelah



terdeteksi, penyerang mungkin menyuntikkan aktivitas ke dalam aliran aktivitas "normal" yang sengaja terlihat seperti aktivitas berbahaya. Objek tersebut dapat menyebabkan otomatisasi loop tertutup memblokir aktivitas berbahaya tersebut sekaligus memblokir aktivitas yang sah. Hal ini dapat mengakibatkan gangguan pada layanan normal. Menarik untuk dicatat bahwa pertimbangan ini tidak hanya berlaku untuk kontrol keamanan. Penyerang mungkin memanipulasi indikator kinerja normal untuk menciptakan hasil yang diinginkan. Misalnya, jika penyerang dapat menciptakan ilusi bahwa vNF sedang berada di bawah beban berat (meskipun sebenarnya tidak), respons loop tertutup dapat menyebabkan lebih banyak sumber daya dialokasikan ke fungsi ini. Karena jumlah sumber daya yang tersedia terbatas, operasi loop tertutup mungkin terpaksa mengelola prioritas dan menonaktifkan beberapa fungsi tambahan demi menghindari gangguan layanan. Mungkin salah satu fungsi tambahan tersebut adalah deteksi penipuan, penghitungan penggunaan, atau deteksi ancaman.

Metodologi Implementasi Umum

Kelincahan dalam aplikasi keamanan sangat penting untuk kesuksesan. Manfaat dari kontrol keamanan yang tertinggal jauh dari teknologi yang dimiliki oleh aktor ancaman hanya terbatas. Oleh karena itu, proses yang digunakan untuk mengembangkan teknologi perlindungan keamanan harus mampu beradaptasi dengan cepat dan efektif terhadap situasi baru. Ilustrasi pada Gambar 10.4 adalah model yang dapat digunakan untuk membantu implementasi solusi dengan cepat terutama dalam konteks analisis ancaman.

Tujuan dari proses ini adalah untuk membuat fungsionalitas keamanan baru ke dalam kondisi yang dapat digunakan sesegera mungkin. Proses ini terdiri dari langkah-langkah berikut:

- **Definisi Tujuan:** Langkah ini untuk mendefinisikan kebutuhan atau persyaratan. Misalnya, anggaplah metode eksploitasi baru telah diidentifikasi, dan ada kebutuhan untuk mendeteksi upaya penggunaan eksploitasi tersebut. Pernyataan tujuan dapat menghipotesiskan data relevan apa yang akan digunakan untuk mendeteksinya, metode umum yang akan digunakan, apa yang akan dilakukan jika eksploitasi terdeteksi, dan bagaimana hasil analisis harus disajikan. Dalam situasi mendesak, proses untuk mengembangkan dan menetapkan tujuan mungkin hanya membutuhkan waktu beberapa menit. Tidak perlu formal, tetapi perlu ada tujuan yang jelas.
- **Riset/Studi:** Tujuan studi riset adalah untuk menentukan seberapa valid pernyataan tujuan tersebut. Ini merupakan analisis satu kali sebelum mengodekan solusi otomatis apa pun. Biasanya akan membantu jika dimulai dengan kumpulan data sampel yang diketahui untuk studi. Data ini harus mencakup data aktivitas nominal serta data yang mewakili jenis aktivitas yang perlu diidentifikasi. Biasanya, data "berlabel" yang sebenarnya tidak tersedia, sehingga verifikasi penuh bahwa metode deteksi akan berhasil mungkin tidak dapat dilakukan. Sintesis data mungkin diperlukan untuk memverifikasi deteksi. Studi yang menggunakan data nominal juga dapat membantu menentukan apakah akan ada banyak deteksi positif palsu (yaitu, peringatan yang tidak mewakili eksploitasi yang sebenarnya). Jika terdapat terlalu banyak positif palsu,



metode deteksi perlu disempurnakan atau direvisi. Dalam situasi mendesak, studi dapat dilakukan dalam hitungan jam. Namun, untuk masalah yang lebih umum, kumpulan data yang lebih besar dan studi yang lebih banyak mungkin diperlukan. Hal ini mungkin membutuhkan waktu berminggu-minggu.

- **Praproduksi:** Setelah studi yang berhasil dilakukan, langkah selanjutnya adalah mengodekan solusi dengan cara yang sangat sederhana. Solusinya mungkin belum sempurna atau diformat dengan baik. Namun, solusi tersebut mungkin terbukti bermanfaat. Langkah ini sangat mirip dengan proses pengembangan agile pada umumnya, yang bertujuan untuk membuktikan solusi lebih awal (atau gagal lebih awal). Lebih lanjut, tujuan di sini adalah untuk membuat solusi tersebut benar-benar digunakan dengan data nyata. Proses ini mungkin melibatkan reproduksi studi satu kali dalam batch kecil yang berulang dan menghasilkan laporan keluaran berkala untuk diperiksa. Dalam banyak kasus, kami telah berhasil menerapkan kemampuan deteksi baru dalam hitungan jam dan dapat memantau ancaman baru 24/7 dengan keyakinan relatif bahwa kami mampu mendeteksi dan merespons ancaman tersebut. Solusi praproduksi dapat tetap beroperasi selama sehari-hari, berminggu-minggu, atau bahkan berbulan-bulan, tergantung kebutuhan. Penyesuaian dan penyempurnaan dapat dilakukan di sepanjang proses untuk lebih menyempurnakan solusi. Sementara itu, jika diinginkan, solusi produksi penuh dapat direkayasa untuk mengakomodasi kebutuhan operasional jangka panjang.
- **Pengembangan Produksi:** Berdasarkan solusi praproduksi, persyaratan untuk solusi produksi dapat dipahami dengan lebih baik. Selalu ada peluang untuk perbaikan (misalnya, hasil yang lebih real-time, peningkatan granularitas dalam analisis, ambang batas yang dapat disesuaikan sendiri, pengumpulan detail tambahan, dan penyempurnaan untuk meminimalkan positif palsu, sehingga operasi menjadi lebih andal). Mungkin beberapa kode dari solusi praproduksi dapat digunakan kembali dalam solusi produksi.
- **Dukungan Siklus Hidup:** Setiap solusi yang diterapkan memerlukan dukungan siklus hidup agar tetap layak. Hal ini khususnya berlaku untuk solusi deteksi ancaman, di mana variasi baru pada eksploitasi ditambahkan atau disesuaikan dalam serangan. Algoritme/metode deteksi juga harus disesuaikan untuk mengakomodasi perubahan tersebut. Lingkungan operasi juga berubah seiring waktu. Aplikasi baru yang sah dapat menimbulkan deteksi positif palsu. Jika solusi produksi dirancang dengan baik, metodenya akan cukup umum untuk mendeteksi variasi serangan baru secara akurat. Namun, hal ini tidak selalu terjadi.





Gambar 10.4 Proses umum untuk implementasi dan adopsi cepat teknologi keamanan baru.

10.3 PENERAPAN AI DAN ML UNTUK KEAMANAN

Kontrol Preventif Berdasarkan Pemodelan Perilaku Aplikasi Normal

Untuk contoh ini, kami akan berfokus pada kontrol kebijakan keamanan jaringan. Umumnya, tampaknya ada kecenderungan untuk menerapkan pembelajaran mesin (ML) untuk mengidentifikasi ancaman (seperti mencari perilaku orang jahat). Memang, kami akan membahas penerapan ML dalam mencari ancaman (seperti analitik ancaman). Namun, setiap program keamanan harus menekankan metode untuk mencegah keberhasilan serangan sejak awal.

Salah satu cara untuk mencegah ancaman adalah dengan mengizinkan hal-hal yang diinginkan terjadi pada aplikasi dan menggunakan lapisan kontrol (misalnya, aturan firewall, izin akses pengguna, izin berkas, fitur perangkat lunak yang dapat diakses) untuk membantu memblokir hal-hal yang tidak diinginkan. Konsep menetapkan aturan firewall yang seketat mungkin tanpa menghambat fungsi aplikasi umumnya diakui sebagai praktik terbaik untuk keamanan. Namun, pertumbuhan abstraksi pengembangan, alat plug-in, dan pustaka pengembangan yang kompleks telah mengisolasi pengembang aplikasi dari fungsi jaringan. Akibatnya, pengembang aplikasi tidak tahu port, protokol, dan alamat mana yang diperlukan agar aplikasi mereka beroperasi. Akibatnya, sesuatu yang mendasar seperti aturan firewall umumnya tidak dirancang dengan baik untuk seketat mungkin.

Dengan ribuan aplikasi di perusahaan dan interaksi antar aplikasi, mustahil bagi administrator firewall untuk mengelola aturan firewall secara efektif dengan cara yang paling terbatas. Ingat kembali diskusi kita tentang virtualisasi keamanan, di mana kita memiliki kesempatan untuk menerapkan firewall dalam mikroperimeter di sekitar beban kerja individu di perusahaan dan dalam infrastruktur jaringan. Akan ada ribuan (mungkin ratusan ribu) mikroperimeter yang tersedia untuk dikonfigurasi. Hal ini semakin menciptakan tantangan penskalaan, yang merupakan kesempatan bagi ML untuk membantu mengamati perilaku, merumuskan kontrol, dan menetapkan kontrol yang didasarkan pada perilaku normal aplikasi.

Kami telah menyatakan sebelumnya bahwa karakteristik penting untuk penggunaan ML adalah memiliki data yang dapat diamati untuk dipelajari. Mengkarakterisasi perilaku normal aplikasi adalah (atau setidaknya seharusnya) aktivitas paling umum yang tersedia



untuk diamati. Peluang terbaik untuk penggunaan ML adalah untuk membuat profil perilaku normal aplikasi, merumuskan dan menerapkan kontrol pencegahan, serta meminimalkan kebutuhan untuk mendeteksi dan merespons ancaman.

Perilaku aplikasi normal jauh lebih mudah diamati, yang merupakan komponen penting ML. Pengamatan ancaman umum mungkin dapat diamati, tetapi ancaman tingkat lanjut tidak mudah diamati. Mengharapkan mesin untuk memahami dan mengantisipasi ancaman tingkat lanjut adalah hal yang lemah dan berisiko. Mustahil untuk memprediksi kapan ancaman baru akan muncul dan apa yang akan coba dilakukan oleh ancaman tersebut. Namun, adalah mungkin untuk "mempelajari" perilaku normal untuk suatu aplikasi, sistem, atau jaringan.

Pembelajaran mesin harus diterapkan untuk menentukan perilaku normal sistem dan menggunakan informasi tersebut untuk memperketat aturan operasi sistem yang diizinkan (memblokir port yang tidak diperlukan, mengontrol alamat IP yang diizinkan ke/dari, mengidentifikasi pola konten paket yang diterima, menerapkan batas volumetrik, membatasi proses yang berjalan, menandai deviasi untuk validasi).

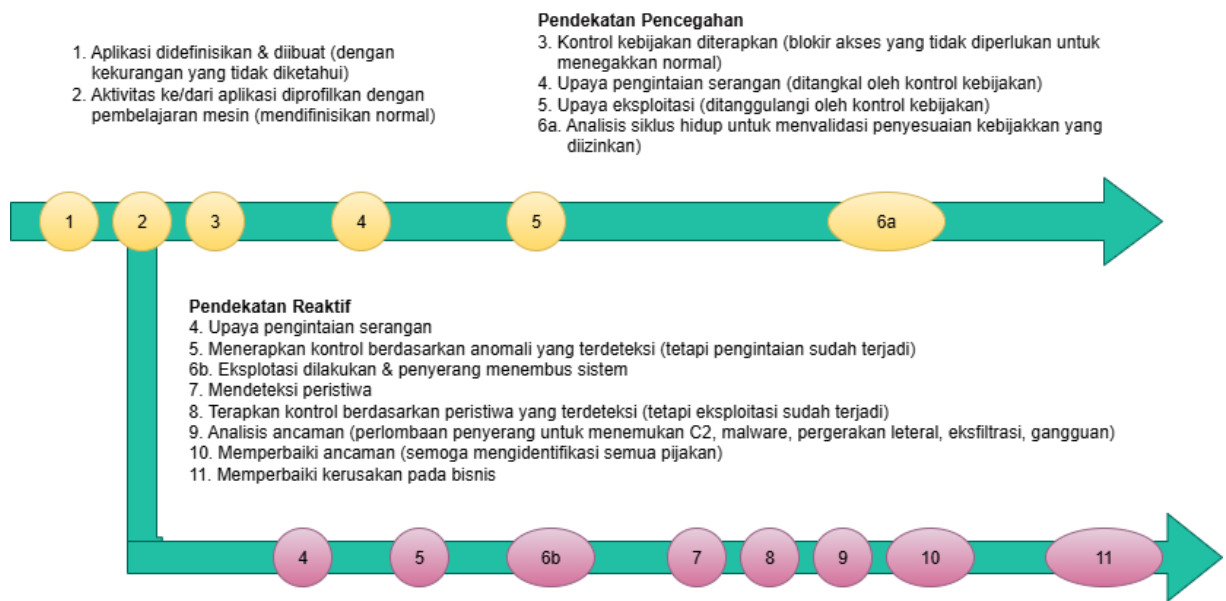
Ada nuansa dalam konsep ini. Misalnya:

1. Selalu ada aktivitas penyalahgunaan yang "normal" dalam sistem apa pun terutama jika sistem tersebut terekspos ke Internet. Dengan kata lain, sistem selalu diserang (misalnya, perintah masuk akan menjadi sasaran tebakan kata sandi). Aktivitas yang "biasanya buruk" tidak boleh dimasukkan dalam model untuk dianggap dapat diterima.
2. Akan ada kasus pengecualian yang merupakan skenario operasional yang valid untuk suatu sistem (misalnya, volume penjualan akan lebih tinggi di musim liburan). Definisi yang baik tentang "dapat diterima" harus memperhitungkan perbedaan yang dapat diterima dari perilaku normal.

Pertanyaannya adalah, bisakah kita membuat proses ML yang cukup cerdas untuk menyaring "biasanya buruk" dan mengizinkan "abnormal valid"? Ini adalah masalah yang lebih mudah daripada mencoba memprediksi serangan baru atau mendeteksi eksploitasi kerentanan yang tidak diketahui (zero day).

Mengkarakterisasi perilaku normal untuk aplikasi otomatis kemungkinan jauh lebih dapat diprediksi daripada mengkarakterisasi perilaku normal untuk perilaku manusia. Manusia (dan aktivitas yang perlu mereka lakukan) kurang dapat diprediksi dan akan berubah di luar kendali manajemen konfigurasi normal.





Gambar 10.5 Pendekatan preventif memiliki lebih sedikit langkah, lebih sederhana, tidak memaksa responden untuk berpacu dengan penyerang, dan memiliki dampak yang lebih kecil terhadap bisnis.

Gambar 10.5 menyajikan dua linimasa komparatif yang mempertimbangkan pendekatan preventif dan pendekatan reaktif terhadap keamanan. Hal ini dimaksudkan untuk lebih memperkuat pembahasan sebelumnya. Secara khusus, linimasa dalam pendekatan reaktif ditentukan oleh penyerang, yang tidak diinginkan. Lebih lanjut, jumlah pekerjaan yang dibutuhkan untuk melindungi dari ancaman jauh lebih sedikit ketika menggunakan pendekatan pencegahan.

Perbedaan mendasar antara keduanya adalah ketika kontrol kebijakan diimplementasikan dalam solusi. Dalam pendekatan reaktif, kita berpacu dengan waktu untuk memblokir peristiwa yang tidak diinginkan yang mungkin sudah terjadi. Dalam pendekatan preventif, kontrol kebijakan diterapkan segera setelah perilaku "normal" teridentifikasi. Meningkatkan perlindungan jauh lebih baik daripada menunggu ancaman menemukan celah dalam perlindungan. Selanjutnya, perilaku ancaman harus diblokir secara inheren.

Kepercayaan Autentikasi Dinamis (Autentikasi Berbasis Risiko)

Ketergantungan pada kata sandi sebagai dasar utama autentikasi merupakan sumber dari banyak kompromi keamanan. Serangan phishing, rekayasa sosial, pelanggaran sistem, dan metode lainnya secara rutin digunakan oleh penyerang untuk mengumpulkan informasi ID pengguna dan kata sandi. "Stuffing" kredensial digunakan untuk melihat apakah kata sandi yang diambil dari satu akun juga digunakan di akun lain oleh pengguna yang sama. Terdapat kebutuhan untuk autentikasi multifaktor, di mana kredensial pengguna tidak dapat dikompromikan hanya melalui situs phishing, malware pencatat keyboard, atau rekayasa sosial. Umumnya diinginkan untuk memverifikasi pengguna berdasarkan sesuatu yang mereka miliki (misalnya, perangkat), sesuatu yang mereka ketahui (misalnya, kata sandi), dan idealnya mencakup sesuatu yang "mereka" (misalnya, biometrik). Namun, sebagian besar metode



cenderung menciptakan beban yang lebih besar bagi pengguna karena pengguna umumnya perlu mengambil langkah tambahan untuk autentikasi.

Konsep industri yang disebut "autentikasi berbasis risiko" sedang berkembang. Dengan bantuan solusi berbasis aturan yang dikombinasikan dengan pembelajaran mesin (ML), manajemen risiko yang dinamis dapat dilakukan. Tujuannya adalah meminimalkan beban pengguna saat melakukan transaksi dengan menilai status autentikasi pengguna dan menyeimbangkannya dengan risiko transaksi.

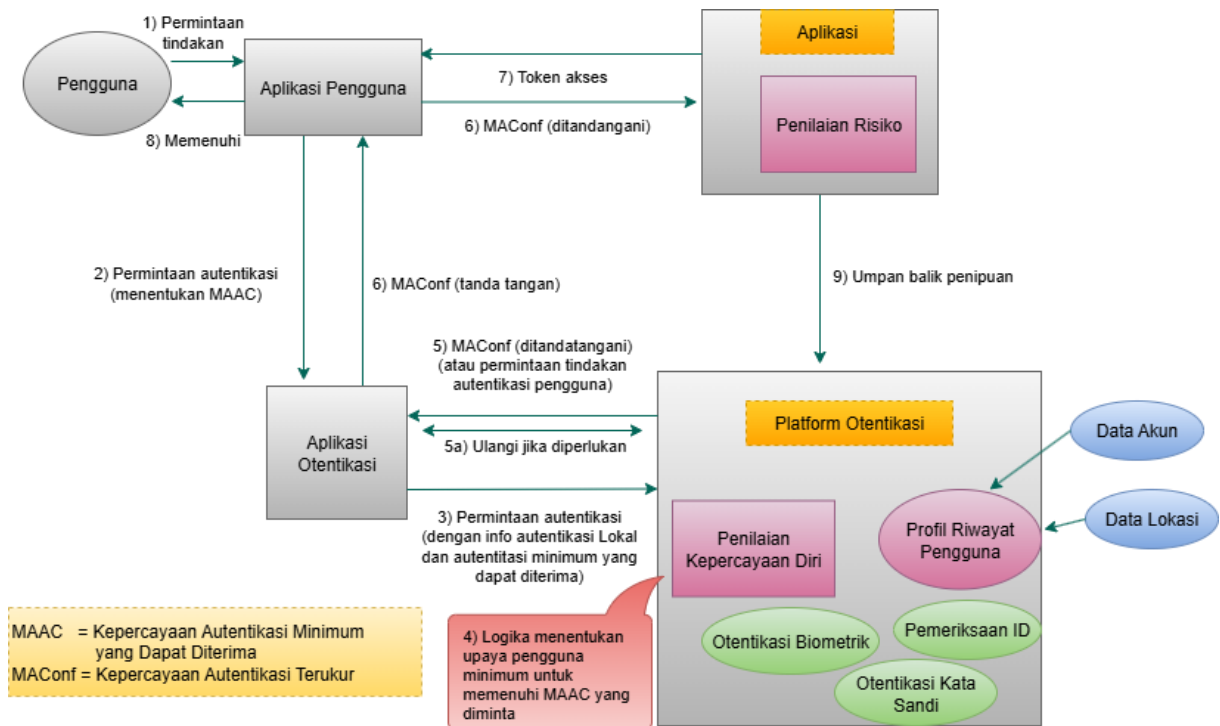
Di AT&T, kami menggunakan aplikasi seluler umum untuk menyediakan autentikasi ke berbagai aplikasi (seluler dan web). Dengan demikian, kami umumnya dapat menyediakan akses masuk tunggal bagi pengguna dengan autentikasi multifaktor berkekuatan tinggi, dan dengan sedikit atau tanpa upaya dari pihak pengguna. Namun, tidak semua aplikasi dan transaksi memiliki kebutuhan yang sama. Kekuatan autentikasi untuk melihat informasi cuaca yang dipersonalisasi mungkin tidak perlu sekuat akses ke aplikasi perbankan, yang memungkinkan transaksi keuangan dalam jumlah besar. Bahkan, mungkin ada kemungkinan adanya tingkatan yang berbeda dalam satu aplikasi (misalnya, melihat saldo bank kurang sensitif dibandingkan melakukan transfer dana).

Tujuan dari solusi ini adalah untuk memberikan keyakinan sebesar mungkin kepada aplikasi tentang kekuatan autentikasi dengan upaya seminimal mungkin dari pengguna akhir. Hal ini dapat mencakup pertimbangan hal-hal seperti perubahan akun perangkat seluler terkini, lokasi perangkat, waktu sejak autentikasi aktif, faktor-faktor lain, dan bahkan mungkin umpan balik dari aplikasi yang menunjukkan kemungkinan telah terjadi masalah. Tujuannya adalah untuk menyediakan fungsi autentikasi terpusat yang memanfaatkan hubungan pengikatan antara perangkat seluler dan jaringan, meminimalkan tindakan khusus oleh pengguna, dan memungkinkan aplikasi untuk mengendalikan nasibnya sendiri terkait risiko autentikasi.

Hal ini tidak perlu dibatasi hanya pada aplikasi seluler. Solusi ini juga dapat digunakan untuk mengautentikasi ke aplikasi web di perangkat lain (misalnya, laptop). Misalnya, jika kita mempertimbangkan proses masuk tunggal standar yang digunakan untuk aplikasi web pada umumnya, prosesnya mungkin seperti ini:

1. Pengguna mengunjungi situs web yang mendukung masuk tunggal dan diminta untuk masuk. Pengguna memilih opsi autentikasi aplikasi seluler dari prompt akses tunggal (yang dapat berupa opsi yang telah dihafal).
2. Pengguna memasukkan ID pengguna ke dalam prompt akses tunggal (yang juga dapat dihafal) dan mengirimkannya.
3. Aplikasi akses tunggal akan meminta aplikasi seluler (pada perangkat seluler pengguna) dengan permintaan autentikasi dan menentukan tingkat kepercayaan minimum yang diperlukan.
4. Aplikasi autentikasi seluler akan merespons kembali aplikasi akses tunggal dengan hasil autentikasi dan tingkat kepercayaan yang terukur.
5. Jika berhasil, pesan autentikasi berhasil akan ditampilkan.
6. Pengguna melanjutkan ke aplikasi.





Gambar 10.6 Manajemen risiko autentikasi memungkinkan penggunaan perangkat seluler untuk meminimalkan upaya pengguna sekaligus membangun kepercayaan autentikasi yang diperlukan.

Ini mungkin tampak seperti banyak langkah, tetapi pengguna tidak perlu melakukan banyak hal, dan tidak perlu memasukkan kata sandi. Mereka cukup menyentuh (atau melihat) perangkat seluler mereka untuk memverifikasi kepemilikan, dan hasilnya adalah autentikasi multifaktor dengan tingkat keyakinan tinggi yang sangat sulit ditembus oleh penyerang. Kredensial tidak dapat direkayasa sosial atau di-phishing. Pengguna akan segera mendapatkan notifikasi setiap kali seseorang mencoba mengakses akun mereka.

Terdapat peluang untuk AI di sini juga. Langkah-langkah dalam proses autentikasi diilustrasikan pada Gambar 10.6. Saat aktivitas autentikasi berlangsung, dan saat aplikasi memberikan umpan balik tentang potensi masalah dengan akun pengguna, metode yang diperlukan untuk autentikasi ke aplikasi dapat ditingkatkan untuk pengguna yang tampaknya berisiko. Misalnya, apakah aplikasi melaporkan kejadian yang mencurigakan? Apakah perangkat berada di tempat yang tidak biasa? Aplikasi autentikasi seluler harus dapat secara dinamis meminta informasi lebih lanjut kepada pengguna untuk memastikan orang yang tepat memegang perangkat mereka dan tingkat kepercayaan autentikasi yang dibutuhkan terpenuhi.

10.4 APLIKASI DALAM ANALITIK ANCAMAN

Deteksi Anomali Jaringan Umum

Penggunaan deteksi anomali pada jaringan merupakan tren yang populer. Pencarian dengan frasa "deteksi anomali perilaku jaringan" menunjukkan hal ini. Kami di AT&T telah berhasil menggunakan ML untuk mengembangkan referensi dasar bagi banyak atribut



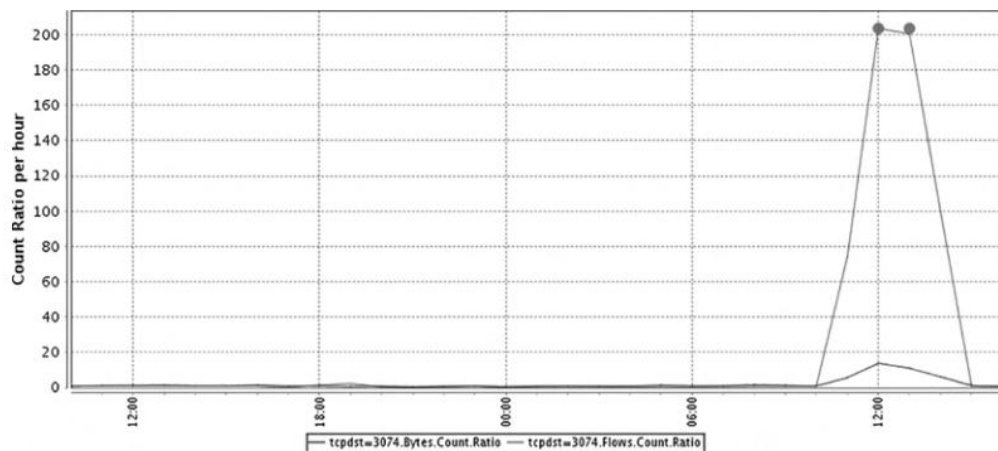
aktivitas jaringan dan untuk mendeteksi anomali pada jaringan. Ini digunakan sebagai sarana untuk mendeteksi peristiwa yang mungkin berdampak pada layanan yang disediakan AT&T kepada pelanggan. Deteksi dini adalah langkah pertama menuju koreksi cepat dan dampak minimal pada layanan. Hal ini terdengar relatif mudah, tetapi dapat menjadi rumit untuk mendeteksi peristiwa yang relevan karena jaringan terus berkembang. Rute berubah, sistem berubah, demografi pengguna berubah, dan aplikasi berubah.

Mendeteksi anomali saja tidak cukup. Anomali yang relevan harus dideteksi di antara berbagai parameter atau atribut aktivitas jaringan (misalnya, volume byte, rasio aliran/koneksi terhadap byte, volume aktivitas pada port tertentu). Interpretasi anomali yang bermanfaat diperlukan untuk merancang respons yang tepat. Sebagian besar anomali bersifat normal (misalnya, aplikasi baru dibuat, perutean pada jaringan berubah, pelanggan baru bergabung dengan jaringan). Namun, terdapat beberapa jenis anomali yang relevan dengan keamanan dan mudah dideteksi serta diinterpretasikan dengan deteksi anomali. Contohnya meliputi:

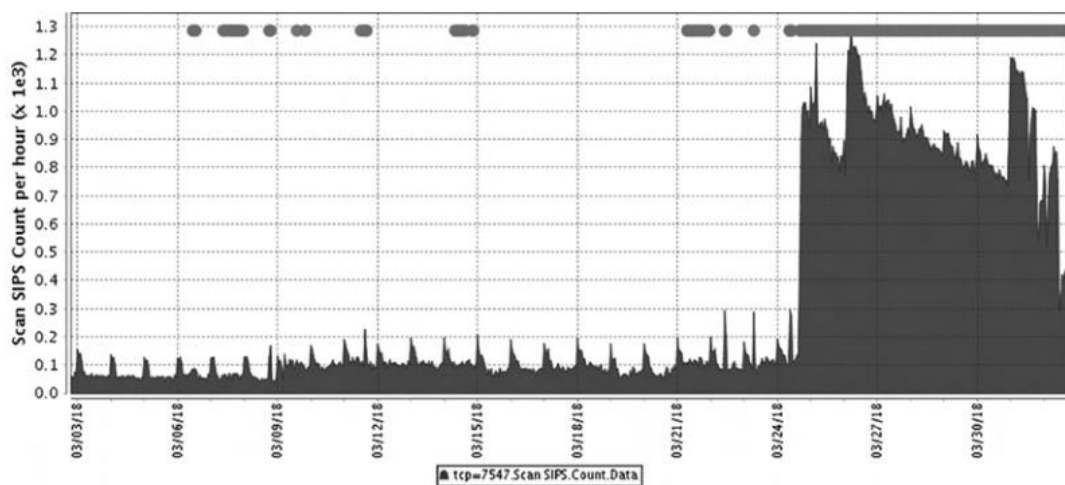
- Serangan Penolakan Layanan Terdistribusi: Ada banyak jenis serangan DDoS, tetapi sebagian besar ditandai dengan banjir lalu lintas menuju satu atau lebih alamat target dari banyak alamat sumber. Berbagai jenis serangan bermanifestasi dengan cara yang berbeda. Misalnya, serangan SYN Protokol Kontrol Transmisi (TCP), yang merupakan banjir paket SYN TCP tanpa koneksi lanjutan, umumnya dikenal sebagai banyak aliran dengan sangat sedikit (seringkali hanya satu paket), dengan bendera SYN yang ditetapkan dan hanya 40 byte lalu lintas. Ada variasi dalam hal ini, tetapi rangkaian aturan yang relatif sederhana ini dapat mengisolasi serangan SYN hampir sepanjang waktu. Contoh deteksi semacam itu diilustrasikan pada Gambar 10.7. Ada banyak jenis serangan DDoS lain yang dapat diidentifikasi dengan jenis aturan yang serupa. Namun, ada variasi serangan baru yang terus berkembang karena penyerang mencoba menciptakan metode serangan yang tidak secara otomatis dikarakterisasi dan/atau lebih sulit untuk dimitigasi (yaitu, membuang paket serangan dan meneruskan paket yang baik).
- Deteksi Worm dan Botnet: Umumnya, worm ditandai dengan meningkatnya jumlah alamat IP sumber yang memindai jaringan pada satu atau lebih protokol dan/atau port IP tertentu. Tujuannya adalah menemukan host target yang rentan untuk diinfeksi dan dilibatkan dalam pencarian host target yang lebih rentan. Analisis ini memerlukan deteksi pola aktivitas yang menunjukkan aktivitas pemindaian (misalnya, alamat IP sumber yang mencoba banyak koneksi ke banyak sekali alamat IP target yang berbeda). Berdasarkan deteksi aktivitas pemindaian, deteksi anomali digunakan untuk mengidentifikasi perubahan jumlah sumber pemindaian dari waktu ke waktu. Umumnya, peningkatan jumlah sumber pemindaian dari waktu ke waktu merupakan indikasi adanya worm internet. Jika peningkatan alamat sumber unik terjadi secara tiba-tiba, hal ini umumnya menunjukkan aktivitas botnet. Contohnya diilustrasikan pada Gambar 10.8, di mana peningkatan tiba-tiba dalam jumlah sumber yang memindai port 7547/tcp merupakan indikasi adanya botnet yang berupaya



memasukkan Peralatan Tempat Pelanggan (CPE) Internet yang rentan seperti modem *Digital Subscriber Line* (DSL) ke dalam botnet.



Gambar 10.7 Perubahan rasio aliran yang terkait dengan serangan DDoS SYN terlihat paling jelas pada grafik, sementara peningkatan rasio byte relatif terhadap normal terlihat tetapi kurang menonjol. Titik-titik di bagian atas grafik menunjukkan alarm yang dihasilkan oleh anomali tersebut.



Gambar 10.8 Deteksi aktivitas botnet yang menargetkan port 7547/tcp (CPE WAN Management Protocol). Grafik menunjukkan jumlah alamat IP sumber unik yang dipindai pada port tersebut. Penanda di bagian atas grafik menunjukkan alarm yang dihasilkan oleh anomali.

Untuk kejadian-kejadian ini, interpretasi menjadi penting. Mendeteksi serangan DDoS tidaklah terlalu sulit atau penting. Mengisolasi karakteristik yang anomali dan dapat digunakan untuk membantu memitigasi serangan di jaringan adalah proposisi nilai yang sebenarnya. Untuk setiap alarm pada contoh sebelumnya, serangkaian lengkap detail "analisis pasca deteksi" dipublikasikan untuk membantu memberikan interpretasi kejadian dan mengidentifikasi entitas spesifik (alamat sumber, alamat target, port, protokol, jumlah) yang



dapat membantu merumuskan respons. Namun, hal ini menghadirkan peluang baru untuk otomatisasi tingkat lanjut. Interpretasi sebaiknya dihasilkan secara lebih otomatis seiring dengan terdeteksinya variasi baru pada ancaman. Hal ini mungkin dapat dilakukan melalui ML yang diarahkan manusia. Atau mungkin beberapa analisis kognitif dapat digunakan untuk mengaitkan kerentanan yang diketahui atau publikasi eksploitasi dengan aktivitas yang terdeteksi.

Secara umum, deteksi aktor tingkat lanjut yang menggunakan deteksi analisis perilaku tingkat jaringan secara eksklusif belum berhasil. Terdapat beberapa tanda yang menunjukkan serangan aktor tingkat lanjut (misalnya, deteksi transfer data besar, aktivitas check-in berkala, atau aliran aktivitas yang sangat panjang). Namun, terdapat banyak sekali aktivitas jaringan yang sah yang memiliki karakteristik serupa. Lagipula, yang menjadikan aktor tingkat lanjut "maju" adalah kemampuannya untuk berperilaku dengan cara yang tampak relatif normal. Oleh karena itu, dua faktor tambahan sangat membantu dalam mendeteksi aktor tingkat lanjut:

- **Deteksi Anomali Berbasis Host:** Ingat kembali pembahasan kita tentang virtualisasi keamanan dan mikroperimeter sebelumnya di bab ini. Alat-alat ini memungkinkan pembuatan profil perilaku normal aplikasi (dan host yang mendukung aplikasi tersebut) dan mendeteksi pola yang tidak biasa. Analisis perilaku berbasis host lokal dapat melengkapi analisis perilaku jaringan untuk mencegah dan mendeteksi aktor tingkat lanjut.
- **Deteksi Berbasis Tanda Tangan:** Tanda tangan deteksi umumnya memiliki dua masalah. Tanda tangan tersebut cenderung terlalu spesifik dan dapat dengan mudah dilewati oleh perubahan kecil pada malware atau perilaku aktivitas, atau terlalu umum dan penuh dengan deteksi positif palsu. Penggunaan tanda-tanda umum bersama dengan pola perilaku secara umum jauh lebih efektif dalam mendeteksi aktor tingkat lanjut.

Seperti yang dapat Anda bayangkan dari diskusi ini, ada beberapa hal yang agak menantang dalam strategi ini sejauh ini:

1. **Banyak Atribut:** Kita harus menentukan atribut apa yang paling berguna untuk mendeteksi peristiwa yang relevan dengan keamanan dalam aktivitas jaringan, mengingat banyaknya atribut yang tersedia, jumlah kombinasi atribut yang dapat diukur, dan jenis anomali yang mungkin muncul dalam deteksi peristiwa yang relevan dengan keamanan. Seiring kita memperkenalkan pola aktivitas tertentu (misalnya, check-in berkala, aktivitas pemindaian), jumlah atribut yang tersedia semakin bertambah. Kita telah membahas beberapa jenis atribut, tetapi jumlahnya sungguh ribuan, yang terlalu banyak untuk dievaluasi secara efektif oleh manusia.
2. **Konteks Penting:** Sudut pandang di jaringan mengubah bagaimana anomali seharusnya diinterpretasikan dan seberapa sensitif deteksi tersebut. Misalnya, setiap aktivitas pemindaian itu sendiri dalam jaringan perusahaan harus dicatat. Namun, pemindaian di internet merupakan praktik umum bagi para peneliti, calon penyerang, dan botnet. Faktanya, masih terdapat cukup banyak gangguan di internet yang terkait dengan

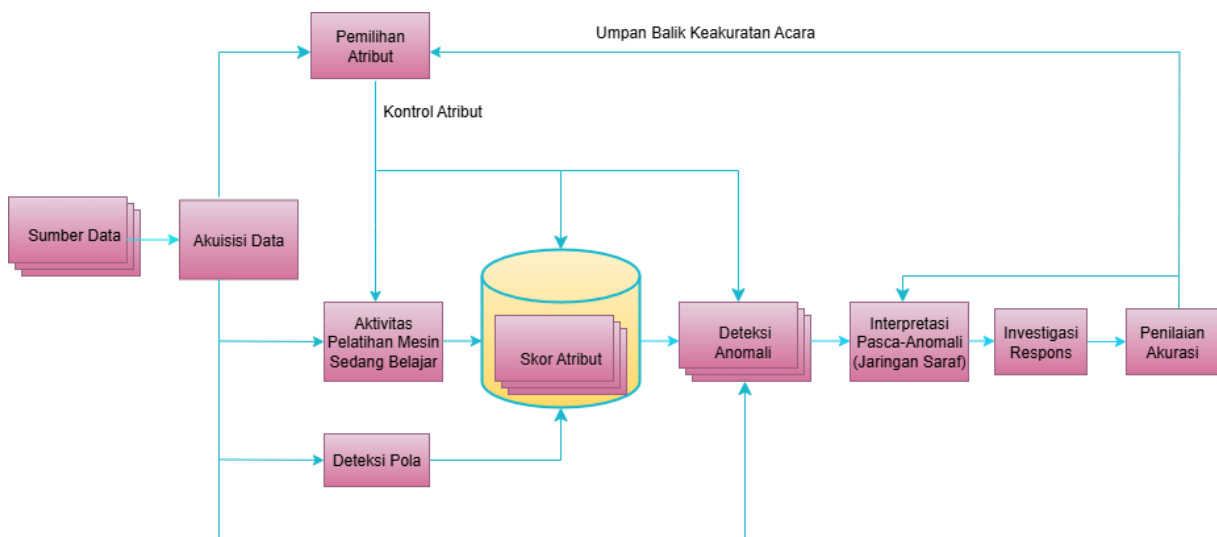


worm Conficker, yang muncul sejak tahun 2008. Metode analisis perlu mengakomodasi perbedaan-perbedaan ini.

3. Tanda Tangan Membutuhkan Banyak Tenaga Kerja: Meskipun bermanfaat, pengembangan dan pemeliharaan tanda tangan yang relevan sulit dan membutuhkan banyak tenaga kerja. Akan bermanfaat jika kita dapat mengurangi kebutuhan akan tanda tangan yang baik melalui peningkatan kesadaran konteks dan analisis perilaku.

Hal ini menunjukkan bahwa analisis yang lebih umum perlu dibuat dan dapat diterapkan pada berbagai set data dengan modifikasi minimal (misalnya, log firewall, log proksi, log autentikasi, data aliran) dalam berbagai konteks, dan agar analisis tersebut dapat beradaptasi dengan kebutuhan lingkungan.

Gambar 10.9 memperkenalkan model fungsional umum untuk deteksi anomali jaringan yang berpotensi dapat diterapkan pada berbagai konteks.



Gambar 10.9 Aplikasi umum pembelajaran mesin dan kecerdasan buatan untuk analisis ancaman.

Setelah data diperoleh melalui Akuisisi Data di sebelah kiri ilustrasi, terdapat empat jalur paralel yang diumpangkan data:

- Pemilihan Atribut adalah jalur non-waktu nyata. Jalur ini menilai data yang tersedia dan memilih atribut yang kemungkinan relevan dengan keamanan. Mungkin terdapat beberapa eksperimen yang terlibat untuk menentukan apakah atribut baru melengkapi pengetahuan yang ada tentang peristiwa keamanan. Penting untuk fungsi ini adalah jalur umpan balik dari Penilaian Akurasi, yang membantu menunjukkan positif-benar (yaitu, deteksi yang baik), positif-salah (yaitu, deteksi yang buruk), dan negatif-salah (yaitu, deteksi yang terlewat). Penilaian akurasi mungkin awalnya perlu dilakukan terutama oleh manusia atau dikaitkan dengan tiket masalah. Informasi kontekstual lainnya dapat diberikan untuk membantu Pemilihan Atribut dalam membuat keputusan (misalnya, informasi tentang aplikasi, infrastruktur, dan penetapan alamatnya).



- **Pelatihan Aktivitas:** Berdasarkan atribut yang diidentifikasi oleh Pemilihan Atribut, Pelatihan Aktivitas digunakan untuk menghasilkan profil aktivitas untuk atribut yang relevan. Misalnya, mungkin perlu melacak jumlah lalu lintas byte pada interval per menit yang terkait dengan sekelompok alamat pada port tertentu. Keluaran dari ini akan berupa profil atau Skor Atribut yang akan disimpan untuk dibandingkan dengan aktivitas di masa mendatang.
- **Deteksi Pola:** Kami telah membahas beberapa pola sebelumnya yang berguna untuk mendeteksi peristiwa. Misalnya, pola aktivitas pemindaian memang berguna. Pola ini bukan tanda tangan, tetapi merupakan indikator perilaku yang kuat. Pola ini juga dicatat sebagai Skor Atribut untuk penentuan dasar dan deteksi anomali.
- **Deteksi Anomali:** Aktivitas data saat ini dimasukkan ke dalam mesin deteksi anomali yang membandingkan aktivitas saat ini dengan aktivitas sebelumnya yang tercatat dalam data Skor Atribut.

Keluaran Deteksi Anomali digunakan untuk memicu analisis selanjutnya.

- **Interpretasi Pasca Anomali** dimulai dengan mengumpulkan detail tambahan terkait anomali yang terdeteksi. Detail lebih lanjut tentang aktivitas historis dapat dikumpulkan. Data mentah terkait peristiwa yang terdeteksi dapat diiris dan dipecah ke berbagai arah untuk menghasilkan perspektif tambahan dari data tersebut. Analisis tambahan lainnya seperti pencarian Whois alamat IP dan/atau geolokasi dapat dilakukan. Nama domain yang terkait dengan alamat IP dapat diselidiki untuk hubungan lebih lanjut. Dan data intelijen ancaman dari berbagai sumber dapat diperiksa untuk hubungan dengan peristiwa yang baru ditemukan. Sebuah mesin diperlukan untuk mengembangkan interpretasi setiap peristiwa—untuk menentukan apakah peristiwa tersebut relevan dengan keamanan, jenis peristiwanya, dan merekomendasikan tindakan perbaikan yang tepat.
- **Investigasi Respons** diperlukan untuk membantu memandu proses pembelajaran. Mengingat kita mencoba menciptakan sesuatu yang umum dan hanya akan mengetahui sedikit tentang konteks dan lingkungan untuk memulai, perlu ada investigasi lebih lanjut terhadap peristiwa untuk memandu proses pembelajaran.
- **Penilaian Akurasi** dimaksudkan untuk merepresentasikan proses manual yang difasilitasi melalui berbagai alat. Seiring dengan semakin banyaknya kejadian yang dipelajari, kebutuhan akan investigasi dan koreksi respons manual akan berkurang dan tingkat deteksi positif-sebenarnya akan meningkat.

Mengisolasi Aktivitas Sistem Nama Domain yang Anomali

Bagian ini mengambil konsep deteksi anomali dan menyelidiki contoh yang lebih spesifik. Seperti contoh sebelumnya, tugas utama dalam merancang analitik ancaman adalah menetapkan garis dasar sebagai model aktivitas "normal". Biasanya tidak ada definisi yang tegas tentang apa yang normal, selain tidak boleh mencakup serangan yang diketahui dan kejadian khusus apa pun dalam jaringan, seperti pemadaman atau pemeliharaan yang tidak terjadwal. Baseline harus mencakup faktor-faktor, seperti pola diurnal, jam sibuk vs. jam tidak sibuk dalam sehari, variasi hari kerja vs. akhir pekan, atau pola musiman lainnya. Baseline juga



harus mempertimbangkan kondisi kesalahan umum dalam jaringan atau aplikasi yang merupakan bagian dari operasi harian berbagai layanan.

Untuk aktivitas jaringan, satu model sederhana terdiri dari pemilihan serangkaian metrik atomik, seperti jumlah paket atau jumlah byte, yang masuk dan keluar dari aplikasi atau jaringan, dan serangkaian statistik dasar seperti rata-rata, median, dan deviasi standar selama periode waktu tertentu (misalnya, 5 menit, 1 jam, 1 hari, atau 1 minggu). Anomali kemudian ditandai ketika salah satu metrik yang diinginkan melebihi rata-rata atau median ditambah atau dikurangi x jumlah deviasi standar. Variabel x dapat digunakan untuk memberi label ukuran anomali: misalnya, $x = 3$ untuk anomali minor, $x = 4$ untuk anomali mayor, dan $x \geq 5$ untuk anomali kritis. Meskipun pendekatan ini sederhana dan umum diterapkan, pendekatan ini tidak terlalu tepat, karena hanya melihat sejumlah kecil statistik dan berasumsi bahwa distribusi metrik yang dipilih adalah normal (yaitu, Gaussian) dan oleh karena itu mengikuti kurva "lonceng".

Model data yang lebih canggih menangkap periodisitas, musiman, dan tren keseluruhan. Model-model ini lebih akurat daripada model "datar" yang dijelaskan sebelumnya, terutama untuk lalu lintas jaringan yang digerakkan oleh manusia, baik di rumah, di tempat kerja, maupun saat bepergian. Misalnya, lalu lintas akses internet yang dihasilkan oleh konsumen bisa tiga hingga empat kali lebih tinggi pada jam malam dibandingkan jam kerja. Meskipun model datar (sepanjang hari) kemungkinan akan salah menandai lalu lintas malam sebagai anomali, model data yang menangkap variasi per jam sepanjang hari akan mengidentifikasinya dengan tepat sebagai normal. Demikian pula, seiring meningkatnya atau menurunnya popularitas suatu layanan, terdapat tren aktivitas jangka panjang yang berubah, yang perlu dicatat sebagai bagian dari baseline. Peningkatan atau penurunan yang sesuai dalam total pelanggan atau kunjungan pelanggan pada musim yang berbeda (misalnya, musim panas vs. kembali ke sekolah) mendorong permintaan jaringan dan aplikasi yang berbeda. Untuk lalu lintas yang menunjukkan karakteristik ini, penting bahwa deteksi anomali didasarkan pada deviasi dari model yang menangkap periodisitas, tren, dan musiman seperti yang dijelaskan sebelumnya.

Menandai titik data sebagai anomali setara dengan menerapkan fungsi keputusan. Dalam pembahasan sebelumnya, keputusan dibuat berdasarkan data satu dimensi, tetapi lebih sering keputusan tersebut harus dibuat pada data yang multidimensi. Misalnya, untuk satu titik waktu, data kita dapat berisi jumlah paket masuk, paket keluar, paket dengan protokol tertentu, paket dengan jenis bendera tertentu, dll. Jadi, keputusan pada titik waktu tertentu terkait klasifikasi data sebagai anomali dapat bergantung pada sejumlah elemen data ini atau kombinasi tertentu dari mereka. Jaringan saraf telah lama digunakan untuk memodelkan kelas data multidimensi, seperti gambar atau suara manusia. Mereka mensimulasikan perilaku neuron manusia yang mengambil masukan dari neuron lain, menggabungkannya, dan menghasilkan keluaran yang dapat diumpankan ke neuron lain. Bergantung pada berbagai masukan, neuron-neuron ini dapat mempelajari pola-pola tertentu dan mengenalinya kembali jika pola-pola itu muncul kembali atau bahkan jika pola-pola itu sedikit berbeda.



Meningkatnya ketersediaan data dalam bentuk terstruktur atau tidak terstruktur dan biaya daya komputasi yang rendah telah membawa kebangkitan teknik ML berdasarkan NN. Sementara NN telah digunakan sejak tahun 1950an untuk pengenalan suara dan analisis gambar, NN juga mendapatkan popularitas dan penerapan dalam tugas-tugas keamanan siber untuk mengklasifikasikan dan mengenali berbagai jenis peristiwa keamanan, seperti pengintaian dan berbagai jenis serangan *denial-of-service* (DoS). Daya tarik NN adalah bahwa mereka dapat menangkap hubungan yang sebelumnya tidak diketahui antara berbagai variabel dalam data berdimensi tinggi. Mereka juga memungkinkan banyak fleksibilitas dalam hal arsitektur mereka untuk berkinerja lebih baik dalam kategori masalah tertentu, dan mereka dapat digunakan untuk klasifikasi data yang diawasi atau tidak diawasi.

Misalnya, jenis khusus NN berulang, yang disebut jaringan Memori Jangka Pendek Panjang, telah ditemukan sangat efektif dalam mengklasifikasikan domain Algoritma Pembuatan Domain yang digunakan oleh botnet untuk berkomunikasi dengan pengontrol mereka. Jaringan saraf tiruan ini, sesuai namanya, dirancang untuk mempelajari pola berdurasi panjang, sekaligus mampu mengabaikan anomali tertentu dalam data pelatihan. Jaringan saraf tiruan ini juga relatif mudah dilatih menggunakan implementasi yang tersedia sebagai sumber terbuka.

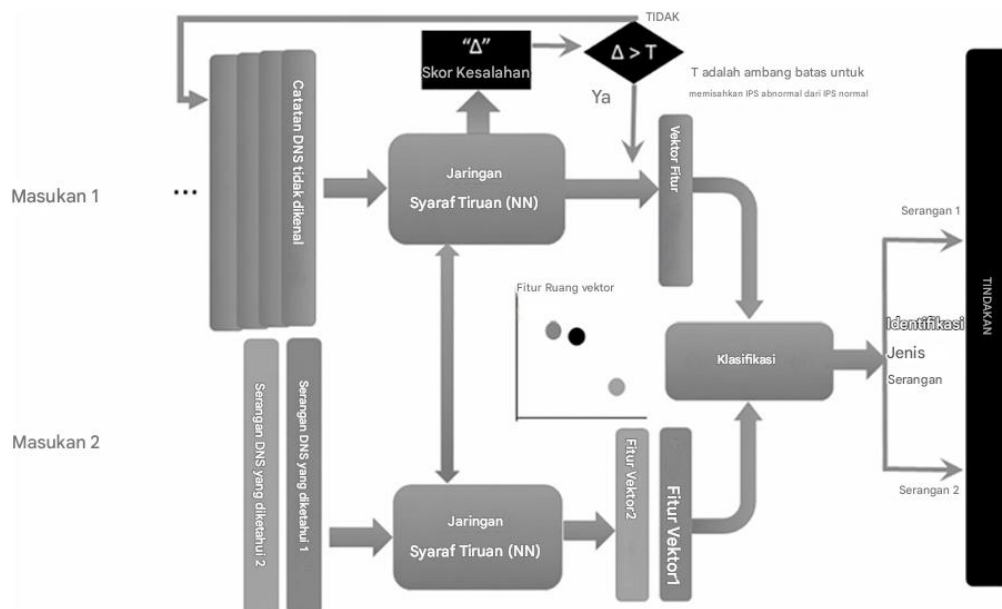
Jaringan NN Autoencoder juga sangat berguna dalam mempelajari fitur-fitur penting dari data multidimensi dengan pelatihan tanpa pengawasan menggunakan sejumlah kecil simpul tersembunyi. Untuk mendeteksi anomali, keluaran direkonstruksi dari lapisan tersembunyi dengan sejumlah kecil simpul, dan kesalahan dihitung antara masukan dan keluaran. Anomali dapat ditandai untuk keluaran dengan kesalahan tinggi, sementara pengelompokan status simpul lapisan tersembunyi dapat digunakan untuk mengklasifikasikan anomali. Hal ini telah diuji di AT&T untuk serangan penipisan sumber daya yang sering diamati dalam lalu lintas Sistem Nama Domain (DNS).

Untuk mengoperasikan jaringan besar secara aman, deteksi dan interpretasi anomali sangat penting. Analis dan operator keamanan perlu memahami sifat, tingkat keparahan, dan apakah mereka perlu menindaklanjuti anomali tersebut. Apakah anomali tersebut merupakan peristiwa tunggal yang singkat atau merupakan bagian dari pola yang berkembang? Misalnya, deteksi satu atau beberapa pemindaian pada port tertentu mungkin berguna untuk kewaspadaan situasional, tetapi peningkatan jumlah sumber pemindaian dapat menjadi indikasi perkembangan infeksi baru, memanfaatkan kerentanan zero-day yang ditemukan di alam liar. Mendapatkan ringkasan data selama dan di sekitar anomali, dan menggabungkan ringkasan beberapa anomali, dapat sangat berguna untuk meningkatkan pemahaman kita tentang status jaringan. Ini pada dasarnya adalah meta-analisis yang dapat meningkatkan kewaspadaan situasional atau mengarahkan tindakan tertentu oleh analis atau operator. Meta-analisis dapat berfokus pada fitur anomali atau dapat berfokus pada fitur ortogonal. Misalnya, jika anomali tersebut merupakan pemindaian pada beberapa target pada port tertentu, meta-analisis dapat berfokus pada jumlah sumber unik pemindaian dari waktu ke waktu atau dapat berfokus pada lalu lintas sumber tersebut yang tidak terkait dengan pemindaian untuk melihat aktivitas lain yang sedang berlangsung.



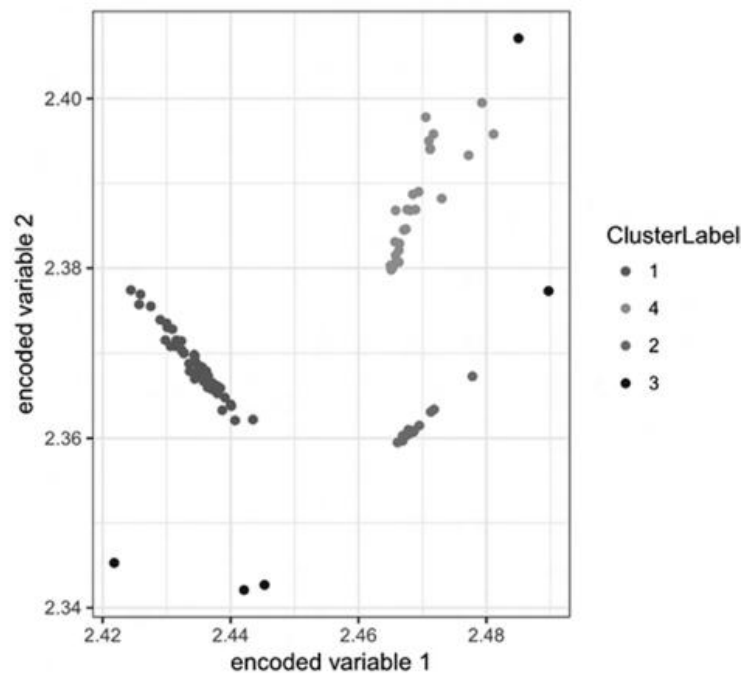
Pembelajaran mesin dapat memfasilitasi interpretasi anomali dan atribusi ke jenis serangan yang diketahui dengan merancang algoritma sebagai Jaringan Saraf Tiruan (NN) auto-encoder dan melatihnya dengan lalu lintas normal. Gambar 10.10 menunjukkan garis besar bagaimana anomali DNS dapat dideteksi dan diklasifikasikan menggunakan NN auto-encoder. Awalnya, lalu lintas normal dapat digunakan untuk melatih NN auto-encoder. Serangan yang diketahui dapat digunakan sebagai masukan ke NN (Masukan 2 pada gambar) dan membuat tanda tangan dalam ruang vektor fitur. Selama fase pengujian, anomali dapat dideteksi ketika skor kesalahan antara masukan dan keluaran yang direkonstruksi tinggi. Ketika NN mendeteksi anomali, vektor fitur yang sesuai dapat dibandingkan dengan vektor fitur dari serangan yang diketahui (misalnya, menggunakan pengelompokan tetangga terdekat K). Algoritme ini dapat digunakan untuk mendeteksi anomali lain yang tidak diketahui, yang dapat disajikan sebagai log atau metadata kepada analis untuk penyelidikan lebih lanjut.

Gambar 10.11 memberikan contoh set data fitur lapisan tersembunyi untuk dua jenis serangan DNS yang diketahui dan satu yang tidak diketahui (tidak berlabel). Dalam contoh ini, set data 30 dimensi awal dikompresi menjadi 2 dimensi pada lapisan tersembunyi tahap encoder Jaringan Saraf Tiruan (NN). Nilai vektor 2 dimensi ditunjukkan pada gambar untuk serangan DoS (sumber tunggal), serangan DDoS, dan anomali tak berlabel. Setiap serangan selanjutnya yang memiliki karakteristik serupa menunjukkan vektor fitur yang dikelompokkan sesuai dengan itu.



Gambar 10.10 Formulasi auto-encoder untuk mendeteksi anomali dan atribusi ke jenis serangan yang diketahui.





Gambar 10.11 Anomali lalu lintas DNS sebagaimana direpresentasikan dalam ruang fitur dua dimensi jaringan saraf auto-encoder.

Analisis Ancaman Internal Analisis Perilaku Peristiwa Pengguna

Umumnya, ancaman internal adalah seseorang dengan niat jahat yang memiliki akses ke aset organisasi. Tujuan mereka mungkin untuk mencuri properti, mengganggu bisnis, atau mengekspos informasi ke domain publik. Beberapa juga menganggap dampak yang tidak disengaja terhadap organisasi yang disebabkan oleh personel dengan akses resmi sebagai ancaman internal. Dengan kata lain, jika tidak ada kontrol yang memadai untuk meminimalkan peluang personel melakukan kesalahan yang tidak disengaja yang berdampak pada bisnis, hal ini juga terkadang dianggap sebagai ancaman internal. Ancaman ini dapat berupa karyawan, mantan karyawan, kontraktor, atau rekan bisnis.

Terakhir, jika penyerang eksternal dapat memperoleh kredensial pengguna yang valid dan masuk ke sistem dengan tampilan pengguna normal, cara untuk mendeteksi ancaman tersebut memiliki tumpang tindih yang signifikan sebagai ancaman internal. Ancaman internal menyalahgunakan hak istimewanya untuk melakukan aktivitas yang tidak seharusnya diizinkan. Kesadaran akan ancaman yang ditimbulkan oleh orang dalam tepercaya telah meningkat pesat dalam beberapa tahun terakhir karena insiden orang dalam yang menarik perhatian publik baru-baru ini (misalnya, Snowden dan Manning), di mana sejumlah besar informasi yang sangat sensitif telah terekspos ke domain publik. Teknologi informasi dan aset digital dapat memungkinkan ancaman internal untuk melakukan aktivitas mereka dari jarak jauh, yang dapat membuat mereka merasa lebih aman.

Organisasi yang mendukung proyek sensitif seperti kontrak Departemen Pertahanan yang menangani informasi rahasia kini diwajibkan memiliki Program Pencegahan Ancaman Internal yang terdefinisi. Aspek-aspek program ancaman internal yang lengkap meliputi kebijakan yang membantu menentukan perilaku dan kewajiban personel yang dapat diterima,



kontrol yang membatasi akses ke sistem sensitif atau kritis, pelatihan kesadaran, aktivitas investigasi, dan akhirnya deteksi.

Penerapan metode Analisis Perilaku Peristiwa Pengguna (UEBA) untuk ancaman internal telah berkembang dari pendekatan reaktif tradisional. Program ancaman orang dalam tradisional mengandalkan seseorang yang melaporkan perilaku mencurigakan atau penemuan pelanggaran sebelum investigasi diluncurkan. Teknologi deteksi baru sedang dikembangkan untuk mengidentifikasi kandidat secara anonim untuk investigasi. Beberapa indikator utama ancaman orang dalam meliputi: kinerja kerja yang buruk, kesulitan keuangan, keuntungan finansial yang tiba-tiba, jam kerja yang tidak biasa, ketidaksepakatan yang tidak biasa dengan kebijakan dan/atau rekan kerja, perjalanan luar negeri yang tidak biasa (terkait dengan spionase asing), dan meninggalkan perusahaan. Penting untuk mempertimbangkan bahwa teknologi tersebut mencoba menggunakan perilaku yang diamati secara pasif untuk menentukan maksud atau tujuan seseorang. Jenis analisis ini tidak sepenuhnya deterministik, jadi selama tidak ada kemampuan untuk membaca pikiran individu, UEBA untuk memprediksi atau mendeteksi ancaman orang dalam akan penuh dengan positif palsu atau tidak akan menemukan banyak hal. Namun, teknologi tersebut jika digunakan dengan benar dapat membantu mengidentifikasi jenis aktivitas ancaman orang dalam tertentu.

Teknologi ini biasanya menggunakan pendekatan berbasis risiko kumulatif untuk mengidentifikasi kandidat/individu yang paling mungkin untuk investigasi lebih lanjut dengan harapan dapat mencegah atau meminimalkan kerusakan yang disebabkan oleh ancaman internal. Mirip dengan yang dilakukan dengan analitik ancaman. Jenis indikator yang mungkin dicari oleh alat deteksi ancaman internal meliputi:

1. Individu dapat diprofilkan berdasarkan perilaku masa lalu mereka, dan analisis digunakan untuk mengidentifikasi perubahan perilaku mereka. Profil pola perilaku jangka panjang dan jangka pendek beserta deteksi anomali dapat digunakan untuk mendeteksi perubahan perilaku individu.
2. Perilaku individu dapat dibandingkan dengan orang lain di organisasi mereka, di mana mungkin terdapat orang lain yang membutuhkan sistem dan akses data yang sama, dan mereka menjalankan fungsi pekerjaan yang serupa atau sama. Ini mungkin mencakup metode lanjutan untuk membuat profil aktivitas suatu kelompok dan menggunakannya sebagai dasar untuk mendeteksi anomali dalam perilaku individu.
3. Aktivitas yang didefinisikan sebagai pelanggaran praktik bisnis yang berlaku dapat diidentifikasi. Metode-metode ini dapat didasarkan pada aktivitas yang dipelajari dari masyarakat dan/atau mungkin didasarkan pada beberapa aturan kebijakan dasar. Deteksi semacam ini umumnya akan lebih bermanfaat jika terdapat program pelatihan kesadaran yang kuat di mana pelanggar kebijakan umumnya akan tahu bahwa mereka melanggar kebijakan.
4. Pola aktivitas yang umumnya terkait dengan ancaman internal dapat dideteksi. Misalnya, seseorang yang mencuri data mungkin menggunakan stik memori USB pada sistem, mungkin mengunggah data dalam jumlah besar ke situs cloud publik, atau mungkin menggunakan metode enkripsi yang tidak biasa. Ini pada dasarnya adalah



aturan dan/atau pola perilaku terprogram yang dapat dibandingkan dengan perilaku individu.

5. Beberapa personel yang bekerja dengan data yang sangat sensitif atau sistem kritis mungkin dianggap sebagai target risiko yang lebih tinggi untuk pemerasan atau penyuapan dan mungkin diperlakukan dengan perhatian khusus. Akibatnya, personel dengan akses ke data sensitif umumnya akan memiliki skor risiko yang lebih tinggi.
6. Tentu saja, jika seseorang melaporkan perilaku mencurigakan yang terkait dengan seseorang, orang tersebut mungkin diperlakukan sebagai orang dengan risiko yang lebih tinggi dibandingkan dengan orang lain. Seperti sebelumnya, laporan pasti akan meningkatkan skor risiko individu. Jenis laporan mungkin memiliki bobot tertentu yang terkait dengan risikonya. Secara umum, praktik yang diterima secara umum adalah bahwa salah satu jenis indikator di atas tidak cukup sebagai dasar untuk memulai investigasi. Skor risiko untuk seorang individu akan diperingkat relatif terhadap fungsi semua perilaku atau karakteristik risiko yang diamati. Seiring dengan semakin banyaknya jenis indikator yang relevan yang dimasukkan ke dalam analisis, kita mungkin akan melihat peningkatan akurasi dalam skor risiko.

Seperti yang dapat dilihat, terdapat banyak aspek dari deteksi ini yang dapat memanfaatkan teknik analisis canggih untuk mengumpulkan indikator dan menetapkan skor risiko pada perilaku individu. Pada tahap teknologi ini, umumnya tidak memungkinkan untuk menetapkan ambang batas deteksi ancaman internal. Namun, individu dengan skor tertinggi umumnya layak untuk dilakukan investigasi tambahan.

Metode spesifik yang digunakan umumnya bersifat hak milik dan berada di luar cakupan bab ini. Semoga deskripsi teknologi deteksi ancaman internal ini membantu membangun dasar bagi kemajuan teknologi analisis untuk mendeteksi ancaman internal. Namun, kita masih belum memiliki cara untuk membaca pikiran para pelaku.

Memahami Kerentanan dan Eksploitasi

Untuk memahami lanskap ancaman keamanan, organisasi perlu mengevaluasi secara akurat kerentanan dan risiko perangkat lunak, firmware, dan perangkat keras untuk setiap peralatan pabrikan dan jenis versi yang terpasang di jaringan mereka. Kerentanan dapat dieksploitasi oleh penyerang dengan membangun dan menyebarkan eksploitasi baik secara jarak jauh maupun lokal di jaringan. Ketersediaan eksploitasi yang diketahui dan kerentanan yang belum terselesaikan menimbulkan risiko yang tinggi namun dapat diperhitungkan. Biasanya, penemuan kerentanan segera diikuti oleh pembuatan eksploitasi di pasar gelap, terutama jika potensi keuntungan dari pembobolan aplikasi atau jaringan dapat menghasilkan imbalan yang tinggi (misalnya, memperoleh akses ke identitas pelanggan dan basis data kontak, data pesaing, desain cetak biru, dll.).

Kecerdasan buatan dapat memainkan peran penting untuk menilai risiko kerentanan yang baru dipublikasikan dan memprediksi terciptanya eksploitasi yang berhasil dengan menelusuri berbagai basis data publik dan menganalisis kerentanan yang diketahui. Ini bisa sangat membantu bagi administrator dan personel operasi yang terbebani dalam memprioritaskan patch mana yang perlu mereka terapkan terlebih dahulu. Penundaan dalam



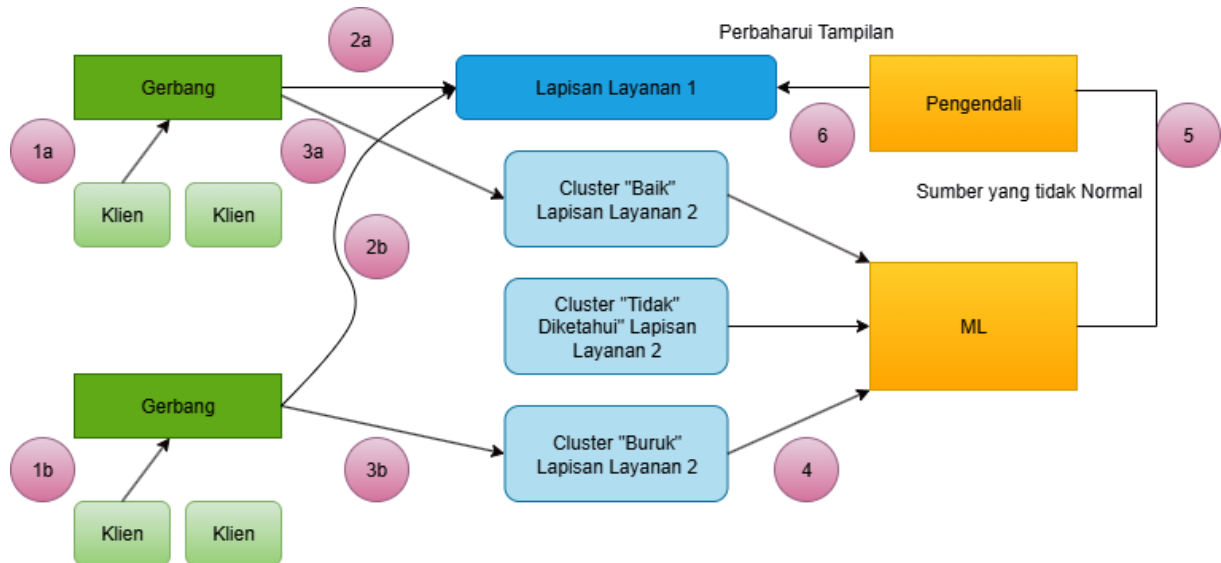
menerapkan patch penting dapat menjadi bencana, karena dapat menyebabkan pelanggaran besar-besaran data pribadi untuk jutaan pelanggan (misalnya, Equifax) yang dapat berdampak parah pada kelangsungan finansial perusahaan yang terkena dampak.

Pendekatan ML dalam keamanan siber dapat membantu mendeteksi anomali, memprediksi hasil, atau menilai risiko. Analisis dapat membantu operasi dengan membuat log atau peringatan jika tingkat keparahan anomali tinggi atau ada prediksi keyakinan tinggi untuk peristiwa berisiko tinggi. Jika pengklasifikasi menentukan bahwa anomali tertentu adalah jenis yang diketahui, dan ada tindakan mitigasi yang dapat diambil, maka pengontrol perangkat lunak dapat menggunakan output modul ML untuk memperbaiki situasi. Ini dapat memfasilitasi patching dan pengujian otomatis di lab, atau konfigurasi ulang dan verifikasi di lapangan. Tindakan otomatis hanya boleh dilakukan jika terdapat tindakan yang terdefinisi dengan baik untuk jenis anomali yang diprediksi dan tindakan tersebut tidak akan menyebabkan masalah perifer lainnya dalam jaringan. Patch yang belum diuji, misalnya, dapat menyebabkan aplikasi lain rusak, sehingga otomatisasi dalam kasus ini hanya boleh diperluas hingga patching di lab dan menjalankan pengujian otomatis, sementara penerapan di lapangan harus ditangani oleh bagian operasi setelah meninjau hasil pengujian otomatis atau menjalankan pengujian regresi tambahan untuk memverifikasi bahwa patch tersebut memecahkan masalah awal tanpa merusak hal lain dalam sistem.

Dalam kasus di mana mitigasi peristiwa keamanan dapat dilakukan dengan konfigurasi ulang, salah satu perhatian utama adalah tidak membahayakan lalu lintas yang sah, meskipun itu hanya sebagian kecil. Dalam bisnis, satu keluhan pelanggan yang besar dapat mengesampingkan perubahan konfigurasi yang bertujuan melindungi layanan untuk semua pelanggan seperti pembatasan kecepatan, karena pembatasan kecepatan transaksi ini dapat memengaruhi throughput jaringan.

Selain itu, memblokir alamat IP yang menghadap Internet berdasarkan penilaian bahwa sumber lalu lintas mungkin berbahaya adalah berisiko, karena sumbernya mungkin merupakan gateway *Network Address Translation* (NAT) di depan banyak pengguna dan perangkat yang berbeda. Gateway NAT memungkinkan banyak perangkat di sisi LAN gateway untuk berbagi satu alamat IP di sisi Internet gateway. Untuk meminimalkan risiko pemblokiran sumber lalu lintas secara tidak sengaja, salah satu pendekatan adalah melayani semua lalu lintas tetapi mendedikasikan kumpulan sumber daya yang berbeda untuk kelas risiko permintaan lalu lintas yang berbeda (jinak, serangan, atau tidak diketahui). Gambar 10.12 menunjukkan contoh di mana algoritma ML yang dikombinasikan dengan pengontrol perangkat lunak dapat menyesuaikan permintaan layanan yang berasal dari gateway yang berbeda ke kumpulan kluster server yang berbeda, sehingga memungkinkan gateway yang berperilaku baik (dan semua klien di belakangnya) untuk mengakses sumber daya khusus.





Gambar 10.12 Menggunakan ML dan pengontrol yang ditentukan perangkat lunak untuk mengkonfigurasi ulang sistem bertingkat guna mengalokasikan sumber daya khusus ke klien terpercaya.

Dalam ilustrasi, langkah-langkahnya adalah sebagai berikut:

1. (a) Permintaan dari klien "baik" atau (b) Permintaan dari klien "buruk".
2. Kueri dirutekan ke lapisan pertama klaster layanan (misalnya, HTTP atau DNS) yang mengalihkan atau merujuk klien ke klaster yang sesuai di lapisan kedua.
3. Klien dirujuk ke lapisan kedua dan ke klaster yang sesuai.
4. Permintaan dianalisis secara berkelanjutan melalui NN dua tahap.
5. Keluaran modul ML menyediakan serangkaian alamat IP sumber yang perlu direklasifikasi.
6. Pengontrol menerapkan skrip konfigurasi ulang (misalnya, playbook Ansible) untuk mengonfigurasi ulang lapisan pertama klaster layanan.

10.5 RINGKASAN

Dalam bab ini, kita telah membahas kebutuhan akan keamanan untuk membantu melindungi jaringan dari serangan, kerusakan, atau gangguan. Kita telah membahas teknologi terkait seperti virtualisasi yang merupakan faktor pendukung untuk otomatisasi fungsi keamanan. Kita telah membahas pertimbangan tentang cara memilih tugas keamanan yang tepat untuk diotomatisasi. Kita juga telah melihat beberapa contoh di mana otomatisasi keamanan dan AI dapat memajukan perkembangan terkini dalam manajemen keamanan di jaringan dan perusahaan. Mungkin pelajaran terpenting di bagian ini adalah mempertimbangkan fakta bahwa kebutuhan akan keamanan didorong oleh motivasi para pelaku ancaman untuk mengeksploitasi sistem. Hingga biaya untuk mengeksploitasi sistem lebih besar daripada keuntungan para pelaku ancaman, peningkatan keamanan akan terus ada. Ini adalah perlombaan senjata. Penggunaan otomatisasi adalah alat yang dapat membantu mendorong kontrol keamanan yang lebih efisien dan hemat bi



BAB 11

KECERDASAN BUATAN UNTUK JARINGAN PERUSAHAAN

11.1 PENDAHULUAN

Kecerdasan buatan (AI) merupakan pendorong utama transformasi digital yang membentuk perusahaan di seluruh sektor ekonomi. Transformasi digital didefinisikan sebagai proses berkelanjutan yang digunakan perusahaan untuk beradaptasi atau mendorong perubahan disruptif pada pelanggan dan pasar mereka (ekosistem eksternal) dengan memanfaatkan kompetensi digital untuk menciptakan model bisnis, produk, dan layanan baru. Transformasi digital memungkinkan perusahaan untuk memadukan pengalaman bisnis dan pelanggan digital dan fisik secara mulus sekaligus meningkatkan efisiensi operasional dan kinerja organisasi. Menurunkan biaya operasional dan meningkatkan pengalaman pelanggan merupakan inti dari transformasi digital. Transformasi digital yang sukses berfokus pada pembuatan, pengumpulan, dan analisis data pelanggan dan operasional sehingga keputusan berbasis data dapat diambil untuk meningkatkan proses. Kecerdasan buatan merupakan inti dari peningkatan proses otomatis ini.

Pada akhir tahun 2019, pengeluaran untuk transformasi digital akan mencapai Rp 170 triliun di seluruh dunia naik 42% dari tahun 2017, menurut laporan baru dari International Data Corporation. Paradigma bisnis baru bermunculan seiring dengan tergesernya bisnis lama oleh model-model baru yang lebih efisien. Belum lama ini, konsumen mengunjungi toko untuk menyewa DVD. Kini, bisnis penyewaan video telah tergeser sepenuhnya oleh bisnis daring seperti Netflix. Bisnis taksi telah terdisrupsi oleh Uber dan Lyft, dan dalam waktu dekat, pengemudi mungkin akan digantikan oleh kendaraan tanpa pengemudi. Pusat perbelanjaan konvensional perlu membangun proposisi nilai yang menarik pelanggan untuk datang langsung daripada berbelanja daring. Ketika pelanggan mengunjungi toko, inventaris harus tersedia dan mudah ditemukan.

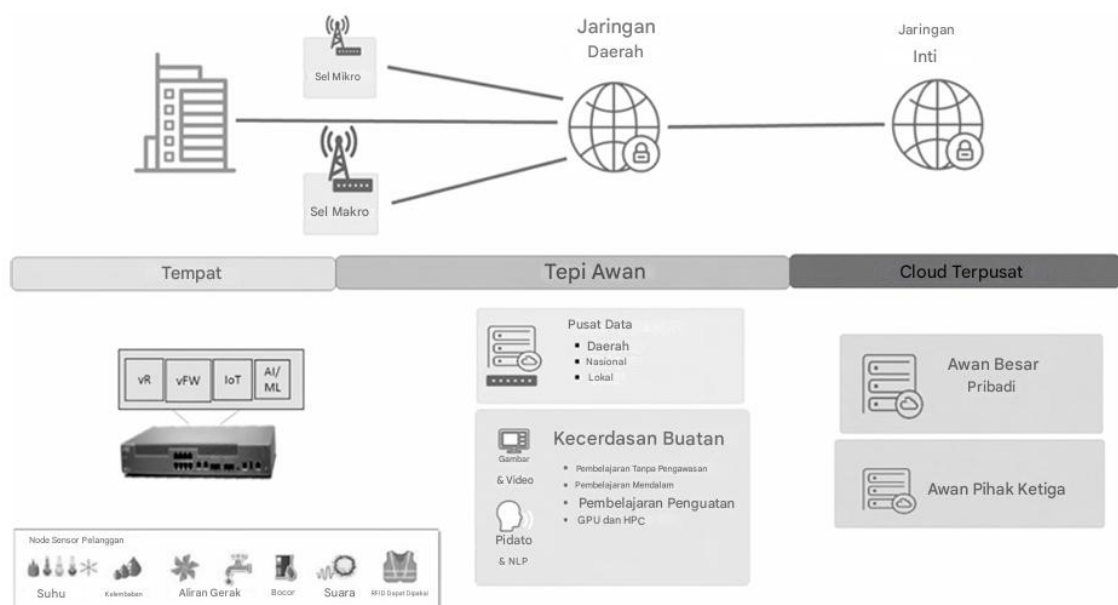
Selain mendorong perubahan mendasar dalam cara perusahaan melayani pelanggan dan meningkatkan kinerja operasional mereka, transformasi digital juga mendorong perubahan dalam jaringan yang melayani perusahaan. Pelanggan ingin menjadi pengemudi dalam desain, skala, dan otomatisasi jaringan mereka. Karakteristik ini memberi mereka kemampuan untuk menjadi gesit dan menawarkan layanan sesuai permintaan kapan pun dan di mana pun. Kecerdasan buatan memungkinkan generasi baru jaringan yang ditentukan perangkat lunak (SDN), yang memungkinkan pelanggan bisnis membangun hampir semua jaringan, mengoptimalkan kapasitas bandwidth mereka, mengurangi biaya transportasi per terabyte, dan merespons peristiwa jaringan secara dinamis. Mereka dapat memanfaatkan kekuatan data jaringan yang akan tersedia bagi mereka untuk lebih memahami pelanggan mereka, meningkatkan efektivitas pemasaran, dan memahami secara internal permintaan untuk memprioritaskan produk/fitur.



11.2 ARSITEKTUR PERUSAHAAN UNTUK TRANSFORMASI DIGITAL

Gambar 11.1 menunjukkan arsitektur jaringan untuk perusahaan yang bertransformasi secara digital di mana data dan pengambilan keputusan berbasis data mendorong proses dan operasi bisnis. Fungsi jaringan yang berada di lokasi perusahaan divirtualisasikan, sehingga lokasi tipikal berisi kotak putih yang berfungsi sebagai awan kecil yang menghosting router virtual, firewall, gateway *Internet of Things* (IoT), dan fungsi AI terbatas. Istilah "kotak putih" mengacu pada perangkat keras komoditas yang dapat digunakan untuk menghosting berbagai aplikasi virtual, seperti router dan firewall. Kotak putih yang terletak di lokasi pelanggan terkadang juga disebut sebagai *Universal Customer Premises Equipment* (UCPE). Fungsi kecerdasan buatan yang dapat dijalankan pada UCPE cenderung membutuhkan latensi rendah, komputasi dan penyimpanan terbatas, serta data eksternal atau jaringan terbatas.

Lokasi perusahaan terhubung ke jaringan regional menggunakan akses tetap atau nirkabel. Jaringan regional akan menjadi tuan rumah pusat data lokal dan regional yang menjadi tuan rumah fungsi jaringan, seperti perutean virtual, bersama dengan pemrosesan untuk fungsi jaringan. Pusat data regional juga dapat menjadi tuan rumah aplikasi AI perusahaan dengan kompleksitas sedang.



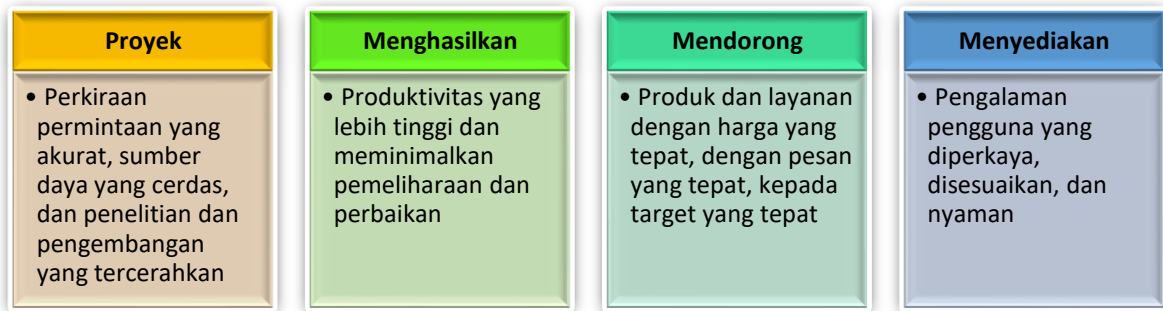
Gambar 11.1 Arsitektur untuk perusahaan yang bertransformasi digital.

Terakhir, jaringan inti akan menyediakan transportasi berkecepatan tinggi serta hosting untuk cloud privat dan pihak ketiga yang besar. Tugas AI intensif, seperti pelatihan model dan penyimpanan data dalam jumlah besar, akan dilakukan di inti, sedangkan model AI akan berjalan secara lokal dan di tepi jaringan. Solusi AI perusahaan yang paling sukses akan menggabungkan data privat yang dikumpulkan perusahaan dengan data yang dikumpulkan jaringan. Solusi kecerdasan buatan yang membutuhkan data perusahaan dan jaringan kemungkinan akan berada di pusat data jaringan regional atau inti yang aman yang menjamin privasi.



11.3 TRANSFORMASI PERUSAHAAN DAN KECERDASAN BUATAN

Perusahaan memanfaatkan arsitektur yang ditunjukkan pada Gambar 11.1 untuk pengambilan keputusan berbasis data guna menciptakan pengalaman pelanggan dan operasional yang efisien dan lancar. Di bagian ini, kami memberikan contoh bagaimana AI mentransformasi perusahaan.



Gambar 11.2 Empat area di mana AI dapat menciptakan nilai.

Contoh Kecerdasan Buatan dalam Perusahaan yang Bertransformasi Secara Digital

Sebuah studi terbaru oleh PWC menilai vertikal industri berdasarkan potensi AI. Vertikal teratas adalah layanan kesehatan, otomotif, jasa keuangan, logistik transportasi, teknologi, produk konsumen, minyak dan gas, serta manufaktur industri. Meskipun terdapat peluang yang berbeda untuk menerapkan AI di seluruh vertikal industri, peluang tersebut dapat dikategorikan berdasarkan area umum yang kondusif bagi transformasi digital.

Empat area di mana AI dapat menciptakan nilai dijelaskan dalam laporan McKinsey terbaru (Gambar 11.2).

1. **Proyek:** Memungkinkan perusahaan untuk memproyeksikan dan memperkirakan dengan lebih baik guna mengantisipasi permintaan, mengoptimalkan litbang, dan meningkatkan sumber daya.
2. **Produksi:** Meningkatkan kemampuan perusahaan untuk memproduksi barang dan jasa dengan biaya lebih rendah dan kualitas lebih tinggi. Meningkatkan produktivitas dan meminimalkan perawatan dan perbaikan.
3. **Promosi:** Mempromosikan penawaran dengan harga yang tepat, dengan pesan yang tepat, dan kepada target pelanggan yang tepat;
4. **Penyediaan:** Memberikan pengalaman pengguna yang kaya, personal, dan nyaman.

Dalam studi serupa, Cisco mengategorikan area yang terdampak AI ke dalam pemanfaatan aset, rantai pasokan dan logistik, keberlanjutan, inovasi, pengalaman pelanggan, dan produktivitas karyawan, yang mencakup jenis area yang sama yang disajikan dalam laporan McKinsey. Apa pun klasifikasinya, pesannya jelas. Transformasi digital bertenaga data mengubah pengalaman dan operasi pelanggan, mengurangi biaya, dan meningkatkan pendapatan. Perusahaan yang paling sukses menggunakan AI untuk membedakan diri dari pesaing mereka.



Di sisa bagian ini, kami menunjukkan bagaimana AI dapat diterapkan di perusahaan di masing-masing dari empat area yang didefinisikan oleh McKinsey: Proyek, Produksi, Promosikan, Menyediakan. Jika memungkinkan, kami menyoroti cara-cara di mana data yang dikumpulkan jaringan dapat dikombinasikan dengan data yang dikumpulkan perusahaan untuk memaksimalkan potensi AI. Persyaratan untuk menjaga privasi dan keamanan mendorong arsitektur AI perusahaan, yang akan memiliki komponen lokal dan jaringan yang bekerja bersama dalam cara pribadi yang aman.

Proyek

"Project" berfokus pada peramalan permintaan yang akurat, pengadaan cerdas, dan peningkatan produk. Hal ini memungkinkan perusahaan untuk memastikan ketersediaan inventaris yang memadai saat dibutuhkan konsumen, tanpa kelebihan inventaris. Dalam proses pengadaan dan manufaktur, hal ini membantu memastikan pemilihan pemasok terbaik berdasarkan kebutuhan biaya dan kualitas, dan dapat digunakan untuk mengidentifikasi peluang peningkatan produk.

Industri ritel sangat cocok untuk peramalan permintaan berbasis AI. Di era belanja daring, pelanggan cepat putus asa karena tidak tersedianya inventaris di toko fisik. Kurangnya inventaris di toko daring mengakibatkan hilangnya penjualan secara langsung. Kecerdasan buatan dapat digunakan untuk menganalisis data penjualan historis guna memprediksi kapan inventaris perlu diperbarui. Data lokasi jaringan, seperti jumlah pembeli di area tertentu toko, jumlah mobil di tempat parkir, atau tren lalu lintas di jalan-jalan sekitar, dapat digunakan untuk mengantisipasi permintaan di masa mendatang. Data ini dapat dikombinasikan dengan demografi populasi untuk mengidentifikasi jenis produk yang kemungkinan besar diminati. Untuk pengalaman berbelanja di toko, peritel dapat menggunakan AI untuk mengalokasikan kembali staf ke bagian-bagian toko yang mengalami lebih banyak lalu lintas atau menambah atau mengurangi personel menggunakan tren historis serta analitik lokasi di toko.

Perusahaan besar yang mendapatkan layanan jaringan dari berbagai vendor dapat menggunakan data penggunaan dan kesalahan jaringan sebagai bagian dari analitik prediktif berbasis AI untuk memperkirakan penggunaan di masa mendatang, membandingkan pemasok, dan menegosiasikan tarif untuk meminimalkan biaya. Secara lebih luas, data jaringan dapat digunakan untuk mengevaluasi kualitas jaringan dan pemasok, untuk menegosiasikan kontrak dan Perjanjian Tingkat Layanan, serta untuk memprediksi kegiatan pemeliharaan di masa mendatang. Pendekatan ini tidak terbatas pada layanan jaringan; pendekatan ini dapat diterapkan pada negosiasi dan kontrak pemasok apa pun, misalnya dalam industri otomotif. Dalam industri otomotif, sensor dapat digunakan untuk mengidentifikasi usia komponen, potensi kegagalan, dan mode kegagalan. Hal ini pada gilirannya dapat digunakan untuk mendorong manufaktur komponen, pemeliharaan, atau pemilihan vendor.

Dalam industri perawatan kesehatan, demografi lokasi dan klaim asuransi dapat digunakan untuk mengidentifikasi orang-orang yang kemungkinan sakit, dan memberikan perawatan pre-emptif atau mempersiapkan peningkatan permintaan (misalnya, ketika musim flu dimulai). Dengan izin privasi yang sesuai, data lokasi dapat ditambahkan untuk



mengidentifikasi individu yang mungkin telah terpapar penyakit menular. Sektor perawatan kesehatan juga mulai melihat eksplorasi dalam diagnostik berbasis data dan perawatan yang lebih personal. Hal ini dapat ditargetkan untuk pencegahan yang lebih efektif, dan dapat menghasilkan manfaat jangka panjang, seperti implan cerdas.

Penyedia layanan dapat menggabungkan data jaringan dengan data layanan pelanggan untuk meningkatkan pengalaman pelanggan. Kecerdasan buatan dapat digunakan untuk mengantisipasi kondisi yang dapat menyebabkan ketidakpuasan pelanggan dan dapat merekomendasikan perbaikan atau secara proaktif menghubungi pengguna yang terdampak. Penggunaan ini tidak terbatas pada data jaringan. Metode yang sama dapat diterapkan di perusahaan mana pun yang berorientasi layanan.

Produksi

"Produksi" berfokus pada peningkatan kemampuan perusahaan untuk memproduksi barang dan jasa dengan biaya yang lebih rendah dan kualitas yang lebih tinggi. Hal ini dapat mencakup peningkatan pemanfaatan aset dengan meminimalkan pemeliharaan dan mengoptimalkan perbaikan, manajemen energi, pemanfaatan ruang yang optimal, otomatisasi pabrik, atau rantai pasok dan logistik untuk meningkatkan manajemen dan pelacakan inventaris. Ada banyak aplikasi AI di area "Produksi", sebagian besar terkait dengan pengurangan biaya. Peluang utama untuk memanfaatkan data jaringan di bidang ini adalah data yang dihasilkan oleh sensor IoT.

Keberlanjutan dan manajemen energi merupakan pengeluaran besar bagi banyak bisnis. Menurut jajak pendapat Konsumsi Energi oleh National Foundation of Independent Businesses, biaya energi merupakan salah satu dari tiga pengeluaran bisnis teratas di 35% usaha kecil. Biaya energi utama untuk perusahaan kecil adalah:

- Mengoperasikan kendaraan, untuk 38% perusahaan kecil
- Pemanas dan/atau pendingin, untuk sepertiga perusahaan kecil
- Mengoperasikan peralatan, untuk seperlima perusahaan kecil

Usaha kecil memiliki daya ungkit yang relatif sedikit untuk mengatasi kenaikan biaya energi. Perubahan harga yang besar untuk mengimbangi kenaikan biaya energi dapat menjadi hal yang sulit, dan usaha kecil tidak mampu melakukan peningkatan peralatan secara berkala. Untuk usaha besar, manufaktur, atau pusat data, biaya energi mungkin bahkan lebih besar. Terdapat peluang yang signifikan bagi perusahaan untuk menggunakan pengukuran pintar (smart metering) guna mengukur konsumsi energi, air, atau gas alam. Data sensor *Internet of Things* (IoT), yang dikumpulkan dari jaringan perusahaan, dapat digunakan untuk meningkatkan prakiraan dan menyederhanakan konsumsi daya serta mengurangi pencurian energi. Di lingkungan perkantoran, AI dapat digunakan dalam kontrol otomatisasi gedung untuk menyesuaikan pengaturan suhu atau pencahayaan secara dinamis guna meminimalkan konsumsi energi. Kantor perusahaan menyediakan dan mengelola berbagai layanan pendukung untuk membantu memastikan produktivitas, kenyamanan, dan keselamatan karyawan. Banyak dari biaya layanan pendukung ini (petugas keamanan, layanan makanan, bandwidth jaringan, pemanas/pendingin ruangan, petugas kebersihan, dll.) berbanding lurus dengan metrik hunian karyawan. Pola musiman dan hari kerja umum yang sudah dikenal akan



digantikan dengan algoritma AI baru yang juga mempertimbangkan hari tertentu: kalender prospek karyawan, ruang rapat yang dipesan, pengaturan tujuan GPS mobil yang terhubung, analitik berbasis lokasi dari data makroseluler lalu lintas, konstruksi, berita, dan peristiwa cuaca. Prinsip yang sama dapat diterapkan pada perusahaan besar atau kecil, pusat data, dan perusahaan utilitas untuk mengoptimalkan konsumsi energi.

Aplikasi penting lainnya di bidang "Produksi" adalah pelacakan inventaris menggunakan sensor IoT. Misalnya, dalam industri pelayaran maritim, sensor membantu melacak lokasi kapal di laut. Dalam skala yang lebih kecil, sensor dapat memberikan informasi status dan suhu masing-masing kontainer kargo. Penginderaan suhu dapat digunakan untuk memastikan suhu yang tepat dipertahankan dalam kontainer pengiriman berpendingin untuk menjaga kesegaran. Atau, jika sistem pendingin mengalami kegagalan, operator dapat segera diberi tahu sehingga perbaikan dapat dilakukan sebelum barang rusak. Dalam lingkungan restoran, sensor IoT dapat menjadi bagian integral dari manajemen mutu dan minimisasi biaya. Misalnya, suhu dan viskositas minyak dapat dikelola agar makanan yang digoreng menghasilkan tekstur yang tepat. Viskositas minyak dapat diperiksa untuk menentukan waktu optimal untuk mengganti minyak goreng, yang selanjutnya dapat digunakan untuk mendorong kebijakan pengisian ulang minyak secara efisien.

Otomatisasi pabrik akan menjadi salah satu penggunaan AI yang paling utama. Robot akan digunakan untuk menggantikan manusia dalam tugas-tugas seperti pengambilan, pemilahan, dan pemuatan stok untuk pengiriman, dan juga akan digunakan untuk perakitan produk. Robot logistik yang dilengkapi kamera dan kecerdasan buatan dapat dilatih untuk mengenali ruang kosong dan menempatkan objek dengan benar. Semua aplikasi ini akan digerakkan oleh data jaringan nirkabel.

Data jaringan dengan latensi sangat rendah dan model AI yang kompleks untuk beroperasi dengan andal. Kecerdasan buatan dapat digunakan untuk mengoptimalkan dan mengadaptasi rute kendaraan guna meminimalkan waktu tempuh dan konsumsi bahan bakar. Teknologi yang sama ini dapat diterapkan pada mobil yang dikendarai manusia dalam armada perusahaan, atau oleh operator jaringan untuk mengoptimalkan pengiriman dan pergerakan truk.

Promosi

"Promosi" menciptakan penawaran dengan harga yang tepat, pesan yang tepat, dan kepada target pelanggan yang tepat. Contoh transformasi digital dengan AI meliputi analitik dan wawasan di dalam toko, interaksi yang dipersonalisasi, efektivitas promosi, pengalaman pelanggan berbasis wawasan, dan kemampuan untuk memonetisasi wawasan pembeli. Berdasarkan izin konsumen, banyak aplikasi AI dalam kategori ini memanfaatkan data jaringan seluler secara ekstensif, seperti riwayat belanja atau penelusuran dan lokasi pelanggan.

Dengan menggabungkan pengetahuan tentang preferensi pelanggan dengan riwayat penelusuran terkini dan lokasi pelanggan di dalam toko, AI dapat digunakan untuk mengirimkan kupon atau iklan bertarget kepada pelanggan saat mereka melewati layar yang berisi informasi yang mungkin ingin mereka beli. Harga produk dapat disesuaikan berdasarkan pola lalu lintas di dalam dan di luar toko dan faktor eksternal, seperti cuaca, waktu, atau hari.



Sephora memungkinkan pelanggan untuk "mencoba" produk mereka. Setelah mengunggah gambar, chatbot pintar akan membantu pelanggan memvisualisasikan berbagai gaya riasan dan produk, memberikan saran yang dipersonalisasi, dan menawarkan beberapa barang yang mungkin ingin dibeli pelanggan.

Di sektor perawatan kesehatan, data jaringan yang dihasilkan oleh sensor kesehatan dan lingkungan dapat digunakan untuk mendorong perilaku sehat (seperti lebih banyak berjalan kaki atau mengurangi makan), untuk mengidentifikasi rekomendasi pengobatan yang dipersonalisasi, atau dalam deteksi dini penyakit.

Di sektor energi, penetapan harga yang disesuaikan dapat dibuat menggunakan data yang dikumpulkan dari perangkat pengukur pintar. Kecerdasan buatan dapat digunakan untuk menggabungkan data ini, serta data demografi dan data penggunaan dari pelanggan lain, untuk menciptakan insentif penetapan harga yang disesuaikan bagi pengguna yang menghemat energi. Industri jasa keuangan sedang direvolusi oleh AI. Banyak kasus penggunaan yang tidak memerlukan data jaringan. Misalnya, industri ini menggunakan layanan robo-advisory bertenaga AI yang merekomendasikan perubahan portofolio yang dipersonalisasi berdasarkan tujuan keuangan klien atau toleransi risiko.

Menyediakan

"Provide" menggunakan AI untuk menciptakan pengalaman pengguna yang kaya, personal, dan nyaman. Pengalaman ini meningkatkan keterlibatan pelanggan, meningkatkan peluang peningkatan penjualan, dan mengurangi gesekan antara pengalaman online dan offline. Aplikasi kecerdasan buatan di bidang ini juga dapat digunakan untuk pelatihan atau memberi pelanggan atau karyawan perusahaan akses ke pembelajaran interaktif yang dipersonalisasi.

Pembahasan fungsi "penyediaan" tidak akan lengkap tanpa membahas layanan pelanggan. Terdapat peluang luar biasa untuk menggunakan chatbot AI guna menangani tugas-tugas rutin seperti menjawab pertanyaan standar atau menangani pengembalian barang. Hal ini membebaskan waktu karyawan untuk menangani tugas-tugas yang lebih kompleks dan dapat mengurangi jumlah karyawan yang dibutuhkan secara keseluruhan. Kecerdasan buatan dapat membantu mengotomatiskan proses layanan pelanggan dengan mengembangkan dan menerapkan program komputer berbasis percakapan yang berinteraksi langsung dengan pelanggan tanpa campur tangan manusia (atau menggunakan perwakilan layanan pelanggan manusia yang didukung oleh teknologi AI).

Di banyak industri, terutama di industri keuangan, penggunaan AI untuk meningkatkan keamanan sangatlah penting. Seiring meningkatnya jumlah pelanggaran keamanan, pengguna akan mengharapkan fitur-fitur seperti pengenalan suara/wajah, identifikasi sidik jari, dan pemindaian iris untuk meningkatkan manajemen identitas. Kamera keamanan yang mengirimkan video ke aplikasi AI yang mengidentifikasi individu yang meminta akses merupakan peluang besar bagi AI. Banyak dari aplikasi ini membutuhkan jaringan canggih. Sebuah aplikasi menarik yang membutuhkan data jaringan terdapat di industri asuransi. Perusahaan asuransi sedang bereksperimen dengan drone yang terbang di atas mobil-mobil yang rusak dan mengambil gambar yang diperlukan, untuk menghindari keharusan mengirim



anggota staf atau mengganggu jadwal pemilik mobil. Aplikasi-aplikasi ini dapat menghasilkan data jaringan dalam jumlah yang signifikan.

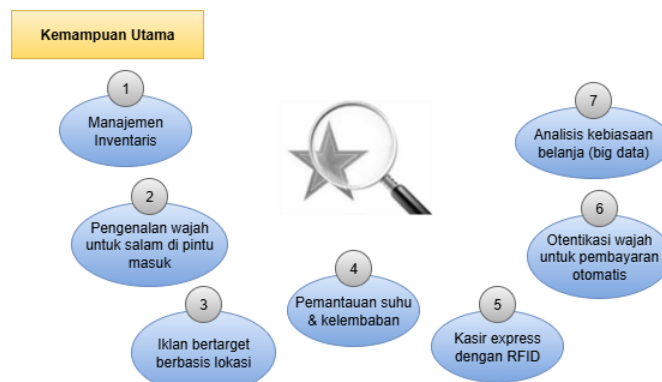
Aplikasi penting AI lainnya adalah pelatihan yang dipersonalisasi. Pelatihan ini dapat berupa pelatihan bagi staf operasi perusahaan tentang cara mengoperasikan peralatan dan layanan jaringan, atau pelatihan yang ditujukan bagi pelanggan. Kecerdasan buatan dapat digunakan untuk memodulasi pelatihan berdasarkan perilaku pengguna. Aplikasi realitas tertambah akan semakin meningkatkan pengalaman pelanggan dan dapat dipersonalisasi serta disampaikan pada waktu yang nyaman bagi pengguna.

Di sektor ritel, penggunaan AI yang menonjol adalah dalam sistem rekomendasi yang diterapkan oleh Amazon dan Netflix yang merekomendasikan barang yang mungkin disukai pelanggan berdasarkan preferensi dan pola belanja mereka. Panera Bread memiliki aplikasi seluler yang memungkinkan pelanggan untuk memesan terlebih dahulu. Toko Rebecca Minkoff memungkinkan pembeli untuk memilih barang dari layar digital dan menempatkan barang tersebut di ruang ganti. Kecerdasan buatan dapat digunakan untuk menyarankan barang tambahan atau menempatkannya di ruang ganti untuk dipertimbangkan oleh pelanggan. Jika aplikasi AI memiliki informasi tambahan tentang pelanggan, seperti usia, preferensi sebelumnya, atau tipe tubuh, informasi ini dapat digunakan untuk menyempurnakan saran.

Contoh Pengalaman Terintegrasi Toko Cerdas

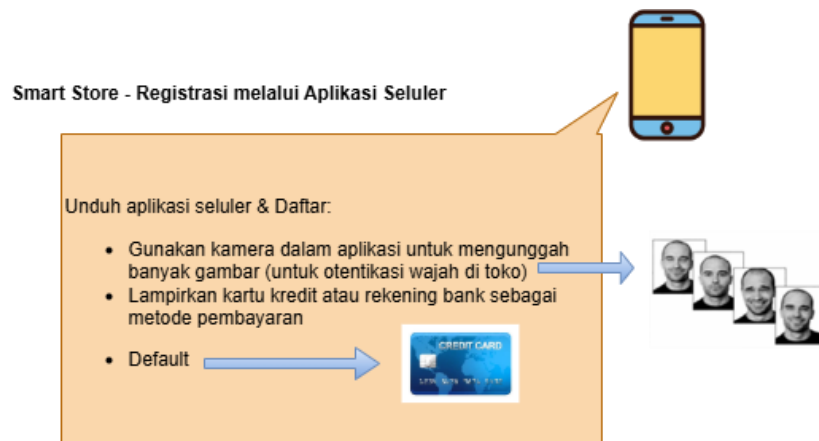
Di bagian ini, kami memberikan contoh untuk menunjukkan bagaimana kapabilitas di bagian sebelumnya dapat diintegrasikan untuk mendefinisikan ulang pengalaman pelanggan, dengan menggunakan "Toko Pintar" sebagai contoh. Kapabilitas utama yang digunakan dalam contoh ini ditunjukkan pada Gambar 11.3.

Pengalaman Smart Store dimulai ketika pengguna mengunduh aplikasi seluler untuk toko tersebut dan mendaftarkan identitas serta informasi pembayaran mereka, seperti yang ditunjukkan pada Gambar 11.4. Profil pengguna dan informasi pembayaran kemungkinan akan disimpan di cloud pribadi, yang dapat diakses secara aman oleh Smart Store saat pengguna masuk atau membayar. Pengguna dapat mengunggah berbagai informasi untuk autentikasi selanjutnya, seperti foto, sidik jari, dan sidik jari. Aktivitas transaksi kartu kredit dapat digunakan untuk mengidentifikasi kemungkinan penipuan menggunakan AI atau teknologi lainnya.



Gambar 11.3 Kemampuan utama yang ditunjukkan di Smart Store.





Gambar 11.4 Pendaftaran Smart Store Melalui Aplikasi Seluler.

Ketika pengguna memasuki Smart Store seperti yang ditunjukkan pada Gambar 11.5, AI digunakan untuk pengenalan wajah menggunakan model yang diimplementasikan secara lokal pada UCPE di toko, yang telah dilatih di cloud publik atau privat di intinya. Algoritma kecerdasan buatan juga dapat menciptakan algoritma autentikasi multifaktor yang menggunakan profil lokasi pelanggan untuk memperkuat autentikasi. Pelanggan dapat diberikan sapaan yang dipersonalisasi dan mungkin rekomendasi penawaran khusus untuk produk yang mungkin mereka minati berdasarkan riwayat pembelian dan preferensi pelanggan serupa. Jika toko tidak memiliki staf, keamanan dapat diatur agar hanya pelanggan terdaftar yang dapat memasuki toko.

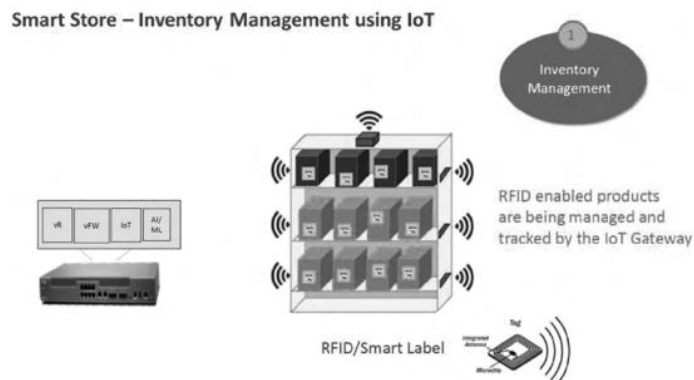


Gambar 11.5 Menggunakan pengenalan wajah dengan AI saat memasuki Smart Store.

Kembali di gudang, inventaris dilacak menggunakan Label Pintar seperti yang ditunjukkan pada Gambar 11.6. Kecerdasan buatan digunakan untuk mengidentifikasi kapan barang perlu diisi ulang dan untuk melacak lokasi barang di dalam gudang. Jika pengambilan dan penempatan dilakukan oleh robot, Label Pintar dapat digunakan untuk menemukan inventaris dan mengidentifikasi di mana harus menempatkan inventaris baru di rak. Dengan melacak kesegaran barang di gudang, pemasok dapat memastikan bahwa barang yang paling lama dikirim terlebih dahulu, sekaligus memastikan bahwa barang yang basi tidak dikirim.



Kebijakan pengiriman juga dapat disesuaikan dengan karakteristik penjualan masing-masing toko.



Gambar 11.6 Manajemen inventaris Toko Pintar menggunakan IoT.



Gambar 11.7 Penggunaan IoT untuk periklanan dan pemantauan.

Kecerdasan buatan juga dapat digunakan untuk meningkatkan kualitas pengalaman berbelanja di toko seperti yang ditunjukkan pada Gambar 11.7. Data demografi dan pemasaran pelanggan dapat digabungkan dengan data lokasi jaringan dan data sensor identifikasi frekuensi radio untuk menayangkan iklan bertarget di dalam toko. Dengan hadirnya belanja daring, pentingnya kualitas produk di dalam toko semakin meningkat. Sensor suhu dapat dipantau untuk terus menyesuaikan diri dengan suhu optimal, dan kesegaran dapat dilacak sehingga barang lama dapat segera dikeluarkan dari rak toko. Karyawan atau pelanggan dapat dengan cepat menemukan barang melalui aplikasi yang melacak lokasi semua barang dagangan, dan karyawan dapat dengan cepat menemukan barang yang tidak pada tempatnya. Kecerdasan buatan dapat diterapkan untuk memprediksi kapan inventaris baru harus ditambahkan, untuk menentukan barang mana yang laku, dan untuk menentukan penempatan inventaris yang optimal di dalam toko.

Terakhir, Gambar 11.8 menunjukkan pengalaman pembayaran. Kecerdasan buatan digunakan untuk autentikasi pengguna, dan riwayat pembelian disimpan di cloud pribadi atau jaringan tempat AI dapat diterapkan untuk tujuan penjualan dan pemasaran di masa mendatang. Chatbot kecerdasan buatan dapat digunakan untuk memproses pengembalian rutin, dan data pengembalian dapat dianalisis untuk menyempurnakan pesanan di masa mendatang.



Transformasi Perusahaan dengan Kecerdasan Buatan

Sebagaimana dibahas di bagian sebelumnya, transformasi digital telah menciptakan peluang baru bagi AI untuk mentransformasi perusahaan, yang menghasilkan peningkatan efisiensi operasional, pengurangan biaya, dan pengalaman pelanggan yang lebih personal. Transformasi digital ini mengubah hakikat bisnis, mendisrupsi cara pelanggan berinteraksi dengan perusahaan. Beberapa kasus penggunaan AI yang dibahas di bagian ini akan ditawarkan melalui aplikasi vertikal yang memperhitungkan kebutuhan unik pelanggan dalam suatu sektor. Aplikasi-aplikasi ini perlu mengintegrasikan data dari berbagai sumber secara mulus data profil pelanggan, data demografi, data pemasaran dan penjualan, data industri, data publik, dan data jaringan. Data jaringan dapat dikumpulkan dari jaringan pribadi pelanggan atau jaringan penyedia layanan, yang mencakup data sensor,



Gambar 11.8 Pengalaman checkout di Toko Pintar.

Data lokasi seluler, data penggunaan jaringan, dan data transaksi pelanggan yang dikumpulkan melalui jaringan. Pengumpulan dan pengelolaan data yang sangat besar yang dibutuhkan oleh aplikasi-aplikasi ini akan menimbulkan tantangan yang signifikan bagi perusahaan, dan aplikasi AI itu sendiri mungkin memerlukan daya komputasi yang signifikan. Perusahaan besar harus memilih antara mengelola aplikasi AI sendiri atau membayar layanan yang menjalankan fungsi-fungsi ini. Perusahaan kecil yang ingin memanfaatkan AI mungkin tidak memiliki keahlian atau sumber daya untuk melakukan investasi AI sendiri.

Cloud akan memainkan peran sentral dalam mendukung AI dalam transformasi digital. Perusahaan akan memanfaatkan private cloud, public cloud, dan provider cloud untuk menyimpan dan memproses data, seperti yang ditunjukkan pada Gambar 11.1. Komputasi intensif sumber daya, seperti pelatihan model AI, dapat diterapkan pada public cloud atau provider cloud yang besar, sementara eksekusi model yang telah dilatih dapat dilakukan di private cloud lokal yang lebih kecil. Penyedia jaringan dan layanan perlu mengakomodasi beberapa cloud, bekerja sama untuk mencapai tujuan transformasi digital.

Transformasi digital juga menuntut penyedia jaringan dan layanan untuk menyediakan layanan yang lebih gesit dan adaptif. Mereka juga perlu memanfaatkan AI untuk memberikan kelincuhan dan kualitas yang dibutuhkan perusahaan agar dapat mengimbangi inovasi yang pesat. Transformasi jaringan perusahaan akan dibahas di bagian selanjutnya.



11.4 TRANSFORMASI JARINGAN PERUSAHAAN DAN KECERDASAN BUATAN

Untuk memfasilitasi transformasi digital suatu organisasi, perusahaan pertama-tama harus mentransformasi jaringan dan arsitekturnya agar lebih fleksibel, aman, dan dinamis untuk menangani kebutuhan proses dan alur kerja digital baru. Tekanan persaingan, peningkatan ekspektasi dalam hal waktu respons, dan kebutuhan akan efektivitas biaya mengharuskan jaringan sangat tersedia dan menyediakan manajemen swalayan bagi administrator untuk merespons kebutuhan bisnis dengan lebih cepat. Persyaratan baru ini mewakili perubahan signifikan bagi jaringan yang ada di dalam perusahaan. Jaringan dan layanan perusahaan di masa depan memerlukan persyaratan pengadaan, implementasi, dan operasi yang sangat berbeda, di mana jaringan sebagai komponen dapat dengan cepat dan mudah diintegrasikan ke dalam kasus penggunaan spesifik transformasi digital.

Transformasi jaringan dimulai beberapa tahun lalu dengan penerapan prinsip-prinsip SDN, virtualisasi, dan cloud. Para *Chief Information Officer* (CIO) perusahaan sedang mengembangkan strategi yang "mendefinisikan perangkat lunak" berbagai aspek arsitektur, proses, dan organisasi jaringan mereka, serta mengadopsi proses operasional baru yang lebih gesit, lebih otomatis, dan memberikan total biaya kepemilikan yang lebih baik.

Kecerdasan buatan dan teknologi terkait, seperti pembelajaran mesin, pembelajaran mendalam, dan chatbot, semakin meningkatkan transisi ini dengan memanfaatkan sejumlah besar data yang dapat dikumpulkan dan diintegrasikan di seluruh komponen perangkat lunak. Bagian berikut akan menjelaskan transisi industri terkini di berbagai bagian jaringan perusahaan dan bagaimana AI akan semakin mendorong transisi ini. Gambar 11.1 menunjukkan arsitektur perusahaan tingkat tinggi. Aspek-aspek arsitektur perusahaan yang paling diuntungkan oleh AI telah dijelaskan. Hal ini juga dibahas pada bab sebelumnya.

- Pusat Data dan Transformasi Cloud
- Transisi Jaringan Area Luas (WAN) ke WAN Terdefinisi Perangkat Lunak (SD-WAN), termasuk virtualisasi jaringan lokal
- Jaringan Area Lokal (LAN) Berkabel dan Nirkabel

Transformasi Data Center dan Cloud dengan AI

Ketika seseorang mempertimbangkan aset infrastruktur teknologi informasi (TI) perusahaan yang paling berharga dan paling mahal, pusat data hampir selalu menempati posisi teratas. Pusat data sering dianggap sebagai tulang punggung infrastruktur TI perusahaan yang menjalankan semua aplikasi perusahaan dan menyimpan data internal atau pelanggan. Gangguan apa pun pada pusat data dapat menghentikan operasional perusahaan. Oleh karena itu, organisasi di seluruh vertikal terus menuntut peningkatan keandalan, pemanfaatan optimal, dan pengurangan biaya terkait operasional pusat data.

Transisi cloud dimulai karena meskipun telah mengadopsi virtualisasi, lingkungan pusat data saat ini masih mahal untuk dioperasikan dan dikelola. Banyak perusahaan mulai menggunakan infrastruktur cloud eksternal untuk beban kerja sementara dan mengadopsi model perangkat lunak sebagai layanan (*software-as-a-service*) untuk banyak aplikasi, alih-alih mendedikasikan komputasi dan penyimpanan yang mahal di pusat data privat. Namun, dari perspektif perusahaan, meskipun menjalankan aplikasi di cloud "off-premise"



menyederhanakan lingkungan TI internal, persyaratan keandalan, pemanfaatan, dan biaya secara keseluruhan tetap berlaku. Hanya saja, mereka kini mewajibkan penyedia cloud untuk memenuhi persyaratan yang sama.

Kecerdasan buatan dan teknologi terkait menjanjikan solusi bagi banyak masalah inti, terlepas dari strategi pusat data perusahaan. Contoh awal AI yang memecahkan beberapa masalah inti ini muncul ketika Google mulai menggunakan teknologi dari perusahaan rintisan AI, "DeepMind", untuk memangkas biaya dan meningkatkan efisiensi perangkat komputasi. Menurut, Google mengurangi konsumsi daya pusat datanya sebesar 15% dan penggunaan daya pendingin sebesar 40% dengan menemukan dan memantau inefisiensi di 120 variabel pusat data, termasuk kipas, pendingin, dan jendela, sehingga menghemat ratusan juta dolar bagi perusahaan selama beberapa tahun.

Saat ini, semua penyedia solusi pusat data, seperti Cisco, Aruba, HP, dan lainnya, memanfaatkan AI dan konsep terkait seperti pembelajaran mesin untuk meningkatkan efektivitas biaya, kemudahan operasional, dan meningkatkan keandalan. Beberapa masalah yang berpotensi dipecahkan oleh AI di pusat data antara lain:

- Peningkatan waktu aktif, keandalan, dan optimalisasi pemanfaatan server komputer dan infrastruktur penyimpanan dengan menghasilkan lebih banyak informasi telemetri bagi algoritma AI untuk menyebarkan beban.
- Pengurangan konsumsi energi dan daya di lingkungan pusat data yang elastis, dicapai dengan menginstansiasi aplikasi dan node server secara on-demand.
- Alokasi sumber daya teknis menggunakan analitik prediktif dan prinsip jaringan berbasis intent yang memungkinkan sumber daya teknis untuk mengomunikasikan intent mereka kepada orkestrator, yang kemudian mengubah intent mereka menjadi kebijakan khusus produk.
- Mengurangi risiko operasional dengan memantau node komputer dan penyimpanan secara berkelanjutan dan mengalihkan lalu lintas ke node lain yang tersedia berdasarkan probabilitas node yang kelebihan beban pada waktu tertentu dalam sehari atau hari dalam setahun.
- Mengurangi risiko keamanan dengan memantau alur aplikasi sambil membandingkannya dengan baseline yang terus berkembang dan memperingatkan pengguna tentang deviasi utama.

Di masa mendatang, kami yakin banyak perangkat AI akan terintegrasi ke dalam solusi manajemen infrastruktur pusat data tradisional yang disediakan oleh vendor. Perangkat AI juga dapat diintegrasikan ke dalam solusi pusat data sebagai layanan bernilai tambah yang memungkinkan perusahaan melakukan operasi, seperti penambangan data otomatis, analitik data, dan keamanan data, untuk aplikasi yang berjalan di node komputasi dan server. Dari perspektif penyedia cloud, langkah serupa dapat memungkinkan mereka menawarkan layanan bernilai tambah baru untuk semakin memperkuat kasus bisnis dalam memindahkan aplikasi tambahan ke lingkungan cloud. Singkatnya, perusahaan akan mendapatkan manfaat dari AI di cloud dalam berbagai cara.



- Perusahaan besar dengan private cloud dapat memanfaatkan teknik AI yang sama yang digunakan oleh operator pusat data untuk meningkatkan efisiensi dan ketahanan menggunakan skala masuk/keluar dan migrasi aplikasi.
- Perusahaan dapat menggunakan AI untuk memperkirakan beban kerja dan memigrasikan beban puncak sementara ke penyedia cloud eksternal, sehingga menghemat biaya kapasitas mahal yang jarang digunakan.
- AI bergantung pada cloud. Perusahaan akan memanfaatkan cloud secara ekstensif untuk penyimpanan dan pemrosesan data, memanfaatkan private cloud di lokasi, cloud jaringan regional, dan cloud pihak ketiga untuk menyeimbangkan kebutuhan biaya, kapasitas, keamanan, dan privasi.
- Perusahaan dapat memanfaatkan perangkat AI saat tersedia sebagai layanan cloud.

Jaringan Area Luas Prediktif Berbasis Kecerdasan Buatan

Didefinisikan Perangkat Lunak

WAN perusahaan pada dasarnya adalah jaringan yang menghubungkan berbagai segmen jaringan perusahaan, seperti lokasi cabang, lokasi kantor pusat, kantor regional, kantor pusat, dan titik akhir seluler. WAN perusahaan tradisional didasarkan pada desain statis dan teknologi WAN khusus atau khusus. WAN yang ditentukan perangkat lunak merupakan solusi baru yang memungkinkan perusahaan memenuhi berbagai persyaratan terkait transisi cloud dan adopsi seluler dengan cara yang hemat biaya. SD-WAN adalah aplikasi spesifik teknologi SDN yang diterapkan pada koneksi WAN, yang digunakan untuk menghubungkan jaringan perusahaan dalam jarak geografis yang luas. SD-WAN bertujuan untuk mengganti router cabang tradisional dengan perangkat yang menggunakan virtualisasi, kebijakan tingkat aplikasi, dan overlay jaringan untuk membuat beberapa tautan Internet tingkat konsumen berperilaku seperti sirkuit khusus. Tujuannya adalah untuk menyederhanakan pengaturan sehingga satu-satunya hal yang perlu dilakukan oleh personel kantor cabang adalah mencolokkan kabel aplikasi untuk "menelepon ke rumah" dan secara otomatis menerima konfigurasinya dari pengontrol pusat.

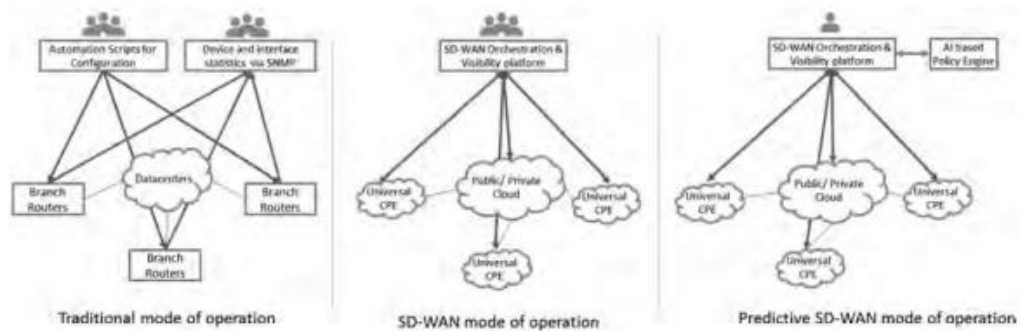
Gartner telah mendefinisikan SD-WAN memiliki empat karakteristik yang diperlukan:

- Kemampuan untuk mendukung berbagai jenis koneksi, seperti *multi-protocol label switching* (MPLS), frame relay, dan komunikasi nirkabel LTE berkecepatan tinggi
- Kemampuan untuk melakukan pemilihan jalur dinamis, untuk berbagi beban dan tujuan ketahanan
- Antarmuka sederhana yang mudah dikonfigurasi dan dikelola
- Kemampuan untuk mendukung jaringan privat virtual (VPN), dan layanan pihak ketiga seperti pengendali optimasi WAN, firewall, dan gateway web.

Gambar 11.9 menunjukkan evolusi konektivitas WAN. Jaringan WAN tradisional bergantung pada router cabang khusus dan peralatan gedung lainnya yang saling terhubung menggunakan koneksi WAN statis, biasanya MPLS atau jalur privat. Dengan SD-WAN, router cabang digantikan oleh UCPE. UCPE adalah bentuk "kotak putih", perangkat keras komoditas yang dapat digunakan untuk menghosting berbagai aplikasi virtual, seperti router dan firewall. (lihat juga Gambar 11.1). Koneksi WAN statis digantikan oleh berbagai jenis koneksi, termasuk



koneksi internet tingkat konsumen yang dikelola untuk menciptakan konektivitas yang andal. SD-WAN prediktif menggunakan AI untuk mengoptimalkan bandwidth di jaringan SD-WAN.



Gambar 11.9 Mode operasi WAN tradisional, SD-WAN, dan SD-WAN prediktif.

Tiga aspek utama solusi SD-WAN dalam menghemat biaya:

- Penyediaan zero-touch: Solusi SD-WAN memungkinkan penerapan WAN edge (UCPE) yang dikontrol cloud tanpa memerlukan pra-penataan router atau teknisi di lokasi, sehingga menghindari biaya yang terkait dengan pengangkutan truk untuk penerapan CPE.
- Independensi transportasi: SD-WAN menggunakan kontrol perangkat lunak untuk menambahkan fitur-fitur baru, seperti transmisi ulang paket, perutean berbasis kualitas jalur, dan duplikasi paket, untuk menjadikan internet atau broadband sebagai transportasi yang dapat diterima untuk lalu lintas kelas perusahaan, alih-alih memerlukan transportasi MPLS untuk keamanan dan Kualitas Layanan.
- Manajemen dan orkestrasi: SD-WAN menyediakan antarmuka kebijakan grafis sederhana untuk memvisualisasikan telemetri jaringan dan model kebijakan berbasis aplikasi yang lebih sederhana, sehingga menghindari kebutuhan akan pekerja berkeahlian tinggi dan alat khusus untuk manajemen kebijakan.

Dengan penggunaan antarmuka kebijakan yang disederhanakan dan analitik data jaringan yang dimungkinkan oleh SD-WAN, administrator jaringan kini lebih mudah merancang dan mengoperasikan jaringan. Admin dapat menggunakan antarmuka grafis sederhana untuk merancang jaringan dan grafik interaktif visual untuk memantau operasi, serta menentukan kebijakan yang jauh lebih terperinci untuk aliran lalu lintas tingkat aplikasi.

Kecerdasan buatan akan semakin mentransformasi solusi SD-WAN yang ada menjadi "SD-WAN prediktif". Meskipun solusi SD-WAN saat ini menyediakan antarmuka kebijakan yang disederhanakan, sebagian besar bersifat statis dan dikontrol secara manual. Meskipun SD-WAN menyediakan telemetri jaringan, penggunaan utamanya terbatas pada pembuatan laporan untuk memahami topologi jaringan, perilaku lalu lintas aplikasi, dan penggunaan.

SD-WAN Prediktif akan mengintegrasikan pembelajaran berkelanjutan ke dalam ekosistem WAN untuk membuka jalan bagi konsep-konsep populer, seperti jaringan yang dapat memulihkan diri sendiri dan melindungi diri sendiri. Banyak penyedia SD-WAN saat ini sedang mengevaluasi bagaimana mereka dapat memanfaatkan AI untuk meningkatkan solusi



SD-WAN mereka guna meraih keunggulan kompetitif. Berikut adalah beberapa contoh bagaimana teknologi AI dapat digunakan untuk SD-WAN perusahaan.

- Menggunakan AI untuk membuat topologi VPN dinamis. Dalam solusi SD-WAN saat ini, VPN dan jaringan akses internet dikonfigurasi secara statis dengan bandwidth tetap oleh teknisi berkeahlian tinggi yang bertanggung jawab untuk membuat dan mengelola desain jaringan secara keseluruhan. Dengan menggunakan AI, dimungkinkan untuk membuat peta jaringan dari jaringan underlay di mana layanan overlay seperti VPN tidak perlu persisten dan dapat dibuat secara dinamis berdasarkan permintaan lalu lintas. Kemampuan ini akan mengurangi kompleksitas desain jaringan secara signifikan.
- Antarmuka kebijakan berbasis maksud. Kebijakan berbasis maksud mengacu pada konsep di mana administrator dapat mendefinisikan dan mengomunikasikan kebijakan yang diinginkan ke platform orkestrasi jaringan tanpa harus memahami topologi jaringan. Secara umum, konsep ini sangat berbeda dari cara pengelolaan jaringan perusahaan saat ini, di mana maksud pengguna diterjemahkan ke dalam desain jaringan oleh para ahli desain jaringan dan terkadang diimplementasikan tanpa validasi yang tepat dari maksud awal. Kemajuan terbaru dalam konsep-konsep seperti pemodelan data jaringan (yaitu, kemampuan untuk mengatur elemen jaringan, antarmuka, atribut, dll. dalam format terstruktur dan sebagai objek yang dapat dialamatkan) beserta kemajuan dalam teknologi chatbot dan Pemrosesan Bahasa Alami memberikan kemampuan bagi administrator jaringan untuk berkomunikasi dengan jaringan menggunakan antarmuka suara atau obrolan, sementara seorang orkestrator mengonversi maksud pengguna menjadi kebijakan yang dipahami oleh orkestrator jaringan.
- Pemecahan masalah jaringan dan analisis akar penyebab. Pemecahan masalah jaringan dan analisis akar penyebab merupakan area lain di mana banyak vendor mencoba menggunakan konsep AI untuk penyelesaian masalah yang lebih cepat. Sebagian besar vendor saat ini mengelola basis data masalah jaringan yang terus berkembang. Dengan menggunakan sumber data ini sebagai sampel pembelajaran, vendor dapat meningkatkan identifikasi akar penyebab secara signifikan.
- Jaringan pertahanan diri. Deteksi dan mitigasi serangan secara otonom. Jaringan perusahaan semakin rentan terhadap serangan siber. Dengan memanfaatkan kekayaan data yang dapat dikumpulkan dan dianalisis menggunakan SDN, AI dapat digunakan untuk mendeteksi dan memitigasi serangan secara otomatis.
- Performa aplikasi. Teknologi AI dapat digunakan untuk melacak berbagai parameter dan karakteristik lalu lintas yang berkaitan dengan performa aplikasi dan untuk secara dinamis memasukkan layanan tambahan seperti akselerasi WAN berdasarkan arus lalu lintas individual.
- Manajemen bandwidth dinamis. Alokasi bandwidth dalam jaringan tradisional bersifat statis, hanya dapat diubah menggunakan proses pemesanan layanan manual. Dengan menggunakan AI yang dipadukan dengan kemungkinan konfigurasi real-time yang diperkenalkan oleh SD-WAN, bandwidth dapat disesuaikan secara dinamis berdasarkan



penggunaan real-time, kondisi jaringan, musim, variasi waktu, dinamika pasar, dan peristiwa eksternal seperti keadaan darurat. Hal ini memungkinkan jaringan beroperasi lebih efisien dan pelanggan mengoptimalkan biaya dan kinerja mereka.

- Jaringan yang dapat memulihkan diri sendiri. Dengan menggabungkan data telemetri jaringan dengan algoritma pembelajaran mesin untuk menentukan kesehatan perangkat jaringan dan probabilitas terkait kegagalan jaringan akibat kegagalan, jaringan dapat menerapkan langkah-langkah untuk mengambil tindakan korektif secara mandiri, seperti memulai ulang perangkat untuk menghindari atau memperbaiki masalah jaringan.

Sebagaimana terlihat dari contoh-contoh sebelumnya, teknologi AI dapat diintegrasikan dalam banyak aspek jaringan WAN perusahaan. Namun, secara umum, terdapat tema umum pada banyak kasus penggunaan yang dijelaskan sebelumnya yang dapat membantu merangkum dampak teknologi AI terhadap lingkungan WAN. Pada dasarnya, AI adalah kemampuan mesin atau platform jaringan untuk memahami dan mengkorelasikan banyak aliran data dan parameter serta memfasilitasi aktivitas yang mahal, rawan kesalahan, atau mewujudkan visi jaringan WAN yang belajar mandiri, menyembuhkan diri sendiri, dan mempertahankan diri sendiri.

Optimalisasi LAN Kabel dan Nirkabel untuk Kecerdasan Buatan

"LAN perusahaan" mungkin merupakan bagian terpenting dari jaringan perusahaan. Hal ini karena jaringan LAN menyediakan konektivitas ke pengguna akhir dan/atau perangkat tempat data dibuat dan dikonsumsi. Jaringan perusahaan lainnya, termasuk WAN dan pusat data, dibangun untuk melayani pengguna akhir, perangkat, dan data yang mereka hasilkan.

Dalam beberapa dekade terakhir, LAN perusahaan, yang terdiri dari jaringan kabel dan nirkabel, telah berkembang secara signifikan dalam hal kinerja, ukuran, dan kompleksitas karena semakin banyaknya pengguna, perangkat, dan aplikasi yang terhubung. Untuk mengatasi peningkatan kompleksitas dan merespons kebutuhan bisnis dengan cepat, beberapa perusahaan juga mempertimbangkan solusi seperti "SDN" yang dirancang untuk mengelola lingkungan LAN perusahaan. Namun, seperti SD-WAN, solusi LAN yang ada saat ini berfokus pada penyediaan antarmuka yang "disederhanakan" untuk manajemen pengguna/perangkat, kapasitas, dan kinerja menggunakan kebijakan statis.

Ketika kita membayangkan masa depan LAN perusahaan dengan perangkat IoT generasi baru, perangkat pintar yang terhubung dengan setiap pengguna, serta perangkat/pengguna yang selalu aktif dan bergerak, tantangan bagi LAN perusahaan meningkat secara signifikan. Misalnya,

- Meningkatnya jumlah perangkat dan "benda yang terhubung" akan membutuhkan lebih banyak sumber daya jaringan, seperti port LAN kabel dan kapasitas nirkabel.
- Ledakan aplikasi yang digunakan oleh perangkat baru, yang masing-masing membutuhkan waktu respons terbaik, memerlukan skema manajemen aliran jaringan yang lebih baru.
- Sifat dinamis perangkat-perangkat ini yang bergabung dengan jaringan hanya ketika dibutuhkan akan menciptakan tantangan bagi penskalaan jaringan.



- Area serangan yang lebih luas dari ratusan jenis perangkat menimbulkan risiko keamanan yang jauh lebih tinggi yang harus dikelola.

Solusi jaringan LAN yang ada di pasaran mungkin hanya dapat mengatasi tantangan sebelumnya dengan menerapkan jumlah port yang lebih banyak atau jaringan nirkabel yang lebih padat. Kami yakin penggunaan AI dan teknologi terkait dapat semakin meningkatkan lingkungan LAN perusahaan besar. Misalnya,

- Penerapan titik akses baru mungkin dapat dihindari jika teknologi AI dapat digunakan untuk meningkatkan daya pancar titik akses "terdekat" yang ada, bergantung pada "populasi pengguna" di dekat titik akses tertentu yang diterapkan untuk melayani area tertentu.
- Algoritma AI baru dapat digunakan untuk mengoptimalkan infrastruktur kabel atau nirkabel yang ada dengan meningkatkan akurasi rasio kelebihan langganan dengan memantau penggunaan bandwidth, jenis aplikasi, dan prioritas.
- Keamanan dapat dikelola dengan lebih baik dengan menerapkan algoritma pembelajaran mesin baru sebagai bagian dari autentikasi perangkat dan pengguna.

11.5 DAMPAK AI PADA JARINGAN DAN ARSITEKTURNYA

Tren industri saat ini menunjukkan bahwa lingkungan perusahaan generasi berikutnya akan dipenuhi dengan lebih banyak perangkat yang dibekali kemampuan kecerdasan buatan (AI). Karena perangkat-perangkat tersebut sangat bergantung pada streaming telemetry dan mesin kebijakan yang terkait, maka wajar jika kebutuhan terhadap throughput, keandalan, keamanan, serta kapabilitas analitik dari jaringan WAN perusahaan akan semakin meningkat. Tuntutan baru ini pada akhirnya dapat mendorong evolusi lebih lanjut dalam jaringan WAN.

Meski demikian, pergeseran arsitektur yang paling signifikan justru dapat berasal dari perubahan pada lingkungan pusat data (*datacenter*) maupun komputasi awan (*cloud*). Dalam arsitektur generasi saat ini, perusahaan umumnya memanfaatkan layanan cloud publik maupun privat yang terpusat untuk menjalankan workload aplikasi. Namun, dalam beberapa tahun terakhir, dengan adanya adopsi konsep seperti virtualisasi fungsi jaringan (*network function virtualization*) dan teknologi cloud sumber terbuka, workload di cabang-cabang lokal perusahaan kini juga dapat merasakan manfaat cloud. Manfaat tersebut mencakup kemudahan penyebaran, elastisitas, serta infrastruktur yang tidak bergantung pada perangkat keras tertentu.

Dengan meningkatnya jumlah perangkat di cabang secara eksponensial, pendekatan cloud terpusat untuk menyimpan dan menganalisis data menjadi semakin mahal. Beberapa pihak di industri mulai mengeksplorasi arsitektur baru, di mana algoritme pembelajaran mesin yang dijalankan di lokasi terpusat dapat mengendalikan dan memberi instruksi kepada perangkat WAN edge untuk mengumpulkan data tambahan atau menyaring data yang tidak diperlukan. Hal ini bertujuan untuk mengurangi kebutuhan penyimpanan dan komputasi di pusat.

Selain itu, karena AI dan teknologi terkait semakin banyak disematkan pada berbagai perangkat yang digunakan dalam aktivitas sehari-hari, kebutuhan pemrosesan berbasis graphical processing unit (GPU) di lokasi cabang akan meningkat secara signifikan. Kemungkinan besar, perangkat WAN edge generasi mendatang (seperti UCPEs) akan



dilengkapi GPU, selain CPU tradisional, guna mempercepat pemrosesan deep learning, analitik, dan aplikasi rekayasa. Beberapa skenario penggunaan yang mendukung terjadinya “desentralisasi” cloud ke tingkat cabang (*branch edge*) antara lain mencakup:

- Menjalankan workload yang sensitif terhadap latensi atau jenis workload lain yang hanya relevan bagi cabang lokal. Dengan cara ini, tim lokal memperoleh kendali lebih besar tanpa harus membuka titik akses tambahan pada workload yang berjalan di cloud publik maupun privat.
- Menjalankan workload untuk melakukan pra-pemrosesan atau merangkum data secara lokal, sehingga dapat meningkatkan keamanan serta mengurangi biaya yang muncul dari penyimpanan data di cloud publik maupun privat.
- Memindahkan berbagai fungsi fisik ke lingkungan virtual, yang memungkinkan struktur biaya lebih efisien, kontrol yang lebih baik dalam desain high-availability, serta kemudahan operasional.
- Membangun edge cloud secara sementara (*on-demand*) untuk kebutuhan khusus, seperti acara, pop-up store, dan lain-lain, sekaligus mengurangi ketergantungan pada konektivitas WAN berkecepatan tinggi yang harus tersedia sepanjang waktu (24 × 7).

Untuk menjawab berbagai skenario penggunaan tersebut, arsitektur perusahaan perlu berevolusi dengan memperluas simpul cloud terpusat hingga ke tingkat branch edge. Perubahan ini akan menggeser arsitektur aplikasi menjadi terdistribusi, yakni terbagi antara simpul cloud terpusat tradisional dan simpul UCPE yang berada di lokasi pelanggan. Aplikasi berbasis kecerdasan buatan akan memanfaatkan pola ini dengan menjalankan aplikasi komputasi berat di jaringan inti (*core network*), sementara aplikasi yang membutuhkan respons waktu nyata (*real-time*) atau mendekati waktu nyata dijalankan di edge. Dengan demikian, jaringan dapat membuat keputusan otonom untuk mengoptimalkan fungsi mana yang sebaiknya dijalankan di lokasi pelanggan, di edge, maupun di core.

Sebagai contoh, model keamanan dapat dilatih di core lalu dijalankan di edge atau di lokasi pelanggan. Sebaliknya, fungsi keamanan waktu nyata dapat dilatih di edge lalu dieksekusi di lokasi pelanggan. Tren menuju komputasi edge ini akan mendorong arsitektur aplikasi untuk didesain ulang secara mendasar agar dapat memanfaatkan kapabilitas baru tersebut, sekaligus mengoptimalkan konsumsi sumber daya jaringan dan memungkinkan latensi rendah. Dari transformasi ini, aplikasi kecerdasan buatan akan menjadi salah satu pihak yang memperoleh manfaat paling besar.

11.6 RINGKASAN

Dalam bab ini telah dibahas berbagai peluang penerapan kecerdasan buatan (AI) di lingkungan perusahaan. Didukung oleh arus transformasi digital, pemanfaatan AI di perusahaan dapat dipetakan ke dalam dua ranah utama: pertama, transformasi model bisnis perusahaan, dan kedua, transformasi jaringan perusahaan agar selaras dengan model bisnis baru tersebut.

Transformasi model bisnis perusahaan sangat dipengaruhi oleh perkembangan *Internet of Things* (IoT), serta kemampuan untuk menyimpan, memproses, dan mengintegrasikan beragam data di cloud. Perubahan pada model bisnis inilah yang kemudian mendorong terjadinya pergeseran mendasar pada arsitektur jaringan perusahaan. Faktor pendorong utama transformasi jaringan meliputi: pergeseran pusat data menuju cloud, transisi WAN



menuju SD-WAN yang mencakup virtualisasi jaringan di lokasi pelanggan, serta transformasi LAN kabel maupun nirkabel. Jaringan perusahaan yang telah bertransformasi ini selanjutnya membutuhkan dukungan AI untuk dapat memenuhi tuntutan fleksibilitas, keamanan, serta kemampuan adaptasi dinamis yang dibutuhkan oleh model bisnis digital masa kini. Seiring berlanjutnya transformasi tersebut, dapat diperkirakan bahwa evolusi jaringan akan terus berlangsung dengan penambahan fungsi-fungsi baru di tingkat network edge. Hal ini ditujukan untuk memenuhi tuntutan ketat dari berbagai aplikasi baru yang semakin kompleks dan menuntut kinerja tinggi.



BAB 12

KECERDASAN BUATAN UNTUK 5G+ DAN IOT

12.1 PENDAHULUAN

Kecerdasan buatan (AI) bukan hanya teknologi menarik untuk meningkatkan akurasi dan kemampuan prediksi pada berbagai permasalahan, tetapi pada akhirnya menjadi kebutuhan utama untuk mengekstraksi kecerdasan dari jumlah data yang sangat besar yang dihasilkan oleh jaringan modern. Saat ini, data yang tersedia tidak hanya berukuran masif, tetapi juga tumbuh dengan sangat cepat. Bahkan, 90% data dunia dihasilkan hanya dalam kurun waktu dua tahun terakhir. Sebagai ilustrasi, jumlah data digital global saat ini mencapai 7,9 zettabyte, dan diperkirakan akan meningkat hingga 44 zettabyte pada tahun 2020, dengan setiap individu menghasilkan rata-rata 1,7 Megabyte data setiap detiknya. Bab ini akan menguraikan motivasi serta faktor-faktor pendorong di balik ledakan data tersebut, sekaligus membahas peran serta solusi yang ditawarkan oleh teknologi 5G, komputasi edge, dan *Internet of Things* (IoT).

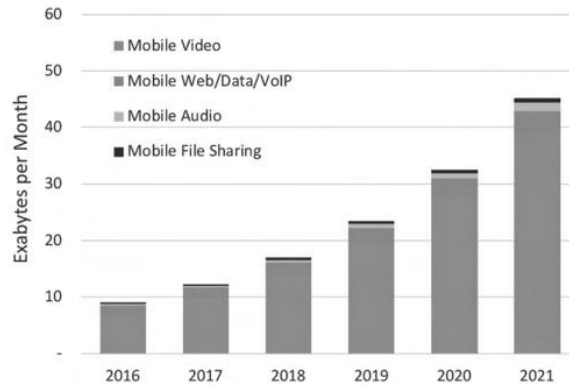
12.2 LATAR BELAKANG DAN MOTIVASI EVOLUSI TEKNOLOGI

Sejak peluncuran iPhone pada tahun 2007 oleh AT&T dan Apple, penggunaan data mengalami lonjakan luar biasa hingga mencapai 250.000% pada periode 2007–2017 untuk AT&T, dan tren serupa juga terlihat di seluruh industri. Pertumbuhan konsumsi data ini terutama didorong oleh hadirnya jaringan dengan kecepatan lebih tinggi, berkembangnya penggunaan smartphone, serta kemunculan berbagai aplikasi yang memudahkan akses dan distribusi konten digital, khususnya video. Menurut laporan AT&T dan Cisco, mayoritas lalu lintas pada jaringan nirkabel modern kini berasal dari video, dengan proporsi lebih dari 60%. Angka ini bahkan diproyeksikan akan meningkat hingga 78% pada tahun 2021, sebagaimana ditunjukkan pada Gambar 12.1. Hal ini menegaskan bahwa video telah menjadi pendorong utama dalam pertumbuhan permintaan data, sekaligus menjadi tantangan besar bagi arsitektur jaringan untuk menyediakan kapasitas, keandalan, dan kualitas layanan yang memadai.

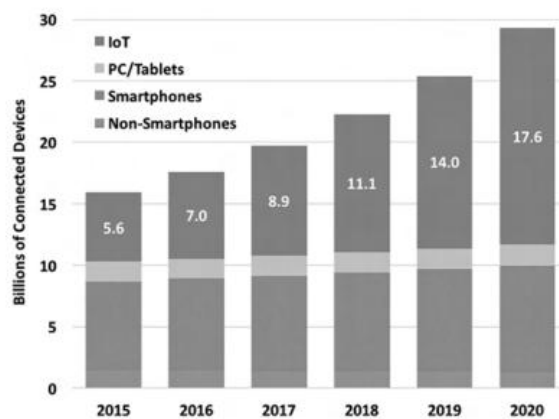
Selain terjadinya pergeseran penggunaan jaringan nirkabel dari layanan suara, teks, dan penelusuran web menuju dominasi video, terdapat perubahan lain yang signifikan terkait komposisi perangkat dalam jaringan nirkabel. Pada tahun 2017, penetrasi ponsel di Amerika Serikat telah mencapai hampir rasio 1:1 per individu (sekitar 81% menurut [4]). Hal ini menandakan bahwa pertumbuhan utama pada masa mendatang tidak lagi berasal dari ponsel, melainkan dari perangkat terkoneksi lainnya. Ericsson memprediksi bahwa perangkat IoT akan menjadi kategori dengan pertumbuhan terbesar dalam beberapa tahun ke depan. Jumlah perangkat IoT diperkirakan meningkat dari 5,6 miliar perangkat pada tahun 2015 menjadi sekitar 17,6 miliar perangkat pada tahun 2020, sebagaimana ditunjukkan pada Gambar 12.2 dan 12.3. Lonjakan ini menegaskan bahwa IoT tidak hanya mengubah pola lalu lintas data, tetapi juga secara fundamental memengaruhi rancangan arsitektur jaringan nirkabel agar mampu menangani skala, keberagaman, dan kompleksitas perangkat yang semakin tinggi.



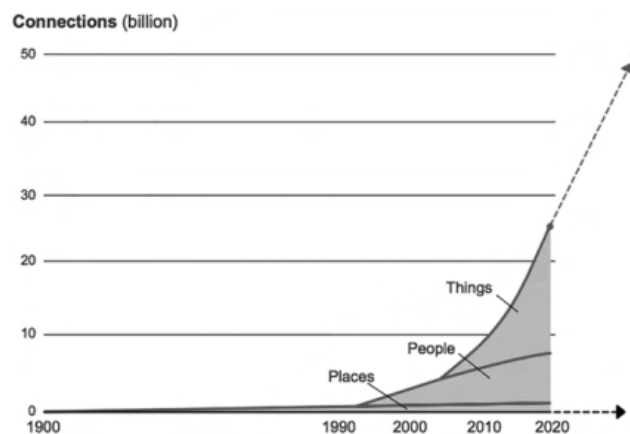
Evolusi arsitektur jaringan nirkabel: Seiring dengan pertumbuhan penggunaan data yang masif dan perubahan tren bauran perangkat, terdapat sejumlah tren arsitektur yang sedang berlangsung seperti yang ditunjukkan pada Tabel 12.1.



Gambar 12.1. Perkiraan IoT dan Perangkat Terkoneksi



Gambar 12.2. Perkiraan IoT dan Perangkat Terkoneksi



Gambar 12.3. Pertumbuhan Perangkat Terkoneksi sebagai Pendorong Utama Pertumbuhan di Masa Depan



Tabel 12.1 Tren Evolusi Arsitektur Jaringan Nirkabel

Karakteristik Jaringan	Saat Ini	2020+
Penerapan situs seluler	Jumlah situs satu digit per km persegi	Puluhan situs per makrosel per kilometer persegi
Konektivitas Backbone	Serat optik dengan kepadatan rendah	Serat optik dengan kepadatan tinggi
Jaringan Inti (Core Network)	Gabungan antara jaringan tradisional dan virtual (NFVs)	Virtualisasi penuh
Jaringan RAN	RAN yang terdesentralisasi, beberapa small cell	Cloud RAN, Komputasi Tepi (Edge Computing)
Perangkat	- Hingga 4x4 MIMO - Agregasi spektrum (hingga 50 MHz) + beberapa pita tak berlisensi	- Banyak antenna (Massive MIMO) - Pemanfaatan spektrum 10x - Agregasi spektrum (5 jalur) + pita tak berlisensi (contohnya: pita 3,5 MHz)

Peningkatan kepadatan situs seluler dengan memanfaatkan banyak small cell pada setiap macrocell kini tengah dilakukan operator untuk mengimbangi lonjakan permintaan lalu lintas data yang sangat tinggi. Konsekuensinya, hal ini mendorong kebutuhan akan tulang punggung serat optik yang lebih padat. Secara paralel, demi memperoleh efisiensi biaya dan operasional, operator juga mulai mengadopsi virtualisasi masif pada *core network* (CN). Sebagai ilustrasi, pada tahun 2017 AT&T telah memvirtualisasi lebih dari 55% infrastruktur core network-nya, dan operator lain pun berusaha mengikuti langkah serupa. Meskipun software-defined radio telah lama menjadi teknologi andalan di industri pertahanan, virtualisasi pada *radio access network* (RAN) diperkirakan akan segera diimplementasikan secara luas di industri telekomunikasi. Konsep virtual RAN kemungkinan akan menempatkan sebagian fungsi di cloud, sementara sebagian lainnya tetap berada di edge sesuai kebutuhan aplikasi.

Kemampuan perangkat pengguna juga semakin berkembang, sehingga memungkinkan pemanfaatan spektrum nonkontigu dengan lebar pita yang lebih luas, baik berlisensi maupun tidak berlisensi, menggunakan berbagai standar antarmuka udara. Sebagian dari evolusi arsitektur ini, seperti virtualisasi, dapat direalisasikan tanpa perlu perubahan standar. Namun demikian, banyak aspek lainnya yang akan semakin dipermudah dan dipercepat melalui evolusi standar menuju 5G+, yang akan dirangkum pada bagian berikutnya.

5G+

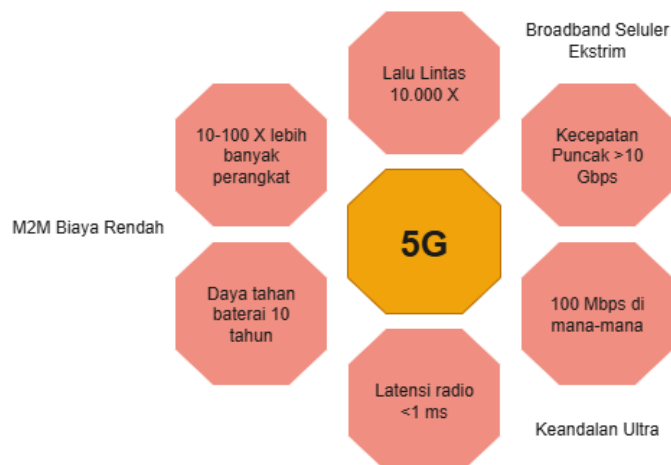
Peningkatan kepadatan situs seluler dengan memanfaatkan banyak small cell pada setiap macrocell kini tengah dilakukan operator untuk mengimbangi lonjakan permintaan lalu lintas data yang sangat tinggi. Konsekuensinya, hal ini mendorong kebutuhan akan tulang punggung serat optik yang lebih padat. Secara paralel, demi memperoleh efisiensi biaya dan operasional, operator juga mulai mengadopsi virtualisasi masif pada *core network* (CN). Sebagai ilustrasi, pada tahun 2017 AT&T telah memvirtualisasi lebih dari 55% infrastruktur core network-nya, dan operator lain pun berusaha mengikuti langkah serupa.

Meskipun software-defined radio telah lama menjadi teknologi andalan di industri pertahanan, virtualisasi pada *radio access network* (RAN) diperkirakan akan segera



diimplementasikan secara luas di industri telekomunikasi. Konsep virtual RAN kemungkinan akan menempatkan sebagian fungsi di cloud, sementara sebagian lainnya tetap berada di edge sesuai kebutuhan aplikasi. Kemampuan perangkat pengguna juga semakin berkembang, sehingga memungkinkan pemanfaatan spektrum nonkontigu dengan lebar pita yang lebih luas, baik berlisensi maupun tidak berlisensi, menggunakan berbagai standar antarmuka udara. Sebagian dari evolusi arsitektur ini, seperti virtualisasi, dapat direalisasikan tanpa perlu perubahan standar. Namun demikian, banyak aspek lainnya yang akan semakin dipermudah dan dipercepat melalui evolusi standar menuju 5G+, yang akan dirangkum pada bagian berikutnya.

Beberapa target kinerja ditetapkan dengan membandingkannya terhadap 4G (LTE), misalnya peningkatan jumlah perangkat hingga 10–100 kali lipat dan lonjakan lalu lintas data hingga 10.000 kali lebih besar. Sementara itu, terdapat pula target absolut, seperti latensi kurang dari 1 milidetik, kecepatan puncak mencapai 10 Gbps, ketersediaan kecepatan minimal 100 Mbps di semua lokasi, serta masa pakai baterai minimal 10 tahun untuk perangkat machine-to-machine (M2M).



Gambar 12.4 Persyaratan kinerja 5G sebagaimana ditetapkan dalam standar IMT-2020.

Beberapa kemampuan baru dalam kategori extreme mobile broadband, seperti kecepatan puncak 10 Gbps, ketersediaan 100 Mbps di semua lokasi, dan lonjakan lalu lintas hingga 10.000 kali lipat, berpotensi membuka peluang baru bagi penerapan fixed wireless yang lebih hemat biaya di wilayah suburban dan perkotaan. Sebelumnya, teknologi ini umumnya hanya digunakan di daerah pedesaan. Dengan adanya peningkatan kemampuan ini, tumbuh optimisme baru. Seperti yang dituliskan dalam laporan ABI Research, jaringan 5G diperkirakan akan mendorong pertumbuhan fixed wireless hingga mencakup 151 juta orang pada tahun 2022, dengan AT&T dan Verizon yang telah memulai implementasi sejak tahun 2017.

Tidak dapat dipungkiri, peningkatan lalu lintas data hingga 10.000 kali lipat dengan hadirnya 5G menimbulkan tantangan besar dalam hal optimasi jaringan, perutean, serta pemanfaatan wawasan data. Sebagai gambaran, operator sebesar AT&T pada tahun 2016 mengelola sekitar 118 Petabyte data per hari. Jika volume ini meningkat 10.000 kali, maka jumlah yang harus diproses mencapai sekitar 1,18 Zettabyte (10²¹ byte) per hari. Sebagai perbandingan, total data global pada tahun 2015 hanya sekitar 7,5 Zettabyte. Dengan skala



sebesar itu, optimasi jaringan maupun analisis data hampir mustahil dilakukan secara terpusat. Oleh karena itu, penerapan solusi komputasi tepi menjadi suatu keharusan.

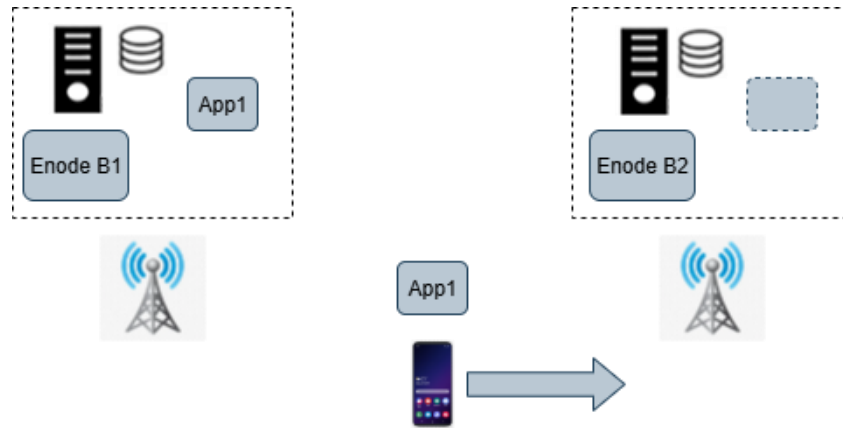
Komputasi EDGE Seluler

Selain kebutuhan akan kemampuan komputasi di tepi jaringan untuk mengolah volume data yang sangat besar demi memperoleh wawasan, pergeseran komputasi ke arah edge juga didorong oleh sejumlah motivasi tambahan. Sebagai contoh, Ahme merangkum beberapa alasan tersebut sebagai berikut:

- **Optimalisasi konten dinamis:** Sama seperti fungsi *Content Distribution Networks* (CDN) saat ini, komputasi tepi dan penyimpanan lokal memungkinkan konten ditempatkan lebih dekat dengan pengguna. Hal ini menurunkan latensi, biaya transportasi data, serta beban server cloud. Bergantung pada jenis konten yang di-cache di edge dan aplikasi yang digunakan pengguna, kecerdasan buatan (AI) dapat membantu mengatur aliran lalu lintas, sehingga penghematan biaya transportasi bisa sangat signifikan.
- **Pelepasan beban komputasi, terutama untuk IoT:** Dengan adanya puluhan miliar perangkat IoT yang diperkirakan akan terhubung termasuk perangkat wearable dengan kemampuan pemrosesan terbatas komputasi tepi memungkinkan aplikasi dengan kebutuhan pemrosesan tinggi, seperti realitas tertambah atau aplikasi pengawasan, berjalan lebih optimal. Misalnya, jika seseorang menggunakan headset VR/AR dengan proses rendering dilakukan di cloud, gerakan kepala yang cepat dapat menimbulkan keterlambatan bila pemrosesan dilakukan di pusat data jauh. Dengan edge computing, hal ini dapat diatasi.
- **Transportasi cerdas:** Dalam ekosistem yang serba terhubung mulai dari ponsel pintar, meteran lalu lintas, lampu lalu lintas, hingga kendaraan lain kemampuan untuk berkoordinasi dan berbagi informasi secara lokal dengan latensi rendah membuka peluang bagi aplikasi transportasi cerdas. Data yang dibagikan dapat langsung dimanfaatkan untuk menghasilkan wawasan yang bermanfaat bagi komunitas setempat.

Gambar 12.5 menunjukkan contoh bagaimana komputasi tepi diintegrasikan ke dalam jaringan nirkabel untuk mendukung aplikasi-aplikasi tersebut maupun aplikasi lainnya. Dalam arsitektur ini, sebuah server *Mobile Edge Compute* (MEC) dengan local cache ditambahkan dan ditempatkan berdampingan dengan eNode-B, serta ditambahkan pula cache tambahan di dalam *Core Network* (CN).





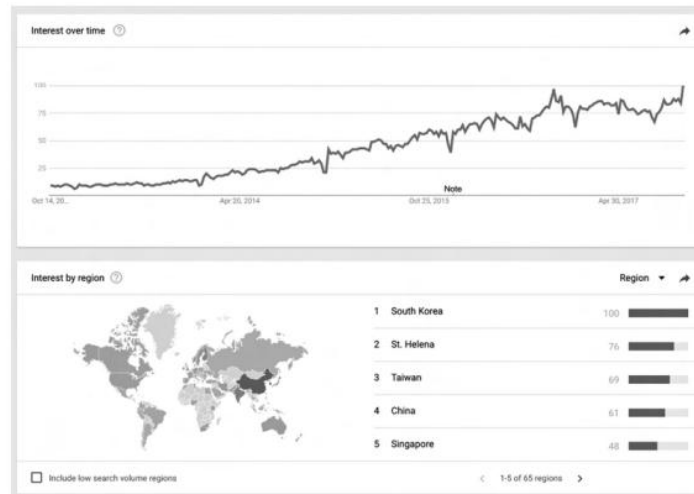
Gambar 12.6 Persyaratan kinerja 5G sebagaimana didefinisikan oleh IMT-2020.

Internet untuk Segala Hal

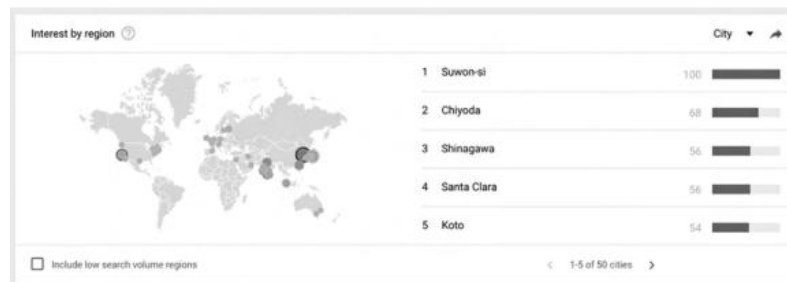
Seperti telah dijelaskan pada Bagian 12.2, pertumbuhan utama jumlah perangkat di jaringan nirkabel masa depan diproyeksikan berasal dari “perangkat yang saling terhubung”, atau yang dikenal sebagai *Internet of Things* (IoT). Seiring dengan meningkatnya penggunaan, wajar jika muncul pula peningkatan kesadaran publik terhadap IoT. Hal ini tampaknya dikonfirmasi oleh data dari Google Trends, yang merekam seberapa sering suatu istilah dicari oleh pengguna dalam hal ini kata kunci “IoT” (lihat Gambar 12.7). Data tersebut menunjukkan adanya peningkatan minat pencarian terhadap IoT sekitar 10 kali lipat sejak periode 2012–2013, dengan tingkat pencarian tertinggi berasal dari Asia Tenggara, khususnya Korea Selatan, Tiongkok, dan Singapura.

Gambar 12.8 Seperti telah dijelaskan pada Bagian 12.2, pertumbuhan utama jumlah perangkat di jaringan nirkabel masa depan diproyeksikan berasal dari “perangkat yang saling terhubung”, atau yang dikenal sebagai *Internet of Things* (IoT). Seiring dengan meningkatnya penggunaan, wajar jika muncul pula peningkatan kesadaran publik terhadap IoT. Hal ini tampaknya dikonfirmasi oleh data dari Google Trends, yang merekam seberapa sering suatu istilah dicari oleh pengguna dalam hal ini kata kunci “IoT” (lihat Gambar 12.7). Data tersebut menunjukkan adanya peningkatan minat pencarian terhadap IoT sekitar 10 kali lipat sejak periode 2012–2013, dengan tingkat pencarian tertinggi berasal dari Asia Tenggara, khususnya Korea Selatan, Tiongkok, dan Singapura.

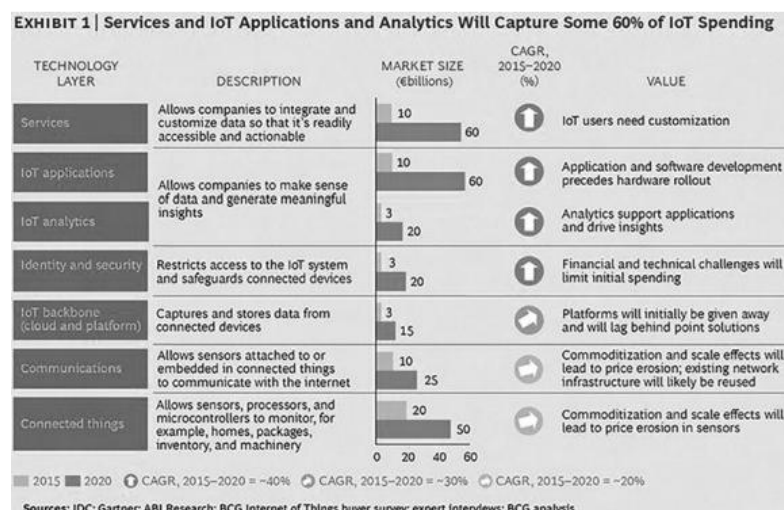




Gambar 12.7 Minat IoT dari waktu ke waktu, dan minat berdasarkan wilayah dari data Google Trends (11/2017).



Gambar 12.8 Area minat kota IoT dari data Google Trends (11/2017).



Gambar 12.9 Analisis IoT, pertumbuhan aplikasi dan layanan, serta potensinya.

- **Maskapai Penerbangan:** Contoh yang sering dikemukakan oleh GE adalah bahwa mesin jet modern mampu mengumpulkan sekitar 500 GB data dalam satu penerbangan, mencakup lebih dari 5.000 parameter dari berbagai sensor yang merekam data suhu, ketinggian, hingga performa mesin. Jika dibandingkan dengan



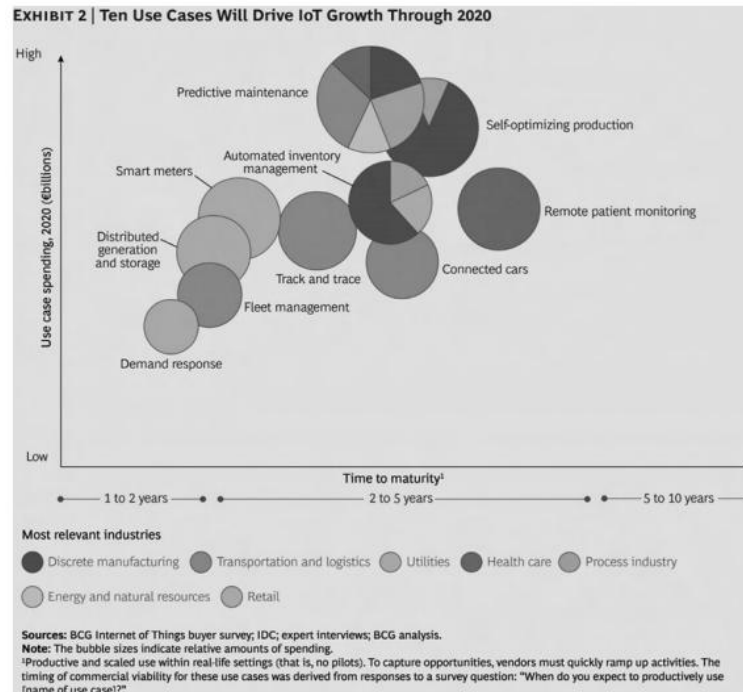
mesin generasi lama yang hanya mampu mengumpulkan sekitar 1 KB data, teknologi ini berhasil meningkatkan keandalan hingga 287 kali lipat.

- **Anjungan Minyak Lepas Pantai:** Perusahaan minyak kini melengkapi peralatan pengeboran yang bernilai tinggi dengan sensor canggih untuk mendeteksi tanda-tanda degradasi serta menggunakan pengenalan visual guna merekomendasikan pemeliharaan preventif. Pendekatan ini membantu menghindari kerugian akibat downtime yang mahal dan memungkinkan inspeksi peralatan tanpa perlu mengirimkan tim teknis secara langsung.
- **Transportasi dan Manufaktur:** Saat ini sudah menjadi hal yang umum bahwa kendaraan dilengkapi dengan beragam sensor. Beberapa model, seperti Tesla, bahkan memanfaatkan sensor tersebut untuk mendukung fungsi mengemudi otonom. Di masa depan, diperkirakan sensor tidak hanya berkomunikasi di dalam kendaraan, tetapi juga berinteraksi dengan sensor di sepanjang jalan maupun dengan kendaraan lain di sekitarnya.
- **Bangunan dan Rumah Pintar:** AT&T telah mengembangkan platform M2X yang memungkinkan jutaan perangkat terkoneksi untuk mengirimkan data ke cloud. Sebagai contoh, produk AT&T “Digital Life” Smart Home memungkinkan berbagai perangkat rumah pintar untuk saling terhubung dan terintegrasi melalui layanan berbasis cloud.
 - Termostat pintar
 - Sensor udara pintar
 - HVAC pintar
 - Air pintar
 - Dapur pintar
 - Keamanan rumah
 - Pemantauan anak
- **Sensor tubuh/perangkat yang dapat dikenakan:** Banyak orang sudah menggunakan jam tangan yang terhubung, gelang olahraga, atau monitor lain yang berinteraksi satu sama lain, ponsel pintar, dan cloud.

Dalam menghubungkan berbagai perangkat IoT, kecerdasan buatan (AI) dipandang sebagai faktor kunci yang memungkinkan berjalannya berbagai aplikasi analitik (Gambar 12.10). Dua penerapan terbesar adalah perawatan prediktif dan produksi yang mampu mengoptimalkan diri. Contoh perawatan prediktif dapat ditemukan pada industri perkeretaapian, di mana kondisi rel serta perkiraan usia pakainya dapat ditentukan dengan memantau getaran rel menggunakan sensor ketika kereta melintas. Sementara itu, produksi yang mampu mengoptimalkan diri merupakan kategori luas yang memanfaatkan sistem pembelajaran mesin berbasis closed-loop. Sistem ini secara berkelanjutan dapat mengevaluasi suatu “fungsi objektif” keseluruhan, lalu melakukan optimasi terhadap berbagai peralatan (misalnya pengaturan, penggunaan, atau penyesuaian) serta proses produksi. Tujuannya adalah untuk memperbaiki potensi cacat, meningkatkan hasil produksi, sekaligus menekan biaya maupun konsumsi energi.



Tidak mengherankan jika para analis memprediksi potensi besar bagi penerapan AI pada IoT. Hal ini karena manfaat IoT skala besar tidak akan bisa dimaksimalkan apabila tumpukan data yang dihasilkan tidak dapat diproses secara otomatis dan cerdas.



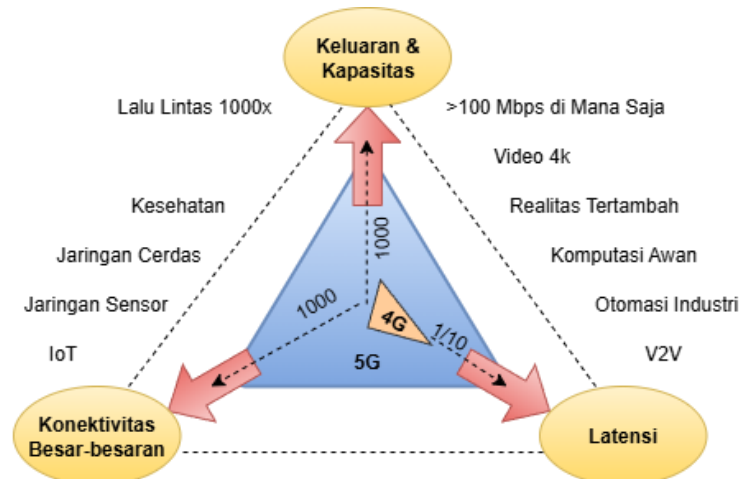
Gambar 12.10 Kasus penggunaan IoT.

12.3 ARSITEKTUR JARINGAN DAN AKSES

Faktor Pendorong Utama bagi Arsitektur Baru

Untuk mewujudkan janji menghadirkan konektivitas di mana saja, kapan saja, bagi siapa saja dan untuk segala perangkat, jaringan masa depan (5G dan generasi setelahnya) harus mampu menjawab tantangan utama yang berkaitan dengan kapasitas, kecepatan data, latensi ujung ke ujung (end-to-end/E2E), jumlah koneksi yang sangat besar, biaya, serta kualitas pengalaman pengguna. Selain itu, jaringan juga dituntut untuk secara bersamaan mendukung berbagai aplikasi dengan karakteristik yang sangat beragam. Beberapa kasus penggunaan yang menjadi contoh adalah: mobile broadband berkecepatan tinggi lebih dari 1 Gbps, kota pintar, rumah pintar dengan jumlah sensor yang masif, hingga aplikasi ultra-reliable dan berlatensi rendah seperti mobil otonom serta IoT yang bersifat kritis terhadap misi. Semua skenario tersebut menunjukkan nilai ekstrem dalam hal kecepatan, kepadatan koneksi, latensi, serta tingkat keandalan. Untuk penjelasan lebih rinci, lihat Gambar 12.11.





Gambar 12.11 Kasus penggunaan 5G.

Teknologi *New Radio* (NR), peningkatan lebar pita spektrum, virtualisasi fungsi jaringan (NFV), jaringan yang ditentukan perangkat lunak (SDN), arsitektur jaringan yang fleksibel, komputasi edge, serta manajemen dan optimasi jaringan berbasis AI merupakan faktor kunci yang memungkinkan terwujudnya 5G. Baik *Radio Access Network* (RAN) baru maupun *Next Generation Core* (NGC) dirancang untuk memenuhi berbagai skenario penggunaan sebagaimana diuraikan dalam. Lebih lanjut, pemisahan fungsi control plane memungkinkan percepatan inovasi dari berbagai pihak baik vendor, operator, maupun penyedia layanan pihak ketiga untuk mengontrol perilaku jaringan secara real-time melalui SDN. Arsitektur 5G juga mendukung konsep *network slicing*, yaitu kemampuan operator untuk membagi sumber daya jaringan agar dapat disesuaikan dengan kebutuhan layanan tertentu. Namun, meningkatnya fleksibilitas jaringan juga menghadirkan kompleksitas tambahan. Jumlah fungsi jaringan dan antarmuka, banyaknya *network slice*, jumlah koneksi, *throughput* data pengukuran, serta kebutuhan kontrol yang sangat detail hingga ke tingkat koneksi individu, semuanya mendorong kebutuhan otomatisasi dalam manajemen dan optimasi jaringan 5G. Proses otomatisasi ini terutama diwujudkan melalui pemanfaatan kecerdasan buatan (AI).

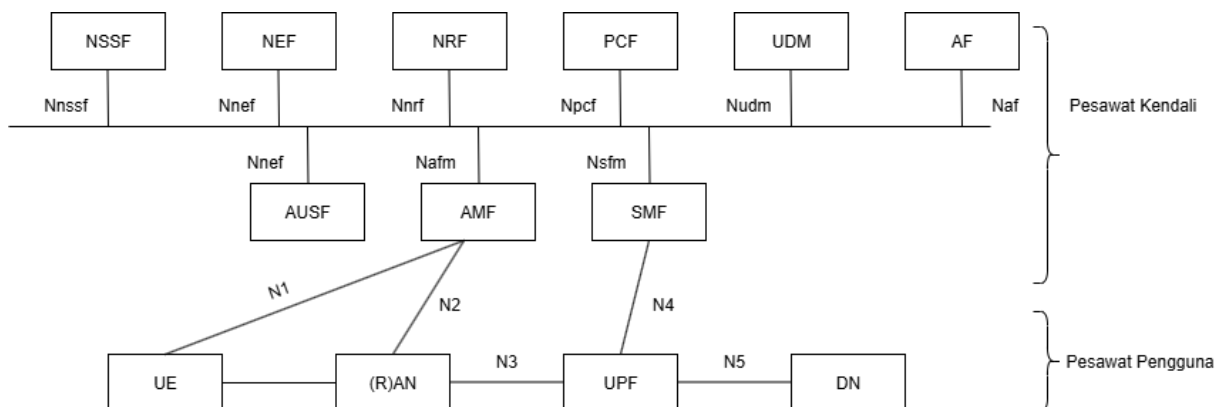
Kemampuan *edge computing* merupakan prasyarat penting untuk mendukung berbagai aplikasi dengan kebutuhan latensi sangat rendah, seperti kendaraan otonom, layanan kesehatan digital (e-Health), serta pengalaman imersif. Aplikasi distribusi konten juga memperoleh keuntungan signifikan apabila ditempatkan pada sisi tepi jaringan (edge), karena dapat mengurangi biaya transportasi data sekaligus meningkatkan kualitas pengalaman pengguna. Fungsi jaringan yang tervirtualisasi seperti *Virtualized RAN* (VRAN), core, dan transport network yang dijalankan pada lingkungan *Multi-access Edge Computing* (MEC) dapat ditempatkan berdekatan dengan aplikasi MEC lainnya. Hal ini memungkinkan pertukaran informasi secara real-time antara jaringan dan aplikasi. Selain itu, akses terhadap pengukuran RAN secara real-time dan informasi lokasi, ditambah kemampuan untuk memengaruhi perilaku RAN melalui *Application Programming Interfaces* (APIs) terbuka, akan mendorong hadirnya solusi inovatif dari pihak ketiga maupun operator. Inovasi ini dapat diarahkan untuk melayani kebutuhan sektor perusahaan maupun konsumen. Konvergensi antara komputasi



awan, informasi jaringan real-time, bandwidth tinggi, serta kedekatan dengan pengguna akhir dan perangkat menjadi karakteristik unik dari MEC. Konvergensi inilah yang menjadikan MEC sebagai salah satu pilar utama dalam arsitektur jaringan era 5G.

Arsitektur Sistem 5G dan Fungsi Jaringan Inti

Dengan mempertimbangkan kebutuhan berbagai kasus penggunaan serta fleksibilitas penyebaran jaringan oleh operator melalui NFV dan SDN, 3GPP telah merancang arsitektur sistem 5G berbasis layanan. Gambar 12.12 menampilkan arsitektur referensi non-roaming sebagaimana didefinisikan dalam 3GPP TS 23.501. Dalam arsitektur ini terdapat pemisahan yang jelas antara fungsi user plane dan control plane. Lalu lintas data pada user plane mengalir dari *User Equipment* (UE) menuju data network (DN) melalui RAN (atau jaringan akses lain) serta fungsi *user plane core network* (UPF), sebagaimana diperlihatkan pada bagian bawah diagram. Sementara itu, bagian lain dari diagram menggambarkan berbagai fungsi control plane dalam core network 5G, seperti *Access and Mobility Management Function* (AMF), *Session Management Function* (SMF), serta *Authentication Server Function* (AUSF). Komunikasi antar fungsi jaringan di dalam control plane didefinisikan sebagai panggilan layanan (service calls). Meskipun interaksi titik-ke-titik antara dua fungsi masih dapat dipandang sebagai proses komunikasi langsung, pada arsitektur berbasis layanan, interaksi tersebut pada dasarnya dianggap melalui kerangka service-based interface yang lebih fleksibel, sehingga mendukung modularitas dan inovasi yang lebih cepat dalam ekosistem 5G. antarmuka berpasangan, konsep panggilan layanan untuk mengakses fungsi jaringan memberikan modularitas dan fleksibilitas yang lebih besar dalam skenario penerapan dengan potensi untuk mendapatkan berbagai fungsi jaringan dari vendor yang berbeda.



Gambar 12.12 Arsitektur Sistem 5G.

Tabel 12.2 Singkatan NGC

AF	Fungsi aplikasi
AMF	Fungsi manajemen akses dan mobilitas
AUSF	Fungsi server otentikasi
DN	Jaringan data



NEF	Fungsi eksposur jaringan
NRF	Fungsi repositori jaringan
NSSF	Fungsi pemilihan irisan jaringan
PCF	Fungsi kontrol kebijakan
SMF	Fungsi manajemen sesi
UDM	Manajemen data terpadu
UDR	Repositori data terpadu
UDSF	Fungsi penyimpanan data tidak terstruktur
UPF	Fungsi bidang pengguna

Fungsi user plane (UPF) dapat ditempatkan secara terpisah dari fungsi control plane dengan tujuan mengurangi latensi ujung-ke-ujung sekaligus memungkinkan transportasi data yang lebih hemat biaya. Misalnya, layanan *Ultra-Reliable Low-Latency Communication* (URLLC) memerlukan dukungan UPF yang ditempatkan di lokasi *Multi-access Edge Computing* (MEC). Demikian pula, ketika konten aplikasi dipindahkan ke sisi edge demi meningkatkan kinerja dan mengurangi biaya, maka pemilihan UPF lokal menjadi penting agar lalu lintas data dapat diarahkan secara optimal dan manfaat dari penempatan edge tersebut benar-benar tercapai. Keputusan strategis dapat dilakukan hingga pada tingkat sesi individu untuk memilih UPF yang paling tepat. Proses ini dapat memanfaatkan algoritme pembelajaran mesin (machine learning/ML) dengan mempertimbangkan berbagai faktor, seperti kebijakan operator, lokasi perangkat pengguna (*User Equipment/UE*), alamat server aplikasi, jenis aplikasi, tingkat mobilitas, beban jaringan, biaya, serta informasi lain yang terkait dengan jaringan, perangkat, dan aplikasi.

Dalam arsitektur sistem 5G, *Unified Data Management* (UDM), *Unified Data Repository* (UDR), dan *Unstructured Data Storage Function* (UDSF) berperan penting dalam menyediakan kemampuan untuk menyimpan serta mengambil kembali data terstruktur maupun tidak terstruktur. Data tersebut mencakup berbagai informasi, seperti kebijakan jaringan, langganan pengguna, otorisasi pengguna, preferensi pengguna, riwayat penggunaan, hingga pola mobilitas yang diperkirakan. Informasi ini dapat diakses secara langsung oleh berbagai fungsi jaringan untuk mendukung pelaksanaan prosedur penting, seperti manajemen akses dan mobilitas, manajemen sesi komunikasi, serta fungsi jaringan lainnya. Dengan demikian, integrasi UDM, UDR, dan UDSF memungkinkan pengelolaan data yang lebih komprehensif sekaligus meningkatkan efisiensi operasional jaringan 5G.

Fungsi *Network Exposure* (NEF) berperan dalam menyediakan akses terhadap data dari UDM dan UDR bagi aplikasi maupun entitas eksternal lainnya. Melalui fungsi ini, aplikasi dapat memanfaatkan informasi jaringan yang tersedia untuk memahami perilaku pengguna (UE), memberikan rekomendasi yang dapat memengaruhi *Policy Control Function* (PCF) dalam pengambilan keputusan routing, serta meningkatkan kinerja aplikasi secara keseluruhan



sekaligus memperbaiki kualitas layanan (QoS) bagi pengguna. Selain itu, operator dapat mengembangkan sendiri atau memperoleh dari pihak ketiga algoritma *Machine Learning* (ML) yang berlangganan informasi dari *Core Network* (CN) melalui NEF (atau secara langsung dari UDM dan UDR). Algoritma ini kemudian dapat digunakan untuk mengusulkan penyesuaian pada konfigurasi *Radio Access Network* (RAN) maupun jaringan inti, sehingga kinerja jaringan dapat dioptimalkan secara dinamis dan berkelanjutan.

Arsitektur 5G mendukung konsep *network slicing*, yaitu kemampuan membagi jaringan menjadi beberapa irisan (slices) yang dapat dikonfigurasi sesuai kebutuhan layanan tertentu. Operator dapat mengatur berbagai irisan jaringan untuk memenuhi karakteristik spesifik, seperti *enhanced Mobile Broadband* (eMBB), *massive IoT* (MIoT), dan *Ultra-Reliable Low-Latency Communication* (URLLC). Dalam arsitektur sistem 5G, terdapat *Network Slice Selection Function* (NSSF) yang berfungsi menentukan irisan jaringan mana yang diizinkan untuk digunakan oleh *User Equipment* (UE) serta melakukan proses pemilihan slice tersebut. Setiap UE dapat dikonfigurasi untuk menggunakan lebih dari satu irisan jaringan, bahkan pada setiap *Slice/Service Type* (SST) misalnya eMBB, MIoT, atau URLLC dapat didefinisikan lebih dari satu irisan. Penerapan sebuah *network slice* harus dilakukan secara menyeluruh, mencakup lapisan *core network*, RAN, hingga *transport network*.

Fungsi inti 5G (5G core functions) dapat dijalankan secara virtualisasi dan ditempatkan baik di cloud milik operator maupun di cloud pihak ketiga dalam bentuk *Virtual Network Functions* (VNF). Apabila perilaku fungsi, protokol antarmuka, serta model data telah distandardisasi, maka VNF dapat disediakan oleh berbagai vendor dan tetap mampu beroperasi secara interoperable melalui mekanisme *service calls*. Lebih jauh, fungsi jaringan virtual ini dapat dipecah menjadi komponen *microservices* atau disusun kembali dari *microservices* tersebut untuk meningkatkan tingkat reuse perangkat lunak, menurunkan biaya pengembangan, serta mempercepat waktu menuju pasar. Selain itu, dengan dukungan platform orkestrasi dan kontrol VNF, seperti ONAP (*Open Network Automation Platform*), proses manajemen siklus hidup dari VNF dapat disederhanakan secara signifikan. Dengan demikian, manfaat dari arsitektur 5G yang fleksibel dapat dimaksimalkan melalui kebijakan otomatis yang didorong oleh analisis berbasis data serta pengendalian hingga pada tingkat UE (*User Equipment*). Uraian lebih rinci mengenai mekanisme kontrol berbasis ONAP dalam jaringan 5G dibahas pada Bagian 12.3.7.

Arsitektur Jaringan Akses Radio Baru

Dalam kurun waktu 25 tahun terakhir, teknologi nirkabel telah mengalami evolusi yang sangat pesat, mulai dari 2G, 3G, hingga 4G, dengan peningkatan kecepatan teoretis downlink (DL) hingga seribu kali lipat (dari sekitar 300 Kbps menjadi lebih dari 300 Mbps). Kemajuan ini mendorong lahirnya smartphone serta ribuan aplikasi mobile yang kini menjadi bagian dari aktivitas sehari-hari. Namun, seperti dijelaskan pada Bagian 12.3.1, berbagai skenario penggunaan baru menuntut peningkatan yang jauh lebih besar dalam hal kecepatan, kepadatan koneksi, keandalan, dan latensi, yakni sekitar 10 hingga 1000 kali lebih tinggi dibandingkan kemampuan jaringan 4G. Dalam konteks ini, *Radio Access Network* (RAN) memegang peran yang sangat penting sebagai inti dari jaringan nirkabel. Oleh karena itu,



dibutuhkan teknologi radio baru serta arsitektur RAN yang lebih maju untuk dapat memenuhi tuntutan kapasitas dan kinerja tersebut.

Teknologi Radio Baru

- **Teknologi antena:**

Untuk meningkatkan kapasitas hingga 100 kali lipat per satuan area, 5G RAN akan bergantung pada penggunaan lebar pita spektrum yang besar, yaitu di atas 50 MHz atau lebih. Di Amerika Serikat, ketersediaan spektrum di bawah 6 GHz sangat terbatas. Oleh karena itu, gelombang sentimeter (cmWave) dan gelombang milimeter (mmWave) dipertimbangkan sebagai solusi untuk menyediakan kapasitas 5G RAN berkat lebar pita spektrum yang lebih luas. Pada pita cmWave, tersedia spektrum hingga beberapa ratus MHz yang berpotensi digunakan, sedangkan pada pita mmWave, ketersediaan spektrum bahkan melebihi 1 GHz. Namun, panjang gelombang yang sangat kecil menimbulkan berbagai tantangan dalam propagasi sinyal *Radio Frequency* (RF). Sinyal mmWave dapat berinteraksi dengan molekul udara dan air, serta dipengaruhi oleh pola iklim, kondisi cuaca, dan vegetasi. Di sisi lain, teknologi phased array antenna yang terdiri dari banyak elemen kecil dapat digunakan untuk mengirim dan menerima sinyal mmWave. Teknologi ini tidak hanya membantu mengatasi tantangan propagasi, tetapi juga memungkinkan penerapan *Massive Multiple Input Multiple Output* (Massive MIMO) untuk meningkatkan kapasitas jaringan secara signifikan.

- **Massive MIMO:**

Melalui spatial multiplexing dan berbagai teknik lainnya, Massive MIMO berpotensi meningkatkan efisiensi spektrum serta kapasitas sel, sekaligus mampu menurunkan energi pancaran dan latensi. Peningkatan efisiensi spektrum dan kapasitas ini dimungkinkan karena sistem Massive MIMO menggunakan jumlah antena yang sangat besar, sehingga dapat melayani hingga 10 kali lebih banyak UE (*User Equipment*) dibandingkan sistem tradisional. Selain itu, penggunaan sinar terarah yang sempit membuat energi lebih terkonsentrasi pada area tertentu sehingga meningkatkan efisiensi transmisi. Salah satu syarat utama jaringan 5G adalah pengurangan latensi hingga 10 kali lipat, dan Massive MIMO dapat memenuhi hal ini. Kemampuan beam-forming serta jumlah antena yang besar memungkinkan Massive MIMO mengurangi terjadinya fading dip yang biasanya muncul akibat masalah multipath. Dengan demikian, kemungkinan menunggu perbaikan kualitas kanal saat terjadi fading dapat diminimalkan, sehingga koneksi menjadi lebih stabil dan responsif.

- **Jaringan Padat (Dense Network):**

Berdasarkan karakteristik propagasinya, cakupan sinyal mmWave memang sangat terbatas. Namun, keterbatasan ini justru menjadi keuntungan, karena small cells berbasis mmWave sangat sesuai untuk menyediakan kapasitas tambahan di wilayah dengan lalu lintas data yang padat. Dalam implementasi 5G RAN, kebutuhan akan kapasitas dan cakupan dapat ditangani secara terpisah. Penerapan *New Radio* (NR) pada pita frekuensi di bawah 6 GHz, bersama dengan pemanfaatan jaringan LTE yang



ada dengan fitur canggih, dapat menghadirkan cakupan luas serta keandalan tinggi untuk mendukung skenario dengan mobilitas tinggi. Sementara itu, lapisan jaringan yang dipadatkan melalui small cells mmWave mampu meningkatkan kecepatan data puncak hingga 10 Gbps, serta throughput di tepi sel lebih dari 100 Mbps. Lapisan mmWave ini sangat efektif untuk memenuhi kebutuhan berbagai use case eMBB (*enhanced Mobile Broadband*) maupun URLLC (*Ultra-Reliable Low Latency Communication*), khususnya untuk kondisi stasioner atau dengan mobilitas rendah.

- **Manajemen interferensi:**

Manajemen interferensi sangat penting untuk RAN yang padat. Seiring radius cakupan menyusut 3 kali lipat, jumlah sel yang dibutuhkan untuk mencakup area yang sama akan meningkat 9 kali lipat. Interferensi co-channel perlu dikelola secara efektif untuk mempertahankan rasio sinyal terhadap derau yang baik. Perilaku mobilitas UE, baik historis maupun yang diharapkan, kapabilitas perangkat, detail tingkat aplikasi, kondisi lingkungan geospasial 3D, dan aktivitas sel di sekitarnya harus dipertimbangkan untuk membuat keputusan penjadwalan transmisi secara bersama-sama dan holistik untuk kluster sel. UE mana yang akan dilayani, berapa banyak sumber daya radio yang akan digunakan, subset titik transmisi mana yang akan digunakan, dan kemiringan port antenna ke bawah harus dikoordinasikan dengan sel di sekitarnya untuk meminimalkan interferensi dan memaksimalkan tujuan gabungan kapasitas jaringan, kinerja UE, dan kewajaran. Mengingat sifat lingkungan RF yang sangat dinamis, interaksi kompleks antara UE, aplikasi, dan jaringan radio, serta pendekatan manajemen interferensi berbasis AI memiliki potensi besar untuk meningkatkan kualitas RAN 5G.

- **Multikonektivitas (Multiconnectivity):**

Perangkat 5G UE harus mampu mendukung berbagai gelombang radio, termasuk Wi-Fi, LTE, dan NR. Bergantung pada skenario penggunaan, satu atau lebih teknologi ini dapat digunakan untuk memenuhi kebutuhan kinerja dan kualitas layanan. Hal ini menjadi sangat penting pada tahap awal penerapan NR, ketika cakupan NR masih terbatas hanya di area tertentu. Dalam kondisi tersebut, cakupan luas dari LTE serta pemanfaatan Wi-Fi untuk mengurangi beban trafik tetap berperan penting untuk menjaga kualitas pengalaman pengguna (QoE). Persyaratan multikonektivitas ini membawa dampak pada opsi konektivitas baik di *Radio Access Network* (RAN) maupun *Core Network* (CN). Namun, perlu dicatat bahwa pemisahan fungsi *Distributed Unit* (DU) dan *Central Unit* (CU) dalam arsitektur logis NR tidak ditentukan oleh kebutuhan multikonektivitas, melainkan oleh pertimbangan teknis lain.

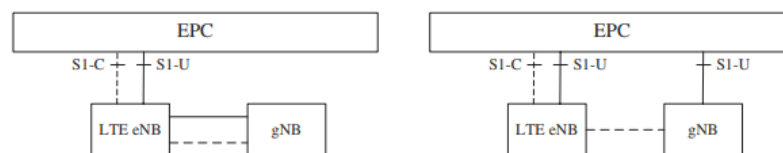
Opsi Konektivitas Jaringan Akses Radio dan Jaringan Inti.

Ada beberapa opsi yang dipertimbangkan untuk menghubungkan RAN ke CN, sehingga operator dapat lebih fleksibel dalam menerapkan 5G di atas jaringan LTE yang ada. Di sisi RAN terdapat tiga jenis node: LTE eNB, eLTE eNB, dan NR gNB. Untuk CN pilihannya adalah LTE EPC dan NGC. Koneksi antara control plane dan user plane bisa dipisahkan. Dalam opsi 3GPP, khususnya opsi 3 dan 3A, keduanya mendukung dual connectivity bagi UE yang mendukung LTE dan NR. Pada kedua opsi, gNB berfungsi sebagai non-standalone dan EPC dipakai sebagai

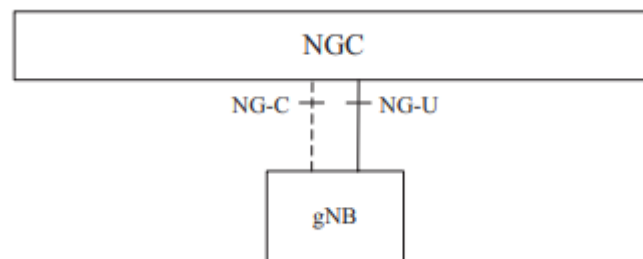


core. Perbedaannya, pada opsi 3A, user plane gNB langsung terhubung ke EPC, bukan melalui LTE eNB.

Pada opsi 2, gNB terhubung langsung ke NGC untuk control plane maupun user plane. Dengan cara ini, perangkat UE yang mendukung LTE dan NR dapat mengirim dan menerima data. Pada tahap awal penerapan 5G, ketika cakupan masih terbatas, opsi non-standalone lebih menguntungkan karena memungkinkan penggunaan kembali LTE EPC untuk mendukung gNB, sehingga dapat mengurangi waktu dan biaya implementasi 5G. Namun, seiring dengan semakin meluasnya penerapan 5G, gNB perlu terhubung langsung ke NGC agar dapat mendukung use case 5G dengan latensi rendah, yang membutuhkan keberadaan *User Plane Function* (UPF). agar lebih dekat atau berlokasi bersama dengan RAN. Arsitektur NGC yang modular dan fleksibel, dengan pemisahan bidang kontrol dan bidang pengguna, memungkinkan skenario penerapan UPF tersebut.



Gambar 12.13 Opsi konektivitas RAN dan CN 3 dan 3A.



Gambar 12.14 Opsi konektivitas RAN dan CN 2.

Arsitektur Perpecahan Logis Jaringan Akses Radio.

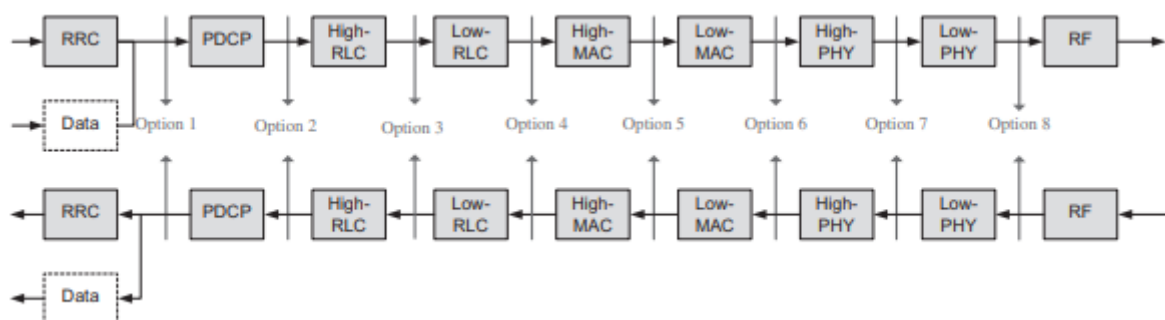
Unit Baseband (BBU) yang menjalankan pemrosesan lapisan 1 hingga lapisan 3 secara tradisional disediakan oleh vendor RAN dalam bentuk perangkat keras dan perangkat lunak terintegrasi yang bersifat proprietary. Biasanya, BBU ditempatkan di lokasi sel untuk makro-deployment, yang membutuhkan ruang, daya, serta konektivitas backhaul. Seiring dengan meningkatnya kepadatan RAN, sentralisasi BBU di pusat (hub) menjadi lebih masuk akal secara ekonomi karena dapat mengurangi biaya operasional dan pemeliharaan. Selain itu, kinerja RAN dapat ditingkatkan jika pemrosesan baseband dapat dikoordinasikan antar sel tetangga. Pendekatan *Coordinated Multipoint* (CoMP) dalam transmisi dan penerimaan mampu meningkatkan cakupan sel, throughput di tepi sel, serta total throughput sistem. Dalam konteks LTE, berbagai skema Downlink dan Uplink CoMP serta keuntungan performanya telah dijelaskan pada. Sementara itu, dengan hadirnya NFV (*Network Function Virtualization*) dan SDN (*Software-Defined Networking*), industri telah mencapai kemajuan besar dalam virtualisasi fungsi jaringan inti dan transportasi. Dengan semakin meluasnya penerapan MEC



(*Multi-access Edge Computing*), sumber daya komputasi, penyimpanan, dan jaringan yang tervirtualisasi tersedia di ujung jaringan. Hal ini membuka peluang unik untuk memindahkan sebagian atau seluruh pemrosesan BBU ke dalam MEC, sehingga meningkatkan efisiensi sekaligus fleksibilitas sistem.

Terdapat berbagai cara untuk membagi fungsi baseband antara Distributed Unit dan Centralized Unit dalam arsitektur NR (*New Radio*). Pada pembagian ini, DU tetap ditempatkan di lokasi sel, sedangkan CU dapat dijalankan di lingkungan MEC (*Multi-access Edge Computing*) sebagai fungsi yang tervirtualisasi. Seperti ditunjukkan pada Gambar 12.15 dari, terdapat delapan opsi pembagian yang memungkinkan untuk mengatur distribusi fungsi antara DU dan CU, sesuai dengan kebutuhan performa dan efisiensi jaringan.

Dari opsi 1 hingga opsi 8, semakin banyak fungsi yang dipusatkan pada Centralized Unit. Hal ini membuka peluang lebih besar untuk melakukan koordinasi antar situs seluler sehingga dapat meningkatkan manfaat jaringan, baik dari sisi kapasitas maupun kinerja. Namun, setiap opsi memiliki persyaratan latensi yang ketat. Proses seperti Radio Resource Control (RRC), *Packet Data Convergence Protocol* (PDCP), dan *Radio Link Control* (RLC) tingkat atas masih dapat mentoleransi jeda waktu yang lebih panjang. Pada opsi 1–3, kebutuhan latensi jaringan transport satu arah berada pada kisaran beberapa hingga 10 milidetik. Sedangkan untuk opsi 4–8, persyaratan menjadi jauh lebih ketat, yaitu dengan latensi satu arah yang harus kurang dari 250 mikrodetik. Selain itu, faktor lain yang sangat penting adalah kebutuhan bandwidth jaringan transport. Sebagai contoh, dalam sistem LTE 100 MHz dengan 256 QAM, 8 lapisan MIMO, dan 32 port antenna, kebutuhan bandwidth koneksi CU ke DU meningkat sangat drastis, dari 4 Gbps pada opsi 1 hingga mencapai 157 Gbps pada opsi.



Gambar 12.15 Fungsi dibagi antara unit pusat dan terdistribusi.

Pada opsi 2, fungsi RRC dan PDCP ditempatkan di Centralized Unit, sedangkan RLC, *Medium Access Control* (MAC), lapisan fisik, dan RF berada di Distributed Unit. CU sendiri dapat dipisahkan lebih lanjut menjadi entitas plane kontrol (RRC dan PDCP *Control Plane* [CP]) serta entitas plane pengguna (PDCP *User Plane* [UP]) (opsi 2-2). Dengan PDCP yang terpusat, dimungkinkan adanya agregasi lalu lintas dan load balancing antara NR dan LTE. Selain itu, pemisahan control plane dan user plane di CU mendukung konsep *Software Defined Networking* (SDN) dan memberi operator peluang untuk beralih ke lingkungan multivendor, dengan tingkat kendali yang lebih besar terhadap operasional jaringan serta manajemen

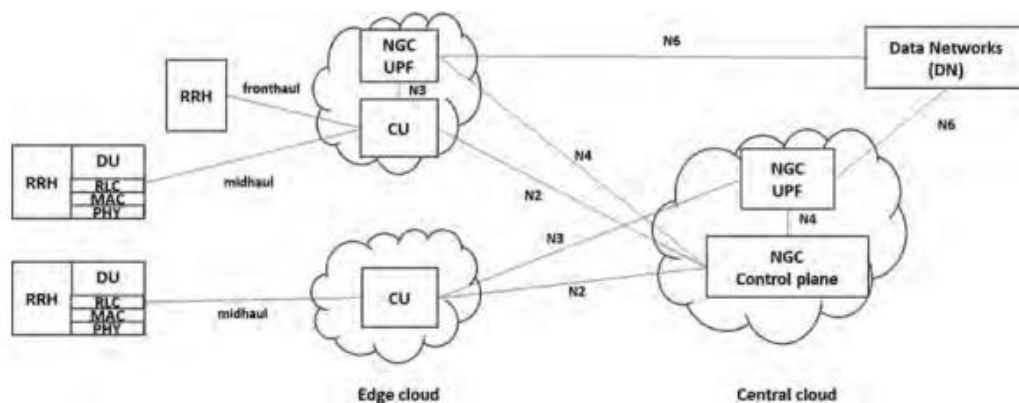


kualitas layanan. Namun, pemisahan UP/CP pada lapisan RLC, MAC, dan lapisan fisik masih membutuhkan analisis dan penelitian lebih lanjut, karena proses transportasi serta penjadwalan baik untuk sinyal kontrol maupun data user plane sangat erat terintegrasi untuk mengoptimalkan pemanfaatan sumber daya radio.

Pemisahan fungsi dapat dilakukan pada berbagai tingkat granularitas. Misalnya, per UE, per bearer, atau per slice. Cara ini memberi operator fleksibilitas lebih besar untuk mengatur fungsi terpusat dan terdistribusi secara detail, sehingga dapat meningkatkan efisiensi dan kinerja RAN. Selain itu, pemisahan ini juga bisa dilakukan secara dinamis seiring waktu. Keputusan yang kompleks mengenai bagaimana, kapan, dan pada level mana pemisahan dilakukan dapat ditentukan oleh algoritme berbasis AI. Algoritme ini mempertimbangkan analitik saat ini maupun prediksi terkait UE, aplikasi, kualitas RAN, kapasitas, serta kinerja jaringan.

Transportasi 5G

Pilihan pemisahan CU/DU memiliki dampak besar terhadap kebutuhan bandwidth transport. Gambar 12.16 menunjukkan contoh skenario implementasi di mana opsi pemisahan 2 dan 8 digunakan secara bersamaan, serta NR gNB memiliki koneksi langsung ke NGC UPF dan control plane.



Gambar 12.16 Skenario penerapan NR dengan opsi split 2 dan 8. gNB terhubung ke NGC secara langsung untuk bidang kontrol dan bidang pengguna.

Penggunaan spektrum dengan lebar pita besar, massive MIMO, serta RAN ultra-padat menimbulkan tantangan besar bagi jaringan transport 5G. Untuk memaksimalkan manfaat dari CoMP, pemrosesan baseband sebaiknya dipusatkan sebanyak mungkin. Namun, koneksi antara DU dan CU (disebut midhaul) harus mampu mendukung laju puncak yang sangat tinggi. Pada opsi 8, semua fungsi baseband dipindahkan ke CU, sehingga koneksi antara DU (yang hanya berperan sebagai *Remote Radio Head* [RRH]) dan CU berubah menjadi fronthaul. Kebutuhan bandwidth fronthaul dapat dengan mudah melampaui ratusan Gbps bahkan mencapai beberapa Tbps. Dalam kondisi ini, diperlukan transport berbasis serat optik dengan *Wavelength Division Multiplexing* (WDM).

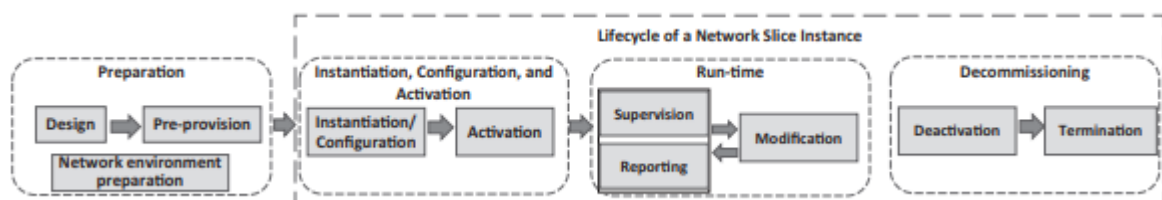
Koneksi backhaul (yakni hubungan dari DU ke CN) juga akan mengalami peningkatan signifikan dalam laju puncak pada jaringan 5G. Untuk menekan biaya backhaul, NGC UPF dapat



ditempatkan di edge cloud agar dapat langsung terhubung ke data network tujuan secara lokal jika memungkinkan. Tautan fronthaul memiliki persyaratan latensi yang sangat ketat, sehingga membatasi jarak fronthaul hanya sekitar 10 hingga 20 kilometer. Selain itu, use case 5G URLLC juga menuntut latensi end-to-end yang sangat rendah. Oleh karena itu, midhaul dan backhaul harus dirancang agar mampu memenuhi kebutuhan keterlambatan keseluruhan. Untuk aplikasi misi kritis, tingkat keandalan tinggi sangat diperlukan. Hal ini menuntut adanya konektivitas ganda atau bahkan tiga lapis antara CU dan DU (meliputi fronthaul dan midhaul), serta antara CU dan CN.

Jaringan transportasi 5G harus mampu mendukung banyak irisan jaringan secara bersamaan. Seperti dijelaskan, beragam use case mendorong jaringan hingga batas maksimalnya pada berbagai dimensi. Network slicing menjadi teknologi penting dalam 5G yang memungkinkan pembagian sumber daya jaringan sesuai dengan kebijakan operator, sehingga dapat memenuhi kebutuhan spesifik dari berbagai kelas use case.

Desain topologi dan pengoperasian berkelanjutan jaringan transportasi 5G dapat memperoleh manfaat besar dari penerapan algoritme kecerdasan buatan (AI). Berbagai tantangan, seperti penentuan jalur peningkatan kapasitas link transport yang optimal, perutean paling efisien, penanganan skenario kegagalan, serta pembagian dan pemanfaatan kapasitas secara tepat, membutuhkan analisis berbasis data yang berkesinambungan.



Gambar 12.17 Siklus hidup NSI.

Pengirisan Jaringan

Network slicing merupakan salah satu fitur utama dalam arsitektur sistem 5G. Teknologi ini memungkinkan operator jaringan membangun jaringan logis dari sumber daya fisik maupun virtual yang ada, untuk melayani pelanggan dengan tingkat perjanjian layanan (*Service Level Agreement/SLA*) yang berbeda-beda. Satu network slice dapat mencakup komponen dari core network, jaringan akses (RAN), atau kombinasi keduanya. Sejumlah kajian, seperti dan, telah menetapkan persyaratan awal yang dibutuhkan untuk mendukung network slicing pada CN dan RAN. Sementara itu, manajemen siklus hidup network slicing dibahas secara lebih rinci dalam.

Untuk mendukung layanan komunikasi dengan karakteristik *SLA (Service Level Agreement)* tertentu, dapat dibuat instance irisan jaringan (*Network Slice Instance/NSI*). Sebuah NSI terdiri atas sub-instance irisan jaringan (*Network Slice Subnet Instance/NSSI*). Setiap NSSI dibangun dari fungsi jaringan (*Network Functions/NF*) baik dalam bentuk fisik maupun virtual dan/atau gabungan dari beberapa NSSI lainnya. Pengelolaan NSI dilakukan



oleh fungsi manajemen irisan jaringan. Proses siklus hidup dari sebuah NSI dijelaskan pada Gambar 12.17 dalam.

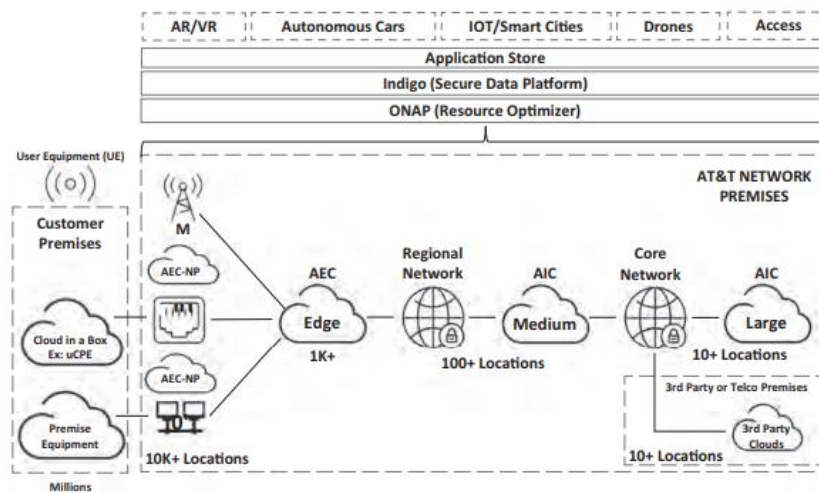
Network slicing menawarkan fleksibilitas, tetapi sekaligus menambah kompleksitas dalam pengelolaan jaringan. Untuk manajemen siklus hidup sebuah NSI (*Network Slice Instance*), algoritme AI dapat dimanfaatkan untuk melakukan konfigurasi otomatis, optimasi saat jaringan berjalan, serta pemulihan mandiri, mirip dengan konsep *Self-Organizing Network* (SON) pada LTE. Karena satu NSI dapat dipakai bersama oleh berbagai layanan dengan kebutuhan SLA yang berbeda, sering kali muncul konflik ketika ada permintaan yang saling bertentangan terhadap fungsi jaringan, misalnya perubahan parameter NF ke arah yang berlawanan. Selain itu, perebutan sumber daya antar-NSI juga perlu dikelola dengan baik. Maka dari itu, penyelesaian konflik berbasis kebijakan seperti kemampuan yang disediakan oleh ONAP dan otomatisasi berbasis AI akan menjadi kunci keberhasilan dalam mengoperasikan jaringan 5G yang mendukung network slicing.

Komputasi Tepi

Kasus penggunaan URLLC (*Ultra-Reliable Low-Latency Communication*) merupakan salah satu tantangan sekaligus peluang utama bagi jaringan 5G dan generasi berikutnya. Untuk mengurangi latensi ujung-ke-ujung serta meningkatkan keandalan jaringan, sumber daya jaringan perlu ditempatkan lebih dekat dengan pengguna dan perangkat. Arsitektur sistem yang terdisagregasi memberikan fleksibilitas dalam menempatkan fungsi *core network* (CN) dan RAN di lokasi yang berbeda guna meminimalkan latensi. Kemampuan edge computing menjadi sangat penting untuk memaksimalkan manfaat fleksibilitas arsitektur 5G dalam memenuhi kebutuhan layanan kelas URLLC.

Edge computing memberikan keuntungan penghematan biaya bagi operator jaringan dengan cara mengurangi lalu lintas backhaul serta mengoptimalkan pemanfaatan aset pusat data. Bagi pengguna akhir, kualitas pengalaman (*Quality of Experience/QoE*) akan meningkat berkat waktu respons yang lebih cepat dan peningkatan keandalan jaringan. Sementara itu, pengembang pihak ketiga akan memperoleh akses ke informasi jaringan secara real-time melalui API jaringan terbuka, yang memungkinkan lahirnya berbagai aplikasi inovatif. Dengan kedekatan pada pengguna akhir dan kemampuan menyediakan layanan nyaris real-time berbasis lokasi, node edge computing memiliki potensi besar untuk mentransformasi pengalaman sehari-hari kita.





Gambar 12.18 Arsitektur edge cloud AT&T (AEC).

Edge computing memberikan keuntungan penghematan biaya bagi operator jaringan dengan cara mengurangi lalu lintas backhaul serta mengoptimalkan pemanfaatan aset pusat data. Bagi pengguna akhir, kualitas pengalaman (*Quality of Experience/QoE*) akan meningkat berkat waktu respons yang lebih cepat dan peningkatan keandalan jaringan. Sementara itu, pengembang pihak ketiga akan memperoleh akses ke informasi jaringan secara real-time melalui API jaringan terbuka, yang memungkinkan lahirnya berbagai aplikasi inovatif. Dengan kedekatan pada pengguna akhir dan kemampuan menyediakan layanan nyaris real-time berbasis lokasi, node edge computing memiliki potensi besar untuk mentransformasi pengalaman sehari-hari kita.

Kemampuan orkestrasi dan manajemen sangat penting untuk mengotomatisasi operasional platform edge computing. Data jaringan berupa *Key Performance Indicators (KPIs)* perlu tersedia secara tepat waktu dan dapat diakses oleh pengembang pihak ketiga melalui lapisan *Application Program Interface (API)*. Fungsi-fungsi ini juga tercermin dalam arsitektur ETSI MEC. Dalam implementasinya, arsitektur edge cloud AT&T mewujudkan hal tersebut melalui ONAP untuk orkestrasi/manajemen dan Indigo untuk dukungan tambahan.

Jumlah node *Edge Computing (EC)* jauh lebih banyak dibandingkan dengan node *Central Cloud (CC)*. Penempatan node EC dioptimalkan dengan mempertimbangkan permintaan layanan, kebutuhan latensi, biaya, kemampuan jaringan, serta faktor lainnya. Untuk memastikan kinerja yang optimal, latensi pemrosesan di EC harus kurang dari 10 milidetik, sehingga latensi ujung-ke-ujung bagi aplikasi yang dilayani melalui EC dapat dijaga tidak lebih dari 20 milidetik.

Edge Cloud merupakan hasil konvergensi antara komputasi awan, analitik waktu nyata, serta kedekatan dengan pengguna dan perangkat. Kombinasi unik ini menghadirkan sumber daya komputasi dan wawasan berbasis data langsung ke tepi jaringan. Dengan kemampuan tersebut, banyak aplikasi berbasis AI dapat diberdayakan. Saat ini sudah terlihat pada mobil otonom, robotika industri, dan operasi jarak jauh. Namun, sebagian besar potensi use case AI berbasis Edge Cloud masih belum ditemukan. Pada akhirnya, Edge Cloud membuka peluang



besar bagi seluruh ekosistem digital dan industri untuk mengembangkan layanan dan inovasi baru.

12.4 KECERDASAN BUATAN UNTUK 5G+

Kecerdasan buatan (AI) meliputi teknologi seperti pembelajaran mesin, pembelajaran mendalam, dan pemrosesan bahasa alami. Pembelajaran mesin menggunakan sejumlah besar data dan algoritma kompleks untuk memahami data, menemukan pola, atau membuat prediksi. Wawasan yang diperoleh dari proses ini memungkinkan otomatisasi dijalankan dengan lebih efektif. Sementara itu, pemrosesan bahasa alami membantu mesin memahami informasi penting dari teks dan ucapan. Contohnya, otomatisasi dapat digunakan untuk menanggapi kerusakan peralatan secara langsung atau menyesuaikan kapasitas jaringan ketika terjadi lonjakan permintaan. AI dan pembelajaran mesin menjadi komponen utama dalam mendukung otomatisasi dan virtualisasi modern.

Teknologi 5G dan IoT diprediksi akan mengubah jaringan dan pola lalu lintas data secara lebih dinamis dan sulit diprediksi dibandingkan kondisi saat ini. Agar jaringan bisa merespons lonjakan permintaan yang tak terduga lebih cepat daripada manusia, AI dan pembelajaran mesin harus diintegrasikan ke dalam lapisan kontrol SDN. Para analis memperkirakan bahwa pada tahun 2020 akan ada sekitar 20 miliar perangkat IoT yang saling terhubung. Volume data yang dihasilkan sangat besar, dan akses terhadap data ini membuka peluang untuk menghadirkan berbagai layanan tambahan yang bernilai lebih tinggi.

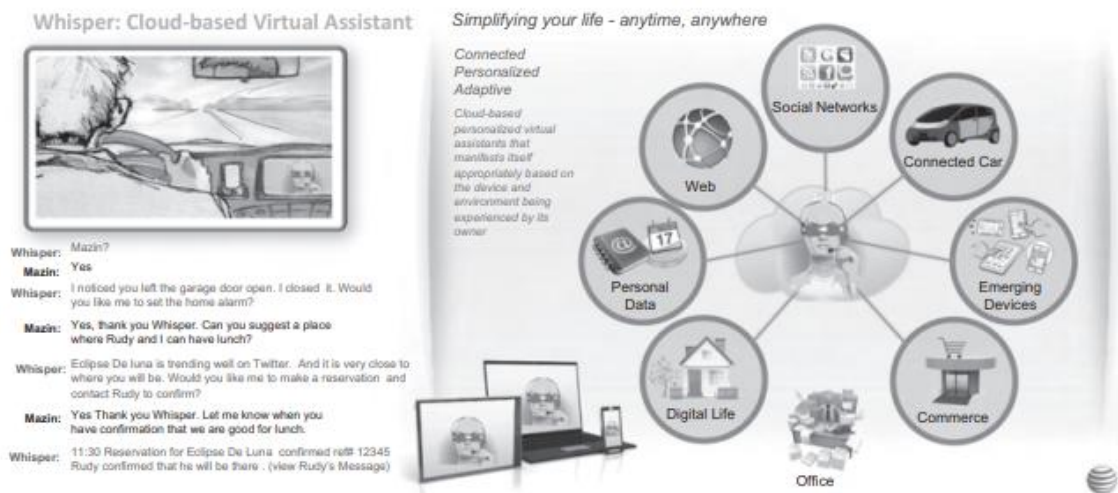
Untuk menangani lonjakan lalu lintas ekstrem dan mengelola jaringan yang kompleks secara real-time, AI dan pembelajaran mesin menjadi kunci. Manusia akan “mengajarkan” jaringan bagaimana merespons situasi serupa di masa depan. Seiring jaringan semakin maju, konsumen juga semakin menuntut layanan yang lebih cepat, berkualitas, murah, dan fleksibel. Dalam kondisi sumber daya terbatas, kemampuan untuk mengatur aliran lalu lintas 5G secara cerdas dari satu titik ke titik lain bergantung pada AI dan pembelajaran mesin agar dapat memprediksi pola lalu lintas, bukan sekadar merespons setelah masalah muncul.

Bagaimana Kecerdasan Buatan Membantu?

AI kini semakin banyak diterapkan dalam berbagai bidang, mulai dari komunikasi, aplikasi, konten, hingga perdagangan, dengan tujuan meningkatkan pengalaman pengguna, mendorong produktivitas, dan membuka peluang bisnis baru yang menguntungkan. Pada Bagian 12.7, beberapa contoh penerapan AI yang menjanjikan akan dibahas. Secara garis besar, AI akan digunakan dalam beberapa kategori berikut:

- Desain dan optimasi berbasis AI dan ML untuk jaringan nirkabel
- Jaringan pembelajaran mandiri dan adaptif
- Analisis akar penyebab, deteksi otomatis, dan mitigasi otomatis
- Desain berbasis pengalaman pengguna untuk ponsel pintar
- Asisten virtual yang dipersonalisasi di EC





Gambar 12.19 Agen personalisasi berbasis cloud dan contoh dialog interaktif.

Salah satu penerapan AI yang penting di industri adalah menciptakan asisten virtual yang dapat disesuaikan secara personal. Dengan menggunakan agen otonom, asisten ini mampu memberikan pengalaman percakapan yang cerdas. Berbeda dari sistem “Pertanyaan dan Jawaban” yang umum saat ini seperti Amazon Alexa, Google Home, atau Apple Siri, asisten virtual yang dipersonalisasi dapat berinteraksi dalam dialog nyata untuk membantu menyederhanakan kehidupan kita kapan saja dan di mana saja. Gambar 12.19 menunjukkan contoh agen virtual berbasis cloud dan ilustrasi dialog kontekstual yang berlangsung dengan Asisten Virtual Terpersonalisasi Cerdas di cloud.

Posisi AI dalam Arsitektur Jaringan Masa Depan

Komponen atau mikroservis berbasis AI akan disebarkan di berbagai titik dalam jaringan, dan penempatannya disesuaikan dengan kebutuhan tiap kasus penggunaan. Seiring infrastruktur seluler menjadi semakin terdistribusi dan tervirtualisasi, baik elemen yang tersebar maupun yang terpusat akan berperan penting. Edge computing akan dimanfaatkan untuk menampung RAN dan Packet Core VNF, sementara layanan terdistribusi seperti Asisten Virtual dapat ditempatkan di Edge Cloud untuk meminimalkan latensi atau memproses data secara lokal. Agar interaksi percakapan antara agen perangkat lunak dan manusia terasa mulus, mikroservis percakapan memerlukan latensi yang rendah.

12.5 KECERDASAN BUATAN DAN OTOMATISASI JARINGAN BARU

Mengapa Otomatisasi?

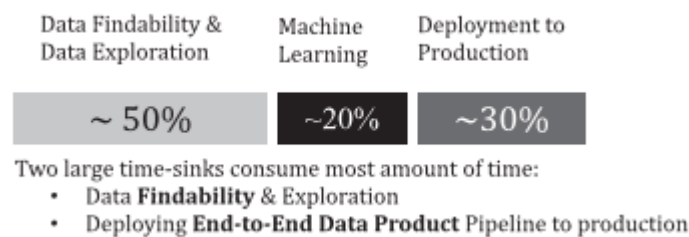
Otomatisasi memiliki beberapa manfaat penting. Pertama, efisiensi secara keseluruhan akan meningkat, yang berpotensi meningkatkan pendapatan sekaligus menekan biaya operasional dan investasi. Agen perangkat lunak atau robot mampu menangani tugas-tugas rutin secara konsisten, sehingga produksi lebih lancar, kesalahan manusia berkurang, dan kualitas output terjaga tinggi. Kedua, penggunaan otomatisasi dan AI memungkinkan teknisi bekerja lebih sedikit dalam kondisi berisiko, sehingga mengurangi kecelakaan kerja. Ketiga,



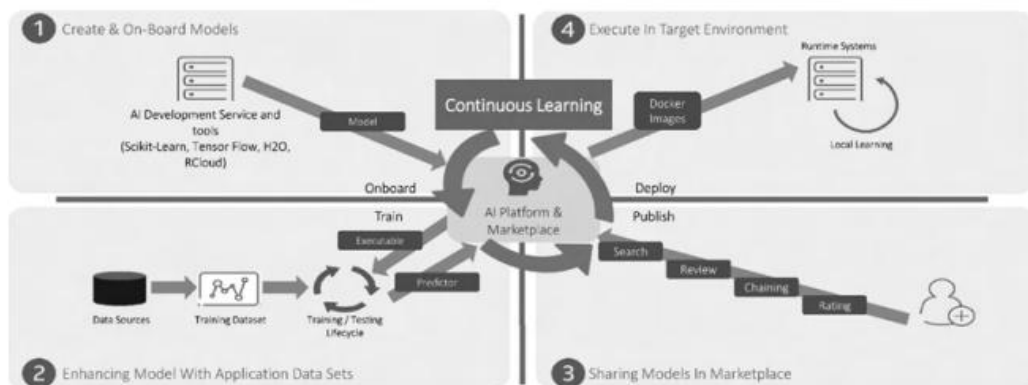
otomatisasi membuka kesempatan bagi pekerja untuk fokus pada pekerjaan yang lebih kompleks, kreatif, dan inovatif.

12.6 KERANGKA KECERDASAN BUATAN DAN OTOMATISASI

Dalam survei internal yang dilakukan di tim Big Data dan Data Science AT&T untuk menilai distribusi waktu dalam pembuatan produk berbasis data, ditemukan bahwa setengah dari waktu mereka digunakan untuk mencari dan mengeksplorasi data, sekitar 20% digunakan untuk pembelajaran mesin, dan sisanya 30% digunakan untuk menerapkan data ke dalam produksi, sebagaimana terlihat pada Gambar 12.21.



Gambar 12.21 Waktu yang dihabiskan pengembang big data dan ilmu data.



Gambar 12.22 Siklus hidup platform dan pasar AI.

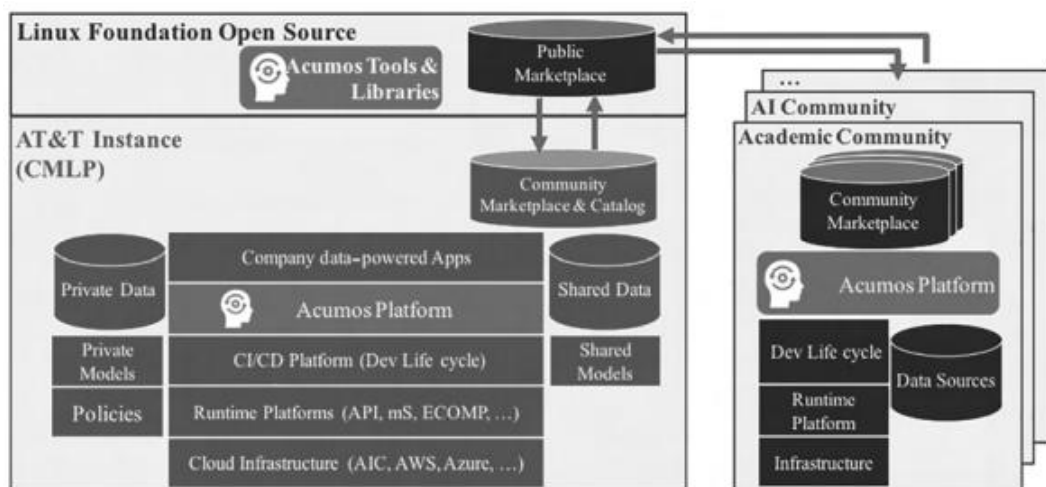
Dalam studi terkait peningkatan proses, terlihat bahwa perlu ada langkah-langkah untuk mempermudah pencarian dan eksplorasi data di tahap awal sekaligus mempercepat implementasi data ke produksi. Solusi yang dikembangkan adalah sebuah kerangka platform AI, sebagaimana ditunjukkan pada Gambar 12.22.

Untuk menangani tantangan terkait waktu yang banyak dihabiskan dalam pencarian data, AT&T menciptakan berbagai alat penemuan data, seperti Data 360, serta Alat Visualisasi Data. Alat-alat ini memungkinkan ilmuwan data dengan lebih cepat menemukan, mengeksplorasi, dan memanfaatkan kumpulan data di data lake untuk kebutuhan analitik dan pembelajaran mesin. Acumos, platform AI yang baru-baru ini diperkenalkan dan dikembangkan bersama oleh AT&T dan Tech Mahindra serta dirilis sebagai open source oleh Linux Foundation, merupakan salah satu solusi yang mampu memenuhi kemampuan yang digambarkan pada Gambar 12.22.



Salah satu fungsi utama platform AI adalah kemampuan untuk membuat dan mengunggah model. Model bisa berasal dari berbagai sumber ML seperti R, Python, dan lain-lain, kemudian diunggah baik sebagai Predictive Model Markup Language atau sumber yang terdokumentasi lengkap dengan pustakanya, yang bisa diterjemahkan menjadi eksekutabel dan dijalankan sebagai mikroservis dalam Docker. Fungsi kedua (Modul 2 pada Gambar 12.22) adalah kemampuan untuk menyempurnakan model menggunakan set pelatihan berbeda dan memperbarui koefisien model. Setelah dikembangkan, model harus mudah ditemukan dan digunakan oleh orang lain. Untuk itu, dibuatlah “marketplace” (Modul 3 pada Gambar 12.22) yang memungkinkan berbagi model dan mikroservis yang dibuat. Terakhir, platform menyediakan kemampuan run-time untuk menjalankan model, memungkinkan pengembang dan ilmuwan data langsung memindahkan model dari pengembangan ke produksi tanpa harus menulis ulang algoritma ML.

Gambar 12.23 menunjukkan secara umum bagaimana platform AI diterapkan, baik secara internal maupun sebagian di platform open source Acumos. Visi yang ingin dicapai adalah agar algoritma yang dikembangkan di platform internal yang dapat diimplementasikan oleh perusahaan di data lake mereka sendiri yang aman maupun algoritma dari marketplace publik, dapat dibagikan bila memungkinkan. Hal ini diharapkan akan menghasilkan solusi yang lebih baik sekaligus mempercepat proses dari pengembangan hingga produksi.



Gambar 12.23 Kerangka platform AI

12.7 KASUS PENGGUNAAN

Di bagian ini, disajikan ringkasan beberapa contoh kasus penggunaan 5G yang berbasis data dan menonjol, yang menunjukkan bagaimana 5G, AI, dan otomatisasi saling berinteraksi dan mendukung satu sama lain.

Evaluasi Kasus Data dengan Metodologi POWER

Di perusahaan yang memiliki banyak data yang akan menjadi hampir semua perusahaan di masa depan, terdapat banyak peluang kasus penggunaan yang bisa dieksplorasi, sehingga perlu dilakukan prioritisasi. Setelah beberapa wawasan pertama dari suatu kasus penggunaan diperoleh, ilmuwan data seringkali kewalahan dengan permintaan



tambahan untuk proyek atau solusi AI lainnya. Tanpa prioritas, beban ini bisa menjadi sangat besar. Dari pengalaman menangani ratusan kasus penggunaan berbasis data, metode penting yang harus dilakukan adalah memastikan lima pertanyaan kunci dijawab sebelum memulai aktivitas apapun. Banyak proyek berbasis data gagal dan menyebabkan hilangnya banyak bulan-person karena pertanyaan-pertanyaan penting ini tidak dijawab sejak awal.

Agar lima pertanyaan kunci yang sangat penting bagi keberhasilan proyek berbasis data dapat tersampaikan dengan jelas, kami mengembangkan sebuah kerangka kerja bernama DATA POWER. Akronim ini menunjukkan tujuan dari proyek yakni analitik yang dapat ditindaklanjuti serta unsur-unsur yang dibutuhkan untuk mencapainya.

- Mengemudi
- Analisis
- Untuk
- Tindakan
- Masalah yang dapat ditindaklanjuti
- Pemilik tindakan
- Data (analitik) apa yang dibutuhkan untuk tindakan tersebut
- Perkiraan waktu penyelesaian (ETA) untuk menghasilkan tindakan tersebut
- ROI yang diharapkan dari tindakan tersebut

Banyak proyek berbasis data yang diajukan bermula dari pertanyaan yang menarik, namun seringkali tidak dapat ditindaklanjuti atau tidak memiliki “pemilik” yang bertanggung jawab untuk mengambil langkah nyata. Selain itu, meskipun idenya bagus, tanpa data yang diperlukan, masalah tersebut tidak dapat diselesaikan (misalnya pertanyaan “Data apa yang dibutuhkan?”). Penting juga untuk memperhitungkan estimasi waktu hingga tindakan dapat diambil (ETA) serta potensi ROI, agar keputusan memilih proyek atau solusi AI yang akan diinvestasikan menjadi tepat di tengah banyaknya opsi yang tersedia.

Merancang masalah atau solusi AI menggunakan kerangka POWER biasanya dapat dilakukan dengan cepat untuk memprioritaskan beberapa proyek secara sederhana. Kerangka ini juga dapat digunakan sepanjang proyek untuk memperjelas elemen “Apa” yang harus dilakukan, termasuk mendefinisikan komponen platform AI seperti dijelaskan di Bagian 12.6. Misalnya, Tabel 12.3 menunjukkan beberapa kasus penggunaan internal AT&T yang diuraikan dalam kerangka POWER, di mana terlihat informasi mengenai masalah, pemilik, perkiraan ETA, dan ROI, serta komponen platform AI yang relevan, termasuk mikroservis yang harus dibuat dan container run-time yang diperlukan. Publikasi dilakukan melalui publish-subscribe message bus yang disebut *Data Movement as a Platform (DMaaP)*.

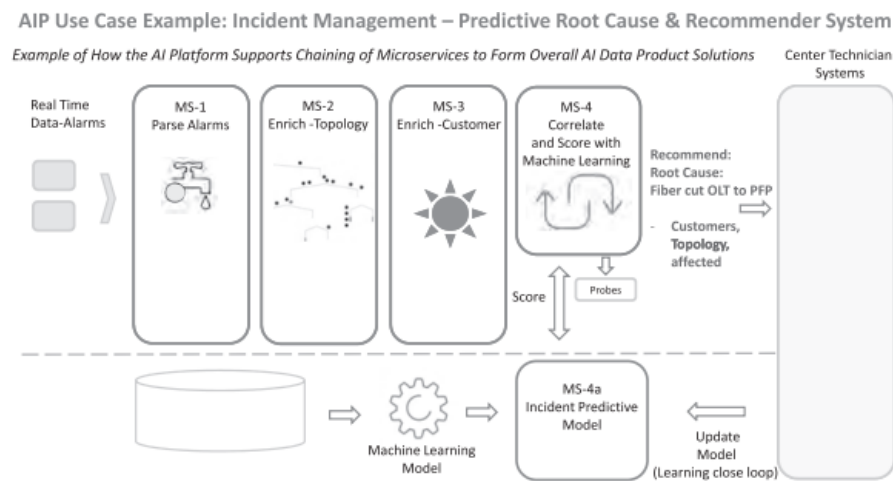
Manajemen Insiden End-to-End untuk Jaringan Eksisting dan 5G

Jaringan apa pun harus mampu menangani kegagalan elemen atau konektivitas, dan mengetahui akar penyebab masalah dengan cepat menjadi sangat penting, terutama ketika jaringan semakin kompleks dan terdistribusi. Gambar 12.24 menunjukkan bagaimana platform AI diterapkan untuk “Manajemen Insiden” atau deteksi akar masalah. Pendekatan ini memanfaatkan data real-time, seperti data alarm, yang dikorelasikan melalui topologi jaringan, diperkuat dengan data dari titik akhir pelanggan, dan dibandingkan secara real-time



dengan model ML historis. Platform AI ini diterapkan secara modular dengan empat mikroservis berantai (MS-1 sampai MS-4) serta penilai prediktif ML (MS-4a). Agar AI dapat belajar, tiket insiden yang dihasilkan dinilai oleh teknisi yang menangani masalah, dan hasilnya dikembalikan ke sistem pembelajaran untuk mengkalibrasi ulang model.

Solusi platform AI yang ditampilkan pada Gambar 12.24 dikembangkan dan dijalankan di platform AI AT&T sebagai proyek pembuktian konsep selama beberapa bulan. Hal ini menunjukkan betapa efektifnya pendekatan platform AI berbasis mikroservis dalam menyelesaikan masalah yang bergantung pada data.



Gambar 12.24 Contoh kasus penggunaan yang dilakukan dalam kerangka POWER

Survei Kasus Penggunaan 5G

Jaringan 5G menawarkan peningkatan signifikan dalam kecepatan data, kapasitas, kepadatan perangkat, sekaligus menurunkan latensi dan konsumsi energi secara drastis. Kombinasi ini membuka berbagai peluang kasus penggunaan yang dapat mengubah kehidupan kita secara signifikan. Mulai dari rumah, kota, dan mobil pintar, hingga teknologi AR/VR, video 3D, dan layanan e-health, 5G akan menjadi infrastruktur yang meresap ke hampir semua aspek kehidupan sehari-hari. Di sektor industri, 5G mendukung aplikasi inovatif untuk memantau, memberikan peringatan, mendiagnosis, dan mengendalikan aktivitas di bidang manufaktur, energi, utilitas, transportasi, smart grid, keamanan, dan keselamatan publik. Integrasi SDN, NFV, dan edge computing pada 5G memberikan fleksibilitas lebih dalam aplikasi, memungkinkan penskalaan sumber daya secara dinamis, menyediakan kemampuan komputasi dekat perangkat, dan menghadirkan data jaringan real-time untuk mendukung pengambilan keputusan berbasis AI.

Bab 13 membahas beberapa kasus penggunaan utama 5G di berbagai sektor industri dan bagaimana AI dapat mendukung implementasinya. Teknologi drone berbasis 5G berpotensi merevolusi layanan pengiriman jarak akhir. Industri otomotif akan berubah secara fundamental dengan hadirnya pengemudian terbantu dan otomatis melalui jaringan 5G yang ultra-andal dan latensi rendah, serta teknologi komunikasi V2X. Di sektor manufaktur, otomatisasi akan semakin maju dengan penggunaan robotik dan pengawasan jarak jauh.



Sementara itu, industri energi dan utilitas akan memanfaatkan perangkat efisien untuk memantau dan mengendalikan infrastruktur transmisi secara terus-menerus, sehingga manfaat smart grid dapat diwujudkan. Di bidang kesehatan, 5G akan mendukung transformasi melalui pemantauan jarak jauh, perawatan berkelanjutan, dan layanan telemedisin.

Integrasi Layanan Pelanggan dengan Platform Sosial

Sistem percakapan Anda tidak hanya dapat digunakan untuk mendorong layanan pelanggan melalui media sosial, tetapi media sosial juga dapat menjadi sumber wawasan yang berharga untuk keputusan layanan pelanggan. Analisis dapat dilakukan terhadap riwayat media sosial pelanggan untuk menentukan kepribadian atau sentimen. Dengan menghubungkannya dengan sistem pemasaran, riwayat sosial dapat mendorong rekomendasi produk yang dapat dimanfaatkan untuk peluang penjualan silang atau penjualan atas selama interaksi pelanggan. Misalnya, jika Anda memiliki akses dan otorisasi untuk menggunakan data dari platform sosial, Anda dapat memahami bahwa pelanggan memiliki sentimen negatif tentang layanan data internasional. Informasi ini dapat digunakan untuk menghasilkan paket data internasional yang andal dan hemat biaya dengan bermitra dengan penyedia global lainnya. Dengan memanfaatkan catatan historis, paket internasional baru ini kemudian dapat dipasarkan kepada calon wisatawan internasional.



BAB 13

AI UNTUK INFRASTRUKTUR DAN LALU LINTAS

13.1 PENDAHULUAN

Seiring dengan berkembangnya jaringan seluler dan smartphone yang menjadi bagian penting dalam kehidupan sehari-hari, selalu muncul ketegangan antara kebutuhan akan kecepatan serta kapasitas jaringan di satu sisi, dan biaya serta ketahanan jaringan di sisi lain. Jaringan masa depan harus mampu menemukan keseimbangan antara tuntutan yang saling bersaing ini, sekaligus mendukung berbagai kasus penggunaan *Internet of Things* (IoT) yang terus berkembang di berbagai sektor, seperti kendaraan otonom, drone, dan kota cerdas. Bab ini membahas peluang pemanfaatan kecerdasan buatan (AI) dalam pemantauan maupun optimalisasi jaringan. Penerapan AI menjanjikan pemanfaatan kapasitas jaringan dan spektrum yang lebih efisien, peningkatan skala jaringan dalam hal jumlah perangkat, serta memberikan peluang untuk meningkatkan pengalaman pengguna.

Bab ini disusun secara bertahap. Pada Bagian 13.2, dibahas tren yang berkembang, kesenjangan yang ada, serta peluang dalam jaringan masa kini. Selanjutnya, Bagian 13.3 menyoroti perkembangan teknologi yang memungkinkan penerapan AI dalam pemantauan dan optimalisasi jaringan. Bagian 13.4 fokus pada strategi pengelolaan lalu lintas data dan peningkatan pengalaman pengguna dengan dukungan AI. Di Bagian 13.5, dibahas kasus-kasus baru pemanfaatan 5G yang mengandalkan AI untuk otomatisasi. Akhirnya, Bagian 13.6 menutup bab ini dengan rangkuman, tantangan, serta mekanisme perlindungan yang perlu diperhatikan guna mengoptimalkan pemanfaatan AI.

13.2 TREN, KESENJANGAN, DAN PELUANG

Tren Penggunaan dan Variasi Trafik Jaringan Seluler

Sebagaimana dijelaskan pada bab sebelumnya, sejak hadirnya smartphone, jaringan seluler menghadapi lonjakan lalu lintas data yang sangat besar. Pertumbuhan ini menuntut operator untuk terus meningkatkan kapasitas jaringan, baik dengan memperbanyak titik akses, memperluas spektrum, maupun meningkatkan efisiensi spektral melalui inovasi teknologi serta strategi optimasi jaringan. Perubahan perilaku pengguna juga menggeser pola lalu lintas, dari dominasi panggilan suara ke konsumsi video streaming, diikuti berbagai aplikasi lain seperti browsing, media sosial, gim daring, dan unduhan file. Namun, tidak semua lalu lintas mudah diidentifikasi, terutama karena penggunaan enkripsi seperti HTTPS, yang menambah tingkat kesulitan dalam pengawasan serta pengelolaan jaringan.

Beragam perangkat dan aplikasi yang digunakan saat ini tidak hanya memperbesar volume lalu lintas data, tetapi juga menghadirkan kebutuhan yang berbeda-beda terhadap sumber daya jaringan. Jika semua aplikasi memiliki pola kebutuhan yang serupa, maka optimasi jaringan akan lebih mudah dilakukan. Namun, kenyataannya setiap aplikasi memiliki karakteristik berbeda. Misalnya, layanan video streaming membutuhkan kapasitas bandwidth tinggi secara konstan namun cukup toleran terhadap latensi, sedangkan aktivitas browsing



lebih peka terhadap latensi meski tidak memerlukan bandwidth besar secara berkesinambungan. Ke depan, aplikasi seperti AR/VR akan membutuhkan kombinasi bandwidth besar dan latensi rendah, sementara perangkat smart grid hanya memerlukan sumber daya minimal. Oleh karena itu, tantangan utama bagi operator seluler adalah bagaimana mengalokasikan dan memanfaatkan sumber daya jaringan secara optimal agar mampu melayani berbagai aplikasi dengan kebutuhan yang beragam secara bersamaan.

Kemajuan teknologi justru membuat tantangan optimasi jaringan semakin kompleks. Saat ini, jaringan nirkabel terdiri atas beragam teknologi yang berjalan bersamaan seperti 3G, 4G, dan Wi-Fi, menggunakan spektrum terpisah-pisah mulai dari 700 MHz hingga 5 GHz dengan lebar pita 5–20 MHz, serta mencakup beragam jenis sel mulai dari makro hingga pico. Dengan hadirnya 5G, tingkat kerumitan jaringan akan semakin meningkat, ditambah dengan munculnya paradigma layanan baru dan beragam kasus penggunaan yang lebih kompleks.

Kesenjangan Analisis Jaringan Saat Ini dan Peluang.

Analitik jaringan hadir dalam berbagai tingkat kompleksitas, mulai dari yang sederhana hingga yang sangat canggih, dengan tujuan utama meningkatkan performa jaringan. Bentuk paling dasar adalah *descriptive analytics*, yaitu analisis deskriptif yang mengubah data peristiwa jaringan menjadi indikator kinerja utama (KPI). KPI ini dipakai untuk menggambarkan kondisi sistem atau layanan secara statistik, sekaligus dipantau dan ditampilkan dalam dasbor, kartu skor, atau platform berbasis web. KPI jaringan yang umum meliputi volume lalu lintas, intensitas penggunaan aplikasi, efisiensi spektral, serta efisiensi energi. Sedangkan untuk kualitas layanan, contohnya adalah tingkat putusnya panggilan, kecepatan unduh data, dan latensi. Pada jaringan seluler yang kompleks, jumlah KPI bisa mencapai ratusan, mencakup lapisan aplikasi hingga lapisan fisik. KPI tersebut biasanya dikelompokkan berdasarkan keterkaitannya dengan pengalaman pengguna, wilayah geografis, atau periode waktu tertentu.

Analisis deskriptif memang mudah dipahami dan menjadi metode yang paling umum digunakan, namun sifatnya terbatas karena hanya menggambarkan perilaku jaringan di masa lalu. Untuk kebutuhan saat ini, operator jaringan memerlukan analisis diagnostik guna menelusuri penyebab suatu perilaku jaringan, serta analisis prediktif untuk memperkirakan kemungkinan hasil dan memprediksi peluang terjadinya suatu peristiwa. Analisis prediktif ini memanfaatkan berbagai pendekatan statistik seperti pembelajaran mesin, pemodelan, penambangan data, hingga teori permainan dengan tujuan membaca pola masa lalu dan kondisi saat ini untuk memperkirakan masa depan. Lebih canggih lagi adalah analisis preskriptif, yaitu analisis yang secara otomatis menggabungkan big data, aturan bisnis, dan pembelajaran mesin untuk memberikan rekomendasi keputusan atau tindakan yang tepat berdasarkan hasil prediksi.

Kecerdasan buatan berperan penting dalam mendukung analisis preskriptif dengan cara mengolah serta menganalisis data guna memperkuat otomatisasi dan menghasilkan wawasan yang bermanfaat. Integrasi AI dengan *software-defined network* (SDN) akan melahirkan jaringan yang sepenuhnya otonom memiliki ketahanan tinggi, mampu memulihkan diri secara otomatis, dan terus belajar dari pengalaman. Jaringan cerdas ini tidak



hanya mengumpulkan serta menganalisis data, tetapi juga mampu mengambil keputusan untuk memperbaiki kondisi kinerjanya. Beberapa bidang utama yang menjadi ruang penerapan potensial AI dan teknologi pendukungnya akan dibahas berikut ini:

1. **Optimasi Otomatis dengan Closed-Loop.** Hingga kini, proses optimasi jaringan masih banyak mengandalkan keahlian teknis insinyur jaringan serta metode manual berbasis uji coba. Karena pada dasarnya optimasi adalah usaha matematis untuk memaksimalkan atau meminimalkan KPI tertentu, pendekatan yang digunakan mirip dengan proses gradient descent dalam pelatihan model machine learning, yaitu perbaikan bertahap melalui penyesuaian kecil pada parameter konfigurasi. Jika perubahan memperburuk kinerja, arah perbaikan akan diganti. Namun, dengan semakin luasnya jaringan dan meningkatnya kompleksitas teknologi, cara ini menjadi tidak efisien, memakan banyak waktu, dan tenaga. Ledakan lalu lintas data serta beragam aplikasi yang harus ditangani, ditambah tuntutan penghematan biaya, membuat kebutuhan akan analitik dan optimasi jaringan yang otomatis serta skalabel semakin mendesak. Untuk itu, diperlukan model pengambilan keputusan multivariat yang mencakup faktor radio, pengguna, aplikasi, dan lalu lintas. Perkembangan jaringan berbasis perangkat lunak serta penerapan AI membuka jalan menuju solusi optimasi yang sepenuhnya otomatis dan dapat berkembang sesuai skala kebutuhan.
2. **Deteksi Anomali dan Identifikasi Akar Masalah.** Saat ini pemantauan jaringan masih mengandalkan sistem alarm. Sayangnya, penggunaan ambang batas alarm yang kaku sering menimbulkan dua masalah: ada alarm yang terlewat, atau justru muncul alarm palsu akibat dinamika lingkungan jaringan seluler maupun kabel, serta variasi penggunaan aplikasi. Untuk menganalisis penyebab masalah, umumnya digunakan panduan pemecahan berbasis pengetahuan. Namun, pendekatan ini cenderung menggunakan aturan umum yang belum tentu tepat, sehingga sering kali perlu disesuaikan dengan jenis jaringan dan aplikasi yang berbeda. Lebih jauh, panduan semacam itu hanya bisa mendeteksi masalah yang sudah dikenal sebelumnya, sehingga masalah baru atau tak terduga sering luput dari perhatian. Dengan mengadopsi teknik machine learning untuk deteksi anomali dan pengenalan pola, operator dapat secara proaktif menemukan potensi masalah, mempercepat analisis akar penyebab, serta terus mempelajari pola lalu lintas baru secara adaptif dari waktu ke waktu.
3. **Optimasi Jaringan Berorientasi Pengalaman Pengguna.** Selama ini, indikator kinerja utama (KPI) jaringan umumnya difokuskan pada performa jaringan dan perangkat, bukan pada kualitas layanan atau pengalaman kelompok pengguna tertentu. Ketika lalu lintas masih didominasi panggilan suara, hal ini tidak menjadi kendala. Namun, seiring berkembangnya aplikasi modern, pendekatan KPI yang hanya berpusat pada jaringan tidak lagi memadai untuk memenuhi ekspektasi pengguna maupun kebutuhan layanan. Oleh karena itu, diperlukan KPI yang lebih spesifik, baik untuk pelanggan maupun layanan, agar optimasi jaringan dapat benar-benar selaras dengan kebutuhan mereka. Meski demikian, tantangan tetap besar, karena interaksi antara layanan, pengguna, dan jaringan sangat kompleks serta belum ada model yang mampu



menjelaskan kaitan langsung antara KPI layanan dengan KPI jaringan. Jika KPI dapat dirancang dengan tepat pada berbagai lapisan, maka model probabilistik berbasis AI berpotensi besar untuk menemukan pola dan hubungan lintas lapisan yang dapat meningkatkan kualitas pengalaman pengguna secara lebih menyeluruh.

13.3 TREN TEKNOLOGI YANG MEMUNGKINKAN JARINGAN CERDAS

Selama ini, para perancang dan operator berusaha mengatur perilaku jaringan agar sesuai dengan kondisi yang terjadi maupun kebutuhan aplikasi. Namun, optimasi tradisional masih cenderung berpusat pada jaringan karena kendala arsitektur. Hambatan pertama terletak pada elemen jaringan yang sebagian besar berupa perangkat tertutup dan berlisensi khusus, sehingga sulit dipahami cara kerjanya serta hanya menawarkan kontrol terbatas. Dukungan terhadap API pun masih minim. Hambatan kedua adalah pemisahan antara infrastruktur jaringan dan komputasi, yang membuat pengumpulan data baik dari sisi jaringan maupun aplikasi sulit dilakukan secara bersamaan, apalagi untuk diproses tepat waktu sehingga bisa memberikan dampak nyata pada optimasi jaringan.

Salah satu hambatan utama adalah keterbatasan dalam memproses data secara langsung pada saat pengumpulan. Karena itu, data dalam jumlah besar harus dipindahkan terlebih dahulu ke pusat penyimpanan dan komputasi, yang membuat pengumpulan data detail dalam skala waktu cepat menjadi sulit dilakukan. Kondisi ini memperlambat pengendalian jaringan tingkat lanjut. Ditambah lagi, proses pemantauan setiap perubahan membutuhkan waktu lama, termasuk kemungkinan untuk mengembalikan perubahan jika hasilnya tidak sesuai. Oleh karena itu, operator lebih sering mengandalkan analisis offline yang panjang sebelum mengambil keputusan. Bahkan ketika perubahan dilakukan, biasanya hanya berupa langkah kecil dan bertahap. Dampaknya, transformasi besar dalam jaringan membutuhkan waktu lama, sehingga evolusi jaringan berjalan jauh lebih lambat dibandingkan pesatnya perkembangan aplikasi dan teknologi komputasi.

Seiring meningkatnya kebutuhan aplikasi modern, perkembangan jaringan, dan tuntutan akan kontrol yang lebih rinci serta mendekati real-time, pendekatan manual tradisional dalam pengelolaan jaringan tidak lagi memadai. Jaringan 5G, misalnya, akan menghadapi miliaran perangkat, lalu lintas yang beragam dengan kebutuhan kinerja berbeda-beda, serta aplikasi-aplikasi dengan persyaratan yang kompleks. Untuk menanganinya, dibutuhkan infrastruktur jaringan berskala besar, mencakup ribuan stasiun sel dan teknologi nirkabel yang beraneka ragam. Dalam kondisi ini, jelas bahwa pengelolaan manual tidak akan mampu mengikuti skala kebutuhan: biaya pemeliharaan akan sangat tinggi, dan respons jaringan akan terlalu lambat. Oleh karena itu, solusi yang diperlukan adalah sistem pemantauan dan pengendalian jaringan yang cerdas, berbasis data, dan mampu bereaksi secara real-time. Namun, tantangan utamanya terletak pada kemampuan mengumpulkan serta memproses data dalam jumlah besar dengan cepat, sekaligus melaksanakan tindakan yang tepat. Faktor biaya juga menjadi pertimbangan penting; tanpa perbandingan manfaat yang lebih tinggi dari biaya, penerapan otomatisasi semacam ini sulit untuk dibenarkan.



Beberapa inovasi terkini memungkinkan tercapainya pengendalian jaringan yang lebih cerdas. Pertama, munculnya jaringan berbasis perangkat lunak melalui penerapan teknologi seperti komputasi awan dan virtualisasi fungsi jaringan. Integrasi ini memungkinkan fungsi jaringan dan komputasi dijalankan pada satu platform, sehingga data dapat dikumpulkan secara detail dari elemen jaringan dan langsung diproses begitu tersedia. Kehadiran edge computing semakin memperluas kemampuan ini, karena memungkinkan pemrosesan data dilakukan secara terdistribusi di banyak lokasi. Dengan demikian, kebutuhan untuk memusatkan data di satu tempat dapat dihindari, sehingga mempercepat akses, pemrosesan, dan pengambilan keputusan terkait pengendalian jaringan. Kedua, berkembangnya arsitektur jaringan 5G yang terdisagregasi, baik pada core maupun RAN, serta penggunaan perangkat white box. Pendekatan ini membuka peluang untuk menghadirkan API yang lebih detail guna pengumpulan data dan pengendalian, sekaligus memberikan keluwesan dalam memperluas kapasitas jaringan serta menempatkan fungsi-fungsi yang membutuhkan latensi rendah lebih dekat dengan pengguna maupun aplikasi.

Penggunaan platform manajemen jaringan, misalnya *Open Networking Automation Platform* (ONAP), menjadi salah satu pendorong utama dalam menghadirkan kontrol jaringan yang lebih cepat dan terperinci. Platform ini mampu menyederhanakan tugas-tugas yang biasanya rumit dan lama, seperti pengumpulan serta analisis data, perumusan kebijakan, hingga penerapan otomatis dengan mekanisme closed-loop. Lebih jauh lagi, keberadaan kerangka kerja ML yang semakin matang di dalamnya membuka jalan bagi penerapan kasus penggunaan yang dulunya sulit diwujudkan. Sebagai ilustrasi, kombinasi edge cloud dengan ONAP memungkinkan terjadinya korelasi langsung antara performa jaringan secara detail dengan kinerja aplikasi, sehingga memungkinkan penyesuaian jaringan secara real-time. Hasilnya, kualitas pengalaman pengguna (QoE) dapat terjaga, sementara efisiensi jaringan tetap optimal.

Teknologi 5G menghadirkan fondasi penting yang akan mempercepat penerapan AI dalam berbagai skenario. Dengan menghadirkan keandalan tinggi, kecepatan transmisi data super cepat, latensi nyaris nol, dan kehilangan paket yang sangat kecil, 5G mampu menciptakan ekosistem ideal untuk aplikasi otomatisasi berbasis AI. Tidak hanya itu, kepadatan koneksi yang meningkat dan kemampuan throughput yang lebih besar membuka jalan bagi pemanfaatan AI dalam skala industri maupun kebutuhan individu. Fitur lain yang signifikan adalah network slicing, yang memungkinkan satu infrastruktur jaringan digunakan untuk melayani berbagai jenis aplikasi dengan kebutuhan yang berbeda secara bersamaan, tanpa mengurangi kinerja keseluruhan jaringan.

Berikut adalah ringkasan perubahan penting dalam arsitektur dan teknologi jaringan 5G:

- Peningkatan throughput dan kapasitas yang drastis
- Latensi berkurang
- Kepadatan koneksi yang sangat besar
- Jaringan inti 5G dan RAN yang terdisagregasi dan terdistribusi
- Cloud edge
- Pembagian jaringan



Dengan adanya tren teknologi tersebut dan kemampuan jaringan 5G yang semakin canggih, operator kini memiliki peluang besar untuk memanfaatkan perkembangan pesat AI/ML guna meningkatkan kemampuan pengendalian jaringan.

13.4 PENERAPAN AI DALAM PEMANTAUAN DAN OPTIMASI JARINGAN

Self-Organizing Networks (SON) menjadi salah satu penerapan nyata otomatisasi kontrol jaringan dalam sistem seluler modern. Dengan teknologi ini, operator tidak lagi sepenuhnya bergantung pada konfigurasi manual, melainkan dapat menyesuaikan parameter jaringan secara otomatis melalui perangkat lunak. Standar yang dikembangkan 3GPP untuk LTE menempatkan SON pada dua tujuan utama: meningkatkan efisiensi jaringan dan mengurangi biaya operasional. Contohnya, fitur seperti optimasi kanal akses acak, penyeimbangan beban, serta peningkatan cakupan dan kapasitas berfungsi untuk mengoptimalkan jaringan. Sementara fitur konfigurasi otomatis seperti manajemen hubungan antar sel, perencanaan Physical Cell ID, penghematan energi, hingga deteksi dan penanganan gangguan sel lebih menekankan pada efisiensi operasional. Meski demikian, karena kendala volume data dan keterbatasan pengolahan data secara real-time, kebanyakan algoritma SON saat ini masih bekerja dengan respons yang lambat (dalam hitungan jam hingga hari), serta hanya melakukan perubahan kecil secara bertahap untuk meningkatkan performa jaringan.

Perkembangan menuju jaringan 5G, dengan skala yang jauh lebih besar dan layanan yang semakin beragam, menuntut adanya kontrol jaringan yang lebih pintar. Integrasi pemantauan dan pengendalian berbasis AI dan ML secara real-time membuka peluang signifikan bagi operator untuk menekan biaya investasi (CAPEX) maupun biaya operasional (OPEX), sambil tetap menjaga dan bahkan meningkatkan mutu layanan. Saat ini, penyedia jaringan bergerak ke arah penerapan centralized SON dengan kemampuan pemantauan dan kontrol dalam hitungan 5–10 detik, serta menuju kontrol yang benar-benar real-time (kurang dari 100 milidetik) untuk skenario yang kritis, terutama dengan dukungan teknologi edge cloud dalam evolusi 5G.

Penerapan machine learning (ML) dalam jaringan membuka peluang besar untuk meningkatkan kinerja sekaligus memaksimalkan efisiensi operasional, sebagaimana akan dibahas pada bagian berikutnya. Namun, perlu dipahami bahwa proses optimasi jaringan hampir selalu melibatkan pertukaran kepentingan (trade-off). ML menjadi sarana yang efektif untuk menyeimbangkan berbagai tujuan optimasi dan mencari hasil terbaik sesuai target yang diinginkan. Misalnya, strategi pengendalian yang diarahkan pada peningkatan efisiensi spektral dapat menekan kebutuhan akan penambahan carrier baru, yang berarti mengurangi beban investasi (CAPEX). Sebaliknya, strategi yang berfokus pada efisiensi energi berupaya menghemat biaya operasional (OPEX) dengan cara menurunkan konsumsi energi, sembari tetap menjaga tingkat efisiensi spektral pada level yang memadai.

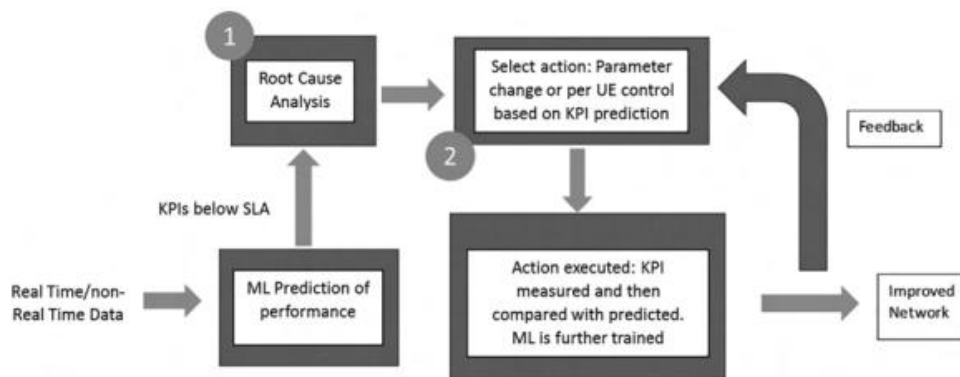
Menerapkan Kecerdasan Buatan pada Kinerja Jaringan

Upaya untuk mengatasi permasalahan performa jaringan umumnya dilakukan dalam dua langkah utama (lihat Gambar 13.1). Pertama, proses diagnosis atau analisis akar masalah

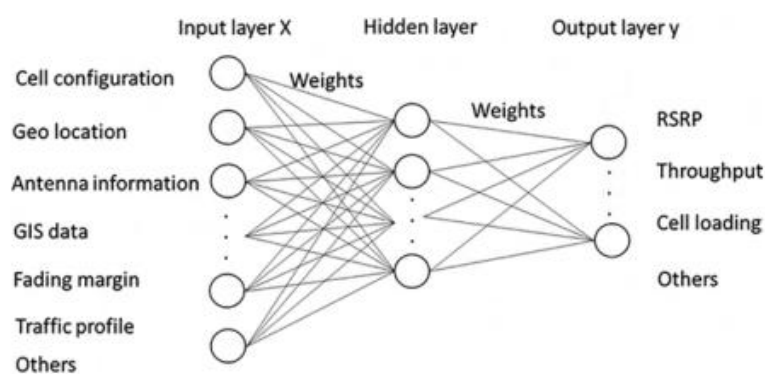


untuk mengetahui sumber gangguan. Kedua, langkah mitigasi atau optimasi guna memperbaiki kondisi jaringan serta meningkatkan kinerjanya.

Pada tahap diagnosis, sistem memanfaatkan analisis berbasis mesin terhadap berbagai indikator kinerja utama (KPI). Apabila hasil prediksi menunjukkan performa di bawah standar SLA, algoritma klasifikasi yang sudah dilatih akan membantu mengungkap penyebab utama, misalnya masalah interferensi, keterbatasan cakupan, atau beban jaringan yang berlebih. Sementara itu, tahap optimasi lebih kompleks karena memerlukan simulasi skenario what-if untuk menentukan langkah yang paling tepat. Tantangan besarnya adalah akurasi prediksi, sebab KPI dipengaruhi oleh banyak faktor dalam sistem nirkabel. Untuk itu, pemanfaatan pengetahuan domain menjadi kunci agar pemilihan variabel tepat sehingga prediksi dapat dilakukan secara cepat dan akurat. Sebagai ilustrasi, Gambar 13.2 menunjukkan penggunaan neural network dalam pemodelan KPI. Dalam konteks ini, persiapan data menjadi tahap krusial yang tidak bisa diabaikan.



Gambar 13.1 Pendekatan peningkatan kinerja dua langkah menggunakan ML.



Gambar 13.2 Jaringan saraf untuk prediksi KPI dalam optimasi Fase 2

Pemodelan memerlukan langkah-langkah krusial seperti memilih data, melakukan pra-pemrosesan, dan mentransformasikan data. Agar hasil yang diperoleh konsisten dan tepat, dibutuhkan waktu, disiplin, pengalaman, serta pemahaman terhadap domain terkait. Pada awal proses pelatihan, alat propagasi yang telah dikalibrasi melalui pengukuran jaringan dapat memberikan data pelatihan yang memadai. Berbagai sumber data, termasuk peristiwa



jaringan, pengukuran dari perangkat pengguna, serta data aplikasi melalui probe, dapat dimanfaatkan untuk kalibrasi. Setelah algoritma berjalan, mekanisme umpan balik digunakan untuk menyempurnakan hasil prediksi. Ringkasan keuntungan dan tantangan AI/ML untuk optimasi jaringan diberikan sebagai berikut:

- ML memiliki kemampuan untuk mengelola kompleksitas jaringan dengan banyak pengguna dan sel hampir secara instan.
- Dibandingkan SON yang berbasis pada pendekatan coba-coba, ML dapat menemukan solusi optimal dengan lebih cepat dan efisien, sekaligus beradaptasi lebih baik terhadap perubahan.
- Teknologi seperti neural networks dan deep learning memungkinkan ML mempelajari hubungan nonlinier yang rumit antara data masukan (misalnya data pengukuran dan counter) dengan KPI, tanpa harus bergantung penuh pada keahlian manusia.
- Kendala utama optimasi jaringan berbasis ML terletak pada kebutuhan data dalam jumlah besar serta kecepatan pengumpulan dan pemrosesan yang sesuai dengan batas waktu optimasi. Sering kali, waktu yang dibutuhkan lebih lama dari jangka waktu yang tersedia.
- Selain itu, hasil optimasi ML sangat dipengaruhi oleh mutu dan ketersediaan data. Meskipun algoritma dapat menangani kekurangan data, membangun model yang tahan terhadap masalah kualitas data tetap merupakan tantangan besar.
- Untuk mampu menangkap dampak lintas-lapisan secara detail, ML membutuhkan model yang benar-benar representatif. Hal ini memerlukan data yang sangat besar dan terperinci, yang tidak mudah diperoleh, serta pembaruan model secara terus-menerus mengikuti perubahan pola penggunaan jaringan dan aplikasi.

Meningkatkan Efisiensi Spektral

Bagi operator jaringan nirkabel, spektrum merupakan aset utama yang paling vital dari sisi investasi. Karena itu, penggunaan spektrum secara optimal menjadi prioritas utama. Walaupun perkembangan teknologi radio terus meningkatkan efisiensi spektral di area tertentu (dalam satuan bps/Hz per luas area), pada jaringan yang sudah ada seperti 4G maupun 5G, masih tersedia peluang besar untuk meningkatkan pemanfaatan sumber daya melalui optimasi di level sistem.

Pada jaringan nirkabel, efisiensi spektral sistem diukur dari total kecepatan transfer data (bit per detik) dibandingkan dengan bandwidth yang dipakai. Berbeda dengan efisiensi spektral per pengguna yang batas maksimalnya ditentukan oleh Teorema Shannon-Hartley berdasarkan nilai SINR, efisiensi spektral sistem lebih ditentukan oleh distribusi SINR di seluruh pengguna dalam jaringan. Selain itu, kinerja efisiensi spektral sistem turut dipengaruhi oleh sejumlah aspek, termasuk teknologi akses radio yang digunakan, struktur/topologi jaringan, pengaturan transmisi, serta pola distribusi lalu lintas data (lihat Gambar 13.3).

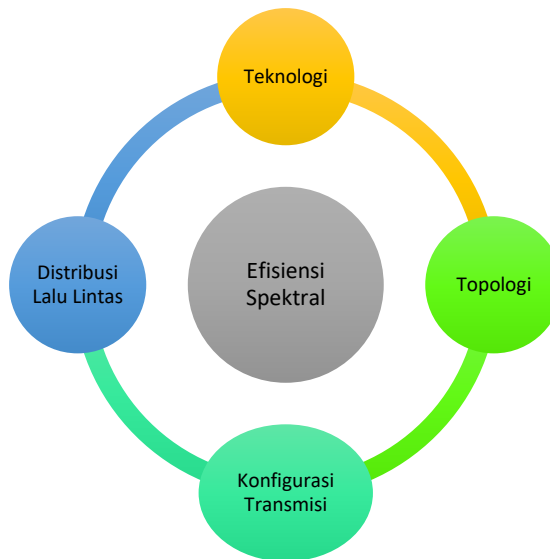
1. **Aspek Teknologi:** Setiap generasi teknologi akses radio menghadirkan metode akses baru yang bertujuan untuk meningkatkan efisiensi spektral. Pada era 3G, digunakan WCDMA dengan pendekatan spread-spectrum. Teknologi ini kemudian digantikan oleh OFDM pada 4G LTE, yang lebih efektif dalam mengatasi gangguan intersymbol



interference akibat efek propagasi multipath. Di jaringan 5G, OFDM dikembangkan lebih lanjut dengan menambahkan elemen penyaringan terintegrasi untuk memperbaiki pengendalian spektrum sinyal transmisi, sehingga efisiensi penggunaan spektrum dapat ditingkatkan.

2. **Aspek Topologi:** Struktur atau susunan jaringan berpengaruh besar terhadap efisiensi spektral, terutama karena tingkat gangguan antar kanal pada sel tetangga sangat ditentukan oleh jarak antar-situs (ISD). Di kawasan perkotaan yang padat, ISD relatif pendek sehingga potensi interferensi lebih tinggi. Dalam konteks jaringan 5G, penggunaan makrosel yang dikombinasikan dengan banyak small cell berdaya rendah membentuk jaringan heterogen (HetNets). Pendekatan ini memungkinkan area cakupan yang lebih fleksibel sekaligus memperkecil radius rata-rata sel, sehingga efisiensi spektral meningkat melalui pemanfaatan kembali spektrum secara lebih intensif, dengan catatan pengelolaan interferensi dilakukan secara efektif. Selain itu, AI dan ML dapat dimanfaatkan dalam pengaturan topologi jaringan, seperti pada strategi perencanaan densifikasi.
3. **Konfigurasi Transmisi:** Efisiensi spektral juga dipengaruhi oleh bagaimana parameter transmisi diatur, termasuk daya pancar, penerapan teknologi MIMO/beamforming, dan strategi adaptasi tautan dengan variasi modulasi serta skema pengkodean. Pada kondisi dengan SINR tinggi, efisiensi transmisi dapat ditingkatkan secara signifikan melalui penggunaan MIMO dengan kapasitas spasial lebih tinggi atau modulasi tingkat lanjut seperti 256QAM. Meski demikian, peningkatan daya pancar bukanlah solusi mutlak, karena justru berpotensi menambah gangguan pada sel di sekitarnya. Oleh sebab itu, penelitian intensif dilakukan dalam bidang pengendalian daya untuk menemukan keseimbangan optimal dalam pengaturan daya. Di sisi lain, AI dan ML dapat dimanfaatkan untuk mendukung proses pengendalian daya tersebut.
4. **Distribusi Lalu Lintas:** Efisiensi spektral sangat dipengaruhi oleh bagaimana lalu lintas jaringan terdistribusi, karena terkait langsung dengan distribusi SINR. Jika sebagian besar lalu lintas berada di area pinggir sel, efisiensi spektral rata-rata akan menurun akibat kualitas SINR yang rendah. Faktor utama yang menentukan pola distribusi ini adalah penempatan menara sel terhadap komunitas pengguna serta kondisi propagasi radio. Keputusan strategis seperti lokasi menara, ketinggian antena, sudut kemiringan, tingkat penguatan, dan pilihan frekuensi pembawa berperan penting dalam membentuk cakupan dan kualitas propagasi sinyal. Pada jaringan heterogen (HetNet), pengaturan daya antara node kecil dan makrosel menjadi faktor penentu apakah pengguna lebih optimal dilayani oleh makrosel atau sel kecil. Salah satu kunci pengendalian di sini adalah load balancing, yang berfungsi mengalihkan beban lalu lintas antar-sel untuk meningkatkan efisiensi spektral maupun pengalaman pengguna, misalnya melalui penyesuaian sudut antena atau pengalihan lalu lintas. Pendekatan AI/ML dapat digunakan untuk mengoptimalkan load balancing.





Gambar 13.3 Dampak sistem efisiensi spektral.

Pengendalian berbasis AI untuk meningkatkan efisiensi spektral di lingkungan dengan banyak pengguna dan sel merupakan bagian dari fungsi *radio resource management* (RRM) dalam RAN. RRM sendiri adalah mekanisme manajemen di tingkat sistem pada eNodeB, yang berfungsi mengoptimalkan parameter penting seperti daya transmisi, skema modulasi/pengkodean, alokasi sumber daya, beamforming, hingga pengaturan handover. Saat ini, operator masih cenderung mengatur parameter RRM secara statis, sehingga hasilnya tidak maksimal dan masih bergantung pada proses manual. Untuk memperoleh kinerja yang lebih baik, diperlukan RRM yang bersifat dinamis, yaitu dengan menyesuaikan parameter secara terus-menerus mengikuti perubahan lalu lintas, variasi aplikasi, distribusi pengguna, maupun target QoS. Di sinilah AI memiliki peran strategis, sebab kemampuannya dalam belajar dari dinamika lingkungan memungkinkan penyesuaian parameter secara adaptif dan cerdas, sehingga menghasilkan pengendalian yang lebih efisien.

Dari berbagai kategori algoritma pembelajaran supervised, unsupervised, reinforcement, hingga deep learning dengan jaringan saraf dua yang paling efektif untuk optimisasi RRM adalah reinforcement learning dan deep learning. Berbagai studi telah banyak mengeksplorasi penggunaan pendekatan ini untuk meningkatkan fungsi serta performa RRM. Seperti yang telah dibahas, peningkatan efisiensi spektral sangat ditentukan oleh tingginya SINR, yang erat kaitannya dengan bagaimana interferensi dikelola secara menyeluruh. Oleh karena itu, bagian berikut akan membahas sejumlah perkembangan penting dalam pengelolaan interferensi menggunakan teknik AI.

Manajemen Interferensi dengan Kecerdasan Buatan

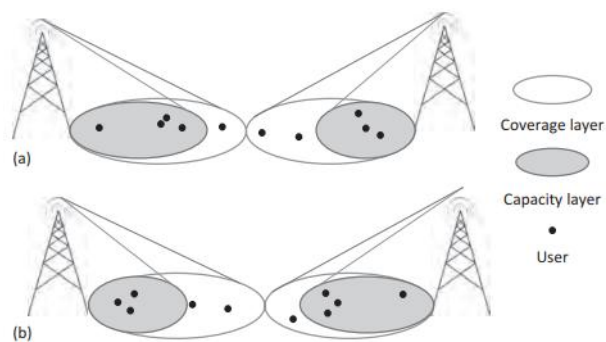
Pengendalian daya merupakan metode penting untuk mengelola interferensi. Penelitian dalam mengusulkan pemanfaatan *reinforcement learning* (RL) guna menyesuaikan daya transmisi dari setiap sel dalam jaringan, dengan tujuan meningkatkan throughput rata-rata pengguna. Dalam pendekatan ini, jaringan seluler diperlakukan sebagai lingkungan RL, sedangkan kondisi digambarkan melalui berbagai fitur hasil pengukuran lokal di tiap sel. Untuk mengoptimalkan hasil, digunakan algoritma Q-Learning yang fungsi-Q-nya dimodelkan



menggunakan jaringan saraf. Reward yang ingin dimaksimalkan didefinisikan sebagai capaian throughput pengguna.

Reinforcement learning juga digunakan dalam fungsi RRM lainnya yang berhubungan dengan manajemen interferensi, terutama pada jaringan HetNet. Salah satu aplikasi pentingnya adalah optimisasi bersama untuk pengaturan sudut antena (tilt), daya transmisi, serta parameter handover antar kelompok sel dalam satu klaster. Tujuannya adalah menemukan keseimbangan ideal antara cakupan jaringan, tingkat interferensi, dan penggunaan daya, agar performa keseluruhan area meningkat. Tantangan muncul karena optimisasi semacam ini dapat menimbulkan area tanpa sinyal yang sulit dideteksi. Meski begitu, kompleksitas masalah dapat dikurangi jika kendala cakupan dibuat lebih longgar dengan cara mempertahankan lapisan cakupan statis, seperti diperlihatkan pada Gambar 13.4.

Reinforcement learning juga digunakan dalam fungsi RRM lainnya yang berhubungan dengan manajemen interferensi, terutama pada jaringan HetNet. Salah satu aplikasi pentingnya adalah optimisasi bersama untuk pengaturan sudut antena (tilt), daya transmisi, serta parameter handover antar kelompok sel dalam satu klaster. Tujuannya adalah menemukan keseimbangan ideal antara cakupan jaringan, tingkat interferensi, dan penggunaan daya, agar performa keseluruhan area meningkat. Tantangan muncul karena optimisasi semacam ini dapat menimbulkan area tanpa sinyal yang sulit dideteksi. Meski begitu, kompleksitas masalah dapat dikurangi jika kendala cakupan dibuat lebih longgar dengan cara mempertahankan lapisan cakupan statis, seperti diperlihatkan pada Gambar 13.4.

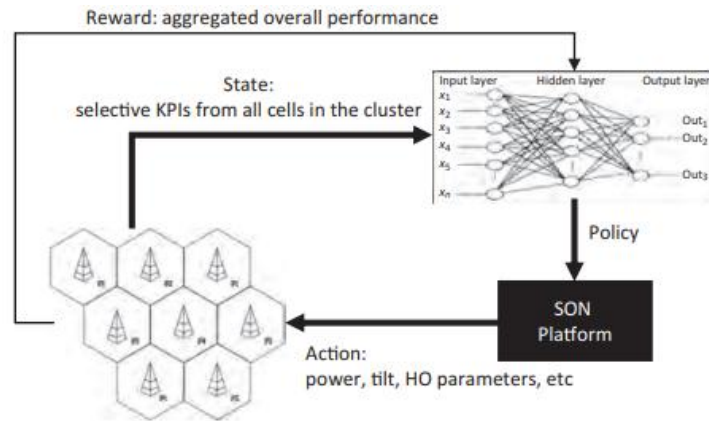


Gambar 13.4 Optimalisasi gabungan lalu lintas dan cakupan. (a) Lalu lintas dan cakupan pada waktu n dan (b) Waktu dan cakupan pada waktu $n+1$.

Dalam pendekatan ini, sekumpulan cell site dipandang sebagai lingkungan bagi algoritma reinforcement learning. Data berupa KPI dari tiap sel seperti beban trafik, tingkat pemanfaatan sumber daya, kualitas kanal rata-rata (CQI), kekuatan sinyal referensi, dan metrik jarak digunakan sebagai state untuk proses pembelajaran. Agar pelatihan berjalan lebih efisien, pengetahuan teknis dimanfaatkan untuk memperkecil kompleksitas ruang state. Seperti yang ditunjukkan pada Gambar 13.5, algoritma RL ini mengandalkan jaringan saraf tiruan yang memetakan kondisi jaringan menjadi tindakan optimal. Reward didefinisikan berdasarkan performa keseluruhan klaster, misalnya peningkatan user throughput. Perubahan nilai reward digunakan untuk menyesuaikan bobot jaringan saraf. Hasil akhirnya berupa



kebijakan yang kemudian diimplementasikan oleh platform SON untuk menyesuaikan konfigurasi sel secara otomatis, sehingga kinerja jaringan pada skala klaster dapat dioptimalkan.



Gambar 13.5 Pembelajaran penguatan mendalam untuk mengoptimalkan konfigurasi sistem.

Jaringan saraf berbasis kecerdasan buatan mampu beradaptasi secara dinamis terhadap perubahan kondisi jaringan, misalnya ketika terjadi variasi pola trafik. Dengan kemampuan ini, sistem dapat menyesuaikan konfigurasi secara otomatis guna menjaga kinerja optimal. Mekanisme adaptif semacam ini menjadi solusi yang ampuh untuk menekan penggunaan sumber daya yang berlebihan sekaligus meningkatkan efisiensi operasional jaringan secara keseluruhan.

Meningkatkan Efisiensi Energi

Efisiensi energi didefinisikan sebagai rasio antara jumlah data yang berhasil dikirim secara andal dengan energi yang digunakan, biasanya diukur dalam bit per Joule. Konsep ini menekankan adanya kompromi antara efisiensi energi dan efisiensi spektrum. Umumnya, peningkatan efisiensi energi dapat dicapai dengan sedikit pengorbanan pada tingkat throughput. Selain memberikan manfaat ekologis, peningkatan efisiensi energi juga bertujuan menekan biaya operasional operator seluler. Jika selama ini desain jaringan lebih berfokus pada peningkatan kinerja seperti efisiensi spektrum, throughput, dan latensi maka era 5G memperkenalkan pergeseran fokus ke arah komunikasi hemat energi, terutama karena lonjakan jumlah perangkat IoT berskala masif yang membutuhkan solusi berbiaya rendah dan efisien.

Berbagai strategi untuk meningkatkan efisiensi energi telah dikaji dalam literatur, dan secara umum dapat dikelompokkan menjadi empat pendekatan utama: pengalokasian sumber daya, perancangan dan penerapan jaringan, pemanfaatan energi terbarukan, serta inovasi pada perangkat keras. Fokus utama dari strategi ini bukanlah pada seberapa besar informasi dapat dikirimkan, melainkan pada seberapa efisien energi digunakan yakni jumlah informasi per Joule energi yang dikonsumsi. Sejumlah algoritma telah dikembangkan untuk tujuan tersebut, meskipun sebagian besar masih menghadapi kendala besar terkait kompleksitas, khususnya saat diterapkan dalam jaringan padat, heterogen, dan rentan interferensi. Salah



satu arah solusi yang menjanjikan adalah dengan memanfaatkan kecerdasan buatan (AI), yang memiliki kemampuan alami untuk mengelola sistem kompleks berskala besar melalui pembelajaran adaptif dan otomatisasi.

Salah satu pendekatan praktis untuk meningkatkan efisiensi energi adalah melalui alokasi sumber daya, khususnya pengaturan daya transmisi, dengan tujuan memaksimalkan rasio informasi per Joule di bawah batasan daya tertentu. Tantangan ini memiliki kemiripan dengan strategi optimalisasi efisiensi spektral, hanya saja tujuan akhirnya berbeda. Di sini, kecerdasan buatan khususnya reinforcement learning dapat dimanfaatkan untuk mempelajari pola lalu lintas dan menyesuaikan penggunaan daya, dengan fungsi penghargaan yang difokuskan pada indikator efisiensi energi. Dalam kasus tertentu, sel tertentu bahkan bisa dimatikan sepenuhnya ketika lalu lintas rendah, kemudian diaktifkan kembali sesuai kebutuhan, sehingga konsumsi energi dapat ditekan secara signifikan. Selain itu, teknik off-loading pada jaringan 5G, seperti komunikasi device-to-device (D2D), memungkinkan perangkat saling terhubung langsung menggunakan kanal seluler dengan tetap berada di bawah kendali base station. Pendekatan ini tidak hanya memanfaatkan kedekatan antarperangkat, tetapi juga mampu mengurangi konsumsi energi yang biasanya lebih besar jika komunikasi dilakukan melalui infrastruktur inti jaringan.

Salah satu pendekatan awal untuk meningkatkan efisiensi energi, yang bahkan dapat diterapkan pada jaringan 4G saat ini, adalah network densification (seperti pada HetNets). Densification merujuk pada paradigma di mana peningkatan kapasitas dan cakupan dicapai melalui penempatan sejumlah besar microcell dan picocell berdaya rendah di jarak antar-sel yang lebih kecil, berbeda dengan model tradisional yang hanya menggunakan beberapa macrocell untuk menyediakan cakupan di wilayah yang lebih luas. Dengan menempatkan small cell secara strategis di dekat pusat populasi, seperti perkantoran, mal, dan kafetaria (terutama yang memiliki infrastruktur backhaul siap pakai), lebih sedikit pengguna yang harus berbagi kapasitas setiap sel, dengan kehilangan jalur dan interferensi lebih rendah, sehingga menghasilkan efisiensi spektral yang lebih tinggi serta throughput pengguna yang lebih baik. Seiring small cell mengalihkan sebagian besar lalu lintas pengguna, macrocell tetap menyediakan cakupan di area terbuka, meskipun dengan pengaturan daya lebih rendah. Skenario ini menyoroti korelasi antara efisiensi energi yang lebih tinggi, pemanfaatan spektrum yang lebih baik, dan pengalaman pengguna. Keuntungan densification akan sangat terasa pada penempatan small cell di area terbatas, seperti dalam gedung, yang strukturnya memberikan isolasi dari interferensi.

Banyak aspek dari perencanaan densification yang dapat dioptimalkan menggunakan AI. Pada tahap awal desain, sangat penting untuk menentukan fasilitas yang layak di lokasi-lokasi potensial untuk menempatkan peralatan radio, seperti tiang listrik, tiang lampu, atau street furniture lainnya yang minim gangguan. Selain lokasi, atribut penting seperti tinggi, bentuk, bahan, dan ketersediaan fasilitas juga harus diperhitungkan. AI berbasis deep learning dan image recognition dapat memanfaatkan data terbuka seperti citra satelit atau street view untuk mengidentifikasi atribut ini secara cepat dan akurat dalam skala besar. Selain itu, aturan lokal mengenai pemasangan peralatan pada tiang utilitas dapat otomatis diekstrak



menggunakan teknologi AI seperti natural language processing. AI juga digunakan untuk mengidentifikasi “zona panas” yang tepat untuk penempatan sel baru, sehingga beberapa tim perencanaan bisa bekerja secara paralel dengan informasi yang sama pada fase awal perencanaan.

Dalam jaringan 5G, densification menjadi pilar penting dalam desain sistem. Penerapan millimeter wave akan memanfaatkan phase-array beamforming bersamaan dengan beamforming di tingkat baseband yang sadar interferensi. Sinar yang sangat terfokus ini meningkatkan efisiensi tautan dan memungkinkan penggunaan ulang spektrum yang padat berkat isolasi spasial dan sudut. Akibatnya, base station millimeter wave berdaya rendah dapat ditempatkan dalam jumlah besar dengan cakupan yang saling tumpang tindih, namun tanpa menimbulkan interferensi antar-sel, sehingga memberikan pengalaman pengguna yang baik sambil menghemat energi.

Gelombang milimeter memiliki perilaku propagasi yang berbeda dibandingkan pita sub-6 GHz tradisional. Panjang gelombangnya yang pendek membuat penyerapan oleh uap air dan oksigen menjadi penting, dan efek difraksi, refleksi, serta hamburan juga berbeda. Bangunan, pepohonan, dan perabot jalanan memengaruhi propagasi, bahkan kehadiran manusia dapat memberikan efek signifikan. Meskipun ray-tracing bisa digunakan untuk simulasi statis, keterbatasan model 3D resolusi tinggi dan objek bergerak menimbulkan tantangan prediksi. Machine learning dapat dilatih untuk mengenali objek di lingkungan RF 3D secara akurat. Operator biasanya menghabiskan banyak waktu untuk mengkalibrasi model propagasi melalui drive test. Dengan jaringan saraf dalam, parameter model RF bisa disesuaikan secara optimal menggunakan pengukuran UE secara terus-menerus, menghasilkan model lokal yang mampu menyesuaikan diri secara dinamis.

Meningkatkan Pengalaman Pengguna

Tujuan utama desain dan optimisasi jaringan adalah meningkatkan pengalaman pengguna. Pengalaman pengguna terkait erat dengan jenis aplikasi yang digunakan, yang masing-masing memiliki kebutuhan QoE yang berbeda. Saat ini, video menyumbang lebih dari separuh lalu lintas seluler, tetapi QoE bagi pengguna masih sering kurang optimal. QoE video ditentukan oleh bit rate yang tinggi, waktu start-up cepat, dan minimnya gangguan freeze. Bagian ini menggunakan video sebagai contoh penerapan AI untuk meningkatkan pengalaman pengguna, dan pendekatan ini dapat diterapkan pada aplikasi Internet lainnya.

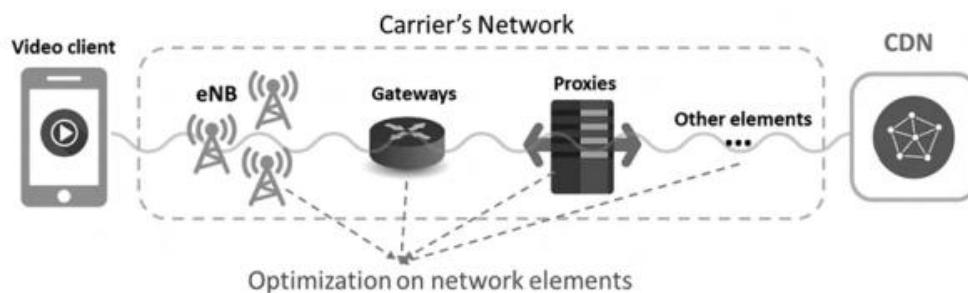
Penerapan pertama AI dalam jaringan seluler adalah untuk klasifikasi lalu lintas, khususnya untuk mengenali lalu lintas video Internet. Metode tradisional menggunakan packet inspection, namun hal ini menjadi sulit karena banyak lalu lintas kini terenkripsi melalui HTTPS atau QUIC, sehingga tanda jaringan seperti HTTP header tidak tersedia. Sebagai gantinya, pola lalu lintas jumlah paket atau byte yang dikirim dalam interval waktu tertentu menjadi fitur utama untuk klasifikasi. AI menawarkan alat yang efektif untuk membedakan lalu lintas video dan non-video berdasarkan pola ini, dan algoritma AI kini sedang dikembangkan untuk menangani lalu lintas terenkripsi.

Selain klasifikasi lalu lintas, AI bisa dimanfaatkan untuk mendeteksi dan mengisolasi masalah jaringan yang memengaruhi QoE video. Beberapa masalah jaringan dapat



menurunkan kinerja video, misalnya konfigurasi yang salah pada mobility gateway atau pengaturan video encoding rate yang tidak sesuai untuk jaringan seluler, sehingga video bisa terhenti saat handover. Menentukan akar penyebabnya memerlukan pengkorelasi antara log aplikasi video pada UE dan data kinerja jaringan dari network probes serta call trace. Saat ini, proses ini masih dilakukan secara manual, yang sangat memakan waktu dan melelahkan.

Solusi AI yang mampu mengukur QoE video secara otomatis dan mengkorelasikan data kinerja jaringan menawarkan pendekatan paling efektif untuk menangani masalah jaringan secara cepat. Sistem ini harus memiliki model kinerja video, baik supervised maupun unsupervised, untuk mendeteksi anomali QoE video secara real time. Model tersebut harus mampu membedakan antara masalah kinerja yang terisolasi, misalnya karena cakupan yang buruk, dengan masalah sistematis seperti kesalahan konfigurasi gateway atau video encoding yang tidak tepat. Dengan kemampuan AI yang terus belajar dan menyempurnakan kalibrasi, model ini akan semakin akurat dan andal. Selain itu, sistem harus memungkinkan drill down dari lapisan aplikasi ke lapisan jaringan yang lebih rendah, seperti TCP atau RAN, untuk menemukan sumber masalah secara tepat.



Gambar 13.6 Optimalisasi video (misalnya, dengan pengaturan lalu lintas) pada elemen jaringan.

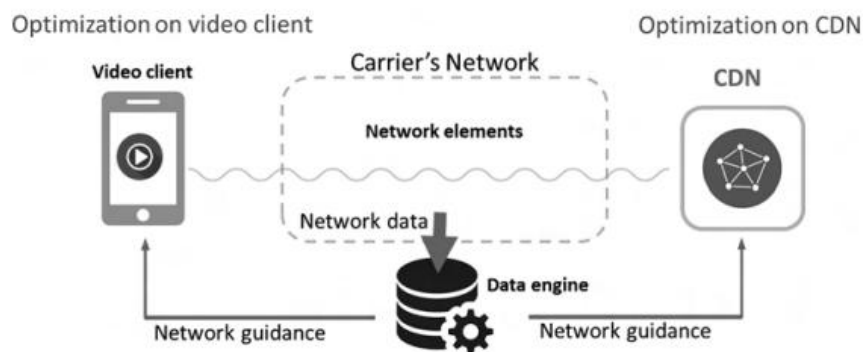
AI dapat diterapkan untuk traffic shaping agar pengalaman pengguna video meningkat. Konsepnya adalah menyesuaikan pola lalu lintas antara video dan non-video sehingga pengalaman video lebih baik. Video, yang sensitif terhadap waktu, sebaiknya dikirim dengan pola lalu lintas yang lebih stabil dan throughput relatif konstan. Tantangan utamanya adalah meningkatkan kinerja video tanpa menurunkan kualitas layanan aplikasi lain serta menyeimbangkan performa di banyak sesi video. Kebijakan traffic shaping statis tidak cukup karena kondisi jaringan sangat dinamis. AI, khususnya reinforcement learning, dapat mempelajari skenario lalu lintas yang berubah-ubah dan menentukan strategi traffic shaping optimal. Operator hanya perlu menentukan kriteria pengalaman pengguna untuk video dan layanan lain, sementara algoritma AI akan menyesuaikan kebijakan secara otomatis.

Penerapan AI pada traffic shaping dapat diperluas ke optimisasi jaringan secara umum. Dalam skema ini, elemen jaringan mencoba menentukan tindakan terbaik, misalnya konfigurasi jaringan, alokasi sumber daya, atau titik operasi dari beberapa opsi yang ada. Setiap tindakan dievaluasi berdasarkan imbalan jangka pendek dan panjang sesuai kriteria pengalaman pengguna. Selama proses pembelajaran, sistem terus memperbarui pengetahuan



untuk menemukan rangkaian tindakan terbaik, sehingga pengalaman pengguna untuk video maupun layanan lain dapat dioptimalkan secara menyeluruh.

Pendekatan lain untuk meningkatkan pengalaman pengguna video adalah dengan menyediakan informasi jaringan misalnya throughput, keterlambatan, dan tingkat kehilangan paket kepada penyedia layanan video OTT, sehingga mereka bisa menyesuaikan aplikasi video mereka. Contohnya, operator dapat menerbitkan panduan throughput untuk membantu aplikasi HLS dan DASH mengoptimalkan algoritma adaptasi video. Biasanya, klien video di perangkat seluler memiliki pandangan terbatas terhadap kondisi jaringan, yang dapat menyebabkan pilihan kualitas video yang tidak optimal. Dengan pandangan jaringan yang lebih lengkap, operator dapat menggunakan algoritma AI, seperti recurrent neural network, untuk memprediksi throughput secara lebih akurat. Klien video kemudian menggunakan prediksi ini melalui API untuk menentukan tingkat video terbaik, sehingga menghasilkan peningkatan kualitas video secara signifikan.



Gambar 13.7 Optimasi video dengan panduan jaringan.

Salah satu upaya yang sedang dilakukan adalah menyediakan panduan throughput dari operator agar aplikasi video dapat menyesuaikan algoritma adaptasi video HLS dan DASH. Klien video seluler menggunakan prediksi throughput ini untuk memilih kualitas video terbaik. Biasanya, pandangan terbatas klien terhadap kondisi jaringan bisa menyebabkan pemilihan video yang kurang optimal. Dengan pandangan jaringan yang lebih lengkap, operator menggunakan algoritma AI untuk memprediksi throughput secara real-time, yang kemudian disediakan ke klien melalui API agar keputusan pemilihan kualitas video lebih tepat. Pengujian awal menunjukkan peningkatan kinerja video yang signifikan, dan recurrent neural network merupakan salah satu alat AI yang digunakan untuk memodelkan prediksi throughput.

ML dan AI dapat digunakan langsung pada algoritma adaptasi video untuk membantu perangkat menentukan kualitas video optimal. Algoritma ini mempertimbangkan bandwidth, jumlah video yang sudah di-buffer, risiko stall, dan faktor lainnya untuk menentukan kualitas terbaik. Penggunaan ML berbasis cloud terbukti mampu memilih kualitas video yang menyeimbangkan berbagai metrik QoE, dan eksperimen menunjukkan performanya setara atau lebih baik dibanding algoritma adaptasi video modern. Dari berbagai contoh di atas, terlihat bahwa AI memiliki peran krusial dalam meningkatkan pengalaman pengguna. Meski



sudah ada banyak aktivitas dan hasil awal yang menjanjikan, penerapan AI untuk memaksimalkan pengalaman pengguna masih belum sepenuhnya terealisasi.

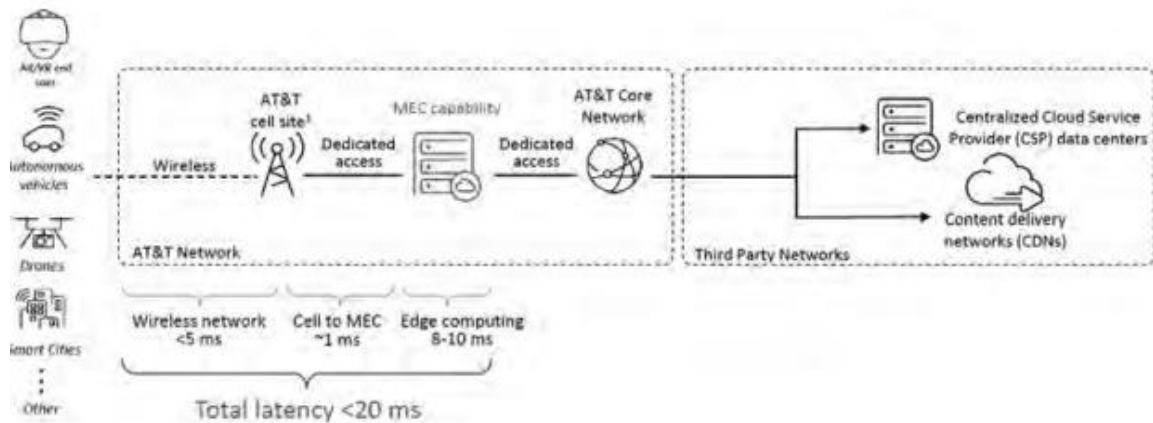
13.5 MENERAPKAN KECERDASAN BUATAN PADA 5G

Perubahan teknologi dan dinamika sosial-ekonomi akan membuat lanskap bisnis 5G setelah 2020 berbeda secara signifikan dibanding masa kini. Next Generation Mobile Network menggambarkan 5G sebagai ekosistem ujung-ke-ujung yang mendukung masyarakat yang sepenuhnya mobile dan terkoneksi, menciptakan nilai baru bagi pelanggan maupun mitra melalui berbagai kasus penggunaan lama maupun baru, dengan pengalaman yang konsisten serta model bisnis yang berkelanjutan. 5G dirancang untuk mendukung beragam aplikasi dengan kebutuhan kinerja yang bervariasi. Untuk itu, konsep network slicing hadir sebagai solusi dalam memenuhi tuntutan layanan dari industri vertikal. Dengan volume data yang terus meningkat dan semakin kompleks, pembelajaran kognitif menjadi pilar penting untuk mengelola dan mengoptimalkan kinerja layanan. Dukungan ini semakin diperkuat oleh komputasi awan, orkestrasi layanan berbasis model, serta edge computing yang memungkinkan pemrosesan dan pengambilan keputusan secara lokal.

Bab 12 telah menjelaskan bahwa edge computing merupakan fondasi utama bagi layanan real-time di jaringan 5G, sebagaimana ditunjukkan dalam Gambar 13.8. Lokasi penempatan edge node berperan penting untuk menekan biaya sekaligus menjaga kualitas layanan. Dimensi, lokasi, dan konfigurasi node ditentukan oleh kebutuhan penyimpanan dan kinerja. Ketika sistem ini diterapkan secara luas, pengaruhnya terhadap jaringan seluler akan besar. Contohnya, perangkat dapat mengalihkan beban kerja ke edge sehingga baterai lebih awet. Perangkat AR/VR juga bisa berjalan mandiri tanpa komputer dengan dukungan mobile-edge compute. Kolaborasi antara edge dan cloud terpusat akan menghasilkan pengalaman yang efisien serta mulus untuk berbagai layanan industri. Secara khusus, edge cloud dirancang bagi aplikasi dengan kebutuhan latensi sangat rendah, yaitu sekitar 10–30 milidetik.

Operator telekomunikasi kini tengah melakukan berbagai uji coba layanan real-time bersama mitra industri dan aliansi, meliputi sektor AR/VR, kendaraan otonom, dan layanan kesehatan digital. Kegiatan ini mendorong kebutuhan akan komputasi real-time yang ditopang oleh jaringan berkecepatan tinggi, integrasi kecerdasan buatan, serta pemrosesan big data di sisi edge. Bagian berikutnya akan menjelaskan bagaimana jaringan otonom berperan dalam mendukung Enhanced Mobile Broadband dengan karakteristik latensi rendah dan kapasitas besar, sekaligus menjadi landasan transformasi besar di sektor otomotif, manufaktur, energi, hingga layanan kesehatan.





Gambar 13.8 Arsitektur komputasi edge.

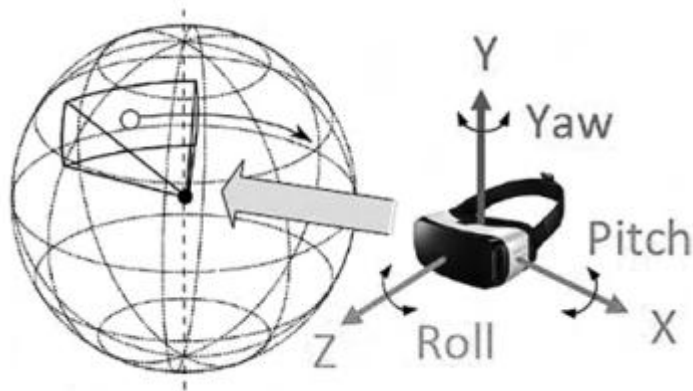
Penerapan Kecerdasan Buatan dalam VR, Video 360°, dan AR

Diperkirakan pasar global untuk VR, video 360 derajat, dan AR akan melampaui Rp775 triliun. pada tahun 2021. Walaupun saat ini perhatian masih banyak terfokus pada perangkat headset dan aplikasi gaming, pengembangan jaringan menjadi faktor krusial untuk mendukung teknologi tersebut. Setiap jenis layanan video membutuhkan kapasitas data yang besar dan latensi ultra-rendah agar dapat memberikan pengalaman imersif yang optimal. Di sini, AI dan machine learning memiliki peranan strategis, baik untuk meningkatkan efisiensi distribusi konten, memindahkan beban komputasi ke cloud, menemukan peluang pemanfaatan komputasi bersama antar pengguna, maupun memastikan keamanan data.

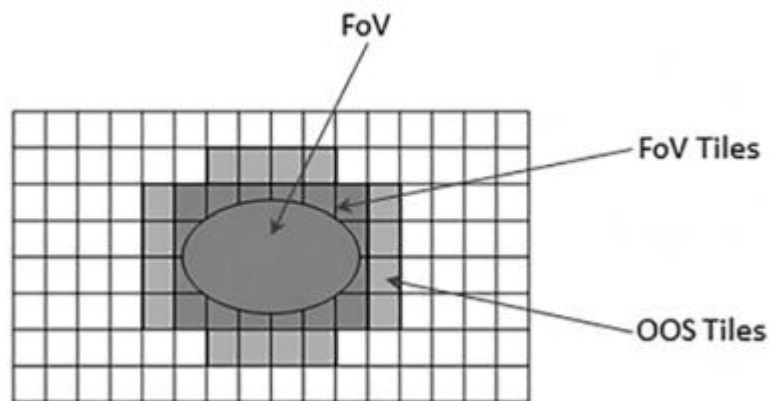
Dalam konteks ini, video 360 digunakan sebagai ilustrasi untuk menunjukkan pemanfaatan AI dalam mendukung distribusi konten yang lebih efisien. Video 360 derajat, atau yang biasa disebut video imersif maupun video sferis, menjadi komponen penting dalam dunia VR karena mampu memberikan pengalaman visual panorama. Ketika dipadukan dengan teknologi video 3D, pengalaman pengguna semakin kaya dan unik. Penonton berada di pusat bola dan dapat mengontrol arah pandangannya secara bebas, sehingga setiap sesi menonton dapat menghasilkan pengalaman yang berbeda. Seperti digambarkan pada Gambar 13.9, pengguna yang memakai VR headset dapat mengubah orientasi dengan melakukan rotasi pitch, yaw, dan roll sesuai dengan sumbu X, Y, dan Z. Pemutar video 360 kemudian menghitung area pandang yang ditampilkan berdasarkan orientasi serta *field of view* (FoV), yakni parameter yang menentukan luas area yang terlihat. FoV umumnya sudah ditetapkan oleh perangkat VR headset, misalnya 110° secara horizontal dan 90° secara vertikal.

Studi pada dua layanan video 360 derajat yang sudah beredar di pasaran menemukan bahwa metode distribusinya masih banyak mengadopsi pola dari video konvensional. Pendekatan ini memang memudahkan penerapan, namun justru membuat streaming video 360 memerlukan bandwidth yang jauh lebih besar. Penyebab utamanya adalah pemutar video perlu mengunduh seluruh konten baik bagian yang dilihat maupun yang tidak dilihat pengguna dalam kualitas tinggi. Hal ini menimbulkan inefisiensi besar pada pemakaian sumber daya jaringan serta meningkatkan konsumsi energi radio pada perangkat pengguna.





Gambar 13.9 Menyesuaikan arah tampilan video 360.



Gambar 13.10 Segmentasi spasial potongan video menjadi petak.

Salah satu cara yang dapat ditempuh untuk menekan penggunaan bandwidth pada video 360 derajat adalah melalui metode berbasis FoV. Konsep utamanya adalah pemutar video memprediksi arah pandangan pengguna, lalu hanya mengunduh bagian konten yang sesuai dengan bidang pandangan tersebut. Untuk mewujudkan streaming adaptif ini, setiap potongan video 360 lebih dulu dipecah menjadi unit-unit kecil bernama tile. Pemutar hanya akan mengambil tile yang masuk dalam area pandangan pengguna, sebagaimana ilustrasi pada Gambar 13.10.

Untuk dapat menebak FoV dengan presisi, pemutar video harus memanfaatkan kombinasi berbagai informasi, misalnya pola gerakan kepala penonton, karakteristik konten video, hingga profil pengguna. Di sinilah AI dan machine learning berperan penting: model dapat dilatih untuk memperkirakan bidang pandang dan arah gerakan kepala pengguna dalam waktu nyata, berdasarkan kumpulan data dari berbagai jenis video dan perilaku tipikal penonton. Prediksi ini juga bisa diperkaya dengan memasukkan faktor tambahan seperti isyarat audio atau elemen yang mampu menarik fokus visual. Karena keterbatasan perangkat pengguna dalam hal data maupun kapasitas komputasi, proses prediksi lebih ideal dijalankan di edge cloud. Dengan pendekatan ini, hasil prediksi dapat diperoleh dengan cepat sekaligus menyesuaikan diri terhadap dinamika penggunaan secara real time.



Penerapan AI juga sangat relevan untuk teknologi video mutakhir lain seperti *Augmented Reality* (AR). Tantangan utamanya adalah memastikan video feed dari lingkungan pengguna bisa diperkaya dengan elemen virtual secara instan dan tanpa gangguan. Karena kapasitas perangkat tampilan terbatas, sebagian besar proses augmentasi harus didukung oleh cloud. AI kemudian dapat berperan dalam mengatur secara adaptif bagian mana dari proses komputasi yang lebih efisien dijalankan di perangkat lokal, dan bagian mana yang sebaiknya dipindahkan ke cloud. Lebih jauh lagi, karena AR selalu berhubungan langsung dengan konteks nyata, ada potensi untuk memanfaatkan kembali hasil komputasi augmentasi antar pengguna. Dengan bantuan AI, sistem dapat mengenali kelompok pengguna yang bisa saling berbagi hasil pemrosesan serta kondisi yang tepat untuk melakukannya. Mekanisme berbagi ini bisa membawa manfaat besar, misalnya dalam situasi darurat di mana petugas lapangan dapat saling bertukar informasi visual secara cepat dan efisien.

AI juga diterapkan pada teknologi video baru seperti AR, di mana tantangannya adalah memproses video feed pengguna dan menambahkan informasi virtual secara instan dan mulus. Karena keterbatasan penyimpanan di perangkat, sebagian besar pemrosesan dilakukan di cloud. AI dapat menentukan secara adaptif bagian proses mana yang lebih baik dijalankan di cloud dan mana yang bisa dilakukan di perangkat lokal. Selain itu, karena AR terkait erat dengan lingkungan nyata, AI dapat membantu memutuskan kapan dan bagi siapa hasil augmentasi yang sudah dihitung dapat digunakan kembali. Hal ini memungkinkan pengguna tertentu, seperti petugas darurat, untuk berbagi informasi secara cepat dan efisien dengan memanfaatkan hasil komputasi yang sama.

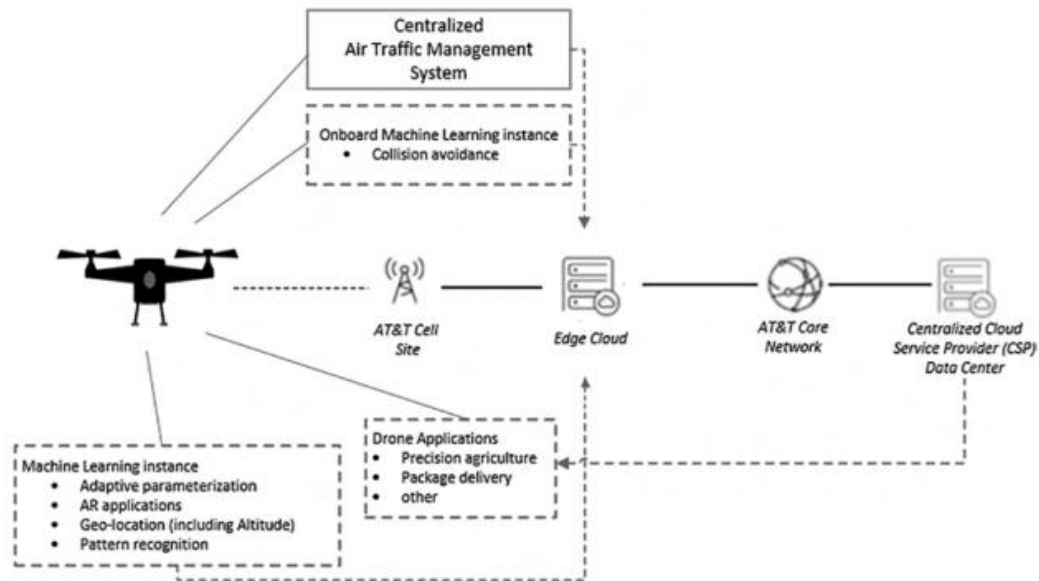
Pesawat Tanpa Awak

Harga teknologi drone yang terus turun membuat penggunaannya untuk berbagai aktivitas harian menjadi semakin ekonomis. Transportasi diprediksi akan menjadi sektor pertama yang terpengaruh secara signifikan. Drone sudah digunakan untuk layanan pengiriman jarak terakhir di berbagai negara, termasuk pengiriman di wilayah pegunungan yang sulit dijangkau. Ke depannya, drone berkemampuan tinggi dapat melakukan pemupukan otomatis di pertanian, memonitor kecelakaan lalu lintas, meninjau lokasi sulit dijangkau, hingga mengantarkan makanan seperti pizza.

Kebanyakan layanan drone akan dioperasikan melalui arsitektur sistem yang mencakup drone, cloud, serta entitas manajemen lalu lintas terpusat, dengan analitik berbasis ML dijalankan pada edge cloud untuk mendukung pemrosesan data secara lokal dan real-time. Sistem manajemen lalu lintas drone dirancang untuk memberikan rute penerbangan otomatis dan fleksibel secara nirkabel, baik saat misi dimulai maupun selama perjalanan berlangsung. Jalur penerbangan ini wajib mengikuti aturan yang berlaku serta diatur agar tidak mengganggu performa jaringan nirkabel. Di dalam arsitektur sistem, kecerdasan buatan akan berperan penting untuk meningkatkan efisiensi dan operasional. Algoritme AI dilatih agar dapat menyesuaikan rute penerbangan dengan kondisi yang berubah secara tiba-tiba, misalnya karena faktor cuaca (hujan, angin) atau situasi darurat akibat aktivitas manusia (seperti kecelakaan atau lalu lintas). Untuk situasi seperti itu, keputusan biasanya harus dibuat oleh pusat pengendali, bukan drone secara langsung. Namun, karena adanya kebutuhan analisis



cepat dan keputusan instan, kemampuan machine learning perlu ditanamkan langsung ke dalam drone. Terutama untuk drone yang beroperasi di area penuh rintangan atau di atas keramaian, sistem harus bisa mengenali perubahan lingkungan yang tidak terduga secara cepat dan mengambil keputusan keselamatan dalam hitungan detik. Sebagian keputusan ini bahkan akan dibuat secara mandiri oleh drone melalui kecerdasan buatan yang terpasang di dalamnya.



Gambar 13.11 Arsitektur sistem drone.

Teknologi swarming menawarkan terobosan menarik dalam aplikasi drone, karena AI memungkinkan sejumlah drone bekerja bersama sebagai satu kesatuan yang terkoordinasi untuk menyelesaikan tugas-tugas kompleks. Dalam sebuah swarm, biasanya ada satu atau beberapa unit pemimpin (pilot) yang terhubung langsung dengan cloud atau sistem manajemen lalu lintas pusat. Unit pemimpin ini bertugas mendistribusikan instruksi melalui koneksi nirkabel D2D kepada drone lain dalam kelompok agar dapat menjalankan misi di lingkungan yang berubah-ubah. Tingkat kemandirian setiap drone dalam kelompok ini berbeda-beda, menyesuaikan dengan kecanggihan perangkat keras yang dimiliki, terutama bagi unit non-pilot. Teknologi swarming memiliki potensi besar untuk berbagai keperluan komersial, misalnya di industri konstruksi maupun aplikasi berbasis AR. Salah satu contoh spektakuler penerapan teknologi ini adalah ketika 1.218 drone terbang secara sinkron dalam upacara pembukaan Olimpiade Musim Dingin 2018 di Pyeongchang, Korea Selatan.

Operator jaringan menghadapi tantangan besar untuk memastikan koneksi radio yang andal antara drone dan sistem pengatur lalu lintas. Jaringan seluler awalnya hanya dirancang untuk pengguna di darat, sehingga kehadiran drone yang berkomunikasi di udara memunculkan masalah baru, terutama terkait pengendalian interferensi pada uplink dari drone ke stasiun basis. Tantangan lain muncul dari kebutuhan latensi yang sangat rendah dalam aplikasi drone real-time. Kehadiran 5G membawa solusi, karena menawarkan koneksi

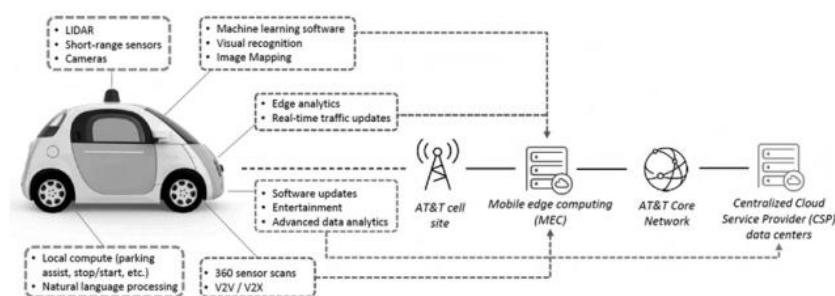


radio dengan latensi ultra-rendah dan dukungan antenna canggih yang mampu menekan interferensi. Teknologi beam steering dengan adaptive beamforming dapat mengurangi potensi gangguan yang ditimbulkan drone saat aktif terhubung. Selain itu, algoritma AI dapat dimanfaatkan di tingkat jaringan untuk mengenali pola interferensi yang berasal dari drone pada kelompok sel tertentu, lalu mengatasinya secara langsung melalui penyesuaian parameter adaptif.

Penerapan AI di Otomotif: Tabrakan & Mengemudi Otomatis

Eksperimen jalan raya pintar pertama dimulai pada 1990an di San Diego, ketika sebuah ruas jalan bebas hambatan dipasang teknologi uji coba. Sejak itu, upaya penelitian berkembang pesat di berbagai belahan dunia, dengan melibatkan ilmuwan, produsen mobil, dan penyedia telekomunikasi, semua berfokus pada terobosan menuju sistem berkendara otomatis sepenuhnya. Di Eropa, proyek Cooperative ITS Corridor sedang dikembangkan dengan integrasi sensor canggih dan jaringan seluler berkecepatan tinggi. Sementara itu, pada 2016 di lintasan balap BMW di Korea Selatan, dua mobil dengan dukungan 5G berhasil bertukar informasi dengan pengemudi. Visi ke depannya adalah menggunakan komunikasi kendaraan-ke-kendaraan berbasis 5G sehingga mobil bisa saling berinteraksi secara langsung tanpa perantara.

AI dan machine learning diperkirakan akan memainkan peran penting dalam transformasi industri otomotif, khususnya dengan dukungan jaringan edge cloud 5G yang sangat cepat dan andal. Penerapan ML di sisi edge memungkinkan sistem kendaraan untuk mengenali pola, menganalisis kejadian teratur maupun tak terduga, serta mengambil keputusan yang lebih cerdas dan aman saat berkendara. Dengan dukungan data dari kamera real-time, sensor kendaraan, radar cuaca, hingga informasi dari mobil lain, mobil masa depan dapat memiliki kemampuan intelligent driving yang jauh lebih aman bagi penggunanya. Teknologi komunikasi seluler V2X (C-V2X) berbasis 5G akan memungkinkan interaksi langsung dan real-time antara kendaraan, perangkat, infrastruktur jalan, bahkan dengan pejalan kaki, sehingga membentuk ekosistem transportasi pintar yang lebih efisien dan aman.



Gambar 13.12 Komputasi edge di dalam mobil dan latensi rendah untuk pengambilan keputusan secara real-time bagi mobil otonom.

Mobil otonom dirancang agar dapat segera merespons situasi mendadak di jalan, seperti lubang besar atau pohon tumbang yang menghalangi jalur. Saat kejadian semacam ini terdeteksi oleh sensor, algoritme machine learning yang tertanam di mobil akan langsung



memberi instruksi untuk memperlambat kecepatan dan menghindari rintangan, sembari tetap menjaga keselamatan kendaraan lain dan fasilitas jalan. Hanya dalam hitungan milidetik, sistem juga akan mengirimkan pesan real-time berbasis C-V2X kepada mobil lain serta infrastruktur jalan di sekitarnya dalam radius tertentu. Pesan tersebut dikirim melalui jaringan seluler 5G berkecepatan tinggi menuju server edge cloud, di mana perangkat lunak ML menganalisis dan menentukan tindak lanjut komunikasi. Hasil analisis ini akan memicu peringatan kepada pihak darurat seperti polisi, pemadam kebakaran, atau petugas jalan untuk segera menuju lokasi insiden, sekaligus memberi notifikasi kepada mobil-mobil terdekat agar menyesuaikan kecepatan dan pola berkendara mereka. Infrastruktur jalan seperti lampu lalu lintas, pagar, atau tiang lampu, serta sistem peta daring juga akan diperbarui agar beradaptasi dengan kondisi terbaru. Mobil yang menerima notifikasi akan melambat untuk mencegah kemacetan, dan secara otomatis diarahkan ke sisi kanan jalan untuk memberi ruang bagi kendaraan darurat.

Contoh tadi hanyalah salah satu dari berbagai kemungkinan penerapan AI, machine learning, dan jaringan 5G dalam solusi kendaraan otonom dan terhubung. Teknologi ini pada akhirnya bertujuan untuk mempermudah kehidupan manusia sekaligus memberikan tingkat keselamatan yang lebih tinggi di jalan raya.

Manufaktur (Otomatisasi Proses, Pengawasan Jarak Jauh, Robotika Awan)

Perkembangan teknologi mendorong robot dan sistem otomasi proses untuk semakin mendominasi industri manufaktur dengan laju yang sangat cepat. Dalam periode 2011–2016 saja, penjualan robot meningkat rata-rata 12% setiap tahun, dan diperkirakan pertumbuhannya akan terus naik sekitar 15% per tahun hingga 2020. Namun, laporan tersebut juga menyoroti sejumlah tantangan penting yang harus dihadapi dalam perkembangan sektor ini, di antaranya:

- Pemanfaatan robot kolaboratif, IoT, dan ML/AI untuk lebih meningkatkan efisiensi dan mendorong peluang baru
- Memungkinkan robot memperoleh dan mengadaptasi keterampilan baru serta merespons masalah tak terduga dalam proses manufaktur

Penerapan Robotic Process Automation sudah membawa perubahan besar dalam industri, tetapi integrasi 5G dengan kecerdasan buatan dan machine learning diprediksi akan melahirkan lompatan revolusioner berikutnya. Kolaborasi teknologi ini mampu menjawab berbagai tantangan utama yang sebelumnya menghambat otomasi industri. Forum Ekonomi Dunia bahkan menamai era baru ini sebagai Revolusi Industri 4.0, dengan fondasi utama pada 5G, AI, dan ML.

5G berperan penting dalam menghadirkan komunikasi berkecepatan tinggi dan berlatensi sangat rendah lintas area dan lokasi. Jika dipadukan dengan teknologi cloud dan edge computing, sistem ini memungkinkan pengumpulan serta analisis data masif langsung dari perangkat di tepi jaringan, seperti sensor, kamera, hingga video. Dengan begitu, robot dapat berkolaborasi secara otomatis dan real-time. Lebih jauh, pemanfaatan teknik AI/ML canggih seperti deep learning memungkinkan sistem mengenali masalah secara proaktif,



memprediksi tren serta pola, mendeteksi anomali dengan lebih akurat, bahkan menyarankan langkah-langkah perbaikan sebelum masalah berkembang lebih serius.

Ada tiga sektor kunci dalam dunia manufaktur yang diperkirakan akan memperoleh manfaat besar dari penerapan teknologi 5G, AI/ML, serta edge computing, sekaligus membuka jalan bagi peluang bisnis inovatif:

1. **Diagnostik dan pemulihan cerdas:** Mesin dan robot modern kini dilengkapi dengan sensor beragam yang merekam data operasional, kondisi sekitar, serta status kesehatan peralatan. Melalui dukungan 5G, edge computing, dan analitik big data, perusahaan dapat menghimpun serta memproses data berskala besar dari jaringan sensor secara instan, bahkan pada fasilitas manufaktur yang tersebar luas. Dengan bantuan algoritma AI/ML mutakhir, sistem mampu membangun model deteksi anomali yang memungkinkan mesin melakukan pemeriksaan kondisi sekaligus pemulihan mandiri. Lebih jauh lagi, teknologi ini membuka jalan bagi komunikasi langsung antar mesin (D2D), sehingga mereka bisa bekerja secara sinkron dalam memantau proses produksi dan menanggapi gangguan dengan cepat. Di masa depan, algoritma self-learning berbasis jaringan saraf tiruan juga berpotensi mendeteksi permasalahan lebih awal dan secara otomatis menyesuaikan alur produksi, misalnya dengan mengalihkan beban kerja ke mesin lain yang masih optimal.
2. **Manufaktur ramping pintar dan pengelolaan rantai pasok/alur kerja:** Integrasi 5G dengan kecerdasan buatan mampu mengoptimalkan efisiensi di sektor manufaktur modern yang kian rumit, terutama dalam pengelolaan proses produksi dan logistik persediaan yang melibatkan skala besar serta lokasi yang tersebar. Karena dipengaruhi oleh faktor internal maupun eksternal, efisiensi di bidang ini bisa menghasilkan penghematan biaya yang besar. Salah satu contohnya adalah pemanfaatan data dari setiap tahap proses produksi mulai dari asal bahan hingga distribusi untuk dijadikan dasar analisis algoritma canggih dalam memprediksi dampak keterlambatan pada jadwal akhir. Dengan demikian, potensi keterlambatan dapat terdeteksi lebih awal dan langkah korektif bisa segera diterapkan. Selain itu, teknologi ini dapat mendorong penerapan konsep smart lean manufacturing di rantai pasok. Dengan menggabungkan data dari banyak sumber, model prediksi yang presisi dapat dibangun untuk menilai kebutuhan nyata pasar. Teknologi otomatis berbasis 5G, termasuk komunikasi antar perangkat (D2D) dan tactile Internet, memungkinkan penyesuaian tingkat produksi secara dinamis agar selaras dengan permintaan aktual.
3. **Pemantauan dan diagnostik infrastruktur jarak jauh:** Banyak sektor industri mengandalkan pengawasan infrastruktur vital mulai dari jalur listrik, menara telekomunikasi, pipa distribusi air maupun gas, bendungan, ketel industri, hingga reaktor nuklir. Namun, inspeksi menyeluruh kerap terkendala oleh tingginya biaya dan sulitnya akses ke lokasi yang berbahaya. Solusinya, penggunaan robot dan drone yang didukung konektivitas 5G memungkinkan pengumpulan data visual maupun sensor secara real-time. Dengan bantuan analisis berbasis edge computing dan model data, tindakan korektif bisa diidentifikasi lebih cepat, bahkan robot atau drone dapat



diarahkan untuk melakukan perbaikan langsung. Lebih lanjut, algoritma canggih juga mampu memandu robot atau drone agar memfokuskan pemantauan pada area tertentu dan mengumpulkan informasi tambahan yang akan mendukung diagnosa teknis oleh insinyur. Dengan cara ini, biaya pengawasan bisa ditekan dan keselamatan operator manusia lebih terjamin karena tidak harus turun langsung ke area berisiko tinggi.

Di ranah otomasi proses dan robotika berbasis cloud, teknologi 5G bersama AI memegang peranan vital dengan memberikan dampak signifikan bagi dunia usaha. Keandalan koneksi yang berkesinambungan menjadi fondasi penting bagi industri modern, sementara sifat nirkabelnya membawa fleksibilitas serta efisiensi biaya dalam menghadapi dinamika perubahan lingkungan industri. Ketika dipadukan dengan edge computing dan teknologi cloud, 5G memungkinkan realisasi respon taktik secara instan yang krusial dalam sektor manufaktur. Lebih jauh, kemajuan machine learning dan AI akan mempercepat otomatisasi, memperkuat kerja kolaboratif, serta mendorong operasional yang bersifat proaktif semua ini membuka peluang dan keuntungan baru bagi industri manufaktur. Selain itu, perangkat pelanggan universal berupa white box yang berfungsi sebagai cloud mini dengan kemampuan menampung virtual router, firewall, IoT gateway, hingga fungsi AI sederhana, akan menjadi elemen penting dalam mendukung berbagai aplikasi dan layanan korporasi.

Energi, Utilitas, dan Jaringan Cerdas

Untuk menjawab tantangan berat dari pertumbuhan populasi dunia dan perubahan iklim, sistem energi kini mengalami perubahan mendasar. Tuntutan akan energi yang murah, aman, dan berkelanjutan membuat jaringan listrik konvensional di berbagai negara mulai digantikan oleh jaringan pintar. Jaringan ini mengintegrasikan teknologi informasi dan komunikasi mutakhir guna meningkatkan efisiensi produksi sekaligus distribusi tenaga listrik. Dalam proses transformasi ini, peran 5G dan kecerdasan buatan menjadi sangat penting sebagai fondasi utama perkembangan smart grid ke depan.

Salah satu persoalan terbesar dalam pengelolaan jaringan listrik adalah menjaga keseimbangan antara permintaan dan pasokan agar tidak terjadi pemadaman. Pada sistem konvensional, hal ini biasanya diatasi dengan mengatur pasokan misalnya meningkatkan kapasitas produksi listrik dengan cepat untuk memenuhi lonjakan kebutuhan. Akan tetapi, masuknya energi terbarukan dalam jumlah besar membuat pengendalian pasokan tidak lagi semudah sebelumnya, seperti pada pembangkit berbahan bakar fosil. Energi dari panel surya maupun turbin angin sangat bergantung pada kondisi cuaca dan musim, sehingga ketersediaannya sering berfluktuasi. Oleh karena itu, smart grid harus memiliki kemampuan adaptif dalam mengatur berbagai sumber daya agar jaringan tetap andal dan pasokan listrik tersedia secara konsisten.

Untuk mengatasi fluktuasi besar dalam pasokan energi terbarukan, smart grid memanfaatkan kecerdasan buatan dan pembelajaran mesin guna mengambil keputusan optimal secara menyeluruh. Dari sisi suplai, sensor pada turbin angin dapat mengumpulkan data historis maupun real-time guna memahami hubungan antara kondisi cuaca dengan kapasitas energi yang dihasilkan. Dengan pola ini, sistem bisa memprediksi secara tepat berapa



besar daya yang mampu dihasilkan turbin dalam kondisi cuaca tertentu, lalu menyeimbangkannya dengan pasokan dari sumber energi lain agar kebutuhan total tetap terpenuhi. Selain itu, smart grid juga memberikan fleksibilitas dari sisi konsumsi energi. Sistem dapat mengatur perangkat atau fasilitas yang konsumsi listriknya tidak terlalu bergantung pada waktu misalnya pendingin ruangan di gudang atau lemari pendingin di supermarket. Saat turbin angin menghasilkan kelebihan energi, sistem bisa memerintahkan lemari pendingin untuk bekerja ekstra dengan menurunkan suhu hingga batas minimal. Namun, ketika permintaan lebih tinggi dari pasokan, konsumsi perangkat tersebut dapat dikurangi atau ditunda, asalkan suhu masih dalam batas aman. Seluruh mekanisme pengendalian dari sisi permintaan ini hanya dapat berjalan efektif berkat adanya infrastruktur telekomunikasi modern dengan konektivitas cepat dan andal, seperti yang ditawarkan jaringan 5G.

Pemadaman listrik telah lama menjadi masalah serius, menimbulkan kerugian ekonomi dalam jumlah besar sekaligus mengganggu kenyamanan masyarakat. Peristiwa ini biasanya terjadi akibat gangguan mendadak, seperti sambaran petir atau kerusakan perangkat energi, dan sering muncul ketika jaringan listrik beroperasi pada kapasitas maksimum. Pada sistem tradisional, standar ketat diterapkan untuk peralatan dan operasional pembangkit demi mencegah pemadaman. Namun, tren baru muncul ketika konsumen rumah tangga ikut serta sebagai penyedia listrik melalui panel surya dan memasok kembali energi ke jaringan. Situasi ini memunculkan tantangan baru: banyaknya pemasok dengan kualitas peralatan beragam membuat sulit menjaga standar seketat pembangkit konvensional. Hal ini meningkatkan risiko kegagalan perangkat dan memperbesar kemungkinan blackout, sehingga pencegahan menjadi semakin rumit. Solusi yang ditawarkan smart grid adalah pemantauan jaringan secara real-time dengan sistem cerdas serta pendekatan proaktif. Di sinilah peran 5G dan AI menjadi sangat penting. Dengan dukungan sensor yang terhubung melalui 5G, data historis dan penggunaan energi terkini dapat dianalisis oleh machine learning untuk memperkirakan kebutuhan energi pelanggan secara presisi. Jika kebutuhan melebihi kapasitas cadangan dari pembangkit atau infrastruktur lain AI dapat segera mengambil keputusan untuk mengisolasi wilayah bermasalah agar gangguan tidak meluas. Selain itu, berkat isolasi gangguan berbasis AI, tim teknisi dapat mempercepat identifikasi masalah serta melakukan perbaikan lebih cepat, sehingga pemulihan layanan berlangsung jauh lebih singkat.

Jaringan 5G akan menjadi fondasi penting bagi smart grid dengan menyediakan sarana pemantauan, pengumpulan, dan pengiriman data maupun kontrol secara real-time. Keberadaan infrastruktur ini menjamin berbagai potensi manfaat dari smart grid benar-benar bisa diwujudkan. Lebih jauh lagi, penerapan software-defined network yang dipadukan dengan kecerdasan buatan membuka peluang penerapan yang luas dalam sistem kelistrikan pintar. Teknologi ini mampu meningkatkan akurasi prediksi pasokan dan permintaan energi, mendukung fleksibilitas penggunaan daya di sisi konsumen, memberikan kendali lebih baik terhadap pola konsumsi, serta membantu mencegah terjadinya blackout. Dengan demikian, kombinasi 5G dan AI akan memungkinkan penyediaan energi yang lebih handal, efisien, ramah lingkungan, serta terjangkau.



Pemantauan eHealth dan Perawatan Berkelanjutan

Saat ini, pemantauan kesehatan telah menjadi bagian penting dalam kehidupan sehari-hari dan terus berkembang dengan pesat. Teknologi LTE-M telah membantu meningkatkan kualitas sinyal di rumah sakit maupun fasilitas medis yang sebelumnya sulit dijangkau. Namun, kehadiran 5G akan membawa dampak yang lebih besar, memperkuat peran jaringan seluler dalam ekosistem layanan kesehatan modern. Konsep precision medicine yang memanfaatkan kecerdasan buatan dan pembelajaran mesin akan mendorong pencegahan penyakit secara lebih dini melalui pemantauan kesehatan real-time, meningkatkan konektivitas antarperangkat medis, serta memungkinkan perawatan berkelanjutan di luar rumah sakit. Selain itu, 5G juga akan mendukung transformasi layanan kesehatan menuju sistem yang lebih terdesentralisasi dan tervirtualisasi, seperti telemedisin, sehingga pelayanan medis dapat lebih personal, efisien, dan merata.

Perangkat wearable, seperti sensor yang dipasang di pergelangan tangan, akan semakin canggih berkat teknologi 5G, memungkinkan pemantauan pasien serta pengumpulan data kesehatan secara terus-menerus. Informasi vital pasien dapat langsung dikirimkan ke dokter tanpa jeda, sehingga kondisi kesehatan dapat terpantau real-time. Dengan dukungan pembelajaran mesin, kondisi pasien bisa diprediksi secara akurat mengurangi kebutuhan kunjungan rutin bagi pasien yang sehat, sekaligus memberikan peringatan dini bagi mereka yang membutuhkan tindakan medis segera. Bagi penderita penyakit kronis seperti asma, diabetes, stroke, maupun gangguan jantung, 5G dan AI membuka jalan untuk perawatan yang lebih efisien. Jika sebelumnya pasien harus sering datang ke klinik untuk pemeriksaan dan penyesuaian terapi, kini hal itu dapat digantikan oleh sistem monitoring real-time. Dokter bisa menyusun rencana pengobatan yang lebih tepat berdasarkan data kesehatan harian pasien, termasuk respons tubuh terhadap obat tertentu. Dengan demikian, konsep precision medicine benar-benar dapat terwujud, menghadirkan perawatan yang lebih personal, efektif, dan tepat sasaran.

Perkembangan telemedisin akan mengalami lompatan besar berkat teknologi 5G. Dengan kemampuan transmisi data yang sangat cepat, latensi mendekati nol, dan stabilitas jaringan yang lebih baik, layanan seperti bedah jarak jauh maupun analisis medis real-time akan lebih mudah diimplementasikan secara aman dan handal, hal yang sulit dicapai dengan jaringan 4G. Kehadiran 5G juga menghapus batasan jarak antara pasien dan fasilitas kesehatan, karena dokter spesialis dapat melakukan tindakan medis dari jauh tanpa harus hadir langsung. Bahkan, pasien yang sulit dievakuasi ke rumah sakit akibat kendala geografis, keterbatasan infrastruktur, atau kondisi cedera yang parah tetap dapat menerima perawatan melalui operasi jarak jauh, baik dengan keterlibatan langsung dokter maupun menggunakan sistem robotik medis lokal yang dikendalikan secara remote.

Pembelajaran mesin berpotensi merevolusi layanan kesehatan, bukan dengan menyingkirkan peran dokter, melainkan dengan memperkuat kemampuan diagnostik dan prediksi hasil perawatan melalui deep learning serta kecerdasan buatan. Bayangkan di masa depan, ketika pasien melaporkan gejala pada dokter, sistem ML dapat langsung mengolah informasi tersebut, menarik data rekam medis, menganalisis hasil MRI atau X-ray, lalu



memunculkan riset medis terbaru yang relevan. Dari situ, sistem dapat merekomendasikan diagnosis dan terapi yang paling sesuai dengan kondisi unik pasien.

Penerapan ML dalam bidang pencitraan medis kini menjadi salah satu arah paling menjanjikan, karena algoritme deep learning mampu mendeteksi pola-pola yang sulit dikenali manusia. Selain itu, AI juga membuka jalan baru dalam penelitian medis, seperti deteksi kanker sejak dini melalui tes darah sederhana, pemanfaatan AR berbasis AI untuk membantu dokter bedah menafsirkan hasil radiologi, atau analitik prediktif yang menyusun pola hubungan antara gejala, prosedur, diagnosis, dan pengobatan pada level individu maupun kelompok pasien. Perkembangan ini masih tahap awal, namun potensinya sangat luas untuk membentuk masa depan layanan kesehatan yang lebih akurat, cepat, dan personal.

13.6 RINGKASAN

Bab ini membahas berbagai penerapan machine learning dan kecerdasan buatan untuk mendukung optimasi jaringan yang bersifat otomatis dan berbasis closed-loop, sekaligus meningkatkan pengalaman pengguna melalui optimasi jaringan yang berorientasi pada layanan. Meskipun sejumlah kapabilitas jaringan sudah tersedia sejak lama, masih ada berbagai keterbatasan teknis yang menghambat operator seluler untuk sepenuhnya mewujudkan potensi jaringan otonom. Pola lalu lintas jaringan dapat berubah sangat cepat seiring hadirnya aplikasi baru dan perubahan karakteristik jaringan. Tidak semua data jaringan tersedia dengan tingkat kedetailan yang sama atau dalam skala waktu yang seragam. Selain itu, data real-time yang sangat penting untuk mendukung pengambilan keputusan optimal sering kali tidak tersedia. Tantangan lainnya adalah ketiadaan model yang jelas antara indikator kinerja utama (KPI) jaringan dan layanan; hubungan sebab-akibat antarparameter jaringan tidak selalu diketahui atau dipahami dengan baik di seluruh lapisan protokol.

Perkembangan teknologi jaringan modern, seperti 5G yang cepat, dapat diskalakan, berlatensi rendah, mendukung beragam perangkat, dengan arsitektur terbuka dan elemen yang terpisah, serta kemampuan pemrosesan data real-time di sisi edge, menjadi fondasi penting untuk memungkinkan pengumpulan sekaligus pengolahan data dalam jumlah besar secara instan. Selain itu, dukungan perangkat lunak AI mutakhir, platform manajemen jaringan seperti ONAP, serta akses langsung ke data dalam skala besar di berbagai lapisan jaringan, mempercepat operator dalam merancang dan mengimplementasikan aplikasi optimasi jaringan berbasis kecerdasan buatan tanpa perlu analisis rumit dari para ahli. Penerapan ML/AI bahkan sudah mulai terlihat pada pengaturan protokol jaringan (contohnya kontrol kemacetan TCP) maupun perilaku aplikasi (seperti penyesuaian kualitas video). Meski demikian, ini baru langkah awal; masih banyak potensi yang bisa digali untuk menjadikan AI sebagai motor penggerak pemantauan dan optimasi jaringan. Hambatan utama hanyalah keterbatasan data dan kreativitas manusia. Bab selanjutnya akan membahas bagaimana AI dapat dimanfaatkan untuk meningkatkan pengalaman serta pelayanan pelanggan.



BAB 14

KECERDASAN BUATAN UNTUK LAYANAN PELANGGAN

14.1 PENDAHULUAN

Dalam era kognitif saat ini, memberikan pengalaman pelanggan yang istimewa bukan lagi pilihan, tetapi kewajiban, terutama karena layanan bergerak menuju sistem jaringan yang semakin otomatis. Teknologi AI memainkan peran sentral, baik dalam menghadirkan pengalaman digital interaktif di sisi pengguna maupun dalam analisis data di belakang layar untuk menghasilkan wawasan cerdas yang dapat membantu agen manusia maupun virtual. Perusahaan yang sukses memanfaatkan AI untuk memandu pelanggan menemukan solusi, mengakses sumber daya yang tepat, sekaligus memperkuat pegawai dan pimpinan dalam membuat keputusan yang lebih akurat. Lebih jauh, AI memungkinkan terciptanya produk serta layanan yang dipersonalisasi sesuai kebutuhan, serta memperluas skala layanan pelanggan melalui sistem cerdas yang terus ditingkatkan dengan pengetahuan terbaik dari organisasi. Bagian ini membahas bagaimana AI mendukung proses pengambilan keputusan yang lebih cerdas guna meningkatkan kualitas pengalaman pelanggan sekaligus memperbaiki layanan pelanggan secara menyeluruh.

14.2 DAMPAK JARINGAN OTONOM PADA PENGALAMAN PELANGGAN

Sejarah Layanan Pelanggan untuk Jaringan

Di era awal jaringan telepon, proses operasional hampir tidak melibatkan teknologi otomatisasi. Semua tugas, baik penyambungan sambungan telepon maupun pelayanan terhadap pelanggan, dilakukan sepenuhnya oleh karyawan perusahaan telekomunikasi. Pada dekade 1950an, Bell System bahkan membuat film pelatihan khusus bagi petugas layanan pelanggan, yang menggambarkan masa itu sebagai “masa keemasan” dalam pelayanan pelanggan. Dalam film tersebut dijelaskan lima tahapan penting dalam menjalin interaksi yang baik dengan pelanggan.

1. Jadilah pendengar yang baik
2. Dapatkan faktanya
3. Lakukan sesuatu
4. Tindak lanjuti hasilnya
5. Yang terpenting, tunjukkan rasa simpati Anda

Berbasis Aturan

Ketika tekanan ekonomi mulai dirasakan oleh perusahaan penyedia layanan, mereka berupaya melakukan efisiensi dengan mengotomatisasi berbagai fungsi pendukung, termasuk layanan pelanggan. Karena mayoritas pertanyaan dari pelanggan disampaikan lewat telepon, bentuk awal otomatisasi diwujudkan melalui sistem *interactive voice response* (IVR). Instruksi sederhana seperti “Tekan 1 untuk masalah tagihan” atau “Tekan 2 untuk gangguan jaringan” menjadi solusi praktis untuk memangkas waktu kerja petugas dalam meneruskan panggilan.



Dengan cara ini, sistem IVR dapat langsung mengarahkan pelanggan ke bagian layanan yang sesuai tanpa campur tangan manusia.

Dalam perkembangan terbaru, teknologi IVR semakin maju berkat integrasi business rules management systems. Sistem ini memungkinkan IVR diprogram menggunakan aturan bisnis tertentu yang bisa mengenali beragam kondisi sekaligus menentukan langkah tindak lanjutnya. Misalnya, jika seorang pelanggan menghubungi karena gangguan jaringan, sistem akan otomatis menjalankan tes jaringan. Berdasarkan hasil tes tersebut, sistem bisa segera melakukan perbaikan atau membuat tiket laporan gangguan, tanpa perlu melibatkan staf layanan pelanggan secara langsung.

Layanan Pelanggan di Era Kognitif

Walaupun sistem IVR berbasis aturan mampu menangani berbagai skenario, kenyataannya sering kali hasilnya belum memuaskan pelanggan. Interaksi dengan IVR biasanya masih mengandalkan tombol nada, atau bila menggunakan suara tetap membatasi pengguna untuk mengucapkan kata-kata tertentu sesuai daftar yang tersedia. Kondisi ini kerap menimbulkan rasa jengkel bagi pelanggan. Saking rumitnya, sampai-sampai muncul situs-situs web yang memberikan panduan agar pelanggan bisa lolos dari “jebakan IVR” dan akhirnya berbicara langsung dengan petugas layanan.

Lebih jauh lagi, keterbatasan sistem berbasis aturan terlihat jelas ketika dihadapkan pada masalah baru yang tidak tercakup dalam daftar kasus yang telah ditentukan. Untuk bisa berfungsi, sistem ini sering kali membutuhkan ribuan aturan yang harus dirancang dan dikelola secara manual oleh pakar. Oleh karena itu, apabila industri ingin mencapai kembali “masa keemasan” layanan pelanggan, langkah yang dibutuhkan adalah inovasi yang melampaui pendekatan konvensional berbasis aturan tersebut.

Bahasa Alami

Dalam video pelatihan layanan pelanggan era 1950an, poin penting pertama adalah kemampuan untuk mendengarkan dengan baik suatu hal yang masih menjadi kelemahan bagi banyak sistem layanan pelanggan saat ini. Bayangkan jika pelanggan bisa menyampaikan keluhan atau pertanyaan mereka menggunakan bahasa sehari-hari, dan sistem layanan pelanggan berbasis AI dapat memahaminya. Lebih jauh lagi, bayangkan sistem itu tidak hanya menanggapi secara natural, tetapi juga mampu menunjukkan empati atau kepedulian terhadap situasi pelanggan.

Pembelajaran

Agen layanan yang bersikap ramah tidak cukup jika tidak dapat menyelesaikan permasalahan pelanggan. Untuk benar-benar menindaklanjuti masalah, dibutuhkan sistem yang dapat mempelajari solusi paling efektif dari pengalaman nyata, bukan hanya sekadar mengikuti aturan tetap yang dibuat oleh beberapa ahli. Sistem semacam ini bisa menganalisis ribuan tiket layanan dan panduan produk yang ditulis dalam bahasa sehari-hari, kemudian menentukan langkah penyelesaian yang paling sesuai untuk setiap pelanggan.

Dengan menyatukan semua konsep dan inovasi ini, kita mencapai tahap yang dikenal sebagai pengalaman digital yang telah direvolusi atau ditingkatkan secara signifikan.



14.3 PENGALAMAN DIGITAL YANG BERTRANSFORMASI

Dalam era pengalaman digital modern, pelanggan memiliki kebebasan untuk berinteraksi dengan perusahaan melalui beragam saluran, mulai dari chatbot, suara, aplikasi mobile, situs web, hingga kios layanan mandiri. Kemajuan besar dalam teknologi AI memungkinkan sistem memahami bahasa manusia secara lebih mendalam bukan hanya kata-kata, tetapi juga emosi, konteks, nada, maksud, hingga konsep yang terkandung. Kemampuan ini membuka peluang luas bagi perusahaan untuk menghadirkan pengalaman pelanggan yang lebih personal, responsif, dan memuaskan.

Melibatkan Pelanggan dengan Chatbots

Dahulu, pelanggan menyelesaikan masalah dengan menelepon layanan pelanggan, di mana petugas menggunakan sistem internal untuk memberikan jawaban. Kini, chatbot hadir sebagai solusi otomatis yang dilatih oleh para ahli menggunakan data crowdsourcing agar mampu memberikan jawaban yang tepat dan spesifik. Tidak seperti mesin pencari, chatbot dirancang untuk memahami bahasa alami sehingga dapat menanggapi pertanyaan sesuai maksud, entitas, dan konteks yang diajukan pelanggan. Contohnya, IBM Watson Assistant dapat diintegrasikan dalam berbagai lingkungan mulai dari mobil, kamar hotel, bandara, hingga layanan perbankan dan telekomunikasi untuk memberikan pengalaman interaktif. Chatbot ini tidak hanya sekadar menjawab pertanyaan, tetapi juga dapat melakukan tugas nyata seperti check-in hotel, mengontrol perangkat di mobil, hingga memberikan informasi layanan, dengan respons yang lebih alami dan kontekstual.

Chatbot bukanlah pengganti manusia, melainkan alat yang memperkuat kemampuan manusia dalam berpikir dan mengambil keputusan. Sistem ini bisa muncul di berbagai platform yang biasa digunakan pelanggan, seperti Facebook Messenger, aplikasi pesan di ponsel, atau chat di situs web. Secara umum, chatbot dibagi menjadi dua kategori utama:

Bantuan Agen

Di sini, agen call center dapat menggunakan sistem AI sebagai pendukung dalam membuat keputusan. Kehadiran chatbot membantu meningkatkan mutu layanan, memperkaya pengetahuan agen, mempersingkat waktu pelatihan, dan mempermudah peningkatan kapasitas layanan saat periode sibuk seperti musim liburan.

Agen Virtual

Agen virtual adalah sistem otomatis yang dapat menjawab pertanyaan pelanggan dan terus belajar dari pengalaman interaksi. Jika agen virtual tidak mengetahui jawaban suatu pertanyaan, pertanyaan tersebut dapat diteruskan ke agen manusia. Contoh terkenal termasuk IBM Watson Assistant untuk industri otomotif dan perhotelan, yang memungkinkan koneksi antara kendaraan, pengemudi, atau tamu hotel dengan dunia di sekitar mereka melalui asisten virtual. Selain itu, Staples Easy Button juga telah dikembangkan menjadi sistem dukungan dan pemesanan virtual, menunjukkan bagaimana agen virtual bisa meningkatkan layanan dan pengalaman pengguna.

Antarmuka Suara

Pengenalan sistem panggilan touchtone pada 1960-an membuka jalan bagi otomatisasi layanan pelanggan. Sebelumnya, pelanggan harus menekan beberapa tombol atau menjawab



melalui suara untuk menghubungi agen yang tepat. Kini, sistem AI yang terlatih dapat memahami maksud pelanggan melalui gabungan pengenalan suara dan bahasa alami. Dengan cara ini, sistem mampu menangkap maksud pertanyaan di antarmuka suara dan mengarahkan permintaan ke agen yang sesuai menggunakan algoritma AI. Sebagai contoh, Verizon menggunakan sistem ini untuk menyalin panggilan masuk, mengenali maksudnya, dan menyesuaikan operasional call center secara real time.

AI dapat berkembang lebih lanjut dengan menyampaikan jawaban langsung kepada pengguna melalui suara menggunakan teknologi text-to-speech, serupa dengan cara kerja chatbot, namun lebih interaktif. Sistem AI ini sama dengan yang digunakan untuk chatbot atau sistem lain yang mendukung pengambilan keputusan. Salah satu contoh awalnya dibuat oleh AT&T Research dengan sistem “AT&T, how may I help you?”

Seperti chatbot, antarmuka suara juga bisa berperan sebagai asisten bagi agen. Panggilan dari pelanggan dibagi menjadi dua jalur: satu untuk agen manusia dan satu untuk sistem AI. Agen call center dapat mempelajari respons AI, yang dilatih berdasarkan pengalaman terbaik call center, untuk membantu menangani situasi tertentu selama panggilan berlangsung.

Interaksi yang Dipersonalisasi, Termasuk Kepribadian

Ekspektasi pelanggan bisa berada di dua ujung spektrum: terpenuhi atau tidak terpenuhi, yang menimbulkan emosi seperti senang, sedih, atau marah. Sistem AI memanfaatkan analisis nada dan sentimen dari jawaban pelanggan untuk memahami emosi mereka. Contohnya, Watson Tone Analyzer dapat mendeteksi nada dalam bahasa lisan maupun tulisan, seperti positif, marah, atau ragu-ragu, sedangkan IBM Watson Natural Language Understanding menilai sentimen sebagai positif, negatif, atau netral. Gabungan analisis ini digunakan dalam model pembelajaran mesin untuk memahami mood dan kepribadian pelanggan. Watson Tone Analyzer juga membantu chatbot dalam membuat keputusan yang lebih tepat berdasarkan perkembangan nada sepanjang percakapan.

Informasi mengenai interaksi individual dan karakteristik kepribadian pelanggan dapat dimanfaatkan untuk melatih AI, sehingga sistem mampu memberikan jawaban yang dipersonalisasi dan meningkatkan kualitas pengalaman pelanggan.

Tindakan Terbaik Berikutnya

Sistem dan proses yang cerdas memanfaatkan analitik untuk mendukung pengambilan keputusan yang lebih efektif. Sistem prediktif berbasis AI atau machine learning dapat menentukan langkah terbaik selanjutnya berdasarkan wawasan pelanggan dan faktor lain untuk tujuan bisnis, seperti meningkatkan retensi pelanggan. Dalam layanan pelanggan, rekaman interaksi call center baik berupa suara maupun teks dianalisis oleh AI. Suara ditranskripsi menjadi teks dan informasi sensitif dihapus. Teks yang telah dibersihkan kemudian dianalisis untuk menentukan nada, sentimen, konteks, kata kunci, dan entitas. Model machine learning yang dihasilkan dapat mengenali situasi serupa di masa depan dan memberikan rekomendasi tindakan terbaik bagi agen call center.

Dalam contoh nyata, seorang pelanggan membeli dua ponsel dalam promosi “beli satu, gratis satu”, membayar satu tunai dan satu lagi dengan cicilan. Seharusnya, ia menerima kartu



Visa senilai harga ponsel setelah menyerahkan struk. Beberapa minggu kemudian, klaimnya ditolak dan pelanggan menanyakan alasannya. Sistem AI mengenali emosi pelanggan yang kesal dan menentukan langkah terbaik berikutnya: menghadirkan ahli layanan pelanggan. Ahli tersebut segera menjelaskan bahwa klaim ditolak karena kedua ponsel harus dibeli dengan cicilan. Setelah itu, transaksi diperbaiki, kedua ponsel ditempatkan pada rencana cicilan, dan klaim diproses dengan benar.

Mencegah Pelanggan Berhenti Berlangganan

Menurut penelitian yang dilakukan di NY EDU, perusahaan telekomunikasi kehilangan sekitar Rp1,007,500,000,000 setiap tahun akibat customer churn (pelanggan berpindah provider). Sekitar 20% pelanggan di perusahaan Telco terbesar berganti penyedia layanan setiap tahunnya. Kecerdasan buatan (AI) dapat mencegah customer churn. Seperti yang dijelaskan sebelumnya, rekaman historis transkrip call-center dapat dianalisis untuk mengidentifikasi parameter yang berhubungan dengan churn. Parameter ini, bersama dengan informasi pelanggan, perilaku pelanggan, serta pesanan dan penggunaan layanan, dapat digunakan untuk membuat model machine learning yang dapat memprediksi kemungkinan pelanggan akan churn.

Setelah model machine learning terlatih diterapkan, sistem AI dapat menyarankan skor churn untuk pelanggan dan langkah terbaik berikutnya berdasarkan skor tersebut dan parameter yang tersedia. Jadi, ketika pelanggan berinteraksi dengan Telco, niat, nada bicara, dan sentimen pelanggan dapat diidentifikasi. Berdasarkan informasi ini serta data pelanggan, sistem interaksi atau perwakilan call-center dapat memperoleh skor churn dan langkah terbaik berikutnya dari model machine learning AI. Beberapa rekomendasi yang mungkin diberikan termasuk meningkatkan transparansi, memberikan diskon tambahan, penetapan harga yang kompetitif, atau menemukan perwakilan layanan yang tepat. AI dapat digunakan untuk memperkirakan kemungkinan pelanggan meninggalkan layanan dan membantu perusahaan meminimalkan kerugian finansial akibat churn.

Sistem Rekomendasi untuk Perawatan Proaktif

Selain menangani churn saat pelanggan menghubungi layanan, sistem rekomendasi berbasis AI bisa bekerja secara proaktif. Dengan memanfaatkan data historis, tren, informasi kompetitor, dan interaksi sebelumnya, model prediktif dibuat dan digunakan untuk menentukan tindakan yang meningkatkan pengalaman pelanggan. Misalnya, memberikan diskon, menawarkan produk baru lebih awal, memperbaiki jaringan, atau menghubungi pelanggan saat ada gangguan layanan.

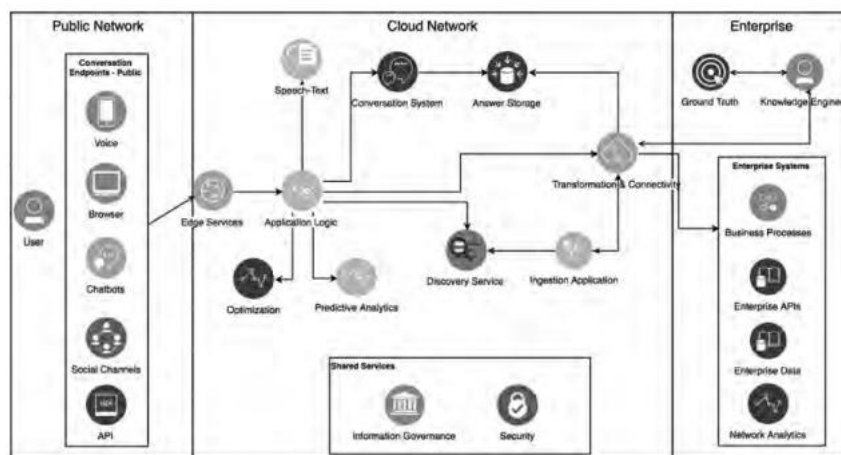
14.4 MEMBANGUN SISTEM AI UNTUK LAYANAN PELANGGAN

Setelah mengetahui bagaimana penyedia layanan mampu meningkatkan pengalaman pelanggan, selanjutnya kita akan membahas elemen-elemen utama yang menyusun sebuah sistem layanan pelanggan berbasis kecerdasan buatan secara menyeluruh, sekaligus meninjau metode yang dipakai untuk mewujudkan sistem tersebut.



Gambar 14.1, yang diambil dari arsitektur referensi percakapan kognitif IBM yang dipublikasikan secara terbuka, menyajikan berbagai konsep yang dijelaskan dalam bagian-bagian terkait.

- **Sistem Percakapan:** Bagian ini berfungsi sebagai pusat utama dari layanan pelanggan berbasis AI. Ia menafsirkan masukan pelanggan, memahami tujuan mereka, lalu merespons baik secara otomatis maupun dengan dukungan staf layanan.
- **Model Prediktif:** Model ini menggunakan analitik prediktif untuk memberikan wawasan tambahan sehingga sistem maupun petugas layanan bisa memberikan tanggapan yang lebih relevan dan personal.
- **Model Optimisasi:** Penerapan teknik optimisasi membantu meningkatkan kinerja organisasi layanan pelanggan maupun operasi jaringan agar lebih efisien.
- **Sistem Perusahaan:** Modul ini, yang digambarkan di sisi kanan diagram, menyumbang data penting tentang pelanggan dan jaringan. Informasi tersebut memperkuat kecerdasan sistem AI sekaligus memperkaya interaksi percakapan.



Gambar 14.1 Arsitektur referensi untuk layanan pelanggan.

- **Omni-channel dan saluran sosial:** Sistem layanan pelanggan modern dirancang untuk menerima masukan dari berbagai kanal, mulai dari panggilan telepon dengan konversi suara ke teks, aplikasi mobile, chatbot internal, hingga interaksi melalui media sosial.
- **Proses pelatihan:** Agar dapat bekerja optimal, sistem AI harus dilatih oleh pakar pengetahuan menggunakan data nyata dari perusahaan yang dipadukan dengan data hasil partisipasi pengguna luas. Pelatihan ini bersifat berkelanjutan karena sistem akan terus diperbaiki berdasarkan masukan dan pengalaman nyata.
- **Fitur Discovery:** Tidak semua pertanyaan pelanggan bisa diantisipasi dalam skenario dialog standar. Dengan adanya fungsi Discovery, sistem memiliki kemampuan untuk melakukan pencarian lebih mendalam sehingga dapat menjawab pertanyaan yang tidak terduga maupun bersifat spesifik.



Pelatihan dan Pemodelan Percakapan

Chatbot, asisten agen, maupun agen virtual pada dasarnya adalah bentuk berbeda dari sistem percakapan, yaitu komunikasi timbal balik menggunakan bahasa alami antara pelanggan dan AI, atau agen manusia yang dibantu AI. Terlepas dari apakah interaksinya berbasis suara atau teks, seluruh sistem percakapan ini memiliki fondasi konsep yang sama, meliputi maksud (intent), entitas, konteks, alur dialog, serta nada percakapan.

Maksud

Dalam sistem percakapan, inti utamanya adalah kemampuan memahami maksud pelanggan atau intent, yaitu apa tujuan, topik, maupun pertanyaan yang mereka sampaikan. Misalnya, ketika pelanggan mengatakan “Sinyal ponsel saya hilang di rumah sejak satu jam lalu,” sistem perlu mengenali hal tersebut sebagai masalah jaringan yang butuh penyelidikan. Tantangannya, variasi bahasa pelanggan sangat luas sehingga sulit jika sistem harus mencocokkan semua kemungkinan ucapan. Cara lain yang pernah digunakan adalah dengan membatasi respons hanya pada kata kunci tertentu, seperti “pembaruan akun” atau “tagihan,” sebagaimana terlihat pada sistem IVR berbasis suara. Namun, metode ini cenderung membuat pengalaman pelanggan menjadi kurang memuaskan, karena pelanggan harus menyesuaikan diri dengan sistem, bukan sebaliknya. Akibatnya, jika sistem gagal memahami pertanyaan, pelanggan mudah merasa kecewa atau frustrasi.

Daripada hanya mengandalkan daftar frasa atau kata kunci, sistem percakapan modern memanfaatkan model pembelajaran mesin berbasis bahasa alami untuk mengidentifikasi maksud pelanggan. Proses pelatihannya melibatkan pengumpulan berbagai contoh ucapan pelanggan yang dipasangkan dengan maksud yang benar. Ketika menerima ucapan baru, sistem akan memprediksi maksud yang paling relevan, disertai dengan tingkat kepercayaan pada setiap hasil. Sebagai ilustrasi, meskipun kalimat “Saya tidak mendapatkan sinyal saat pergi kerja” tidak ada dalam data pelatihan, sistem tetap dapat memahami bahwa pelanggan menghadapi masalah jaringan.

Sama seperti model pembelajaran mesin lainnya, akurasi dalam mengenali maksud pelanggan akan semakin tinggi apabila sistem dilatih dengan jumlah data yang lebih besar. Bagian berikut dari bab ini akan menguraikan metode dan strategi dalam menyusun kumpulan data ucapan yang efektif agar sistem percakapan dapat berfungsi lebih optimal.

Entitas

Walaupun intent berfungsi sebagai dasar percakapan dengan pelanggan, sering kali dibutuhkan pemahaman lebih detail dari apa yang mereka maksud. Misalnya, permintaan “Saya ingin paket TV dengan CNN” dan “Saya mau menonton ESPN” sama-sama bisa digolongkan ke dalam intent permintaan paket video. Namun, agar sistem dapat memberikan rekomendasi yang relevan apakah paket berita atau paket olahraga diperlukan informasi tambahan mengenai saluran TV yang diminta. Jika setiap saluran dipisahkan ke dalam intent sendiri, hasilnya justru tidak efektif karena kalimat pelanggan terlalu mirip untuk bisa dibedakan secara jelas.

Sebagai pendekatan lain, daripada membuat banyak intent khusus, kita cukup menggunakan satu intent umum untuk permintaan paket video, lalu mendeteksi saluran TV



yang disebutkan dalam ucapan pelanggan. Detail spesifik dalam ucapan ini disebut entities. Konsep entities mirip dengan parameter dalam sebuah fungsi, di mana ia berfungsi sebagai variabel yang membantu sistem memberikan respons yang tepat meskipun tindakannya serupa untuk berbagai permintaan pelanggan.

Berbeda dengan intent yang dilatih melalui contoh ucapan, entities biasanya ditentukan secara deklaratif. Artinya, sistem diberikan daftar entitas yang mungkin muncul, lengkap dengan sinonim atau variasi nilainya. Dalam beberapa platform, data pelatihan bisa menandai bagian tertentu dari kalimat pelanggan sebagai entity, misalnya pola “Saya ingin menonton [saluran],” di mana [saluran] sudah ditentukan dari daftar yang terbatas, seperti CNN, ESPN, berita lokal, atau komedi. Selain itu, sebuah ucapan pelanggan dapat memuat lebih dari satu entitas sekaligus.

Nada dan Emosi

Walaupun sistem percakapan yang dilengkapi dengan intent dan entities bisa berfungsi dengan baik, memberikan pengalaman pelanggan yang luar biasa memerlukan pendekatan tambahan. Salah satunya adalah mendeteksi nada atau emosi dalam ucapan pelanggan. Kini tersedia banyak model pelatihan yang dapat melakukan klasifikasi nada dan emosi, sehingga sistem mampu mengenali sentimen dari pernyataan pelanggan. Baik dalam bentuk teks langsung maupun transkripsi suara, model tersebut akan menghasilkan identifikasi nada tertentu beserta tingkat keyakinannya. Sebagai gambaran bagaimana model ini dibangun, IBM Watson Tone Analyzer dapat dijadikan contoh.

Dengan memanfaatkan kumpulan percakapan layanan pelanggan yang telah diberi label nada, dilakukan pelatihan model machine learning agar mampu memprediksi nada pada percakapan baru. Model ini menggunakan berbagai jenis fitur, seperti n-gram, kosakata dari beragam kamus, penggunaan tanda baca, serta adanya rujukan kepada orang kedua dalam kalimat. Algoritma yang dipilih adalah Support Vector Machine. Karena sekitar 30% sampel memiliki lebih dari satu nada sekaligus, masalah ini diperlakukan sebagai klasifikasi multilabel, bukan multikelas. Setiap nada dilatih secara terpisah menggunakan pendekatan One-Vs-Rest. Dalam tahap prediksi, sistem menetapkan nada tertentu apabila probabilitasnya minimal 0,5. Untuk nada-nada dengan data pelatihan yang tidak seimbang, penyesuaian bobot pada fungsi biaya dilakukan agar hasil pelatihan lebih optimal.

Selain melalui teks, nada dan emosi pelanggan juga bisa dikenali lewat analisis suara secara langsung. Teknik ini memanfaatkan aspek-aspek akustik seperti tinggi-rendah nada, karakter suara, volume, kecepatan berbicara, serta intonasi untuk mengidentifikasi kondisi emosional pembicara.

Dalam konteks layanan pelanggan, deteksi nada sangat penting karena dapat memengaruhi strategi penanganan. Jika pelanggan menunjukkan tanda-tanda frustrasi, sistem atau agen perlu menyesuaikan pendekatan, bahkan mungkin melakukan eskalasi agar pelanggan tetap merasa diperhatikan. Sebaliknya, nada juga bisa dipakai untuk menilai kemampuan agen dalam menunjukkan empati atau kesopanan. Pada sistem otomatis, analisis



nada memungkinkan pemilihan respons yang lebih tepat sesuai emosi pelanggan, misalnya mengucapkan “Terima kasih atas kesabaran Anda” kepada pelanggan yang kesal, atau “Kami senang Anda puas dengan layanan kami” kepada pelanggan yang bahagia. Sebagai contoh nyata, MetLife memanfaatkan ikon animasi yang memberikan umpan balik langsung kepada agen call center tentang nada pelanggan, dan strategi ini terbukti mampu meningkatkan kepuasan pelanggan serta memperpendek durasi panggilan.

Konteks

Selain memperhatikan nada percakapan, menambahkan informasi konteks tentang situasi pelanggan dapat membuat sistem lebih tepat dalam memberikan solusi. Konteks ini bisa berupa banyak hal. Sebagai contoh, dari riwayat komunikasi pelanggan dengan call center, sistem dapat mengenali apakah keluhan yang muncul berhubungan dengan masalah lama atau berulang. Konteks lain bisa berupa kondisi cuaca yang berpotensi mengganggu sinyal parabola. Semakin beragam dan lengkap konteks yang tersedia, semakin tinggi tingkat akurasi model pembelajaran mesin dalam memahami masalah dan menyajikan penyelesaian yang sesuai.

Untuk memperoleh informasi konteks, sistem percakapan umumnya harus terhubung dengan sistem eksternal. Namun, proses ini dapat memakan waktu, sehingga perlu dicari keseimbangan antara jumlah panggilan ke sistem luar dan kecepatan respons kepada pengguna. Salah satu solusi praktis adalah menyimpan sebagian konteks agar dapat digunakan kembali tanpa harus memanggil ulang.

Dialog

Alur percakapan antara sistem dan pelanggan, dari satu tahap ke tahap berikutnya, dikenal sebagai dialog. Elemen-elemen penting seperti intent, entitas, nada bicara, serta konteks akan memengaruhi bagaimana respons dipilih dan bagaimana sistem memperkirakan respons balik dari pengguna. Setiap sistem percakapan dapat mendefinisikan dialog dengan cara berbeda. Umumnya, dialog didasarkan pada kondisi tertentu terutama intent dan kemudian diperhalus dengan memanfaatkan entitas, nada, serta konteks. Ada sistem yang merancang dialog dengan model alur visual, ada pula yang memakai pendekatan state machine. Alternatif lain adalah hanya menyediakan API yang sesuai, sementara detail logika dialog dikembangkan sesuai kebutuhan oleh pihak operator.

Perancangan alur dialog sebaiknya melibatkan pakar pengalaman pengguna agar interaksi terasa lebih natural dan efektif. Sistem percakapan juga perlu memiliki karakter yang sesuai dengan konteks dukungan pelanggan, sekaligus menjaga nada komunikasi agar tetap tepat sasaran. Memberikan variasi respons untuk situasi yang sama dapat menghindarkan kesan monoton atau kaku. Sebaliknya, jika alur dialog dirancang terlalu kaku, hasilnya tidak akan jauh berbeda dengan menu IVR lama berbasis touch-tone. Untuk menciptakan percakapan yang lebih dinamis, sistem modern biasanya memanfaatkan model pembelajaran mesin guna menentukan tahapan dialog berikutnya.

Membuat Model Prediktif

Setelah sistem percakapan terbentuk, ada bentuk kecerdasan buatan lain yang dapat mendukung layanan pelanggan, yaitu predictive modeling. Pendekatan ini menjadi komponen penting karena memungkinkan sistem memprediksi perilaku pelanggan, seperti kemungkinan



berhenti menggunakan layanan, atau memberikan saran langkah selanjutnya yang paling sesuai. Model prediktif biasanya menghasilkan tiga jenis keluaran: keputusan biner (misalnya ya/tidak untuk churn), klasifikasi (misalnya rekomendasi produk yang tepat), atau nilai numerik (seperti estimasi penggunaan data bulanan melalui regresi).

Proses membangun model prediktif mirip dengan pelatihan intent pada sistem percakapan karena keduanya menggunakan pendekatan supervised learning. Dataset pelatihan biasanya memuat berbagai parameter yang berpotensi memengaruhi hasil, beserta label hasilnya. Sebagai contoh, dalam analisis customer churn, tiap baris data mewakili profil pelanggan dengan atribut tertentu, seperti jenis paket yang digunakan, lama menjadi pelanggan, rata-rata pengeluaran bulanan, hingga kode pos, serta status apakah pelanggan tersebut akhirnya berhenti berlangganan. Data tersebut kemudian diproses menggunakan alat predictive modeling, yang bisa secara otomatis menentukan algoritma terbaik atau memberi kebebasan bagi pengembang untuk memilih model yang dianggap paling sesuai.

Salah satu hal krusial dalam pengembangan model prediktif adalah tahap evaluasi. Biasanya, sebagian data historis dipisahkan khusus untuk menguji performa model, bukan untuk proses pelatihan. Setelah model selesai dibuat, data uji ini digunakan untuk melihat tingkat keakuratannya dalam memprediksi hasil yang diinginkan. Jika model menunjukkan akurasi rendah pada data evaluasi, berarti pengembang harus mengevaluasi ulang metode atau pendekatan yang digunakan dalam pembuatan model.

Jika model sudah mencapai akurasi yang memadai, langkah penerapan yang disarankan adalah menjadikannya sebagai microservice. Fungsi microservice ini adalah menerima input relevan dari sistem lain dan memberikan output berupa skor probabilitas dari hasil prediksi. Dengan pendekatan ini, model prediktif dapat dengan mudah diintegrasikan ke berbagai sistem, termasuk sistem percakapan. Contohnya, jika saat awal interaksi model memprediksi dengan tingkat keyakinan tinggi bahwa pelanggan berpotensi churn, maka alur percakapan bisa diarahkan untuk mengambil langkah-langkah pencegahan. Sepanjang interaksi, sistem percakapan akan menyesuaikan alurnya secara dinamis berdasarkan konteks dan status pelanggan.

Membuat Model Optimasi

Optimisasi merupakan salah satu penerapan AI yang penting dalam layanan pelanggan. Pada dasarnya, optimisasi bertujuan mencari hasil terbaik, apakah itu keuntungan maksimal, kualitas layanan paling baik, atau biaya serendah mungkin, dengan tetap memperhitungkan batasan tertentu seperti biaya dan ketersediaan sumber daya. Penerapannya sangat luas, mulai dari penempatan antena, perencanaan kapasitas jaringan, penugasan teknisi, hingga distribusi fungsi jaringan virtual (VNF) di berbagai wilayah. Dalam konteks layanan pelanggan, optimisasi bisa membantu menyusun jadwal kerja dan pembagian tugas agen agar kepuasan pelanggan meningkat. Misalnya, karena agen berpengalaman tidak mungkin menangani semua panggilan, sistem optimisasi dapat memberikan rekomendasi agar pelanggan dengan prioritas tertentu (berdasarkan risiko berpindah ke kompetitor, pengaruh sosial, atau nilai jangka panjang) ditangani oleh agen yang paling ahli.



Proses penyelesaian masalah optimisasi berlangsung melalui tahap pemodelan hingga pelaksanaan. Walaupun bentuk pemodelannya bisa berbeda sesuai perangkat yang digunakan, intinya tetap harus memuat batasan, prioritas, dan tujuan dalam format yang bisa dipahami sistem optimisasi. Selanjutnya, mesin optimisasi akan memproses model itu untuk menghasilkan solusi. Dalam penerapannya, selalu ada kompromi antara kecepatan waktu perhitungan dengan kualitas atau tingkat optimalitas solusi yang dicapai.

Integrasi dengan Sistem yang Ada

Sistem layanan pelanggan operator tidak bisa berjalan sendiri tanpa dukungan sistem lain. Integrasi dengan BSS dan OSS menjadi keharusan, baik untuk menyediakan data bagi model AI, melayani permintaan percakapan secara langsung, maupun menangani transaksi seperti pemesanan dan pembaruan akun. Untuk memastikan pengalaman pelanggan yang lancar, diperlukan rancangan integrasi yang hati-hati. Salah satu pendekatan terbaik adalah memanfaatkan microservices dari BSS/OSS serta menerapkan caching yang efektif, sehingga sistem layanan pelanggan tetap aman dari gangguan, terutama pada sistem lama yang masih mengalami downtime terjadwal.

Penggunaan Omni-Channel dari Layanan Pelanggan Terlatih

Sistem layanan pelanggan yang dibangun dengan arsitektur yang baik termasuk fitur percakapan, analitik prediksi, dan optimisasi bisa diterapkan di berbagai kanal interaksi dengan pelanggan. Untuk komunikasi berbasis teks, satu sistem percakapan inti dapat digunakan sekaligus di banyak platform, mulai dari media sosial, SMS, hingga aplikasi chat milik operator. Agar bisa berjalan, biasanya diperlukan lapisan adaptor pada tiap kanal untuk menyesuaikan panggilan API kanal tersebut dengan panggilan yang dipahami sistem percakapan utama.

Sistem percakapan tidak hanya terbatas pada teks, tetapi juga bisa diperluas ke layanan suara dengan bantuan voice gateway yang mengubah suara menjadi teks dan sebaliknya. Dengan begitu, sistem dapat mendukung agen virtual maupun agen pendamping. Meski demikian, pengalaman pengguna bisa berbeda tergantung jenis antarmukanya. Chat biasanya lebih ringkas dan penuh singkatan, terutama di perangkat seluler. Selain itu, chat juga bisa melibatkan pertukaran gambar, baik berupa foto dari pelanggan atau ilustrasi panduan dari agen, sesuatu yang tidak mungkin dilakukan di sistem berbasis suara.

Integrasi Layanan Pelanggan dengan Platform Sosial

Sistem percakapan tidak hanya berfungsi melayani pelanggan lewat media sosial, tetapi juga bisa memanfaatkan media sosial sebagai sumber informasi penting. Dari riwayat aktivitas sosial pelanggan, analisis dapat dilakukan untuk memahami karakter atau sentimen mereka. Jika data ini dihubungkan dengan sistem pemasaran, hasilnya bisa berupa rekomendasi produk yang mendorong penawaran tambahan atau peningkatan layanan selama interaksi. Contohnya, dengan izin pemakaian data media sosial, perusahaan bisa mengetahui keluhan pelanggan terhadap layanan data internasional. Informasi tersebut dapat dijadikan dasar untuk membuat paket data internasional yang lebih baik dan terjangkau, misalnya lewat kolaborasi dengan penyedia global. Berdasarkan riwayat pelanggan, paket baru itu bisa dipasarkan langsung kepada segmen yang sering bepergian ke luar negeri.



14.5 MENINGKATKAN LAYANAN PELANGGAN OPERATOR

Di dalam perusahaan telekomunikasi, hanya sedikit staf yang benar-benar ahli dalam berinteraksi dengan pelanggan atau memiliki wawasan mendalam tentang produk dan layanan. Mereka biasanya dikategorikan sebagai staf dukungan level 2 atau level 3. Sementara itu, mayoritas staf berada di level 1 yang berperan memberikan layanan umum kepada pelanggan. AI yang dilatih dengan pengetahuan para ahli dapat membantu membimbing staf level 1, sehingga beban staf ahli berkurang. Selain itu, AI ini bisa digunakan sebagai sarana pelatihan, baik bagi karyawan baru maupun karyawan lama ketika ada pembaruan. Dengan cara ini, AI mampu memperluas layanan pelanggan berkualitas tinggi dengan menyalurkan pengetahuan pakar kepada staf level 1.

Crowd-Sourcing untuk Pelatihan Layanan Kecerdasan Buatan

Memberikan layanan pelanggan yang unggul membutuhkan kombinasi berbagai keahlian mulai dari pengetahuan tentang pelanggan, cara berinteraksi, strategi menangani masalah, pemahaman produk, hingga pemasaran dan penjualan. Karena itu, siapa yang melatih sistem AI menjadi faktor krusial, sebab kualitas pelatihan menentukan seberapa besar kepercayaan pelanggan terhadap hasilnya. Dalam dunia AI, hal ini sering disebut dengan transparansi. Sama seperti open source yang berkembang lewat crowdsourcing, pelatihan AI terbaik juga sebaiknya melibatkan banyak pihak. Tidak cukup hanya satu orang, sebab tidak ada yang menguasai semua aspek layanan pelanggan. Yang dibutuhkan adalah kontribusi dari berbagai pakar baik agen layanan pelanggan berpengalaman, spesialis produk, maupun tim pemasaran dan penjualan. Dengan begitu, sistem AI dapat dilatih secara lebih komprehensif, akurat, dan bermanfaat. Forbes pernah menekankan bahwa seperti halnya kita mencari dokter atau pengacara karena kualifikasinya, kita juga harus mengetahui asal-usul pelatihan sistem kognitif, karena hal itu menentukan fungsinya. Crowdsourcing memungkinkan perusahaan mengumpulkan sumber daya terbaik lintas departemen untuk melatih AI, sehingga sistem lebih andal dan dipercaya. AI yang dilatih oleh para pakar terbukti memberikan hasil lebih baik, dan yang lebih penting, mampu menyebarkan keahlian para ahli ke semua agen, sehingga kualitas layanan pelanggan meningkat secara konsisten.

Pelatihan Berkelanjutan Lingkaran Tertutup untuk Layanan Pelanggan

Sistem AI pada tahap awal hanya memiliki pengetahuan terbatas, yang bersumber dari riwayat percakapan dan catatan pusat layanan pelanggan sebelumnya. Sistem ini mula-mula dilatih oleh para ahli. Namun setelah diimplementasikan, AI tidak hanya bergantung pada crowdsourcing dan data historis, melainkan juga harus mampu belajar secara terus-menerus. Konsep ini dikenal dengan closed-loop system, di mana setiap umpan balik dari interaksi pelanggan baru, perubahan preferensi, kondisi pasar, produk baru, maupun kampanye pemasaran akan memperkaya kemampuan sistem. Proses belajar ini bisa terjadi secara otomatis maupun dengan bimbingan, tergantung apakah sistem belajar dari data penggunaan atau dari data baru yang diberikan. Dengan adanya pembelajaran berkelanjutan ini, akurasi respons AI meningkat, keyakinan pada jawabannya lebih tinggi, dan hasil akhirnya adalah layanan pelanggan yang semakin efektif.



Memperluas Basis Pengetahuan

Sistem AI yang dibahas sebelumnya merupakan sistem yang sudah melalui proses pelatihan sehingga mampu menjawab pertanyaan tertentu dengan akurasi tinggi. Meski begitu, tidak semua pertanyaan bisa dijawab oleh sistem ini. Situasi tersebut dikenal sebagai *short tail*, yaitu kondisi ketika hanya tersedia sejumlah kasus pelanggan yang sudah diketahui bersama solusinya. Sebaliknya, *long tail* atau perluasan pengetahuan terjadi ketika informasi mentah atau tidak terstruktur diproses, diberi keterangan tambahan, lalu diperkaya untuk menemukan jawaban dari berbagai dokumen. Dengan cara ini, jika sistem AI tidak dapat memberikan jawaban langsung untuk suatu masalah atau pertanyaan, maka ia akan mengacu pada corpus yang juga telah dilatih. Melalui pemahaman bahasa alami (*Natural Language Understanding*), pencarian dipersempit hingga menghasilkan beberapa alternatif jawaban yang tingkat kepercayaannya tinggi.

Membangun Korpus

Untuk membangun corpus, seorang pakar (SME) terlebih dahulu menentukan sumber-sumber informasi, lalu memilih beberapa dokumen representatif untuk melatih sistem. Dokumen ini dilengkapi dengan metadata seperti sentimen, entitas, konsep, dan kata kunci. Jika diperlukan, model khusus juga bisa dibuat untuk mengatur ontologi. Setelah itu, sistem diuji untuk memastikan relevansi hasil, kemudian dilakukan serangkaian kueri berulang agar akurasi meningkat. Seiring dengan penggunaan terus-menerus, corpus menjadi semakin akurat. Ketika sistem sudah mampu memberikan jawaban yang tepat untuk kueri atau pertanyaan penelitian, hasil tersebut dapat dipindahkan ke kategori *short tail* atau sistem AI yang sudah sepenuhnya terlatih. IBM Watson sendiri menyarankan agar dilakukan pelatihan minimal dengan 49 kueri, dan bisa lebih banyak, tergantung kebutuhan.

Menetapkan Fondasi Model

Walaupun metadata standar dalam *natural language understanding* sudah mampu mencakup banyak jenis informasi, setiap industri tetap memiliki metadata dan hubungan spesifik yang berbeda. Karena itu, diperlukan model khusus yang disesuaikan untuk melatih dan membangun corpus. Model dasar tersebut kemudian dapat dimanfaatkan untuk melatih konten yang relevan dengan kebutuhan industri tertentu, terutama dalam konteks pengembangan pengetahuan lebih lanjut.

14.6 RINGKASAN

Sistem AI membutuhkan proses pelatihan yang dilakukan oleh para ahli dari berbagai disiplin. Setelah dilatih, sistem ini harus terus belajar agar semakin akurat dalam merespons kebutuhan pelanggan. AI dengan kemampuan pembelajaran berkelanjutan memungkinkan layanan ditingkatkan ke level yang lebih tinggi dengan meniru strategi para ahli terbaik. Tidak hanya itu, AI juga mampu menemukan langkah-langkah proaktif untuk meningkatkan kualitas produk maupun layanan. Dengan dukungan ini, pengambilan keputusan menjadi lebih baik, sehingga layanan pelanggan dapat berkembang dari sekadar baik menjadi luar biasa. Pada bab selanjutnya, akan dibahas bagaimana teknik AI ini bisa dimanfaatkan untuk membangun gambaran jaringan masa depan.



BAB 15

BATAS BARU AI UNTUK JARINGAN OTONOM

15.1 PENDAHULUAN

Jaringan telah mengalami transformasi besar dari jaringan kabel suara sederhana menjadi jaringan nirkabel seluler yang sangat terhubung. Batas berikutnya adalah jaringan otonom yang digerakkan oleh kecerdasan buatan (AI). Sejumlah teknologi akan terintegrasi secara mulus untuk memberikan nilai tambah bagi pengguna jaringan. Evolusi menuju jaringan otonom akan membutuhkan evolusi proses dan manusia. Banyak tantangan dalam transisi ini tidak hanya bersifat teknis tetapi juga sosial.

Jaringan otonom akan dibangun menggunakan sejumlah teknologi yang banyak di antaranya akan dibuat di komunitas sumber terbuka. Algoritma dan model pembelajaran mesin (ML) dan AI semakin canggih. Hal ini, dikombinasikan dengan kemajuan teknologi komputasi, terutama Unit Pemrosesan Grafis atau perangkat keras aplikasi, seperti Unit Pemrosesan Tensor, serta penyimpanan data yang jauh lebih cepat dan lebih murah, menciptakan peluang yang sangat baik untuk menerapkan AI di sejumlah area dalam jaringan, termasuk perencanaan, operasi, dan keamanan jaringan. Hal ini secara struktural akan berdampak pada jaringan operator dan perusahaan. Seperti halnya kemajuan teknologi lainnya, ada oknum-oknum jahat yang ingin memanfaatkan perangkat terbaru untuk tujuan yang tidak etis atau ilegal. Sangat penting untuk mengevaluasi secara cermat risiko keamanan baru yang ditimbulkan oleh kemajuan ini.

Jaringan yang ditentukan perangkat lunak (SDN) memungkinkan optimalisasi program jaringan. Jaringan SDN akan dirancang dengan fokus yang kuat pada rencana pengumpulan data jaringan yang terperinci. Kumpulan data ini tidak hanya memungkinkan analisis akar penyebab reaktif tetapi juga optimalisasi prediktif proaktif terhadap sumber daya dan kinerja. Platform Otomasi Jaringan Terbuka (ONAP) memungkinkan operator dan bahkan pengguna akhir untuk membuat layanan baru menggunakan portal. Terdapat antarmuka pemrograman aplikasi (API) terbuka untuk menghubungkan berbagai fungsi jaringan virtual (VNF) guna menciptakan layanan baru secara optimal dengan cepat. Teknologi blockchain akan menghadirkan efisiensi yang signifikan dengan sistem pencatatan terdistribusi. Potensi ini untuk mengubah pelacakan aset, kontrak, penagihan, dan kompensasi entitas secara fundamental secara efisien, bahkan untuk nilai tambah yang kecil.

Banyak dari kemajuan ini berasal dari inovasi pesat proyek sumber terbuka seperti OpenStack, ONAP, Kubernetes, Acumos, dan banyak lainnya. Standar de facto akan diciptakan di mana kontribusi akan terus mengembangkan solusi, tidak seperti perdebatan spesifikasi standar terperinci apriori di masa lalu. Organisasi yang memanfaatkan teknologi sumber terbuka secara efisien akan memiliki keunggulan kompetitif. Layanan mikro memungkinkan setiap layanan diimplementasikan secara optimal dan independen untuk memberikan fungsionalitas yang terfokus. Ini meminimalkan duplikasi pengembangan fungsionalitas yang sama.



Teknologi cloud terutama diterapkan di pusat data besar. Batasan berikutnya yang dimungkinkan oleh teknologi SDN/virtualisasi fungsi jaringan (NFV) adalah komputasi tepi (edge computing). Komputasi dan penyimpanan data bergerak lebih dekat ke pengguna akhir untuk aplikasi yang sangat sensitif terhadap latensi. Jaringan 5G akan memiliki bandwidth yang jauh lebih tinggi dan latensi yang jauh lebih rendah, yang akan sangat penting untuk kasus penggunaan seperti kendaraan otonom.

Terakhir, kemajuan teknologi ini akan mengharuskan orang-orang di industri untuk terus mempelajari keterampilan baru dan mengikuti perubahan paradigma dalam perencanaan, desain, operasi, dan keamanan jaringan. Regulasi juga harus berkembang untuk mengimbangi perubahan ini. Bagian 15.2 merangkum bagaimana jaringan telah bertransisi selama beberapa dekade terakhir. Bagian 15.3 menjelaskan bagaimana berbagai komponen berevolusi untuk memungkinkan jaringan otonom. Bagian selanjutnya membahas evolusi proses dan keterampilan yang dibutuhkan seiring transformasi jaringan dengan kemajuan teknologi. Bagian 15.7 menyajikan beberapa kasus penggunaan layanan jaringan otonom berbasis AI untuk menggambarkan bagaimana semuanya saling terhubung. Kami menutup buku ini dengan peluang bagi komunitas untuk memungkinkan transformasi tersebut.

15.2 FUNGSI DASAR JARINGAN

Jaringan didefinisikan sebagai sekelompok orang, tempat, atau perangkat yang saling terhubung. Jaringan biasanya digambarkan secara grafis sebagai sekumpulan simpul yang saling terhubung oleh tautan. Semakin banyak tautan antar simpul, semakin kaya interkoneksi dalam jaringan dan semakin tinggi nilai jaringan bagi para pesertanya.

Jaringan terjadi secara alami dan buatan manusia. Gambar 15.1 menunjukkan bagaimana sungai-sungai di dunia saling terhubung melalui laut dan samudra, memungkinkan orang untuk mengarungi perairan. "Jaringan" inilah yang menghubungkan wilayah-wilayah terpencil dan memunculkan perdagangan dunia. Jaringan yang kaya ini memungkinkan barang dan jasa diperdagangkan secara bebas, dan meningkatkan taraf hidup di seluruh dunia. Keterkaitannya yang kaya menjadikannya sangat berharga bagi para visioner awal yang melihat apa yang dapat dilakukannya, meskipun mungkin membutuhkan waktu berbulan-bulan untuk memindahkan barang-barang sederhana dari satu bagian dunia ke bagian lain.

Contoh di sini untuk "terhubung dengan kaya" adalah Terusan Panama. Terusan Panama adalah jalur air buatan manusia di Panama, dengan panjang sekitar 80 kilometer, yang menghubungkan Samudra Atlantik dengan Samudra Pasifik. Terusan ini, yang selesai dibangun pada awal tahun 1900an, merupakan jalur penghubung utama bagi perdagangan internasional. "Antarmuka" baru ini menghubungkan Timur dan Barat dengan erat dan menjadikan perdagangan lebih sederhana dan lebih berharga bagi para peserta jaringan ini.

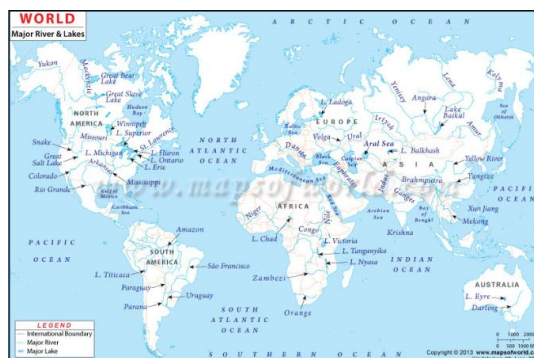
Tujuan jaringan komunikasi awal adalah untuk membuat dunia terasa lebih kecil dan memudahkan serta terjangkau untuk terhubung dengan orang-orang, pertama di wilayah sekitar dan kemudian di seluruh dunia. Teknologi yang tersedia saat itu belum memungkinkan koneksi otomatis ke siapa pun di seluruh dunia, sehingga orang-orang yang melakukannya.



Operator telepon memungkinkan orang-orang untuk terhubung di sekitar dan di seluruh dunia.

Gambar 15.2 menunjukkan operasi telepon awal yang memungkinkan jaringan telepon menyediakan koneksi di seluruh dunia bagi siapa pun yang memiliki telepon. Dampaknya sangat besar bagi kehidupan masyarakat dan perekonomian. Komunikasi langsung dapat terjadi hampir di mana pun di dunia. Jaringan telepon juga memiliki dampak politik, dengan "Hotline" Washington-Moskow pada tahun 1963 antara Gedung Putih dan Kremlin. Jaringan komunikasi modern memiliki miliaran koneksi setiap harinya, dan jutaan perangkat serta elemen jaringan diperlukan untuk memastikan koneksi tersebut terjadi secara lancar, tepat waktu, dan sangat andal. Panggilan telepon rumah bukan lagi satu-satunya alat komunikasi. Komunikasi seluler, SMS, media sosial, perutean paket protokol Internet (IP), termasuk video, dan banyak bentuk lainnya kini banyak digunakan dalam jaringan komunikasi. Seiring berkembangnya metode komunikasi, alat untuk melacak, memelihara, dan mengelola opsi-opsi baru ini pun berkembang.

Alat yang digunakan untuk mewujudkan hal tersebut merupakan alat mutakhir dalam ilmu data. Gambar 15.3 adalah contoh alat yang digunakan dalam AT&T dan tersedia dalam sumber terbuka yang disebut Nanocubes. Alat ini memungkinkan miliaran titik data dilihat dan dimanipulasi dalam peramban web standar.

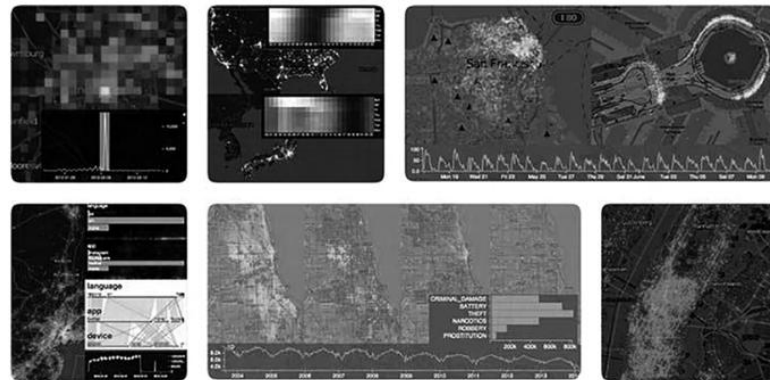


Gambar 15.1 Peta Sungai-sungai di Dunia.



GAMBAR 15.2 Pusat Pengalihan Telepon Internasional dari tahun 1940an dan 1950an.





GAMBAR 15.3 Contoh visualisasi orang, tempat, dan perangkat yang terhubung.

Grafik contoh menunjukkan peta panas geografi, koneksi perangkat, dan konsentrasi jenis perangkat dalam jaringan, tetapi penggunaannya sangat umum.

Pada intinya, jaringan adalah tentang koneksi. Jenis koneksi, baik perairan dunia, telepon, media sosial, ponsel, situs kencan, atau hubungan profesional, menentukan jenis jaringan, tetapi semuanya memiliki beberapa kesamaan. Semakin kaya koneksi dan semakin banyak partisipan dalam jaringan, semakin berharga jaringan tersebut, baik bagi produsen (pemilik jaringan) maupun konsumen (pelanggan). Atribut lain yang dimiliki masing-masing jaringan adalah semakin kompleks jaringan ini, semakin sulit untuk dikelola. Diperlukan alat baru. Seiring berkembangnya perangkat ilmu data yang memungkinkan kompleksitas jaringan modern tumbuh secara andal dan terjangkau, fase pertumbuhan berikutnya akan membutuhkan lebih banyak lagi. Di fase inilah AI dan ML akan berperan.

Evolusi Jaringan Komunikasi

Komunikasi dapat berlangsung satu arah (dari pengirim ke penerima), yang dapat dibatasi waktu (seperti sinyal radio) atau tak terbatas (seperti lukisan di dinding gua); komunikasi dapat berlangsung dua arah, seperti percakapan atau sinyal antara dua orang yang berdiri berdampingan, atau lebih jauh seperti sinyal asap, warna bendera, dan drum awal yang digunakan untuk menarik perhatian pada hal-hal penting yang memengaruhi masyarakat.

Komunikasi manusia dimulai ratusan ribu tahun yang lalu, dengan ucapan yang tak lebih dari sekadar gerutuan. Simbol menyusul sebagai cara untuk menyampaikan informasi yang lebih kaya yang sulit disampaikan melalui ucapan (sebuah gambar bernilai seribu kata, atau gerutuan). Seiring berkembangnya metode komunikasi ini, kebutuhan untuk mengirimkan kata-kata dan simbol tersebut melalui jarak yang lebih jauh pun meningkat. Istilah telekomunikasi berarti komunikasi jarak jauh dengan beberapa sarana akses, baik kabel, udara, atau bahkan asap.

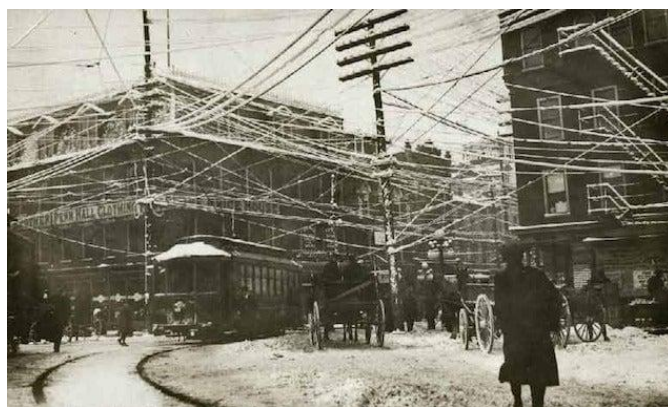
Telekomunikasi modern dimulai pada awal abad ke-20. Telegraf ditemukan di Inggris pada tahun 1830an oleh Sir Charles Wheatstone dan Sir William Fothergill Cooke, tetapi Samuel Morse-lah yang mengembangkan kode yang digunakan untuk menstandarisasi sinyal pada telegraf, yang memungkinkan penyebaran, pelatihan, dan penerimaan media dalam skala yang lebih luas. Kabel telegraf transatlantik pertama dikirim dan diantar pada tahun 1866, yang memungkinkan telekomunikasi transatlantik untuk pertama kalinya.



Telepon konvensional merupakan peningkatan besar bagi telegraf, karena orang-orang dapat mendengar suara satu sama lain, dan merasakan emosi manusia lain. Alexander Graham Bell menemukan telepon pada tahun 1876. Layanan telepon komersial pertama didirikan pada tahun 1878 dan 1879, masing-masing di Amerika Serikat dan di Inggris, di kota New Haven dan London. Perusahaan telepon lokal menjamur di seluruh Amerika Serikat. Seiring telepon menjadi semakin populer, hal ini segera menjadi tidak terkendali. Gambar 15.4 menunjukkan bagaimana kabel dengan cepat menguasai lanskap kota seiring berbagai perusahaan berusaha melayani setiap rumah.

Alexander Graham Bell mendirikan perusahaan American Bell Telephone pada tahun 1877, yang kemudian menjadi perusahaan *American Telephone and Telegraph* (AT&T) pada tahun 1899. Tujuannya adalah menyediakan layanan lokal jarak jauh (bel regional dan AT&T jarak jauh) dan memproduksi peralatan untuk itu, serta menciptakan teknologi baru apa pun yang dibutuhkan, semuanya di satu tempat, satu perusahaan. Struktur ini pada dasarnya bertahan hingga tahun 1984, ketika AT&T dipecah dan didivestasikan menjadi perusahaan jarak jauh dan "bayi" Bell dipisahkan menjadi perusahaan terpisah. Hingga tahun 1984, selain telepon noda sentuh dan fitur-fitur baru untuk bisnis, lanskap telekomunikasi relatif tidak berubah.

Pada saat divestasi tersebut, ada pekerjaan yang sedang dilakukan di Bell Labs untuk mengembangkan mode komunikasi baru yang disebut telepon seluler. Seluler adalah istilah yang digunakan karena adanya kesadaran bahwa frekuensi dapat digunakan kembali jika dipisahkan oleh jarak yang cukup dan ditempatkan pada kisi heksagonal (teselasi); oleh karena itu, lahirlah telepon seluler. Satu-satunya masalah adalah "handset" awal memenuhi bagasi mobil. Ketika ditanya berapa banyak yang bisa dijual, sebuah perusahaan konsultan terkenal menawarkan "10.000, paling banyak." Ini bukanlah skala yang tepat untuk AT&T; sehingga, para pionir awal mengadopsi karya dari Labs untuk memulai telepon seluler, dan AT&T membeli kembali bisnis tersebut dengan harga sekitar 13 triliun pada tahun 1994 melalui pembelian McCaw Cellular.



Gambar 15.4 Kabel telepon di Kota New York tahun 1880an.

Kira-kira dalam rentang waktu yang sama, teknologi penting lainnya mulai berkembang. Komputer pribadi, yang kebetulan juga dipasarkan pada tahun 1984, mulai



diadopsi secara luas di rumah-rumah; ini adalah perangkat data serbaguna yang saling terhubung melalui modem. Jaringan komputer pribadi dan papan buletin elektronik memungkinkan versi pertama Internet untuk dibayangkan. Komunikasi protokol internet juga berkembang pada saat itu, memungkinkan adanya protokol standar untuk menghubungkan semua komputer pribadi ini di rumah-rumah.

Bagian terakhir yang hilang adalah aplikasi IP yang setara dengan telepon: perangkat sederhana atau kemampuan untuk menyediakan konektivitas tanpa batas ke konten web. Peramban web menyediakan kemampuan ini. Oleh karena itu, badai penyelarasan teknologi yang sempurna yang menciptakan ledakan web dot.com telah terjadi. Semua teknologi ini dibutuhkan untuk mewujudkannya, dan dalam rentang waktu lima tahun atau kurang, semuanya bersatu untuk menghadirkan Internet ke dunia. Internet baru ini memberi tekanan pada jaringan komunikasi yang dibangun terutama untuk mengirimkan suara. Banyak orang mendapatkan saluran kedua tetapi masih banyak lagi yang dibutuhkan. Batasan fundamental laju bit yang dapat diberikan modem telepon telah tercapai. Jelas; cara membangun jaringan komunikasi harus berubah untuk menangani transisi teknologi ini.

Transisi Jaringan: Suara ke Data, Sirkuit ke Paket, Rumah ke Seluler

Ada banyak perubahan fundamental dalam desain dan penerapan jaringan komunikasi area luas. Sejak awal, Jaringan Telepon Umum dirancang untuk memberikan suara, tepatnya suara sirkuit-switched; itulah aplikasi utama untuk jaringan tersebut. Segala sesuatu yang lain bersifat sekunder. Seiring berkembangnya Internet, lalu lintas data menjadi semakin penting sebagai aplikasi yang membutuhkan status kelas satu.

Sekitar tahun 2000, tergantung negaranya, lalu lintas data melampaui lalu lintas suara. Laju pertumbuhan ini begitu pesat, pada tahun 2007, lalu lintas suara hanya mewakili 10% dari lalu lintas di jaringan publik. Pertumbuhan data ini memerlukan perubahan fundamental dalam cara jaringan dirancang. IP mengambil alih jaringan Layanan Telepon Biasa sirkuit-switched; Router menggantikan switch; dan mencari tahu cara melakukan kualitas layanan dan suara melalui IP (VoIP) menjadi penting, karena suara hanyalah SATU aplikasi pada jaringan data IP baru ini. Semua konversi ini harus dilakukan dengan cara yang tidak memengaruhi keandalan pelanggan, memungkinkan aplikasi lama dan baru untuk hidup berdampingan, dan menyediakan layanan yang hemat biaya untuk mendukung permintaan kecepatan data yang terus meningkat. Meskipun ini mungkin tampak sederhana dan mudah bagi banyak orang yang tidak terlibat dalam detailnya, sebenarnya dibutuhkan sejumlah besar keterampilan teknik dan operasional untuk melakukannya. Teknologi dan perkakas baru untuk jaringan IP diperlukan untuk membuat keandalannya cocok dengan jaringan suara (5–9 detik, atau 99,999% andal, sekitar lima menit waktu henti setahun). Versi awal router jauh dari keandalan ini, dan dibutuhkan rekayasa bersama dan pengerasan substansial dari versi awal untuk membuatnya "siap jaringan."

Untuk mengimbangi laju data baru yang dihasilkan jaringan IP, diperlukan perangkat ilmu data yang lebih skalabel. Seiring peningkatan lalu lintas sepuluh kali lipat atau lebih dari level sebelumnya, perangkat tersebut harus mencapai batas skala baru. Big data berakar dari telekomunikasi. Misalnya, saat tulisan ini dibuat (Mei 2018), terdapat lebih dari 200 Petabyte



yang mengalir melalui jaringan AT&T pada hari tertentu. Jika hanya 0,1% dari data tersebut yang diambil sampelnya untuk alasan manajemen dan kualitas jaringan, itu masih berarti lebih dari 200 Terabyte yang perlu diproses setiap hari; itu berarti hampir 2 miliar transaksi diproses per detik, dan terus tumbuh 30%–40% per tahun. Semakin banyak transaksi yang dapat ditangani secara otomatis atau yang dapat dipelajari selama pemrosesan, semakin murah pula pengelolaan jaringan tersebut. Di sinilah AI/ML akan berperan dalam manajemen jaringan.

Peralihan dari lalu lintas telepon rumah ke lalu lintas seluler merupakan peralihan mendasar lainnya. Lalu lintas seluler melampaui lalu lintas telepon rumah pada tahun 2009. Jenis lalu lintas memiliki profil yang serupa, baik telepon rumah maupun seluler, tetapi mekanisme akses dan profil investasinya sangat berbeda. Sekali lagi, jaringan seluler awalnya dibangun terutama untuk suara, dan harus direarsitektur ulang untuk data, seperti dalam kasus telepon rumah. Ini adalah prestasi rekayasa lain yang membutuhkan keunggulan operasional dan ketajaman teknis, karena pergeseran lalu lintas mendikte kebutuhan investasi yang berbeda. Perangkat dan teknologi yang baru dan lebih baik diperlukan dalam mobilitas, yang memungkinkan item seperti kualitas dan kapasitas airlin, penggunaan lokasi seluler, handoff lokasi seluler, dan metrik perangkat/handset untuk dipantau.

Saat kita memasuki era baru di mana lebih banyak "benda" akan berada di jaringan, baik seluler maupun telepon rumah, daripada manusia, teknologi harus mengambil lompatan lain. *Internet of Things* (IoT) harus dilayani dengan andal, tetapi dengan titik biaya yang lebih rendah dan skala yang lebih besar, mungkin hingga skala 10X. Efektivitas biaya dan skalabilitas seperti ini membutuhkan penggunaan AI/ML di mana metrik dapat ditangkap, diproses, dan ditindaklanjuti secara otomatis. Lebih penting lagi, metrik tersebut harus tangkas dan dapat dipelajari, sehingga perangkatnya perlu menjadi lebih cerdas setelah setiap tindakan. Di sinilah AI/ML akan berperan untuk memenuhi kebutuhan baru IoT ini. Seiring bertambahnya jumlah dan jenis mesin yang terhubung ke internet, kebutuhan akan lebih banyak alamat IP pun meningkat. Untungnya, IoT sudah cukup canggih untuk menjadi IPv6 asli, yang memungkinkan lebih dari 10^{38} perangkat unik yang terhubung. Beberapa pihak telah menyatakan kekhawatiran bahwa pengalamatan IPv6 terlalu unik, sehingga tidak memberikan privasi yang memadai. Gateway, NATing, firewall, dan teknologi lainnya membantu meringankan beberapa kekhawatiran tersebut.

Transisi Baru Menuju Internet of Everything

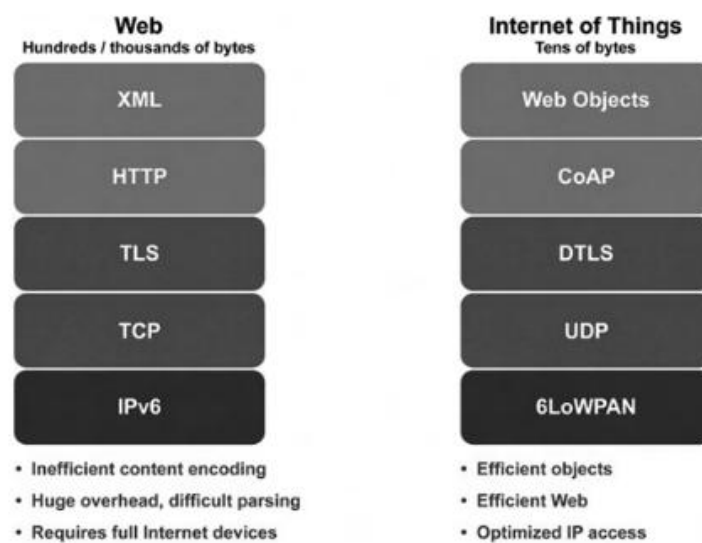
Seiring kita memasuki era di mana lebih banyak benda yang mungkin berada di jaringan daripada manusia, kita perlu menentukan apakah IP yang memunculkan internet masih merupakan pilihan terbaik. Telah terbukti sangat tahan lama selama 40 tahun keberadaannya dan lebih dari 20 tahun penggunaan di bidang jaringan area luas, tetapi apakah masih merupakan pilihan terbaik, atau sudah habis masanya?

Titik akhir yang menghubungkan orang ke internet dan perangkat tertanam yang terhubung ke internet memiliki kebutuhan yang sangat berbeda. Orang terhubung ke aplikasi, email, layanan web, halaman web, e-retail, video, dll. "Hal-hal" mengumpulkan, mengagregasi, memproses, dan menyampaikan informasi, data, serta metrik sistem yang digunakan untuk memantaunya. Meskipun protokol TCP/IP merupakan protokol *de facto* yang digunakan dalam



jaringan data saat ini, terdapat perbedaan penting yang diperlukan untuk memproses data bagi hal-hal yang terhubung ke internet.

Perangkat *Internet of Things* (IoT) berdaya rendah, biasanya perangkat nirkabel, dengan jangkauan terbatas. Gateway IoT biasanya diperlukan untuk koneksi ke Internet yang lebih luas. Perangkat IoT menggunakan HTTP/HTTPS untuk mengirimkan informasi data *eXtensible Mark-up Language* (XML) atau JavaScript Object Notation ke hub/gateway IoT, yang meneruskannya ke server web. HTTP/HTTPS adalah model klien-server utama yang digunakan untuk web. Fungsi keamanan dilakukan di server dan gateway IoT, karena beban keamanan untuk klien terlalu tinggi dalam hal daya pemrosesan pada perangkat IoT.



Gambar 15.5 Web versus IoT untuk Tumpukan Jaringan.

Extensible Messaging and Presence Protocol (XMPP) adalah contoh lain protokol yang dapat digunakan secara efektif untuk pengiriman data IoT. XMPP dimulai sebagai protokol pesan instan dan informasi kehadiran, tetapi berkembang menjadi VoIP, telepresensi, dan pengiriman data XML.

Ada pilihan lain, seperti *Constrained Application Protocol* (CoAP) dan *Message Queuing Telemetry Transport* (MQTT), tetapi memiliki skala atau ekstensibilitas terbatas untuk penerapan di area yang luas. Gambar 15.5 menunjukkan perbedaan tumpukan jaringan untuk Internet berbasis manusia versus berbasis benda. Meskipun saat ini, jumlah manusia melebihi jumlah benda di Internet, dan tidak ada kebutuhan mendesak untuk tumpukan jaringan baru, berdasarkan transisi yang telah dibahas sebelumnya, transisi teknologi ini kemungkinan akan terwujud dalam beberapa tahun mendatang.

15.3 MENYATUKAN SEMUANYA

Integrasi Teknologi untuk Jaringan Otonom

Sebagaimana komputer pribadi, peramban web, jaringan IP, konten web, modem, dan mesin pencari yang semuanya bersatu untuk menyediakan unsur-unsur yang diperlukan dan



memadai agar Internet dapat terwujud, badai teknologi serupa sedang terjadi di ranah jaringan. Badai teknologi baru ini memunculkan jaringan baru yang ditentukan perangkat lunak dan berbasis AI.

Jaringan yang ditentukan perangkat lunak membawa pemrograman jaringan, fleksibilitas, dan kecepatan penyebaran ke tingkat yang lebih tinggi. Virtualisasi fungsi jaringan memungkinkan perangkat keras umum off-the-shelf untuk menyediakan elemen jaringan yang sebelumnya hanya tersedia dalam kotak khusus yang terintegrasi; NFV memungkinkan perangkat keras dipisahkan dari aplikasi perangkat lunak. OpenStack, Kubernetes, ONAP, dan teknologi lainnya memungkinkan penyedia layanan komunikasi untuk membuat cloud jaringan yang sebenarnya, yang mampu memenuhi persyaratan throughput, latensi, jitter, dan kualitas layanan yang diperlukan untuk memberikan kinerja jaringan tingkat operator dari fungsi virtual ini.

Upaya sumber terbuka seperti ONAP dalam Linux Foundation menyediakan komunitas, lingkungan pengembangan kode, untuk platform operasi jaringan tingkat sistem. Banyak orang telah menunjukkan kemampuan DevOps loop tertutup di SDN; misalnya di ONAP, ada proyek, yang dikenal sebagai Platform Manajemen Otomasi Loop Tertutup, yang dapat membawa kemampuan DevOps loop tertutup ke tingkat berikutnya, dan di sinilah AI/ML berperan; Sistem ini dapat menciptakan sistem yang memungkinkan penggunaan loop tertutup tersebut untuk menginformasikan model ML secara offline dan menambahkan instans model secara in-line, menjadikannya bagian dari keputusan dalam loop tertutup, sambil tetap belajar secara offline.

Kini, sistem ini menyediakan sistem pembelajaran loop tertutup yang dapat mengubah keputusannya berdasarkan model ML. Alih-alih selalu membuat keputusan yang sama dalam loop tertutup, sistem ini dapat mengubah keputusannya berdasarkan informasi yang ada tentang parameter model, seperti waktu, kedekatan dengan jendela pemeliharaan, atau waktu sejak pembaruan perangkat lunak terakhir.

DevOps berbasis ML jenis ini hampir memasuki tahap produksi; akan segera ditingkatkan skalanya. Kemampuan ini akan transformatif bagi cara jaringan dijalankan. Kemampuan ini akan mendorong tingkat keandalan baru dengan efektivitas biaya yang lebih baik. Kemampuan ini akan memungkinkan penyedia layanan untuk menanamkan pembelajaran terbaik mereka ke dalam model-model ini dan memiliki cara sistematis untuk menerapkan pembelajaran tersebut di lapangan. Ini adalah masa yang unik dalam jaringan, menyaksikan semua teknologi ini bersatu, mentransformasi industri dan cara jaringan dibangun di masa mendatang. Hal ini juga menetapkan standar masuk yang lebih tinggi, karena penyedia layanan dengan keterampilan teknis terbaik akan membedakan diri dan menjauhkan diri dari penyedia layanan lain yang tidak memiliki keterampilan tersebut.

Inovasi Cepat dengan Modularitas, SDN, dan CI/CD

Teknologi kecerdasan buatan/pembelajaran mesin itu sendiri tidak akan memungkinkan inovasi cepat yang diperlukan untuk transisi berikutnya dalam jaringan. Sebagaimana bagian sebelumnya menjelaskan secara singkat teknologi lain yang dibutuhkan dalam badai sempurna bagi AI/ML untuk mendisrupsi industri jaringan, bagian ini akan



membahasnya lebih mendalam, merinci kerangka kerja yang tepat yang perlu diterapkan agar AI/ML efektif.

Teknologi pertama yang menjadi bagian dari kerangka kerja baru ini adalah membuat antarmuka ke fungsi/elemen jaringan terbuka. Dapat dikatakan, ini adalah atribut terpenting yang diperlukan untuk transisi ini. Tanpa antarmuka yang terbuka, data tidak dapat dikumpulkan, tidak ada data tidak ada wawasan, tidak ada yang dapat mendorong parameter dalam model ML. Tanpa antarmuka terbuka, semuanya akan berantakan. Langkah perusahaan skala web untuk membuat switch dan router mereka sendiri sebenarnya bermuara pada kebutuhan antarmuka terbuka untuk mendorong otomatisasi yang dibutuhkan agar dapat mengikuti skala bisnis mereka. Prinsip yang sama berlaku di sini prinsip keterbukaan.

Dengan keterbukaan menjadi langkah pertama yang diperlukan, mendasarkan SDN pada model adalah langkah berikutnya. Setiap layanan dan elemen jaringan yang unik tidak dapat menjadi upaya pengkodean khusus. Setiap fungsi jaringan dan setiap layanan harus diabstraksikan sebagai model sehingga penggunaan kembali dapat diterapkan untuk layanan atau aplikasi berikutnya. Pendekatan pembangunan "Lego Block" ini memastikan pengiriman layanan dan aplikasi yang cepat, namun kerangka kerja umum untuk abstraksi ini perlu ada. Model seperti *Yet Another Next Generation* (YANG) untuk layanan jaringan, *Topology and Orchestration Specification for Cloud Applications* (TOSCA) untuk abstraksi layanan, dan HEAT untuk templat cloud OpenStack menyediakan abstraksi pemodelan tersebut.

Tidak satu pun dari teknologi ini yang diterapkan di greenfield, yang berarti bahwa aplikasi dan layanan SDN baru harus terhubung ke sistem pendukung operasional (OSS) dan sistem pendukung bisnis (BSS) yang ada. Jika OSS/BSS tidak diurus, OSS/BSS yang ada dapat menjadi faktor pembatas dalam mewujudkan manfaat SDN dan cloud jaringan. Layanan mikro adalah teknologi yang membebaskan OSS/BSS yang ada, memungkinkannya menjadi lebih modular dan lebih gesit dalam beradaptasi dengan SDN. Layanan mikro memungkinkan aplikasi dibangun secara modular dari modul perangkat lunak yang lebih kecil (seperti kernel). Setiap modul mendukung fungsionalitas tertentu dan menggunakan API yang terdefinisi dengan baik untuk berkomunikasi dengan layanan modular lainnya. Arsitektur layanan mikro adalah varian dari arsitektur berorientasi layanan, yang menciptakan aplikasi baru yang terdiri dari kumpulan layanan yang terhubung secara longgar. Manfaat menguraikan aplikasi yang lebih besar menjadi modul yang lebih kecil, seperti kernel, adalah sebagai berikut:

1. Meningkatkan modularitas.
2. Membuat aplikasi lebih mudah dipahami, dikembangkan, dan diuji.
3. Meningkatkan penggunaan kembali di kemudian hari.
4. Memungkinkan tim yang berbeda dalam proyek yang sama untuk bekerja lebih mandiri.
5. Memungkinkan pengiriman dan penerapan berkelanjutan.

Setiap sistem pengiriman perangkat lunak yang ingin mengoptimalkan kecepatan inovasi harus didasarkan pada integrasi berkelanjutan dan pengiriman berkelanjutan (CI/CD). Integrasi berkelanjutan adalah teknik pengembangan perangkat lunak di mana pembaruan kecil pada pohon kode ditambahkan secara berkala, sehingga deteksi dan perbaikan kesalahan lebih



mudah diperbaiki. Tujuan lingkungan CI adalah untuk mendapatkan pemahaman cepat tentang kode yang berfungsi dan menyederhanakan proses debugging, terutama antar tim pengembangan yang berbeda. Perangkat lunak integrasi berkelanjutan mengotomatiskan proses dokumentasi perangkat lunak dan menguji kode secara berkelanjutan dengan cara yang serupa. "Berkelanjutan" dapat berarti setiap hari untuk beberapa tim, atau setiap jam, tergantung pada kode dan tim yang terlibat. Integrasi berkelanjutan juga memungkinkan manajer proyek untuk memiliki gambaran yang lebih baik tentang kemajuan penyelesaian kode, karena komitmen dan pengujian terjadi begitu cepat.

Pengiriman berkelanjutan merupakan perluasan dari CI. Sementara CI menangani bagian buat/bangun/uji dari siklus pengembangan perangkat lunak, CD berfokus pada validasi komitmen sebagai kandidat kode yang berfungsi dengan baik untuk rilis tertentu. Tujuan penting CD adalah untuk membuat waktu antar penerapan yang berbeda sesingkat mungkin. Putaran umpan balik yang pendek memungkinkan pengujian untuk menemukan bug lebih cepat dan membuat penentuan masalah tersebut lebih mudah bagi pengembang. Jika CD sedang berlangsung dan pengujian dilakukan lebih awal dan sering sebuah konsep yang disebut "shift left" pengembang dapat mengerjakan perbaikan sebelum mereka beralih ke aspek lain dari proyek pengembangan. Hal ini dapat membantu meningkatkan produktivitas dan mempercepat pengiriman, karena kode dan perubahannya masih segar dalam ingatan pengembang, dan mereka belum sepenuhnya beralih ke bagian kode lainnya.

Bagaimana Kecerdasan Buatan Membuat Transisi Jaringan Berbeda

Dengan kerangka kerja teknologi yang dijelaskan di bagian sebelumnya, pertanyaannya adalah bagaimana AI/ML dapat memfasilitasi transisi jaringan untuk SDN dan menghadirkan teknologi tersebut secara hemat biaya dengan pendekatan DevOps, yang membuka jalan bagi transisi yang andal dan hemat biaya ke IoT.

Seiring IoT semakin nyata dengan penerapan awal, pertanyaannya adalah "bagaimana membuatnya berskala?" Perangkat Internet of Things menghasilkan banyak data, data yang beragam, dari berbagai perangkat: kamera, sensor, mobil, perangkat medis, perangkat industri, dll. Data ini dapat digunakan untuk melayani masyarakat dengan lebih baik (misalnya, jalan yang lebih baik dengan lalu lintas yang lebih sedikit, rumah yang lebih aman), untuk melayani satu orang dengan lebih baik (misalnya, pemberian insulin otomatis yang optimal), atau untuk melayani bisnis dengan lebih baik (misalnya, pemeliharaan prediktif pada mesin). Data yang disediakan dari perangkat IoT ini memberikan wawasan tentang kinerja operasional dan pemadaman; masing-masing memberikan peluang untuk belajar dan membuat keputusan yang lebih baik di masa mendatang.

Tantangan rekayasa yang harus dipecahkan adalah bagaimana menganalisis dan bereaksi terhadap semua data ini dengan tujuan meningkatkan kinerja, menyediakan layanan yang lebih baik, mendapatkan informasi yang lebih lengkap tentang situasi, dan melakukannya dengan cara yang dapat ditingkatkan secara hemat biaya. Tidak cukup banyak orang untuk mengatasi masalah ini, meskipun itu hemat biaya, yang sebenarnya tidak. Sebaliknya, diperlukan sistem yang dapat belajar dari data itu sendiri, sambil mengumpulkan, menganalisis, dan bereaksi secara real-time; inilah sistem AI/ML.



Seiring dengan meningkatnya skala data dan perangkat, tanpa sistem AI/ML, terdapat risiko bahwa potensi untuk membuat jalan lebih aman, mengelola kesehatan dengan lebih baik, dan meningkatkan efisiensi bisnis tidak tercapai karena analisis yang buruk dan keputusan yang buruk. Hal ini perlu dilakukan dengan benar; taruhannya terlalu tinggi.

Pembelajaran mesin dapat melihat triliunan titik data dari miliaran perangkat dan hanya bereaksi terhadap yang membutuhkan perhatian. Menentukan apa yang "bermakna" akan bergantung pada perangkat, lingkungan, situasi, waktu, dll. Di sinilah model ML yang tepat berperan, dan di sinilah perhatian rekayasa perlu difokuskan. Baik aplikasinya ritel, perawatan kesehatan, manufaktur, atau otomatisasi jaringan, kemampuan untuk menganalisis data waktu nyata dari miliaran sumber dan hanya bereaksi terhadap yang "penting" adalah yang dibutuhkan; model membantu menentukan apa yang berada dalam jangkauan dan apa yang berada di luar tingkat operasi normal. Kesamaan, korelasi, dan abnormalitas harus diidentifikasi dan ditanggapi berdasarkan data streaming waktu nyata. Realisasinya di sini adalah bahwa janji IoT terletak pada kemampuannya untuk mendapatkan wawasan dalam aliran data yang besar dan mengalir yang dihabiskannya.

Kecerdasan buatan/pembelajaran mesin memiliki kemampuan untuk memitigasi tren dalam jaringan saat ini pertumbuhan kompleksitas. Meningkatnya tingkat kompleksitas jaringan ini menimbulkan peningkatan biaya operasional. Seiring bertambahnya jumlah teknologi nirkabel (4G, 5G, mmWave, LAA, Wi-Fi, dll.) dan jumlah pita frekuensi yang digunakan, kompleksitas pengoperasian dan pemeliharaan jaringan ini serta memenuhi target kinerja menjadi semakin sulit. Biaya operasional perlu diturunkan seiring melambatnya pertumbuhan pendapatan. AI/ML berpotensi membantu operator menyediakan layanan pada tingkat keandalan yang diharapkan, tetapi dengan biaya yang jauh lebih rendah.

Di AT&T dan AT&T Labs, para peneliti menggunakan AI/ML untuk menangani tugas-tugas yang lebih kompleks, seperti mengoptimalkan peluncuran jaringan 5G mereka. Menara seluler tradisional biasanya ditempatkan di dekat pusat kota dan membentuk jaringan yang tidak sempurna, yang menyebabkan kesenjangan jangkauan. Menara seluler tradisional juga mahal untuk dibangun dan dirawat, serta menimbulkan tantangan terkait kepemilikan properti dan real estat. Sel kecil lebih murah dan lebih ringkas, serta dapat dipasang di gedung-gedung dalam kota dengan jaringan yang jauh lebih baik. Peran mereka adalah mengulang sinyal dari menara seluler utama agar lebih dekat dengan pengguna akhir. Dengan menganalisis data pelanggan seluler, sistem AI/ML yang terkalibrasi dengan baik membantu menemukan lokasi ideal untuk membangun sel kecil dan memastikan kekuatan sinyal 5G yang maksimal bagi pelanggan. Infrastruktur merupakan investasi yang sangat besar, bahkan untuk sel kecil sekalipun, sehingga pemodelan tren dan pertumbuhan penggunaan yang akurat merupakan kunci keberhasilan. Tren demografis dapat menyebabkan area yang sebelumnya kurang dimanfaatkan tiba-tiba menjadi generator trafik yang tinggi. Meskipun model statistik berguna untuk mengidentifikasi tren pergerakan pelanggan secara keseluruhan, teknik AI dan ML menghasilkan proyeksi masa depan berdasarkan data terkini.

Seperti yang dikatakan oleh Chris Volinsky, Asisten Wakil Presiden (AVP) Riset Big Data di AT&T Labs, "video menyumbang lebih dari separuh lalu lintas seluler kita. Lalu lintas video



tumbuh lebih dari 75% dan ponsel pintar menyumbang hampir 75% lalu lintas data kita hanya pada tahun 2016. Kami memperkirakan pertumbuhan lalu lintas video akan melampaui pertumbuhan data secara keseluruhan pada tahun 2020. Kita perlu memvisualisasikan miliaran titik data secara spasiotemporal.” Tidak ada alat yang dapat melakukan tugas ini, sehingga orang-orang di tim Volinsky membangun dan membuat alat visualisasi sumber terbuka, seperti Nanocubes, alat visualisasi data yang dapat memetakan jutaan koneksi ponsel individual dan perangkat yang terhubung ke menara telepon seluler. Umumnya alat ini dicontohkan dengan penggunaannya di luar perusahaan untuk mengkarakterisasi penggemar olahraga secara real-time dan untuk menganalisis tingkat dan riwayat kejahatan.

Volinsky, yang memimpin tim pemenang penghargaan Rekomendasi Netflix pada tahun 2009, lebih lanjut menyatakan, “Algoritma dan perangkat bukanlah hambatan dalam hal pemecahan masalah. Tantangannya terletak pada data dan alur data. Pendekatan AI modern yang haus data membutuhkan sumber data terpusat, tetapi mengumpulkannya dari berbagai jaringan dengan standar yang unik bukanlah tugas yang mudah. Tidak ada pakar dunia dalam pengolahan data [pengumpulan, pengurutan, verifikasi kualitas, dll.]. Untuk berhasil, Anda harus memahami secara organisasi cara mengakses data dalam silo yang berbeda, secara teknis cara mengintegrasikannya, dan memastikan formatnya selaras. Ini bukan hal yang dipelajari orang di sekolah pascasarjana.”

15.4 PILAR TRANSFORMASI: TEKNOLOGI, PROSES, DAN MANUSIA

Kemajuan dalam teknologi komputasi dan jaringan telah memungkinkan masyarakat yang terhubung di seluruh dunia. Populasi global mendapatkan manfaat dari konektivitas yang disediakan oleh jaringan nirkabel. Di banyak negara terbelakang dan miskin di seluruh dunia, konektivitas telah memberikan peluang ekonomi bagi ratusan juta orang yang bahkan tidak terbayangkan beberapa tahun lalu. Terdapat berbagai solusi perbankan seluler yang memungkinkan bisnis baru dan pinjaman mikro di mana perbankan tradisional tidak tersedia. Internet menyediakan akses ke materi pendidikan dari lembaga akademik dan perusahaan swasta yang memungkinkan peluang pendidikan kelas dunia bagi penduduk di pelosok dunia.

Dalam ekonomi maju, internet dan media sosial yang dikombinasikan dengan digitalisasi proses bisnis mendisrupsi banyak industri dengan model bisnis baru, misalnya, transportasi Uber, perhotelan Airbnb, dan banyak lainnya. Proses dalam berbagai aspek organisasi dan bisnis perlu berkembang untuk mendapatkan manfaat dari kemajuan teknologi dan tidak tertinggal oleh pesaing yang mengadopsi dan beradaptasi dengan teknologi baru.

15.5 EVOLUSI PROSES DI INDUSTRI TELEKOMUNIKASI

Industri telekomunikasi telah menjadi pilar pertumbuhan ekonomi selama lebih dari satu abad. Industri ini mengembangkan serangkaian proses yang tangguh untuk menyediakan layanan yang sangat andal. Industri telekomunikasi telah melakukan pekerjaan yang sangat baik dalam membangun proses untuk mengelola kompleksitas. Umumnya, produk telekomunikasi dirancang untuk memenuhi tujuan waktu penghentian layanan maksimal kurang dari lima menit dalam setahun, atau persyaratan keandalan lima sembilan (99,999%).



Pendekatan ini menghasilkan jaringan yang sangat andal selama bertahun-tahun; namun, pendekatan ini juga menciptakan budaya yang sangat menghindari risiko dalam industri telekomunikasi.

Industri ini didominasi oleh beberapa pemain utama; masa depan industri ini sebagian besar dikendalikan atau setidaknya dikelola oleh beberapa pemain dominan ini. Siklus pengembangan produk yang umum dimulai dengan serangkaian persyaratan yang terdefinisi dengan baik yang memprediksi keadaan teknologi dan persyaratan pelanggan tiga hingga lima tahun ke depan dengan tingkat keyakinan yang tinggi. Proses pendefinisian persyaratan berlangsung berbulan-bulan. Setelah persyaratan dikunci, terdapat siklus pengembangan produk selama berbulan-bulan yang diikuti oleh pengujian fungsional dan regresi selama berbulan-bulan. Dari konsep hingga pengenalan produk membutuhkan waktu 20–36 bulan atau bahkan lebih dalam beberapa kasus. Terdapat fleksibilitas yang terbatas untuk mengubah persyaratan atau proses selama siklus tersebut. Budaya menghindari risiko dibangun dengan prinsip "jangan gagal".

Dunia TI memanfaatkan konvergensi beberapa perangkat keras dan sistem operasi standar. Hal ini memungkinkan konsep CI/CD. Industri TI mengembangkan serangkaian proses untuk meningkatkan produktivitas dengan memanfaatkan hal ini. Sejumlah tugas dan proses berulang dihilangkan atau diotomatisasi untuk meningkatkan efisiensi dan kecepatan manajemen siklus hidup produk. Misalnya, di pusat data besar, hanya ada satu administrator sistem untuk ribuan server dibandingkan dengan satu administrator sistem untuk puluhan server di lingkungan TI tradisional; kemampuan perusahaan web untuk memperkenalkan fitur atau layanan baru beberapa kali sehari.

Di dunia yang bergerak cepat dan dinamis saat ini, perubahan mendasar pada budaya "fast fail" diperlukan untuk mengimplementasikan dan mengadaptasi solusi jaringan secara mendekati waktu nyata (hampir real-time). SDN/NFV memungkinkan disagregasi produk terintegrasi, yang memungkinkan perubahan lebih cepat pada fungsi dan layanan jaringan. Hal ini juga memberi penyedia layanan kemampuan untuk memperkenalkan produk dan layanan baru yang dapat dikelola oleh pelanggan akhir. Misalnya, AT&T memperkenalkan layanan Network on Demand di mana pengguna akhir dapat mengonfigurasi bandwidth yang mereka inginkan secara real-time. Layanan ini memanfaatkan SDN/NFV dan otomatisasi proses bisnis.

Transformasi jaringan akan berdampak pada berbagai proses, mulai dari perencanaan dan perancangan produk dan layanan hingga perancangan, pembangunan, dan pengoperasian jaringan. Tim desain produk akan memiliki akses ke data pelanggan yang luas. Data ini, ditambah dengan kumpulan data lainnya, akan memungkinkan manajer produk untuk mempersonalisasi layanan dan harga bagi masing-masing individu. Industri ini akan beralih dari paket yang dapat disesuaikan ukurannya menjadi layanan telekomunikasi yang dipersonalisasi, yang akan memiliki layanan konektivitas dasar yang terintegrasi dengan berbagai layanan bernilai tambah lainnya, seperti keamanan, pemantauan rumah, rumah pintar, IoT industri, dan banyak lagi layanan yang dipersonalisasi untuk setiap pelanggan.

Untuk memberikan layanan yang dipersonalisasi yang dapat dioptimalkan dan dikonfigurasi oleh pengguna akhir secara hampir real-time, diperlukan BSS yang gesit dengan



API northbound untuk memungkinkan integrasi yang mulus dengan sistem pelanggan akhir. Jaringan yang mendukung SDN/NFV akan mengonfigurasi jaringan secara terprogram untuk memberikan layanan ini. Bagi pelanggan akhir, jaringan akan tersedia dengan lancar.

Di masa lalu, pesanan layanan jaringan dibagi menjadi beberapa perintah kerja yang didistribusikan di antara banyak kelompok kerja. Terdapat sejumlah langkah manual dan entri data untuk menyelesaikan pesanan layanan. Penyelesaian pesanan akan memakan waktu sehari-hari atau berminggu-minggu. Proses yang menyebabkan gesekan sehingga mengurangi efisiensi dan kecepatan akan diotomatisasi dalam jaringan SDN/NFV. Durasi penyelesaian pesanan akan berkurang dari hitungan hari menjadi menit atau lebih baik. Pengiriman layanan akan menjadi proses yang sangat otomatis dan tanpa gesekan.

Jaringan telah mengalami otomatisasi, mulai dari peralihan mekanis ke peralihan elektro-mekanis, peralihan digital, jaringan optik yang dapat memulihkan diri sendiri, dan jaringan yang dioptimalkan sendiri. Namun, sebagian besar solusi otomatisasi ini dicapai menggunakan aturan statis yang telah ditentukan sebelumnya. Dengan hadirnya SDN/NFV dan kemampuan pemantauan serta analisis data otomatis terkait, beserta teknik ML dan AI, operator jaringan akan memiliki kemampuan untuk secara mandiri menggali lebih dalam ke node jaringan guna memantau dan mengoptimalkan kinerja jaringan serta jaminan layanan.

15.6 DAMPAK TERHADAP KARYAWAN DI INDUSTRI INI

Dampak evolusi teknologi terhadap masyarakat selalu pahit sekaligus manis. Kemajuan teknologi menciptakan peluang baru yang menarik dan membuka batas baru bagi umat manusia. Mesin berbahan bakar dan mobil memungkinkan orang bepergian jauh lebih nyaman dengan kecepatan yang jauh lebih tinggi daripada kereta kuda. Meskipun menciptakan sejumlah peluang ekonomi baru dan memungkinkan munculnya industri-industri baru, hal itu mengganggu bisnis transportasi yang dibangun di sekitar industri kereta kuda para pengemudi kereta kuda, produsen, dan orang-orang di semua kegiatan ekonomi pendukung lainnya dalam industri tersebut.

Evolusi teknis di bidang AI tidak berbeda. Evolusi ini akan menciptakan peluang ekonomi baru di banyak bidang, dan akan menghilangkan lapangan kerja di bidang-bidang tertentu, sama seperti evolusi teknis lainnya. Laju perubahan mungkin jauh lebih cepat dan besarnya perubahan mungkin jauh lebih besar, tetapi evolusi ini pada dasarnya tidak berbeda.

Kemajuan jaringan telekomunikasi sungguh merupakan keajaiban selama beberapa dekade terakhir. Bilik telepon umum tersedia di hampir setiap sudut kota. Telepon seluler adalah barang mewah yang hanya mampu dibeli oleh para eksekutif perusahaan senior dan orang-orang kaya. Ponsel genggam pertama berukuran sebesar batu bata. Sebelumnya, telepon seluler harus dipasang di bagasi mobil (Gambar 15.6). Telepon seluler generasi pertama berbasis teknologi analog.





Gambar 15.6 1946 Bell System memperkenalkan layanan telepon seluler komersial pertama, yang disebut Sistem Telepon Seluler (MTS).

Dalam beberapa tahun, telepon seluler menjadi kecil dan cukup ringan sehingga dapat dibawa di saku. Telepon seluler generasi kedua (2G) yang berbasis pada Global System for Mobile dan teknologi *Code Division Multiple Access* (CDMA) diperkenalkan. Standar-standar baru ini membuat keterampilan jaringan seluler generasi pertama menjadi kurang berharga. Hal ini diikuti oleh teknologi generasi ketiga (3G) dan generasi keempat (4G) dengan durasi siklus pengenalan teknologi yang semakin pendek untuk setiap generasi baru. Terjadi perubahan signifikan dalam teknologi antarmuka udara, dan evolusi dari circuit switching menjadi packet switching untuk data seluler. Para pemangku kepentingan harus mengikuti perkembangan teknologi agar relevan dalam industri ini. Pembelajaran berkelanjutan selalu dibutuhkan dalam teknologi dan menjadi semakin penting dalam beberapa tahun terakhir.

Seiring dengan diperkenalkannya kemajuan baru dalam industri telekomunikasi, keterampilan yang dibutuhkan pasti akan berubah. Orang-orang perlu memperoleh keterampilan baru. Keluar dari zona nyaman merupakan tantangan tersendiri. Namun, tidak bijaksana untuk menunggu perubahan terjadi sebelum mengambil tindakan. Pembelajaran berkelanjutan untuk mendiversifikasi keahlian merupakan suatu keharusan. Seseorang perlu memiliki pemahaman yang lebih luas tentang bisnis dan teknologi, alih-alih fokus yang sempit di satu bidang. Banyak organisasi telah memulai program untuk meningkatkan keterampilan tenaga kerja mereka. Terlepas dari peran dan posisi seseorang dalam suatu organisasi, seseorang harus berupaya untuk terus memperoleh keterampilan baru.

Industri telekomunikasi telah memiliki proses standarisasi yang terdefinisi dengan baik yang dipimpin oleh beberapa organisasi standarisasi seperti 3GPP (*Third Generation Partnership Project*). Meskipun sebelumnya telah memberikan manfaat yang baik bagi industri, kolaborasi sumber terbuka telah mampu mendorong inovasi lebih cepat di industri ini. Kemampuan untuk berpartisipasi, memanfaatkan solusi sumber terbuka, dan berkontribusi pada komunitas sumber terbuka akan menjadi keterampilan yang berharga bagi orang-orang di industri teknologi dan telekomunikasi. Solusi terbuka akan terus berkembang



pesat. Memanfaatkan dan mengembangkan solusi terbuka secara efisien akan menjadi keunggulan kompetitif.

Seiring AI semakin terintegrasi di seluruh industri, keterampilan seputar ML, analisis data, dan wawasan bisnis akan menjadi semakin berharga. Akan ada banyak alat, algoritma, dan data yang tersedia; kemampuan untuk membangun solusi secara efektif dan efisien guna menciptakan nilai ekonomi akan menentukan pemenangnya. Untuk memulai perjalanan pembelajaran seumur hidup yang mengasyikkan ini:

1. Rangkul perubahan
2. Pahami implementasi teknologi
3. Evaluasi keterampilan Anda saat ini
4. Identifikasi apa yang ingin Anda lakukan
5. Ambil tindakan untuk meningkatkan keterampilan

Tantangan Utama Regulasi dan Privasi

Sejumlah teknologi, komputasi, konektivitas, sensor, dan analitik data, secara bersamaan telah mengalami peningkatan kinerja dan pengurangan biaya yang signifikan selama beberapa tahun terakhir. Peningkatan teknologi ini telah memungkinkan model bisnis baru di berbagai industri. Data yang dihasilkan oleh jejaring sosial dan sensor di berbagai industri telah menciptakan sumber informasi yang berharga. Sekumpulan data berkualitas tinggi merupakan hal mendasar bagi keberhasilan AI. Banyak algoritma AI telah ada sejak lama. Penggunaan praktis algoritma AI terbatas karena akses ke data akibat tingginya biaya produksi, penyimpanan, dan pemrosesan data. Baik teknologi maupun regulasi telah menjadi hambatan bagi ketersediaan data di masa lalu.

Jejaring sosial, data yang dihasilkan sensor, dan digitalisasi berbagai catatan pribadi dan literatur telah menciptakan jumlah data yang belum pernah terjadi sebelumnya. Hal ini, dikombinasikan dengan daya komputasi dan solusi penyimpanan data berdensitas tinggi dan berbiaya rendah dengan konektivitas berkecepatan tinggi, dapat menyediakan akses ke data di seluruh organisasi dan negara. Teknologi yang dibutuhkan untuk menyediakan akses universal ke data sudah ada saat ini. Tantangan yang lebih besar adalah kepatuhan terhadap persyaratan keamanan, peraturan, dan hukum. Misalnya, rekam medis sedang didigitalkan di AS. Akses yang luas ke rekam medis sangat berharga dalam penemuan obat baru dan pengoptimalan penyembuhan untuk banyak penyakit. Namun, terdapat peraturan privasi yang ketat untuk membatasi akses ke rekam medis individual. Seiring semakin banyaknya data yang tersedia, keputusan sulit harus dibuat untuk menyeimbangkan manfaat dan risiko berbagi data agar memberikan manfaat maksimal bagi masyarakat.

Banyak peraturan dan undang-undang yang statis dan lambat beradaptasi dengan kemajuan teknologi yang pesat. AI akan menciptakan serangkaian masalah baru yang harus ditangani, misalnya, jika mobil otonom menyebabkan kecelakaan karena kegagalan algoritma, lalu siapa yang akan bertanggung jawab? Pemilik, produsen mobil, pengembang perangkat lunak. Contoh lain, jika sebuah bot otomatis mulai merekam percakapan secara otomatis, yang mungkin ilegal menurut hukum suatu negara, bagaimana masyarakat mengatasinya? Selain itu, akan terjadi evolusi berkelanjutan dalam jaringan dan industri otonom yang menuntut



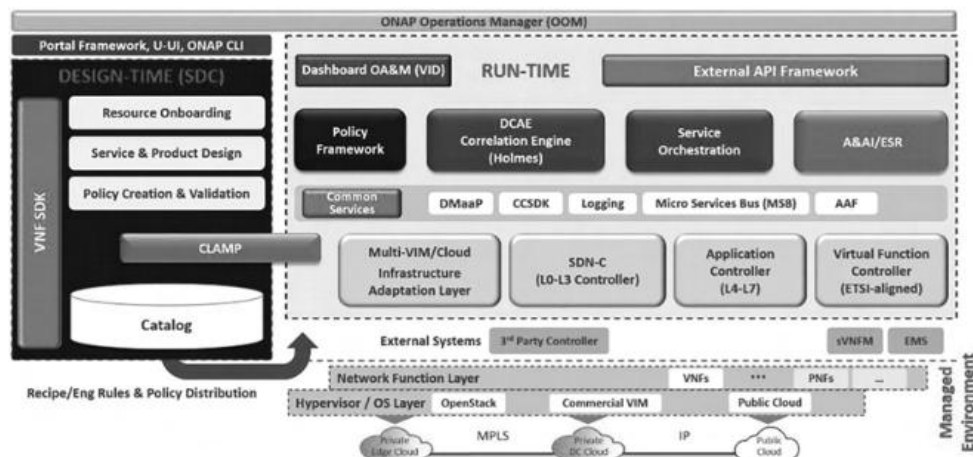
regulasi dan hukum yang dinamis. Batasan antarindustri akan semakin kabur. Apakah Apple sebuah perusahaan teknologi atau ponsel atau media atau keuangan? Apakah Amazon sebuah perusahaan ritel, media, atau teknologi? Regulasi yang mengatur berbagai layanan dan industri harus berevolusi ke dunia baru ini dengan batasan industri yang tidak begitu jelas.

15.7 KASUS PENGGUNAAN JARINGAN OTONOM

Operasi Jaringan Otonom

Bagian ini mengilustrasikan bagaimana otomatisasi mentransfer operasi jaringan. Bagian ini menyentuh manajemen siklus hidup lengkap dari desain, perencanaan, penerapan, dan jaminan layanan jaringan. Proyek ONAP dari Linux Foundation telah mengembangkan sistem operasi jaringan untuk otomatisasi jaringan (Gambar 15.7).

Seorang manajer produk layanan jaringan menggunakan model desain dan pembuatan layanan (SD&C) untuk merancang produk jaringan baru menggunakan modul layanan yang dikatalogkan dalam SD&C. Seorang manajer produk dapat menggabungkan berbagai layanan jaringan, yang terdapat dalam katalog, untuk membuat sebuah produk. Ia dapat memvalidasi produk tersebut dalam sandbox. Setelah produk divalidasi, produk tersebut dapat diterapkan secara otomatis di jaringan produksi melalui *Master Service Orchestrator* (MSO). MSO secara otonom mengorkestrasi infrastruktur, jaringan, dan berbagai fungsi jaringan untuk menerapkan layanan tersebut.



Gambar 15.7 Arsitektur ONAP.

Modul Analisis dan Peristiwa Pengumpulan Data (DCAE) mengumpulkan data dari node jaringan. DCAE menjalankan analitik pada data yang dikumpulkan dan mengidentifikasi jika terdapat anomali dalam kinerja jaringan. Anomali ditentukan menggunakan kebijakan dari manajer kebijakan.

Jika DCAE mengidentifikasi anomali, sistem dapat secara otomatis meluncurkan probe virtual untuk mengumpulkan data tambahan dari node dan antarmuka guna mengidentifikasi akar penyebab masalah. Data tambahan ini dianalisis untuk secara otomatis mengambil tindakan guna mengatasi masalah. Misalnya, jika sistem mengidentifikasi bahwa node (VNF)



berjalan pada kapasitas yang lebih tinggi daripada tanda air berkapasitas tinggi yang diizinkan, sistem dapat secara otomatis menyediakan instansiasi VNF tambahan agar penggunaan berada dalam batas yang diizinkan. Jika sistem tidak dapat mengatasi masalah, sistem akan terintegrasi dengan sistem manajemen kesalahan untuk menghasilkan alarm dan memberi tahu personel teknis.

Data dari DCAE disimpan dalam danau data untuk melatih algoritma ML guna memprediksi masalah jaringan di masa mendatang dan merekomendasikan optimasi kinerja. Solusi ML sumber terbuka akan tersedia (misalnya, proyek Linux Foundation, Acumos, memungkinkan pengguna untuk mengakses, memodifikasi, dan menambahkan berbagai algoritma ML). Kekuatan komunitas yang dipadukan dengan akses ke aliran data berkualitas yang berkelanjutan akan menjadi pendorong yang kuat bagi operasi jaringan. Salah satu contoh solusi tersebut adalah penggunaan data DCAE dari simpul jaringan untuk memprediksi probabilitas kegagalan mesin virtual yang menjalankan VNF. Data ini tidak hanya mampu memprediksi kegagalan, tetapi juga mengidentifikasi penyebab kegagalan. Hal ini memungkinkan peningkatan berkelanjutan untuk menghindari kegagalan di masa mendatang.

Di sisi bisnis, manajer produk akan dapat menghubungkan pengalaman pelanggan dengan metrik kinerja jaringan. Hal ini dapat dikombinasikan dengan kumpulan data lain, seperti data dari sensor untuk pelanggan industri atau media sosial untuk konsumen, untuk memberikan pengalaman pengguna yang lebih personal. Hal ini juga akan memungkinkan manajer produk untuk menciptakan fitur-fitur bernilai tambah baru guna menawarkan layanan-layanan baru yang menghasilkan pendapatan.

15.7.2 Pemrosesan Bahasa Alami untuk Mengurangi Denda Perjanjian Tingkat Layanan (SLA) Banyak layanan perusahaan operator telekomunikasi yang disertai perjanjian kontraktual untuk memenuhi tingkat kualitas layanan tertentu. Terdapat liabilitas finansial yang terkait dengan perjanjian ini. Jika perjanjian tingkat layanan (SLA) tidak dipenuhi, penyedia layanan harus membayar denda, yang dapat berdampak material pada laba bersih.

Biasanya, durasi penghentian layanan dimulai sejak pelanggan melaporkan masalah, dan berakhir saat masalah terselesaikan. Mungkin terdapat beberapa interaksi antara pelanggan dan penyedia layanan untuk menyelesaikan masalah tersebut. Terkadang, ketidaktanggapan pelanggan dapat menyebabkan penundaan yang tidak perlu dalam penyelesaian masalah. Hal ini dapat mengakibatkan denda yang lebih besar yang dibayarkan oleh penyedia layanan, bukan karena kesalahan mereka.

Solusi berbasis pemrosesan bahasa alami (NLP) dapat membantu operator mempercepat penyelesaian tiket masalah dan mengidentifikasi porsi tanggung jawab pelanggan dari denda tersebut. Setelah tiket pelanggan dimasukkan ke dalam sistem, tiket tersebut secara otomatis diklasifikasikan dan diteruskan ke tim resolusi yang sesuai. Tim resolusi dan respons pelanggan, serta log dari berbagai sistem, diproses secara otomatis menggunakan NLP untuk menghasilkan laporan berbasis fakta dari awal masalah hingga penyelesaian lengkap. Linimasa terperinci yang didukung oleh bukti yang tak terbantahkan yang dihasilkan secara otomatis membantu salah satu klien kami menghemat lebih dari Rp. 150 juta dalam satu tahun.



Aspek NLP yang lebih penting dari tiket manual adalah kemampuannya menghasilkan data berharga untuk memberikan gambaran lengkap tentang suatu masalah dari pelanggan akhir hingga node jaringan. Hal ini menghasilkan data yang dapat digunakan untuk melatih algoritma ML guna memprediksi masalah di masa mendatang dan mengambil tindakan korektif sebelumnya. Lagipula, polis asuransi terbaik adalah yang tidak terpakai.

Mobil Otonom

Menurut *Society of Automotive Engineering* (SAE), terdapat enam tingkatan untuk mencapai mobil otonom sepenuhnya.

- **Level 0—Tanpa Otomatisasi:** Performa penuh waktu oleh pengemudi manusia dalam semua aspek tugas mengemudi dinamis, bahkan ketika ditingkatkan oleh sistem peringatan atau intervensi.
- **Level 1—Bantuan Pengemudi:** Eksekusi spesifik mode mengemudi oleh sistem bantuan pengemudi, baik kemudi maupun akselerasi/deselerasi, menggunakan informasi tentang lingkungan mengemudi dan dengan harapan bahwa pengemudi manusia melakukan semua aspek yang tersisa dari tugas mengemudi dinamis.
- **Level 2—Otomatisasi Sebagian:** Eksekusi spesifik mode mengemudi oleh satu atau lebih sistem bantuan pengemudi, baik kemudi maupun akselerasi/deselerasi, menggunakan informasi tentang lingkungan mengemudi dan dengan harapan bahwa pengemudi manusia melakukan semua aspek yang tersisa dari tugas mengemudi dinamis.
- **Level 3—Otomatisasi Bersyarat:** Performa spesifik mode mengemudi oleh Sistem Mengemudi Otomatis dalam semua dengan aspek tugas mengemudi dinamis dengan harapan bahwa pengemudi manusia akan merespons tepat permintaan intervensi.
- **Level 4—Otomatisasi Tinggi:** Performa spesifik mode mengemudi oleh Sistem Mengemudi Otomatis dalam semua aspek tugas mengemudi dinamis. tugas mengemudi, bahkan jika pengemudi manusia tidak merespons permintaan intervensi dengan tepat
- **Level 5—Otomatisasi Penuh:** Kinerja penuh waktu oleh Sistem Mengemudi Otomatis dari semua aspek tugas mengemudi dinamis dalam semua kondisi jalan dan lingkungan yang dapat dikelola oleh pengemudi manusia

Agar mobil mencapai Level 4 (sebenarnya level kelima, karena dimulai dari Level 0), sebagaimana didefinisikan sebelumnya, beberapa transisi harus terjadi. Menurut Michael Ronen, kepala Tim Auto 2.0 Goldman Sachs Investment Banking, transisi tersebut meliputi hal-hal berikut:

1. Peralihan teknologi mesin → Mesin pembakaran internal digantikan oleh kendaraan listrik (lebih sederhana, lebih ringan, dan lebih sedikit komponen).
2. Mobil menjadi superkomputer yang terhubung → Agar dapat mengikuti semua sensor, radar, analitik, ML/AI, komunikasi internal (IP/Ethernet), dan eksternal (V2X) kendaraan serta konektivitas internet seluler dan berkelanjutan, kendaraan harus menjadi pusat data roaming mini.



3. Perubahan model bisnis → Mobil beralih dari sesuatu yang Anda miliki, rawat, dan gunakan (aset) menjadi sesuatu yang Anda panggil dan gunakan sesuai permintaan (layanan). Mobil tanpa pengemudi memungkinkan hal ini menjadi hemat biaya.
4. Ekosistem menyesuaikan → Dampaknya terhadap industri pendukung (asuransi, keuangan, perumahan, parkir, Penyedia Layanan Internet (ISP), real estat, penyedia konten, dll.) sangat signifikan.

Meskipun mobil yang sepenuhnya otonom (Level 5) masih jauh, waktu untuk mobil semi-otonom sudah sangat dekat, mungkin hanya beberapa tahun lagi. IoT baru ini (mobil, dalam hal ini) menunjukkan bagaimana menghubungkan perangkat sehari-hari ke jaringan mengubah apa yang dapat kita lakukan dengannya. “Mobil yang terhubung, yang berkomunikasi satu sama lain dan dengan dunia yang lebih luas, tidak hanya akan mengurangi kecelakaan dan memperlancar lalu lintas. Mobil-mobil ini akan memiliki dampak yang kuat di luar industri otomotif. Perusahaan asuransi, misalnya, akan memiliki cara baru untuk memantau perilaku pengemudi, memberi penghargaan kepada pengemudi yang baik, dan mendistribusikan biaya kepada pengemudi yang buruk. Perusahaan berbagi tumpangan dapat menghubungkan mobil yang tidak digunakan dengan pelanggan yang membutuhkannya dengan lebih baik.

Pada bulan Mei 2017, Negara Bagian New York membuka perizinan dan kesempatan bagi perusahaan mobil yang ingin menguji kendaraan otonom di negara bagian mereka. Audi mengajukan dan menerima lisensi untuk menguji model A7 mereka, mobil otonom SAE Level 3, seperti yang ditunjukkan pada Gambar 15.8. Menurut Audi, “untuk mencapai otonomi Level 4 pada tahun 2020, Audi akan mengandalkan NVIDIA untuk membantu kecerdasan buatan dan daya pemrosesan yang diperlukan untuk membuat berkendara otonom menjadi aman saat mengemudikan kendaraan seberat 4.000 pon pada kecepatan 70 mph.

Dr. Raj Rajkumar, seorang profesor George Westinghouse di Departemen Teknik Elektro dan Komputer di Carnegie Mellon University, telah menciptakan kendaraan prototipe SAE Level 4, sebuah Cadillac SRX otonom. Prototipe SRX ini menunjukkan kemajuan dalam teknologi dan kepraktisan mengemudi otonom. Kendaraan tersebut telah digunakan untuk mendemonstrasikan teknologi tersebut kepada para pembuat kebijakan di Capitol Hill di Washington DC, mengemudi sendiri di sekitar Capitol Hill dan di sepanjang jalan raya I-395 ke/dari Washington DC. Mobil tersebut adalah prototipe canggih yang diisi dengan sensor, radar, pemrosesan komputasi, dan perangkat lunak AI, semuanya tersembunyi dari pandangan, mencoba untuk mewakili seperti apa tampilan dan kinerja mobil otonom SAE Level 4 di masa depan.

Kecerdasan buatan/pembelajaran mesin memainkan peran penting dalam evolusi mobil otonom dan dalam menjadikannya sebuah kepraktisan untuk penggunaan di jalan raya. Akan ada sejumlah aplikasi AI dalam mobil otonom: pengenalan suara dan wajah untuk membuka dan menyalakan kendaraan, optimasi rute, persepsi dan klasifikasi objek, serta konversi keluaran algoritma untuk menggerakkan sinyal kontrol ke aktuator guna mengendalikan kendaraan. Akan ada hierarki komputasi dengan komputasi lokal, edge, dan cloud untuk optimasi kinerja dan biaya. Latensi respons, bandwidth data, dan keandalan



jaringan akan mendorong penempatan optimal berbagai fungsi untuk mendukung kendaraan otonom.



Gambar 15.8 Mobil otonom Audi A7 SAE Level 3.

Seperti transisi apa pun, ini akan memakan waktu. Mobil otonom awal, pada berbagai level, akan berada di jalan pada saat yang sama; semua level dari Level 0 hingga Level 5 kemungkinan akan terwakili pada waktu tertentu. Ini berarti bahwa perangkat lunak AI/ML harus memperhitungkan pengemudi manusia yang dengannya hanya ada sedikit atau tidak ada interaksi/komunikasi; mobil semi-otonom yang tidak ada interaksi/komunikasi; dan mobil semi-otonom yang dengannya ada pembagian informasi. Kebutuhan untuk mengumpulkan data secara berkala, memperbarui peta, memperbarui perangkat lunak, mengubah model ML, dll. membuat otomatisasi model operasional menjadi penting, terutama ketika mobil otonom berskala hingga puluhan juta mobil. Matriks "kompatibilitas" mobil otonom ini membuat transisi terjadi lebih lambat, karena ini akan menjadi brownfield, yang berarti bahwa semua mobil di jalan tidak akan beralih dalam semalam. Hal ini juga membuat AI/ML menjadi lebih penting, karena perilaku mobil lain di jalan lebih sulit diprediksi, yang berarti lebih banyak parameter diperlukan dalam model ML, yang berarti lebih banyak perhitungan diperlukan untuk membuat keputusan cerdas, yang berarti bahwa algoritma AI harus lebih kompleks.

Pendanaan ventura untuk mewujudkan mobil otonom sangat dramatis. Salah satu contohnya adalah pembelian Mobileye oleh Intel senilai Rp 150 triliun. Mobileye memproduksi sensor, perangkat, dan perangkat lunak AI untuk mobil otonom. Intel menciptakan armada 100 mobil self-driving (SAE Level 4) dan mengumumkannya di CES 2018. Menurut Intel, kendaraan tersebut akan menggabungkan "visi komputer, penginderaan, fusi, pemetaan, dan kebijakan mengemudi" Mobileye dengan "platform komputasi terbuka dan keahlian Intel di pusat data. Platform mobil otonom akan mencakup teknologi komunikasi 5G dari Intel untuk menghadirkan sistem 'mobil-ke-cloud' yang lengkap." Armada tersebut akan terdiri dari berbagai merek mobil dan jenis kendaraan untuk menunjukkan fleksibilitas dan kemampuan beradaptasinya. Koneksi ke jaringan seluler ini berbeda dibandingkan dengan sistem sebelumnya, tetapi menyediakan koneksi bandwidth tinggi yang sangat dibutuhkan ke sistem data dan informasi terkini di cloud, tentang mobil itu sendiri, pembaruan perangkat lunak, kondisi jalan, cuaca, area kerja, dll.

Menurut Intel, faktor terbesar dalam adopsi mobil otonom oleh konsumen adalah Kepercayaan. Kepercayaan terhadap teknologinya. Solusi Otomotif Intel dirancang untuk menghadirkan sistem lapangan dengan lebih cepat; solusi ini mencakup komputasi dalam



kendaraan, alat pengembangan perangkat lunak, konektivitas siap 5G, platform pusat data yang tangguh, dan kemajuan terbaru dalam AI. Jaringan nirkabel akan menjadi bagian integral dari solusi mobil otonom. Terdapat empat jenis komunikasi V2X yang didefinisikan dalam 3GPP: kendaraan ke kendaraan (V2V), kendaraan ke infrastruktur (V2I), kendaraan ke jaringan (V2N), dan kendaraan ke pejalan kaki (V2P). Transmisi V2V dan V2P umumnya disiarkan antara kendaraan dan orang untuk memberikan informasi tentang kecepatan, arah, dan lokasi kendaraan.

V2I dan V2N mendukung pertukaran informasi dua arah antara kendaraan dan jaringan nirkabel. Terdapat penyebaran luas stasiun pangkalan seluler yang dapat diperluas sebagai unit sisi radio (RSU). RSU dapat menghosting aplikasi kendaraan otonom dan melakukan pemrosesan lokal untuk meminimalkan penundaan jaringan. Ini akan ditambah dengan aplikasi berbasis cloud backend. Konfigurasi hibrida ini dengan pemrosesan mendekati waktu nyata di edge dan pemrosesan backend yang ekstensif untuk pembelajaran berkelanjutan akan memenuhi persyaratan kinerja latensi yang ketat untuk aplikasi kendaraan otonom dan memanfaatkan data masif yang terus-menerus dihasilkan oleh kendaraan untuk mengoptimalkan operasi kendaraan otonom.

Salah satu keunggulan utama komunikasi V2I adalah pembelajaran kolektif kendaraan dapat dimanfaatkan. Pembelajaran dari setiap kendaraan dapat didistribusikan secara otonom dan berkelanjutan ke semua kendaraan lain yang mungkin mendapatkan manfaat dari pembaruan tersebut. Studi USDOT memperkirakan bahwa teknologi ini dapat mengatasi hampir 80% dari semua kecelakaan kendaraan yang ditargetkan. Banyak analis mengatakan bahwa mobil adalah platform pintar berikutnya, dan jika itu benar, maka seperti halnya ponsel pintar, konektivitas, khususnya konektivitas seluler 5G, akan sangat penting bagi keberhasilan dan kegunaan mobil otonom pintar. Konektivitas 5G ini juga akan memainkan peran penting dalam menginformasikan algoritma AI yang menyediakan kecerdasan pada kendaraan ini. Jaringan 5G akan menyediakan tautan data yang sangat tinggi dengan latensi rendah, yang merupakan dua persyaratan ketat, namun krusial, bagi kendaraan otonom. Selain itu, penerapan 5G akan lebih padat daripada penerapan stasiun pangkalan 3G/4G. Stasiun pangkalan 5G ini akan menyediakan kemampuan komputasi tepi seluler untuk pemrosesan dengan penundaan jaringan minimal untuk aplikasi yang sensitif terhadap waktu.

Kecerdasan buatan dan pembelajaran mesin (ML) telah digunakan dalam banyak aplikasi konsumen dan industri. Seiring kemajuan yang terus kita capai dalam teknologi inti komputasi, penyimpanan, jaringan, dan sensor, ML akan diterapkan pada semakin banyak kasus penggunaan di berbagai industri. Jaringan generasi mendatang akan menjadi pendorong dan penerima manfaat dari hal-hal ini.

Meskipun laju perubahan dalam jaringan mungkin telah berubah selama 100 tahun terakhir, perubahan tersebut merupakan fakta yang konstan dalam industri ini. Industri ini berada di ambang evolusi menarik lainnya dengan bagian-bagian jaringan otonom yang mulai menyatu dalam lingkungan SDN/NFV. Peluang untuk berinovasi tidak terbatas bagi mereka yang bersedia melangkah maju untuk belajar dan berpikir inovatif.



DAFTAR PUSTAKA

- Abnili, M. Z., & Azad, N. L. (2021). On-line situational awareness for autonomous driving at roundabouts using artificial intelligence. *Journal of Machine Intelligence and Data Science (JMIDS)*, 2.
- Ahammed, T. B., Patgiri, R., & Nayak, S. (2023). A vision on the artificial intelligence for 6G communication. *Ict Express*, 9(2), 197-210.
- Ali, M. H., Jaber, M. M., Alfred Daniel, J., Vignesh, C. C., Meenakshisundaram, I., Kumar, B. S., & Punitha, P. (2023). Autonomous vehicles decision-making enhancement using self-determination theory and mixed-precision neural networks. *Multimedia Tools and Applications*, 1-24.
- Andronie, M., Lăzăroiu, G., Iatagan, M., Uță, C., Ștefănescu, R., & Cocoșatu, M. (2021). Artificial intelligence-based decision-making algorithms, internet of things sensing networks, and deep learning-assisted smart process management in cyber-physical production systems. *Electronics*, 10(20), 2497.
- Atakishiyev, S., Salameh, M., Yao, H., & Goebel, R. (2024). Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions. *IEEE Access*.
- Bale, A. S., Rajani, M., Taj, N., Amala, G., & Lal, K. (2024, March). Connecting Autonomous Engineering Domains with the Shared Language of Deep Neural Networks. In *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA)* (pp. 1-6). IEEE.
- Bathla, G., Bhadane, K., Singh, R. K., Kumar, R., Aluvalu, R., Krishnamurthi, R., ... & Basheer, S. (2022). Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities. *Mobile Information Systems*, 2022(1), 7632892.
- Bendiab, G., Hameurlaine, A., Germanos, G., Kolokotronis, N., & Shiaeles, S. (2023). Autonomous vehicles security: Challenges and solutions using blockchain and artificial intelligence. *IEEE Transactions on Intelligent Transportation Systems*, 24(4), 3614-3637.
- Bharadiya, J. P. (2023). Artificial intelligence in transportation systems a critical review. *American Journal of Computing and Engineering*, 6(1), 35-45.
- Bistron, M., & Piotrowski, Z. (2021). Artificial intelligence applications in military systems and their influence on sense of security of citizens. *Electronics*, 10(7), 871.
- Burri, T., & Trusilo, D. (2021). Ethical artificial intelligence: An approach to evaluating disembodied autonomous systems. *Rain Liivoja and Ann Väljataga (eds), Autonomous Cyber Capabilities in International Law*, 51-66.
- Camacho, J., & Wasielewska, K. (2022, April). Dataset quality assessment in autonomous networks with permutation testing. In *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium* (pp. 1-4). IEEE.



- Chamola, V., Hassija, V., Sulthana, A. R., Ghosh, D., Dhingra, D., & Sikdar, B. (2023). A review of trustworthy and explainable artificial intelligence (XAI). *IEEE Access*, 11, 78994-79015.
- Dang, W., Huang, R., Yu, Y., & Zhang, Y. (2024). *Autonomous Driving Network: Network Architecture in the Era of Autonomy*. CRC Press.
- Dhar Dwivedi, A., Singh, R., Kaushik, K., Rao Mukkamala, R., & Alnumay, W. S. (2024). Blockchain and artificial intelligence for 5G-enabled Internet of Things: Challenges, opportunities, and solutions. *Transactions on Emerging Telecommunications Technologies*, 35(4), e4329.
- Durlik, I., Miller, T., Kostecka, E., Zwierzewicz, Z., & Łobodzińska, A. (2024). Cybersecurity in autonomous vehicles—are we ready for the challenge?. *Electronics*, 13(13), 2654.
- Facchini, S. D. (2022, July). Decentralized Autonomous Organizations and Multi-agent Systems for Artificial Intelligence Applications and Data Analysis. In *IJCAI* (pp. 5851-5852).
- Fang, H., Yu, P., Tan, C., Zhang, J., Lin, D., Zhang, L., ... & Meng, L. (2024). Self-healing in knowledge-driven autonomous networks: Context, challenges, and future directions. *IEEE Network*, 38(6), 425-432.
- Garikapati, D., & Shetiya, S. S. (2024). Autonomous vehicles: Evolution of artificial intelligence and the current industry landscape. *Big Data and Cognitive Computing*, 8(4), 42.
- Groshev, M., Guimaraes, C., Martín-Pérez, J., & de la Oliva, A. (2021). Toward intelligent cyber-physical systems: Digital twin meets artificial intelligence. *IEEE Communications Magazine*, 59(8), 14-20.
- Gumbs, A. A., Frigerio, I., Spolverato, G., Croner, R., Illanes, A., Chouillard, E., & Elyan, E. (2021). Artificial intelligence surgery: How do we get to autonomous actions in surgery?. *Sensors*, 21(16), 5526.
- Hasan, M., & Mullick, T. (2021). Blockchain and artificial intelligence enabled autonomous smart manufacturing consortium. In *IISE Annual Conference. Proceedings* (pp. 920-925). Institute of Industrial and Systems Engineers (IISE).
- Hu, S., Liang, Y. C., Xiong, Z., & Niyato, D. (2021). Blockchain and artificial intelligence for dynamic resource sharing in 6G and beyond. *IEEE Wireless Communications*, 28(4), 145-151.
- Ipp, E., Liljenquist, D., Bode, B., Shah, V. N., Silverstein, S., Regillo, C. D., ... & EyeArt Study Group. (2021). Pivotal evaluation of an artificial intelligence system for autonomous detection of refractable and vision-threatening diabetic retinopathy. *JAMA network open*, 4(11), e2134254-e2134254.
- Ismail, L., & Buyya, R. (2022). Artificial intelligence applications and self-learning 6G networks for smart cities digital ecosystems: Taxonomy, challenges, and future directions. *Sensors*, 22(15), 5750.
- Kök, I., Okay, F. Y., Muyanli, Ö., & Özdemir, S. (2023). Explainable artificial intelligence (xai) for internet of things: a survey. *IEEE Internet of Things Journal*, 10(16), 14764-14779.
- Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., ... & Aspuru-Guzik, A. (2022). On scientific understanding with artificial intelligence. *Nature Reviews Physics*, 4(12), 761-769.



- Lahmeri, M. A., Kishk, M. A., & Alouini, M. S. (2021). Artificial intelligence for UAV-enabled wireless networks: A survey. *IEEE Open Journal of the Communications Society*, 2, 1015-1040.
- Latif, A., & Chung, H. (2023). Fire detection and spatial localization approach for autonomous suppression systems based on artificial intelligence. *Fire Technology*, 59(5), 2621-2644.
- LazaroIU, G., Androniceanu, A., Grecu, I., Grecu, G., & Neguriță, O. (2022). Artificial intelligence-based decision-making algorithms, Internet of Things sensing networks, and sustainable cyber-physical management systems in big data-driven cognitive manufacturing. *Oeconomia Copernicana*, 13(4), 1047-1080.
- Lee, H., Lee, S. H., & Quek, T. Q. (2022). Artificial intelligence meets autonomy in wireless networks: A distributed learning approach. *IEEE Network*, 36(6), 100-107.
- Liu, B., Yu, L., Che, C., Lin, Q., Hu, H., & Zhao, X. (2023). Integration and performance analysis of artificial intelligence and computer vision based on deep learning algorithms. *arXiv preprint arXiv:2312.12872*.
- Lv, Z. (2023). Generative artificial intelligence in the metaverse era. *Cognitive Robotics*, 3, 208-217.
- Lv, Z., Chen, D., Feng, H., Wei, W., & Lv, H. (2022). Artificial intelligence in underwater digital twins sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 18(3), 1-27.
- Markouh, O., Adadi, A., & Berrada, M. (2023). Blockchain and explainable ai for trustworthy autonomous vehicles. In *Explainable Artificial Intelligence for Intelligent Transportation Systems* (pp. 241-254). CRC Press.
- Nagy, M., LăzăroIU, G., & Valaskova, K. (2023). Machine intelligence and autonomous robotic technologies in the corporate context of SMEs: Deep learning and virtual simulation algorithms, cyber-physical production networks, and Industry 4.0-based manufacturing systems. *Applied Sciences*, 13(3), 1681.
- Nazir, S., Dickson, D. M., & Akram, M. U. (2023). Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks. *Computers in Biology and Medicine*, 156, 106668.
- Palade, V., & Deo, A. (2022). Artificial Intelligence in Cars: How Close Yet Far Are We from Fully Autonomous Vehicles?. *International Journal on Artificial Intelligence Tools*, 31(3), 2241005.
- Radanliev, P., & De Roure, D. (2023). Review of the state of the art in autonomous artificial intelligence. *AI and Ethics*, 3(2), 497-504.
- Rezwan, S., & Choi, W. (2022). Artificial intelligence approaches for UAV navigation: Recent advances and future challenges. *IEEE access*, 10, 26320-26339.
- Saborío-Taylor, S., & Rojas-Ramírez, F. (2024). Universal design for learning and artificial intelligence in the digital era: Fostering inclusion and autonomous learning. *International Journal of Professional Development, Learners and Learning*, 6(2), ep2408.
- Sarkar, N. I., & Gul, S. (2023). Artificial intelligence-based autonomous UAV networks: A survey. *Drones*, 7(5), 322.



- Sharma, S., Islam, N., Singh, G., & Dhir, A. (2022). Why do retail customers adopt artificial intelligence (AI) based autonomous decision-making systems?. *IEEE Transactions on Engineering Management*, 71, 1846-1861.
- Shen, Y., Shao, J., Zhang, X., Lin, Z., Pan, H., Li, D., ... & Letaief, K. B. (2024). Large language models empowered autonomous edge AI for connected intelligence. *IEEE Communications Magazine*, 62(10), 140-146.
- Sheth, A., Roy, K., & Gaur, M. (2023). Neurosymbolic artificial intelligence (why, what, and how). *IEEE Intelligent Systems*, 38(3), 56-62.
- Shon, T. (2021). In-vehicle Networking/Autonomous vehicle security for internet of Things/Vehicles. *Electronics*, 10(6), 637.
- Soumya, K. V., Sujitha, S., Kanaujia, S., Agarwalla, S., Sameer, S., & Manzoor, T. (2023, February). Silent surveillance autonomous drone for disaster management and military security using artificial intelligence. In *2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM)* (pp. 1-4). IEEE.
- Thangavel, K., Spiller, D., Sabatini, R., Amici, S., Sasidharan, S. T., Fayek, H., & Marzocca, P. (2023). Autonomous satellite wildfire detection using hyperspectral imagery and neural networks: A case study on australian wildfire. *Remote Sensing*, 15(3), 720.
- Tyagi, R., Sharma, S., & Mohan, S. (2022, March). Blockchain enabled intelligent digital forensics system for autonomous connected vehicles. In *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)* (pp. 1-6). IEEE.
- Umoga, U. J., Sodiya, E. O., Ugwuanyi, E. D., Jacks, B. S., Lottu, O. A., Daraojimba, O. D., & Obaigbena, A. (2024). Exploring the potential of AI-driven optimization in enhancing network performance and efficiency. *Magna Scientia Advanced Research and Reviews*, 10(1), 368-378.
- Vermesan, O., Bröring, A., Tragos, E., Serrano, M., Bacciu, D., Chessa, S., ... & Bahr, R. (2022). Internet of robotic things—converging sensing/actuating, hyperconnectivity, artificial intelligence and IoT platforms. In *Cognitive hyperconnected digital transformation* (pp. 97-155). River Publishers.
- Wang, S., Ruiz, M., & Velasco, L. (2024). Context-based e2e autonomous operation in B5G networks. *Sensors*, 24(5), 1625.
- Wang, Y., Du, S., Xin, Q., He, Y., & Qian, W. (2024). Autonomous driving system driven by artificial intelligence perception fusion. *Academic Journal of Science and Technology*, 9(2), 193-198.
- Xie, J., Zhang, Y., Qin, Y., Wang, B., Dong, S., Li, K., & Xia, Y. (2025). Is human-like decision making explainable? Towards an explainable artificial intelligence for autonomous vehicles. *Transportation Research Interdisciplinary Perspectives*, 29, 101278.
- Yan, X., Du, J., Wang, L., Liang, Y., Hu, J., & Wang, B. (2024, September). The Synergistic Role of Deep Learning and Neural Architecture Search in Advancing Artificial Intelligence. In *2024 International Conference on Electronics and Devices, Computational Science (ICEDCS)* (pp. 452-456). IEEE.



- Zhixin, L., Yubo, H., Tianding, Z., Yueming, W., Haoyuan, Y., Wei, Z., & Yang, W. (2023, July). Discussion on the application of artificial intelligence in computer network technology. In *2023 2nd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS)* (pp. 51-55). IEEE.
- Zuo, Y., Guo, J., Gao, N., Zhu, Y., Jin, S., & Li, X. (2023). A survey of blockchain and artificial intelligence for 6G wireless communications. *IEEE Communications Surveys & Tutorials*, 25(4), 2494-2528.



Artificial Intelligence dalam Jaringan Otonom

Dr. Ir. Agus Wibowo, M.Kom, M.Si, MM

BIO DATA PENULIS



Penulis memiliki berbagai disiplin ilmu yang diperoleh dari Universitas Diponegoro (UNDIP) Semarang. dan dari Universitas Kristen Satya Wacana (UKSW) Salatiga. Disiplin ilmu itu antara lain teknik elektro, komputer, manajemen dan ilmu sosiologi. Penulis memiliki pengalaman kerja pada industri elektronik dan sertifikasi keahlian dalam bidang Jaringan Internet, Telekomunikasi, Artificial Intelligence, Internet Of Things (IoT), Augmented Reality (AR), Technopreneurship, Internet Marketing dan bidang pengolahan dan analisa data (komputer statistik).

Penulis adalah pendiri dari Universitas Sains dan Teknologi Komputer (Universitas STEKOM) dan juga seorang dosen yang memiliki Jabatan Fungsional Akademik Lektor Kepala (Associate Professor) yang telah menghasilkan puluhan Buku Ajar ber ISBN, HAKI dari beberapa karya cipta dan Hak Paten pada produk IPTEK. Sejak tahun 2023 penulis tercatat sebagai Dosen luar biasa di Fakultas Ekonomi & Bisnis (FEB) Universitas Diponegoro Semarang. Penulis juga terlibat dalam berbagai organisasi profesi dan industri yang terkait dengan dunia usaha dan industri, khususnya dalam pengembangan sumber daya manusia yang unggul untuk memenuhi kebutuhan dunia kerja secara nyata.



YAYASAN PRIMA AGUS TEKNIK

PENERBIT :

YAYASAN PRIMA AGUS TEKNIK
Jl. Majapahit No. 605 Semarang
Telp. (024) 6723456. Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

ISBN 978-634-7227-48-5 (PDF)



9

786347

227485