Cybertwin Based Cloud Native Networks

Quan Yu, Dandan Liang, Meng Qin, Jiacheng Chen, Haibo Zhou, Jing Ren, Ying Li,
Jun Wu, Yue Gao, Wei Zhang

***Abstract*—With the emerging applications of the Internet of things, artificial intelligence, and satellite communications, the future network will be featured as the Internet of everything around the globe. The network heterogeneity, applications cloudification, and personalized user services demand a revolutionary change in the network architecture. With the rapid development of cloud native technology, the new network should support heterogeneous networks and personalized quality of services for users. In this paper, we propose a Cybertwin-based cloud native network (CCNN) that merges the radio access network (RAN), the IP bearer network, and the data center network and is based on the cloud native data center network using Kubernetes as a network operating system for unified virtualization of computing, storage, and network resources, unified scheduling and allocation, and unified operation and management. Then, we propose a fully decoupled RAN architecture that can flexibly and efficiently utilize the resource for personlized user services. We also propose a Cybertwin-based management framework built on Kubernetes for integrated networking, computing and storage resource scheduling. Finally, we design an immunology-inspired intrinsic security architecture with zero trust security system and adaptive defense system. The proposed CCNN is a new network architecture expected to address future generation communications and networks challenges.**

***Keywords*—6G, network architecture, Cybertwin, cloud native, network security**

Manuscript received Jun. 23, 2023; revised Aug. 10, 2023; accepted Sep. 08, 2023. This work was supported in part by the Key Area Research and Development Program of Guangdong Province under Grant 2020B0101110003, and in part by the major key project of Peng Cheng Laboratory and the Basic and Frontier Research Project of PCL. The associate editor coordinating the review of this paper and approving it for publication was L. Bai.

Q. Yu, D. D. Liang, M. Qin, J. C. Chen, H. B. Zhou, J. Ren, Y. Li. Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: yuq@pcl.ac.cn; liangdd@pcl.ac.cn; qinm01@pcl.ac.cn; chenjch02@pcl.ac.cn; haibozhou@nju.edu.cn; renjing@uestc.edu.cn; liy02@pcl.ac.cn).

H. B. Zhou. School of Electronic Science and Engineering, Nanjing University, Nanjing 210008, China (e-mail: haibozhou@nju.edu.cn).

J. Ren. School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611730, China (e-mail: renjing@uestc.edu.cn).

J. Wu. School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: wujun@fudan.edu.cn).

Y. Gao. Intelligent Networking and Computing Research Center and School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: gao_yue@fudan.edu.cn).

W. Zhang. School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney 2052, Australia (e-mail: weizhang@ieee.org).

I. Introduction

A. Evolution of Internet

In the 1960s: to establish a robust enough network, packet switching technology was developed during this period, laying the foundation for current network communication. The Advanced Research Projects Agency (ARPA) of the U.S. Department of Defense initiated a project to build a wide-area packet switching network called ARPAnet. In 1969, computers from the University of California, Los Angeles, Stanford University, the University of California, Santa Barbara, and the University of Utah were connected, achieving distributed computing and remote access[1].

In the 1970s: scientists envisioned numerous packet-switched networks, such as ARPAnet, satellite network (SATnet), and packet radio network (PRnet). Faced with the challenge of connecting these different networks together, Vinton Cerf and others proposed the core concept of the Internet[2]. ARPA organized an engineering team led by Cerf to develop and validate the Internet protocol (IP) and transmission control protocol (TCP). This period was dominated by monolithic applications, which typically ran as a whole system on mainframe computers. Therefore, these applications were primarily local applications with minimal network requirements, which by today's standards, are almost negligible. Additionally, network interconnections and protocols were proprietary.

In the 1980s: with the advent of the IBM personal computer (PC) and the shift from mainframe computing to personal computing, the client-server application architecture emerged. In addition to text and images, the data exchanged by applications also included audio and video. The network also began to become complex, although according to today's network traffic standards, the bandwidth requirements for the network

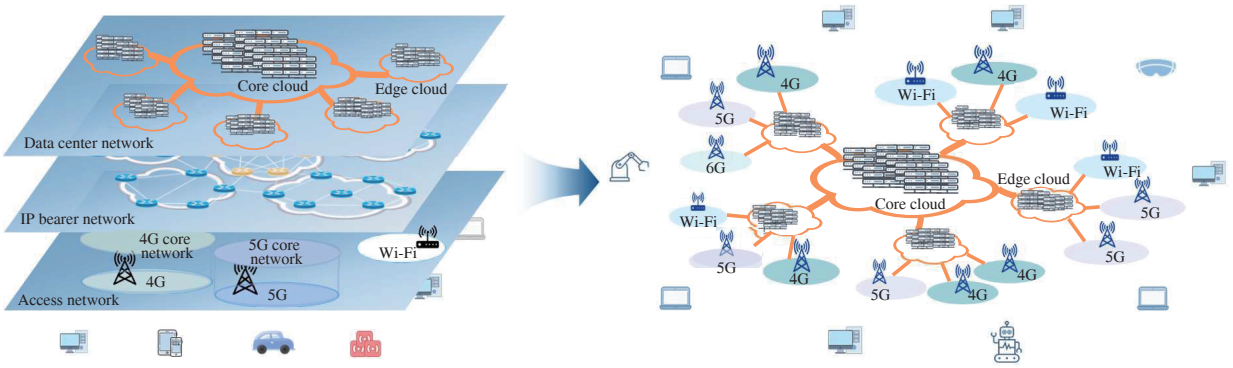


Fig. 1 (a) The current Internet architecture versus (b) Cloud native network architecture

in this era are still negligible, and the interconnection speed can reach up to 100 Mbit/s. And mostly proprietary, such as Ethernet, token ring, and fiber distributed data interface (FDDI) are the most popular LAN technologies. Faced with the challenge of scalability, computer scientists tend to adopt a hierarchical design, and design schemes for hierarchical routing, hierarchical naming, and hierarchical management. Since most upper-layer network protocols are not designed for large-scale or cross-enterprise use, TCP/IP wins in the market because of its openness, cross-platform compatibility, flexibility and scalability.

In the 1990s: the World Wide Web was put forward, bringing the Internet into millions of households and businesses worldwide. Tim Berners-Lee invented the hyper text transfer protocol (HTTP), connecting information from anywhere, allowing individuals with diverse perspectives to collaborate and transform immature ideas into creativity[3]. With the subsequent birth of web applications, applications ventured beyond corporate walls, allowing clients to connect to servers from anywhere, thus significantly increasing the scale of operations that application servers had to manage. Since a single server instance in most enterprises couldn't handle the load from all clients, and a single server instance could result in a single point of failure, the approach of running multiple server instances with a load balancer in the front-end also became popular. Servers themselves were divided into multiple units: a web front-end, applications, and a database or storage. The Internet also evolved from being a communication network centered on connectivity and transmission to an information network centered on content and collaboration.

In the 2000s: in 2007, the iPhone debuted with the posture of "reinventing the mobile phone", signifying that any user on the internet could widely utilize network applications in a mobile environment. Given the global and open nature of the internet, cybersecurity issues also began to take center stage[1]. The internet also gradually shifted from being PC-centric to being human-centric.

In the 2010s: streaming video and social media have led to application architectures increasingly adopting the "microservices" architecture and have also promoted the rise of containers and their orchestration software, Kubernetes[4]. The capabilities of cloud native software[5] have increased communication between servers, supporting newer types of applications like web search. There has been a historic shift in application communication from a client-server model to a server-server model. These applications demand the network to offer higher bandwidths (10 GbE Ethernet is common) and highly distributed communication patterns[6].

In the 2020s: the Apple Vision Pro brings us into the era of the metaverse, which is a milestone in the deep interaction between the real world and the digital world. The metaverse interfaces augment the real world by combining aspects of artificial intelligence, human-computer interaction, and computer vision to create applications that understand and interact with their surroundings. New mixed-reality applications demand higher low-latency and security requirements from the internet.

B. Trend of Cloud Native Network

As shown in Fig. 1(a), the current Internet is a combination of three networks to complete the task: the bottom layer is the access network, such as 4G, 5G, Wi-Fi and other wireless access networks; the middle layer is the earliest Internet core-IP bearer network; the top layer is the data center network that almost includes information and cloud services. These three networks operate independently. However, with the rapid growth of business requirements and the development of cloud native application software technology, the microservices applications deployed in the data center are becoming more and more concentrated, and the typical business model of the network has changed from traditional end-to-end data transmission to end-to-cloud mass sensor data collection and processing and distribute applications. In this way, cloud service providers are becoming larger and larger, such as Google Cloud, Amazon Cloud, Alibaba Cloud, and Tencent Cloud. However, the independence of the access net-

work, bearer network, and data center network has exacerbated problems such as the difficulty of adapting transmission and information services and the inability to guarantee the quality of user services. To address this issue, technologies such as content delivery networks[7], edge computing[8], and cloud-network integration[9] have rapidly developed in recent years. However, these remedial methods have not fundamentally changed the essence of the independence of the three networks. If all services are on the cloud in the future and the cloud centers are directly connected through optical cables, the bearer network will become the user's access. If the wireless base stations (BSs) accessed by users are directly deployed on the edge cloud, then there will be no need for a bearer network in the network architecture.

With the development of cloud native technology, the trend of the three-network integration on dedicated networks at home and abroad is becoming clearer, as shown in Fig. 1(b). In 2020, China Mobile took the lead and jointly formed a project technical committee with 13 domestic and foreign telecom operators, equipment providers, and cloud service providers. Under the Linux foundation, they launched the industry's first network cloud native platform as a service (PaaS) open-source project, XGVela. This platform extracts the general capabilities of the upper business layer into PaaS services deposited on the platform side, achieving lightweight agile business development, and accelerating the creation of new services[10]. In 2020, China Telecom went further in network with cloud native software technology. Its Tianyi Cloud 5G core network adopted a cloud native architecture, possessing advantages such as flexible deployment, simplicity and efficiency, and security and reliability. Based on business needs, it can flexibly choose user-plane function devices and deployment methods, thereby providing diversified and differentiated 5G dedicated network services to enterprise customers[11]. In 2020, China Unicom conducted research and technical standard formulation on the lightweight 5G core network solution[12].

Since June 2021, global operators such as International Telephone and Telegraph Corporation (AT&T), Dish, Telefnica Germany, and Swisscom have announced the migration of their 5G core networks to Microsoft cloud (Azure) or Amazon cloud (Amazon web services (AWS)). Telecom equipment providers Ericsson and Nokia also announced alliances with Google Cloud and AWS to jointly build cloud native 5G core network solutions[13-15]. Taking Amazon as an example, in July 2020, foreign equipment provider Ericsson provided Telefnica Spain with a 5G core network and business orchestration components based on cloud native software architecture. Amazon provided Telefnica Spain with cloud solutions and data privacy protection, among others[13]. Based on the cloud native technology, Telefnica Spain can provide enterprises with the basic functions of a cloud-based 5G core network, transforming the 5G core network from hardware-centric technology to a broad software solution. In this way, enterprises only need to equip themselves with a 5G wireless access network loaded with the corresponding antenna, without needing to set up physical core network infrastructure on-site. In 2021, Amazon proposes to provide all software and hardware for building and running a dedicated network according to user needs (network performance, number of devices, etc.), including network hardware devices that adapt to software, embedded-subscriber identity module (eSIM) cards, and citizens broadband radio service (CBRS) services[14]. Amazon stated that AWS Private 5G simplifies deployment, enabling customers to quickly deploy their own 5G, and AWS Private 5G does not charge any upfront deployment or equipment fees, and customers only pay for the network capacity and throughput they order. The development of these enterprise networks demonstrates the trend of the convergence of the three networks. However, the current network cloudification is all within the enterprise network, which still cannot truly solve the problem of future network transmission and service matching.

As AWS states that they can provide 5G as convenient as Wi-Fi. Only by weakening or eliminating the core network of the cellular system can we achieve the integration of heterogeneous networks. However, removing the core network of the cellular system immediately raises issues such as how to maintain business continuity in fast-moving scenarios, how to securely authenticate network access, and how to ensure the quality of transmission services. Thus, given the challenges of mobility, security, and transmission reliability brought by multi-operator, multi-heterogeneous access, and multi-modal transmission (4G/5G/Wi-Fi/Li-Fi/satellite, etc.) on the edge side of the cloud native network, neither cloud service providers nor telecom operators have ready-made solutions. As a result, cross-operator user and resource management have become a significant challenge in the architectural design of the cloud native network.

C. Future Network Challenges

With the rapid development of Internet of things (IoT), big data, AI, satellites, and communications technologies, the future network will be featured as connectivity of everything, new network architecture and personalized quality of service (QoS)-aware intelligent services. Firstly, the so-called "connectivity of everything" refers to the interconnection of tens of billions, or even trillions, of people, machines, and things. Secondly, network heterogeneity means that the network not only needs to cover densely populated areas but also remote regions. The network should not only serve people but also support data collection from billions of sensors. Therefore, an integrated network characteristic that spans space, air, and ground in a heterogeneous manner is an inevitable trend for

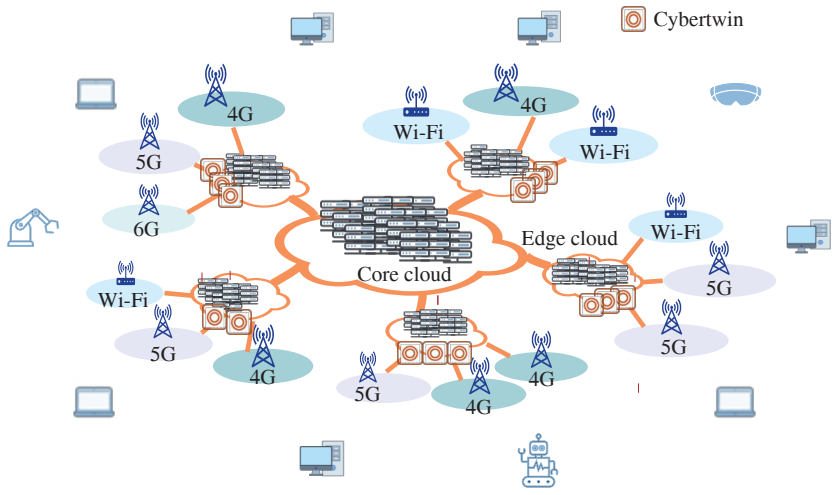


Fig. 2 Cybertwin based Cloud Native Network architecture

future networks. Thirdly, applications cloudification: traditional communication networks primarily achieve a single connection from A to B. In contrast, future communication networks will mainly focus on the massive collection, aggregation, processing, and fusion of sensor data, as well as its distribution and utilization after being empowered by AI. Future network design will face the following six challenges.

- Scalability, that is, the convergence and cooperation of multi-dimensional heterogeneous resources in the time domain, air domain, frequency domain, and code domain of heterogeneous networks.
- Mobility, that is, to maintain business continuity in fast-moving scenarios.
- Usability, that is, the service quality guarantee for user-oriented services.
- Security, that is, network security and protection of personal privacy. The security of the future network is not only related to personal privacy security, but also related to national security.
- Manageability, that is, the efficient operation and maintenance of the integration of transmission, computing and storage. The future network will be more diverse and complex. Simple, efficient and flexible operation and maintenance management is becoming more and more urgent for operators.
- Economic, that is, network infrastructure with low energy consumption and low cost.

II. Cybertwin Enabled Cloud Native Networks

In this paper, we propose a Cybertwin based cloud native network (CCNN) in Fig. 2 that converge the three networks of the data center network, the Internet IP bearer network, and the wireless access network. CCNN is based on the cloud native data center network, using Kubernetes as a network operating system for unified virtualization of computing, storage, network and other resources, unified scheduling and allocation, and unified operation and maintenance management. Utilizing the proposed Cybertwin to realize efficient collaboration and integration of various heterogeneous transmission methods and on-demand quality assurance of various heterogeneous application services. The CCNN breaks the traditional logic of constructing the three networks and relies on the core cloud and edge cloud to build the IP bearer network for the Internet. At the same time, there is no longer a need for the complicated core networks for various wireless accesses.

In the CCNN as shown in Fig. 2, various information services and communication services scattered across the network are rapidly integrated and organized for use. Service instance combinations are built flexibly and efficiently according to task requirements, utilizing various communication resources and transmission means, providing users in the network with low-latency and high-reliability transmission, supporting a diverse range of application services.

The concept of the Cybertwin was first proposed in Ref. [16] to address challenges related to mobility, security, and availability in cloud native networks. A Cybertwin maps the person entity/end-user device/non-person entity/organization in the physical space to the real-name agent in the cyberspace, has functions such as mobile agent, transmission agent, and security agent. It serves as a foundational service operating on both edge and core clouds, primarily fulfilling logical functions like those in 4G/5G core networks, to support the efficient and reliable operation of digital twins and other application services.

The necessity of Cybertwin can be summarized as follows: (1) when a user accesses the Internet, the user's Cybertwin performs security authentication, which solves many network security problems; (2) the user's behavior data in both physi-

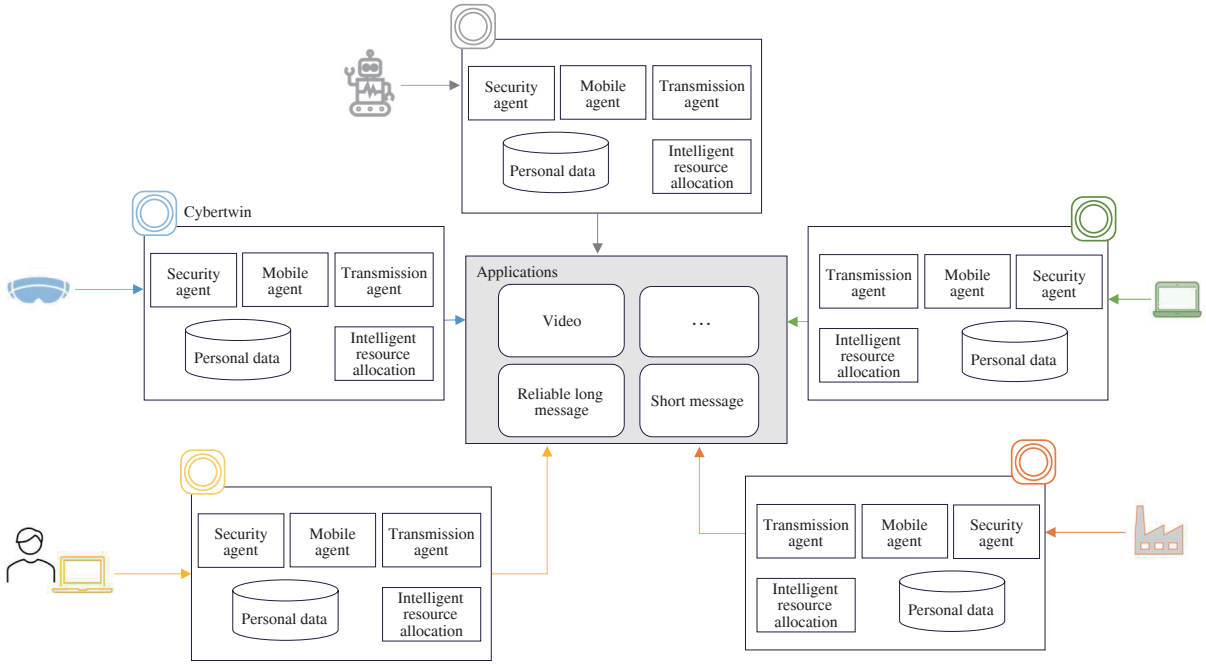


Fig. 3 Cybertwin functions diagram

cal space and cyberspace are fully recorded by their Cybertwin, ensuring the protection of user data resources; (3) at the same time, based on the user's own personal data, intelligent agent function of Cybertwin can provide truly personalized services; (4) Cybertwin integrates application needs and Multi-party resource providers negotiate available resources flexibly, realizing the efficient matching and flexible scheduling of cloud-edge service resources and edge-end transmission resources for ensuring personalized QoS-aware Intelligent Services.

As shown in the Cybertwin function diagram in Fig. 3, Cybertwin is the only entrance for each user to access the Internet. The end entity obtains network resource services through the Cybertwin to support personalized QoS-aware Intelligent Services. The functions in the proposed cloud native network can be summarized as:

- Security agent: when the end entity is connected to the Internet, the security agent function of the Cybertwin performs security authentication and authorization, addressing the security problem of the network; the security agent function includes the Cybertwin of person entity/end-user device/non-person entity/organization authenticating and authorizing access to their cyberspace. By verifying the credentials provided by users, devices, or services, it determines their legitimacy. By examining the roles, permissions, policies, etc. of users, devices, or services, it determines whether they can access specific network resources or services, ensuring that only legitimate users or devices can access network resources or services.
- Mobile agent: the mobile agent of the Cybertwin service quickly launches and establishes mobile policies for transmission scheduling according to user needs. Through the mobile agent, users can freely switch between different networks and transmit aggregation to ensure service continuity.
- Transmission agent: Cybertwin, based on users' personalized integrated communication service demands, establishes multi-link connections from the terminal entity to its Cybertwin and multi-path connections from the Cybertwin to the cloud through the transmission agent function. This ensures reliable network transmission of the proposed network.
- Personal data: the data agent service of the Cybertwin is a complete record of the status and behavior of person entity/end-user device/non-person entity/organizations in both physical and cyber spaces. It includes functions such as: (1) digital asset rights confirmation for person entity/end-user device/non-person entity/organization, managing, protecting, and operating their digital assets; (2) privacy protection: filtering, cleaning, scrambling, and encrypting data of person entity/end-user device/non-person entity/organizations.
- Intelligent resource allocation: the resource collaboration service of the Cybertwin in cloud native networks involves unified automatic scheduling of network service resources based on user business needs. This includes: (1) training universal AI models tailored to person entity/end-user device/non-person entity/organizations based on their data, obtaining personalized business preferences and inference capabilities; (2) the always-online capability of person entity/end-user device/non-person entity/organizations, along

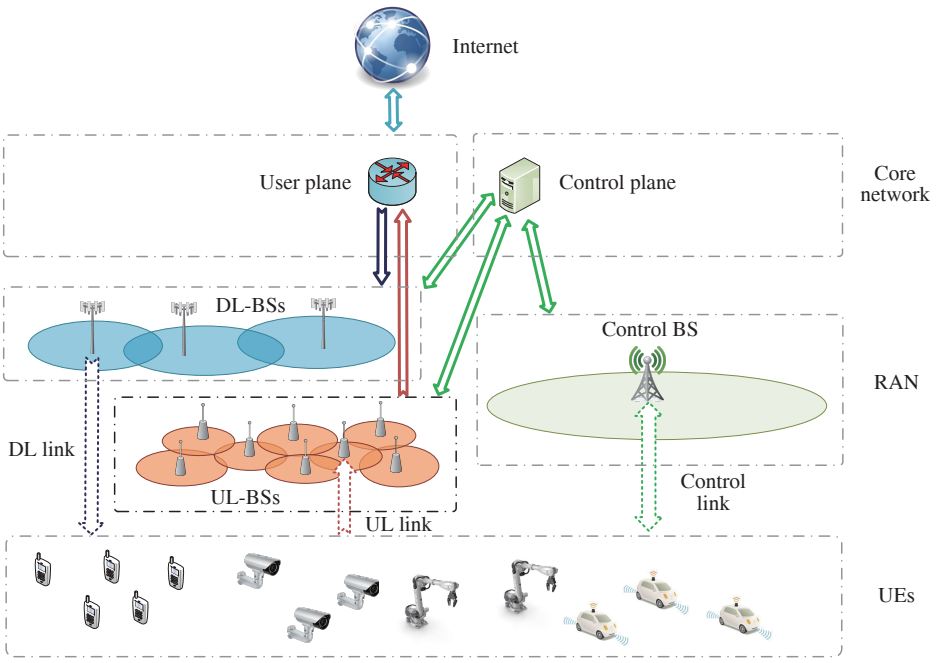


Fig. 4 The fully-decoupled RAN

with virtual-real interaction, integration, and division of labor capabilities; (3) negotiating, obtaining, and matching communication network resources and network service resources to ensure reliable transmission based on personalized services needs.

III. Fully-Decoupled Radio Access Network

The radio access networks (RANs) are responsible for the last-mile wireless access to the cloud native network. As the state-of-the-art RAN, the fifth mobile communications network (5G) can achieve the transmission rate of several hundreds Mbps in practical environments. However, 5G still needs to address some fundamental challenges. First, the costs of 5G become formidable, due to the dense deployment of BSs and higher energy consumption of each BS. Therefore, both capital and operational expenditures should be reduced. Second, high-quality spectrum resources have been exhausted, so resource utilization efficiency should be improved. Last, users' quality-of-experience (QoE) should be guaranteed so as to meet the diversified requirements of various services.

Despite the above challenges, the next-generation mobile communications network (6G)[17] should also take into account the emerging characteristics of future wireless networks. For example, satellite communications[18] are envisioned to be integrated with 6G and serve as the complement to the current BS-based terrestrial networks. Furthermore, to realize Internet of everything, trillions of devices need to be connected. Moreover, the uplink may consume more communication resources instead of downlink in certain scenarios for data collection. Data processing and AI-assisted content generation[19] will also consume huge computing and caching resources at the edge and cloud.

To deal with the above issues, we have proposed a novel fully-decoupled RAN (FD-RAN)[20] for 6G. In the following, we first introduce the architecture of FD-RAN and highlight its unique features. Then, we present the key technologies employed in FD-RAN.

A. Architecture

In general, the most significant features of FD-RAN are: 1) physical separation of control signalling and data transmission; and 2) physical separation of uplink and downlink data transmission. The major components of FD-RAN are illustrated in Fig. 4 and are introduced below.

- **Control BS (C-BS)**. C-BS handles all the control signalling with user equipments (UEs) through the bi-directional control link. C-BS carries out the control plane functions for UEs, including authentication, access management, mobility management, resource allocation indication, etc. A UE will always need to establish the secure control plane connection with C-BS for accessing the network. Typically, C-BS controls a wide area like C-RAN, so that the other BSs can serve UEs in a cooperative fashion with the help of C-BS.
- **Uplink BS (UL-BS)**. UL-BSs only have the signal reception function, thus they are utilized for uplink data reception for UEs. UL-BSs are usually light-weight since the have less antennas and energy consumption. However, UL-BSs can

be deployed densely, so as to be closer to UEs, whose transmit power is typically lower. Furthermore, the transmitted signal of a UE can be received by multiple UL-BSs through the receive diversity technique.

• **Downlink BS (DL-BS)**. DL-BSs are responsible for downlink services, namely transmitting data to UEs. Compared with UL-BSs, DL-BSs are typically equipped with massive antennas to enable high-capacity multiple-in multiple-out (MIMO) transmission. The energy costs of DL-BSs are also much higher, so the deployment of DL-BSs has higher requirements compared with UL-BSs. Similar to uplink, various cooperative transmission techniques can also be exploited in downlink.

• **Core network**. All the above three types of BSs are connected with the same core network, which utilizes cloud native deployment. On the one hand, these BSs are managed and controlled by the control plane. On the other hand, the UEs' data processed by UL-BSs (DL-BSs) are forwarded to (retrieved from) the Internet via the data plane.

The fundamental advantage brought by FD-RAN is flexibility. In FD-RAN, since the uplink and downlink are totally separate, we can regard uplink and downlink as individual networks, which have their own spectrum resources and deployments. Then, it is no longer required to find and use a pair of spectrum bands for both uplink and downlink as in frequency division duplex (FDD) systems, or consider the time slot allocation and synchronization of uplink and downlink as in time division duplex (TDD) systems. Instead, the spectrum used for uplink or downlink can be flexibly aggregated or released, without any impact to each other. Besides, resource cooperation on all dimensions including space, time, frequency, power, etc. can be achieved flexibly and efficiently within the uplink and downlink networks. UEs can also be flexibly associated with different UL-BSs and DL-BSs, instead of being tightly coupled to the same BS as in 5G. The deployment of UL-BSs and DL-BSs can also be separately considered, based on the statistic distribution of uplink and downlink traffic, respectively. This can save infrastructure costs compared with deploying traditional fully functional BSs in many scenarios. What is more, de-activation of UL-BSs and DL-BSs will not influence each other, thus sleeping mechanism of BSs can be freely utilized to save energy costs.

To summarize, the design objectives of FD-RAN include:

• Simplicity and flexibility of network architecture;

• Efficiency and economy of resource utilization;

• Personalization and QoE-guarantee of services provision.

B. Key Technologies

Considering the unique features of FD-RAN, several key technologies need to be further developed so as to realize FD-RAN and its benefits. This subsection introduces the no-feedback transmission at the physical layer, and the flexible resource cooperation at the MAC layer, respectively.

1) No-feedback transmission: In 5G, channel feedback is required to realize MIMO transmission. For example, in downlink, BSs first transmit pilots to UEs. Since pilots are known to UEs, these symbols can be used for channel estimation. Then, UEs calculate the channel state information (CSI) and feedback to the BSs. In 5G standards, CSI contains the precoding matrix indicator (PMI), rank indicator (RI) and channel quality indicator (CQI), which jointly determine how data signals will be transmitted through multiple antennas. However, due to the physical decoupling of UL-BSs and DL-BSs, CSI cannot be feedbacked directly.

To this end, we propose the no-feedback transmission for FD-RAN. In general, the basic idea is to use UE's location as information and infer the CSI. To realize the inference, historical channel data are required. Each channel data sample contains the channel coefficients at a specific time and location. For a DL-BS, a mapping from an arbitrary location within its coverage area to an appropriate CSI needs to be generated and stored at the DL-BS. The UE first informs the DL-BS its location through the C-BS. Then, DL-BS queries the CSI for the UE's current location from the mapping and use it to carry out MIMO transmission. The uplink is similar, where UEs utilize the mapping to determine their transmission parameters at the current locations.

Obviously, it is not trivial to generate the mapping with high performance. Generally, we have two categories of solutions, namely codebook-based and non-codebook-based. In the codebook-based solution, the optimal CSI can be derived from each sample of historical channel data. Then, we only need to find the representative CSI for each location with channel data. Since the values of CSI are all discrete, the representative CSI can be simply obtained from the statistic of all CSI at the location. For those locations without historical channel data, we can use interpolation to find the corresponding CSI from the CSI of nearby locations, or adopt a classification neural network to determine the CSI. In the non-codebook-based solution, we can directly train an end-to-end neural network, which takes the location as input, and outputs the corresponding precoding matrix for the location through the feedforward process. Note that the precoding matrix does not necessarily belong to a codebook, meaning that it has the potential to overcome the loss caused by discretization of precoding matrix. However, such neural networks are not easy to train since the dimension of inputs is too small compared with the dimension of outputs. Advanced neural network architectures and training techniques should be considered.

2) Flexible resource cooperation: In 5G, a UE is still served by a single BS. However, we believe that such a simple

service provision pattern cannot meet the diversified and personalized requirements of users in the upcoming 6G. Therefore, it is indispensable for the network to enable flexible resource cooperation on all resource dimensions and among multiple BSs.

Owing to its unique decoupling paradigm, FD-RAN can pool all the wireless resources and BSs, and employ a powerful resource scheduler to satisfy users' QoE. In order to fully utilize wireless resources, multiplexing on different domains should be considered, which are enabled by corresponding physical layer techniques. Besides, to increase the upper-bound performance of a single UE, cooperative transmission or reception techniques with multiple BSs should be used. For the downlink of FD-RAN, we mainly consider single-user MIMO with spatial multiplexing, multi-point transmission on the same or different frequency, and will integrate multi-user MIMO with multi-point transmission. Since FD-RAN can always evolve by upgrading the BSs, emerging physical layer techniques such as non-orthogonal multiple access (NOMA) and rate-splitting multiple access (RSMA)[21] may also be considered in the future. For uplink, we mainly consider receive diversity. Note that all the physical layer techniques used in FD-RAN are no-feedback transmission.

The most challenging part of flexible resource allocation in FD-RAN is how to implement an efficient resource scheduler in a relatively large-scale network comprising multiple BSs controlled by the C-BS. Also, we need to consider the possible utilization of different physical layer techniques. Note that uplink and downlink networks can use their own schedulers without having to consider each other. Normally, in the extensive literature on wireless resource allocation, an optimization problem is formulated. However, the problem is usually intractable and cannot be solved easily, making the optimization-based schedulers infeasible in practical networks. On the other hand, heuristic algorithms are easy to implement, but the performance cannot be guaranteed, especially when the solution space is very large as in FD-RAN. Therefore, we will turn to AI-based solutions, considering the recent advance of AI, especially large foundation models. Although AI has its own limits, such as the requirement on massive training data, generality to other environments, corresponding solutions are also being developed in such a rapidly growing field.

IV. Cloud Native Technologies for Resource Orchestration

A. Cloud Native Principles

The cloud-native network aims to apply cloud-native technologies to build scalable, agile network architectures. This new design approach is guided by key principles[22]:

- Use lightweight containers. Cloud-native applications compose autonomous microservices in lightweight containers to enable fast, efficient scaling. It can improve the utilization of cloud-native network resources.
- Implement loosely coupled microservices. Microservices in the cloud-native network are loosely coupled, discovering each other through service registries. This enables independent scaling to meet changing demands.
- Use APIs for service interactions. Cloud-native services interact via lightweight APIs, enabling each service to leverage the optimal language and frameworks for its function.
- Leverage DevOps processes. Cloud-native applications have independent lifecycles managed via DevOps. This allows multiple continuous integration (CI)/continuous delivery (CD) pipelines to coordinate deployment and management of cloud-native apps.
- Incorporate security across lifecycles. Cloud-native security calls for robust and seamless intrinsic security embedded across all stages of application and infrastructure lifecycles.

B. Operating System for CCNN

In order to support flexible, on-demand allocation of multi-dimensional resources and address evolving demands in increasingly complex network systems, in this paper we propose a Cybertwin-based management framework for integrated scheduling of networking, computing, and storage resources built on Kubernetes as shown in Fig. 5. This unifies virtual abstraction and management of multi-dimensional resources using containerized plugins[23]. Key components of the operating system include: virtual resource models, control plane, and data plane, which are elaborated in the following.

In the cloud-native network operating system, there are mainly three types of cloud resources: computing resources, storage resources, and networking resources. For instance, computing and storage resources can be managed through CI/CD pipelines and reasonably allocated to users for meeting the personalized resource needs of different users through containers and orchestrators. The cloud-native network operating system can extend the system resources through interface methods (CRI, CNI, CSI). This enables connecting to different backends to implement user service requirements with distributed abstraction and unified scheduling of networking-computing-storage resources.

C. Cloud-native network operating system control plane

The control plane enables global orchestration of heterogeneous networking, computing, and storage resources. It provides personalized policy configuration and flexible resource allocation tailored to individual cloud-native service needs.

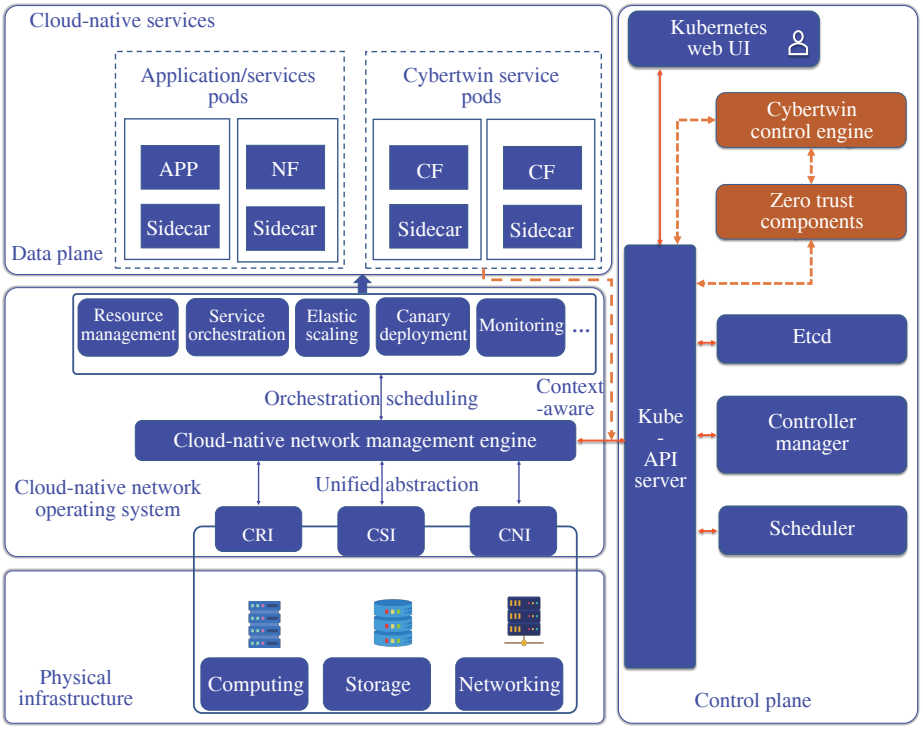


Fig. 5 A Cybertwin-based networking-computing-storage integrated scheduling framework

The master node of the cloud-native network cluster is responsible for managing the resources, providing the resource data access entry for the cloud native network. It mainly includes four components: API server, controller manager, scheduler controller, and Etcd, which are related to the actual work nodes[24].

- API server acts as the sole gateway for HTTP operations on cluster resources. It handles requests to get, create, update, or delete resources, serving as the primary control point. The API server enables unified management spanning networking, computing, and storage resources.
- Controller manager automates Kubernetes objects and workloads. It utilizes specialized controllers tailored to each resource type, handling provisioning and lifecycle management. For example, a ReplicaSet controller manages pod scaling while a Node controller monitors server health. The scheduler places workloads based on resource availability, constraints, and policies. It binds pods to optimal nodes, considering factors like capacity, affinity, and anti-affinity rules.
- Scheduler utilizes configurable filters and scoring algorithms to support different placement strategies for different users resource requirements.
- Etcd provides a consistent key-value store for cluster data and configurations. It serves as the backend database for the API server, scheduler, and controller manager. etcd enables state sharing and coordination between component.

In particular, we propose two new components to enable the implementation of the unified security policies across all cloud-native applications and services, as follows.

- Zero trust is a security model that requires strict identity verification for every access request. By building zero trust into the control plane, all network traffic and resource access can be secured through robust authentication and authorization.
- Cybertwin control engine manages identity, access control, logging, auditing, and other security functions. It interacts in real-time with the zero trust module to obtain user identities, validate access tokens, and retrieve security policies. Based on this security information, the Cybertwin engine configures and pushes out network and access rules to user Cybertwin services. For example, it allows creation of microsegmentation policies to restrict lateral movement between workloads. It also enforces role-based access control for users resources.

By centralizing security functions in the control plane, organizations gain consistent policy enforcement. Security configuration no longer needs to be manually replicated across each application and environment. Granular visibility into all network activity also improves threat monitoring and response.

In summary, the cloud-native network operating system control plane enables unified security across distributed cloud environments. Zero trust principles are applied through security methods such as software-defined perimeter (SDP), strong

identity controls and microsegmentation. Real-time coordination between the Cybertwin engine and zero trust modules automates and optimizes policy configuration. This architecture represents a significant evolution beyond traditional network security approaches.

D. Cloud-native network operating system data plane

The data plane executes the actual workload based on the management and scheduling policies defined by the control plane. The workloads are abstract resources in Kubernetes, including stateless loads (Deployment), stateful loads (StatefulSet), jobs (Job), cron jobs (Cronjob), etc. Pods are the logical management units in the cloud-native network operating system. Node provides the runtime environment to execute Pods assigned by the control plane schedulers. It associates with the Master management nodes, owning a name and IP, system resource information, containers runtime service, maintaining kubelet and load balancer (kube-proxy). Each Node runs the following set of key processes.

• Kubelet: it is responsible for tasks like creating, starting, and stopping containers corresponding to Pods, communicating with the API Server, and managing applications running on the worker nodes. In addition, Kubelet also feeds back the running status of applications to components on the master nodes. Based on the cloud-native network management and control engine, it obtains personalized service resource demands through the users Cybertwin service. And it delivers the multi-dimensional resource management and scheduling policies from the control plane to the worker nodes on the data plane. The working nodes perform personalized resource scheduling according to different users service requirements.

• Kube-proxy: an important component that implements communication and load balancing mechanisms for Kubernetes Service, which is responsible for proxying incoming request traffic to multiple instances of an application.

• Container runtime: responsible for creating and managing containers. It works with Kubelet to run application services in the cloud-native network based on provided interfaces.

The data plane provides a homogeneous runtime environment governed by the intelligent control plane. This enables application portability across the infrastructure. The reliable data plane resources in turn allow the control plane to optimize placement and orchestration. The combination enables automated management of massively distributed cloud-native applications.

In particular, the data plane of the cloud-native network operating system runs the actual network services. Telecom operators are deploying many of their communication services in a cloud-native fashion on the data plane. This provides end users with heterogeneous and diversified access options.

The Cybertwin concept acts as a proxy for users (person entity/end-user device/non-person entity). Cybertwins are deployed on the data plane as fundamental cloud services, which handles communication and mediates service access on behalf of its physical twin. Although they run on the data plane, the control plane manages the Cybertwins centrally. For instance, a Cybertwin could broker network connectivity for a mobile device user. It perceives the personalized needs of the human, such as bandwidth requirements. It then interacts with components in the control plane to facilitate the appropriate network service delivery.

The Cybertwin provides an abstraction layer that hides the complexities of the underlying infrastructure. Users (person entity/end-user device/non-person entity) simply interface with their corresponding Cybertwin to obtain services tailored to their personalized needs. The Cybertwin model also enables advanced functionality through data plane and control plane integration. The Cybertwin leverages data plane visibility into human behavior, device telemetry, and object interactions. It feeds pertinent details to the control plane.

Equipped with rich contextual data, the control plane scheduling engine can make smart decisions aided by Cybertwin services. It can orchestrate services and resources in a way that aligns to the personalized needs of each user. For example, the scheduling engine could direct network bandwidth to a Cybertwin that reports a high-definition video streaming session from its human twin. Or it could assign additional computing power to a Cybertwin managing a machine twin performing intensive batch data processing.

In essence, Cybertwins bridge the physical world their twins operate in and the virtual cloud environment. The control plane utilizes the data plane visibility Cybertwins provide into individual needs. It in turn orchestrates the network and resources accordingly. The cloud-native network operating system provides unified abstraction and management of multi-dimensional resources. It orchestrates service containers based on personalized user requirements. Leveraging service mesh and end-to-end service monitoring technologies, the cloud native network enables resource isolation, data isolation, and elastic scaling for cloud services. This ensures secure, reliable, and efficient operations across cloud-native applications.

E. Key Benefits

The key characteristics and benefits of this designed resource management framework and network operation system are:

• Decouple applications from physical infrastructure. The cloud-native network operating system provides a uniform abstraction layer above heterogeneous infrastructure and data center resources. Resources from across data centers are pooled and made available to applications via flexible, dis-

Category	Innate immune system		Adaptive immune system	
Composition	Skin/mucosa	Phagocytes, dendritic cells, neutrophils, NK cells, complement	T cells (killer, helper, regulatory, memory)	B cells, antibodies, memory B cells
Functions	(1) Block the invasion of pathogens (2) Destroy pathogens by secreting mucus	(1) Destroy pathogens (2) Phagocytose pathogens (3) Recognize, analyze, and present pathogen characteristics to activate adaptive immunity	(1) Respond to dendritic cell antigen presentation signals, adaptively select proliferating T cells. (2) Control macrophages to attack or directly kill infected cells. (3) Generate immune memory	(1) Respond to antigen presentation signals, B cells randomly mutation (2) Produce antibodies to precisely attack invading pathogens (3) Generate immune memory
Features	Non-specific immune capacity, acquired through innate inheritance, is a long-term evolutionary and inherited trait of biological populations		Specific immune ability, acquired by learning and can form immune memory	

Fig. 6 Functions of the biological immune system

tributed scheduling. This standardized runtime environment shields applications from underlying infrastructure differences and enables seamless deployment and migration.

• Automated application service deployment. The cloud-native network operating system leverages containerized applications, decoupling them from underlying infrastructure dependencies. Applications are deployed through declarative YAML manifests, standardizing and automating the deployment process. This enables efficient large-scale deployment of containerized microservices across the cloud-native environment.

• Automated application service orchestration. The cloud-native network operating system has the ability to automatically manage and orchestrate containers, which can automatically allocate and schedule resources for each container. This enables complex applications composed of many microservices to be efficiently deployed across multiple servers in the cloud-native environment. In pursuit of adhering to the cloud-native security principles, specifically the incorporation security across lifecycles, we introduce an intrinsic security architecture rooted in the concept of Cybertwin, drawing inspiration from the mechanisms observed in biological immune systems.

• Elastic scaling. The cloud-native network operating system can continuously monitor application load and demands. It will provide millisecond-level elastic scaling to rapidly provision or deprovision resources based on real-time application requirements and infrastructure capacity.

• Self-healing. The cloud-native network operating system can continuously monitor the running status of applications in the cloud-native network. When an application fails, it will automatically delete and recreate the container and related components. Also, when faults occur on the host where the application is located, it automatically performs service migration, ensuring stable operation of application services.

V. Intrinsic Security Architecture Based on Cybertwin for Cloud-Native Networks

A. Immune System

Through hundreds of millions of years of evolution, the biological immune system has developed a set of delicate mechanisms to maintain a dynamic balance between survival metabolism and defense against invasion in the body[25]. As shown in Fig. 6, the complex biological immune system can be divided into two major categories: innate immunity and adaptive immunity, which together contains four lines of defense. The innate immune system includes the isolating protection provided by skin and mucous membranes, as well as the non-specific protection offered by phagocytes and other cells. The adaptive immune system includes the adaptive immune responses of T cells, and precise response of antibody generated by B cells experiencing random mutations. It should be noted that innate immunity is the non-specific immunity inherited from birth, while adaptive immunity is the specific immunity, which is the ability to learn and form memory.

By drawing on the mechanism of the biological immune system, we propose an immunology-inspired intrinsic security architecture, which can provide intrinsic security for cloud-native networks. As shown in Fig. 7, the immunology-inspired intrinsic security architecture is mainly divided into two parts: zero trust security system and adaptive defense system, in terms of function realization.

• Zero trust security system includes two main functions. One is physical or logical isolation, i.e., achieving the isolation of users and services through the separation of the data plane and control plane, SDP, physical firewalls and other equipment; the second is dynamic authorization of access to network resources based on the results of continuous moni-

Category	Zero trust architecture		Adaptive defense functions	
Composition	Data plane and control plane separation, SDP (load balancer, virtual firewall), physical firewall	User attribute assessment, authorization for access to access (ABAC)	Distributed real-time security response	Adaptive precision response
Functions	Physical isolation or logical isolation (underlay/overlay VPN)	Dynamic authorization of access to network resources based on the results of continuous monitoring and trust level scoring for the user, equipment, environment, service	Blocking network and service being attacked using the dynamic configuration ability of container networks	Dynamic security defense based on GPT/RLHF learning, i.e., adaptive optimization of trust level scoring and access control policies in zero trust architectures
Features	Strategies, methods and means are universal and universal (responding to all changes with constancy)		Distributed collective perception and decision-making; pre-training generative basic model based on network log big data, model fine-tuning based on RLHF, continuously provide adaptive access control strategies (i.e., network antibodies) to the security policy engine	

Fig. 7 Immunology-inspired intrinsic security mechanism

toring and trust level scoring for the user, equipment, environment, service.

• Adaptive defense system also includes two main functions. The first one includes fast response, e.g., blocking network and service being attacked using the dynamic configuration ability of container networks, based on the network security situation obtained through distributed collaborative sensing; the second one is dynamic security defense based on generative pre-trained transformer (GPT)/reinforcement learning with human feedback (RLHF) learning, i.e., adaptive optimization of trust level scoring and access control policies in zero trust architectures.

Overall, zero trust security provides general and pervasive defense methods to all attacks by assuming that all objects are untrustworthy and authenticating all traffic[26-27]. Meanwhile, the adaptive defense architecture emphasizes real-time response and self-adaptation to specific attacks, and continuously generates adaptive access control policies (i.e., network antibodies) to the security policy engine.

The immunology-inspired security intrinsic architecture has the characteristic of "responding to all changes with constancy". Although it is not known what kind of attack has been encountered, as long as risks or hazards are perceived, the immunology-inspired intrinsic security architecture can resist and mitigate the attack by adjusting access control strategies. Once the risks or hazards are eliminated or reduced, it can be considered that a network antibody matching the attack characteristics (that is, specific access control strategy) has been found.

B. Zero trust security system based on Cybertwin

We realize the immunology-inspired security intrinsic architecture based on cloud-native network concept and the Cybertwin. The reason is to consider the integration of cloud-native software development, testing, deployment, and operation (Kubernetes, CI/CD DevOps) as a convenient and low-cost implementation condition for immunology-inspired intrinsic security architecture. The testing environment can be seen as parallel adjoint network (PAN)[28], and the production environment can be seen as a protected network, forming an integrated development-security-operation (DevSecOps). Meanwhile, considering that Cybertwin is a real-name proxy of a people/machine, object, or organization in the cyberspace, possessing functions such as mobile agents, transport agents, and security agents, it has a natural ability to perform immunology-inspired security functions. Thus, we have implemented a cloud-native network security architecture based on Cybertwin[16]. It includes two functional parts: a zero trust security system based on Cybertwin and the adaptive defense functions.

As shown in Fig. 8, the blue box represents the basic module of the proposed zero trust security system based on Cybertwin, including:

• Access control gateway: the execution point that authorizes access to network resources, serving as the interface between the control plane and data plane. It is typically deployed on a SDP.

• Access authorization decision module: the decision point of whether to allow access to network resources. It calculates a trust score for the network agents or Cybertwin of users (the unified combination of persons, devices, environments and services) based on the access rules provided by the security policy engine, the verification results provided by the identity verification module, and the security risk assessment provided by the security situation analysis module. This trust cost is used to determine access control decisions that match the security requirements of the access request task. It is worth noting that the access authorization is temporary and time sen-

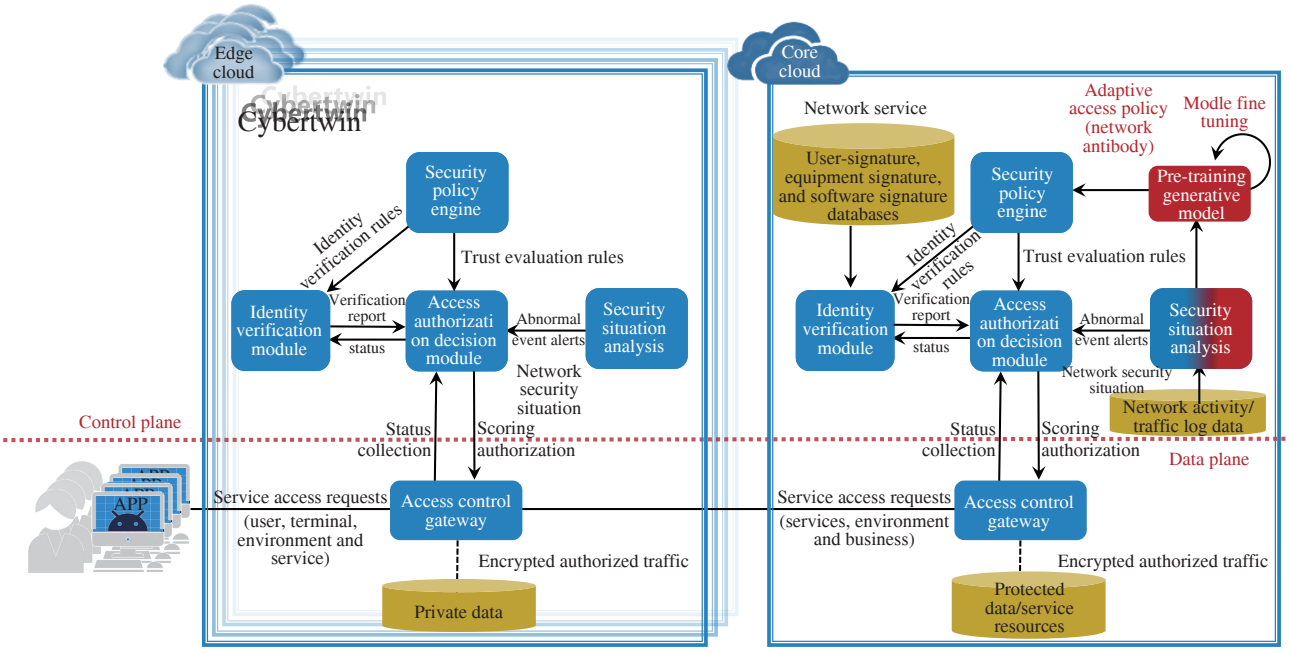


Fig. 8 Cloud-native network security architecture based on Cybertwin

sitive.

• Security policy engine: manager of security requirements and access policies. It determines the methods of identity authentication and access authorization rules according to the security requirements of tasks and provides execution strategies for the identity authentication engine and trust evaluation engine.

• Identity verification module: module for identity verification. It verifies the true identities of users, equipment, and application software based on user-signature databases, equipment signature databases, and software signature databases.

• Security situation analysis: network security situation and risk analysis. It conducts continuous status monitoring, risk assessment, and user portrait based on network activity/traffic log data, such as time attributes, space attributes, behavior attributes, etc. If there exists any abnormal situation, it will report to the access authorization module in real time, and the access authorization decision can downgrade, suspend or cancel the corresponding access authorization.

As shown in the left box in Fig. 8, users and terminals first send requests to the access control gateway of Cybertwin. The security policy engine then determines the identity authentication method and access authorization rules according to the security requirements of the access request task. The identity authentication engine sends an authentication report to access authorization decision module point according to the identity authentication method. If the identity is verified, the users are scored and authorized by the access authorization decision module based on the trust evaluation rules and the security risk assessment provided by the security situation analysis module. It is worth noting that this process is executed in a distributed manner, meaning that the Cybertwin of each entity authenticates the physical entity itself. Once the user (or terminal) has obtained authorization from Cybertwin, they can access local private data.

If a user wants to access a network service in the core cloud, Cybertwin sends a service access request to the access control gateway corresponding to the network service. The identity authentication module authenticates the Cybertwin of the user according to the user, device, software/service signature database and identity authentication rules, then sends the report to the access authorization decision module. After the Cybertwin has been authenticated, it is scored and authorized by the access authorization decision module based on the trust evaluation rules and the security risk assessment provided by the security situation analysis module.

Existing zero trust under the enterprise network also uses these five modules to authenticate and authorize business access requests. Because zero trust under the enterprise network only has one control plane, it makes centralized authorization decisions. The Cybertwin-based zero trust is distributed authentication and authorization due to the distribution characteristics of Cybertwin. Cybertwin acts as the users agent in the edge cloud, storing the users private data, and accessing the services in the core network on behalf of the user. Cybertwin not only authenticate and authorize users but also authenticate and authorize services that access user data to protect user privacy.

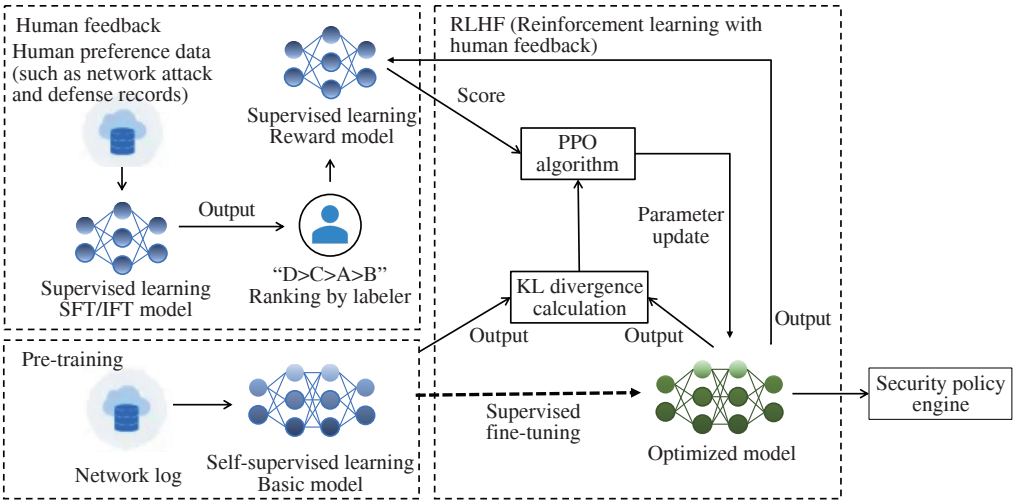


Fig. 9 Fine-tuning and strategy generation based on human feedback

C. Adaptive Defense Functions

Traditional defense techniques are no longer sufficient to address evolving and unknown cybersecurity threats. Therefore, there is an urgent need to deploy adaptive defense techniques, which can dynamically generate security policies based on security situational awareness and historical records. In our architecture, we introduce two different types of defense approaches: (i) real-time security response based on distributed decision-making and (ii) security policy generation based on pre-trained models.

- Real-time Security Response

To sense security situations, we deploy black-box and white-box probe systems. Black-box probes simulate access from outside and monitor response indicators, while white-box probes record critical operation information from inside. The data collected by both black-box and white-box probes are leveraged as the data source for security situational awareness. In order to obtain situational awareness and decision-making capabilities in distributed, decentralized systems, we introduce the Cellular Automata model[29] to our framework. Cellular Automata continuously updates group awareness and decision-making strategies according to state information. Our framework generates response policies based on situational information in real time. When anomalies are detected, the control node will automatically remove and migrate infected containers.

- Security Policy Generation

In addition to the real-time response, the proposed architecture also updates security policies (i.e., authentication and authorization policies) based on historical logs. We follow the pretrain and fine-tuning paradigm and leverage human feedback to train the policy generation model, as shown in Fig. 9. Initially, we train a Transformer-based foundation model on massive unlabeled log data in an unsupervised manner[30]. The foundation model extracts and learns semantic features among log entries. Then, we train a small-scale model using supervised learning to generate some security policies. These policies are scored by human security experts or by certain evaluation metrics. After that, a reward model is trained to automatically rate security policies. We fine-tune the foundation model with reinforcement learning (RL), which obtains scores from the reward model and optimizes the current model. Finally, after the fine-tuning stage, the security policy generation model is acquired, which provides better security policies for the zero trust security modules.

VI. Conclusion

In this paper, we have proposed a CCNN that merge the RAN, the IP bearer network, and the data center network. Based on the cloud native data center network, the CCNN uses Kubernetes as a network operating system for unified virtualization of computing, storage, network and other resources, unified scheduling and allocation, and unified operation and maintenance management. Moreover, we have proposed a fully-decoupled RAN architecture that can flexibly and efficiently utilize the resource for the personlized user services. We have also proposed a Cybertwin-based management framework built on Kubernetes for integrated networking, computing and storage resources scheduling. We have designed an immunology-inspired intrinsic security architecture with a zero trust security system and adaptive defense system. The proposed CCNN is a new network architecture to address the 6G challenges.

References

[1] CLARK D. Designing an internet[M]. Cambridge: MIT Press, 2018.

[2] CERF V, KAHN R. A protocol for packet network intercommunication[J]. IEEE Transactions on Communications, 1974, 22(5): 637-648.
[3] BERNERS-LEE T, CAILLIAU R, LUOTONEN A, et al. The world wide web[J]. Communications of the ACM, 1994, 37(8): 76-82.
[4] Kubernetes[EB].
[5] DAVIS C. Cloud native patterns[M]. Greenwich: Manning Publications, 2019.
[6] DUTT D. Cloud native data center networking[M]. Sevastopol: O'Reilly Media, 2020.
[7] GHAZNAVI M, JALALPOUR E, SALAHUDDIN M, et al. Content delivery network security: a survey[J]. IEEE Communications Surveys and Tutorials, 2021, 23(4): 2166-2190.
[8] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile edge computing: a survey[J]. IEEE Internet of Things Journal, 2018, 5(1): 450-465.
[9] Recommendation ITU-T. Computing power network-framework and architecture: Y.2501[S]. [S.l.:s.n.], 2021.
[10] XGVela[EB].
[11] CHEN C., LU G, ZHOU L, et al. Cloud native-based lightweight 5GC product and key technologies[J]. Telecommunication Science, 2020, 36(12): 89-95.
[12] LI M, TONG J, LIU Q. Research on 5G core network evolution solutions based on cloud native[J]. Information and Communications Technologies, 2020, 14(01): 63-69.
[13] AWS. 5G network evolution with AWS[R]. 2020.
[14] AWS. AWS private 5G[R]. Seattle: Amazon, 2021.
[15] WEISSBERGER A. Google cloud, Nokia partner to accelerate cloud native 5G readiness for communications providers[R]. [S.l.:s.n.], 2021.
[16] YU Q, REN J, FU, Y, et al. Cybertwin: an origin of next generation network architecture[J]. IEEE Wireless Communications, 2019, 26(6): 111-117.
[17] WANG C X, YOU X H, GAO X Q, et al. On the road to 6G: visions, requirements, key technologies, and testbeds[J]. IEEE Communications Surveys and Tutorials , 2023, 25(2): 905-974.
[18] CUI H X, ZHANG J, GENG Y H, et al. Space-air-ground integrated network (SAGIN) for 6G: requirements, architecture and challenges[J]. China Communications, 2022, 19(2): 90-108.
[19] XU M R, DU H Y, NIYATO D, et al. Unleashing the power of edge-cloud generative AI in mobile networks: a survey of AIGC services[J]. arXiv:2303.16129, 2023.
[20] YU Q, ZHOU H B, CHEN J C, et al. A fully-decoupled RAN architecture for 6G inspired by neurotransmission[J]. Journal of Communications and Information Networks, 2019, 4(4): 15-23.
[21] MAO Y J, DIZDAR O, CLERCKX B, et al. Rate-splitting multiple access: fundamentals, survey, and future research trends[J]. IEEE Communications Surveys and Tutorials , 2022, 24(4): 2073-2126.
[22] ARUNDEL J, DOMINGUS J. Cloud native DevOps with kubernetes: building, deploying, and scaling modern applications in the Cloud[M]. Sevastopol: O'Reilly Media, 2019.
[23] BURNS B, GRANT B, OPPENHEIMER D, et al. Borg, omega, and kubernetes[J]. Communications of the ACM, 2016, 59(5): 50-57.
[24] BERNSTEIN D. Containers and cloud: from LXC to docker to kubernetes[J]. IEEE Cloud Computing, 2014, 1(3): 81-84.
[25] OWEN J, PUNT J, STRANFORD S, et al. Kuby immunology (7th Edition)[M]. New York: W.H. Freeman and Company, 2013.
[26] KINDERVAG J. Building security into your network's DNA: the zero trust network architecture[R]. Cambridge: Forrester Research Inc., 2010.
[27] ROSE S, BORCHERT O, MITCHELL S, et al. Zero trust architecture-NIST technical series publications[R]. Gaithersburg: National Institute of Standards and Technology, 2020.
[28] YU Q, REN J, ZHANG J, et al. An immunology-inspired network security architecture[J]. IEEE Wireless Communications, 2020, 27(5): 168-173.
[29] SCHIFF, J. Cellular automata: a discrete view of the world[M]. Hoboken: John Wiley and Sons, 2011.
[30] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of Advances in Neural Information Processing Systems 30 (NIPS 2017). New York: Curran Associates Inc., 2017.

About the Authors

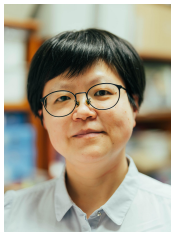
Quan Yu (SM'16-F'22) received the B.S. degree in Radio Physics from Nanjing University, Nanjing, China, in 1986, the M.S. degree in Radio Wave Propagation from Xidian University, Xi'an, China, in 1988, and the Ph.D. degree in Fiber Optics from the University of Limoges, Limoges, France, in 1992. He is currently a Research Professor with the Peng Cheng Laboratory, Shenzhen, China. He is an Academician of the Chinese Academy of Engineering (CAE) and the Founding Editor-in-Chief of the Journal of Communications and Information Networks. He was elevated to Fellow of the IEEE in 2022. His current research interests include the architecture of wireless networks and cognitive radio.

Dandan Liang (S'09-M'14) received the Ph.D. degree in Wireless Communications from the University of Southampton, UK, in 2013. Upon completion of the Ph.D. she conducted research as a Research Fellow at the University of Surrey, UK. In 2017 she joined the research centre of Huawei Technologies in Shenzhen, Shenzhen, China, working as Senior Engineer of wireless communications and standardization of Wi-Fi. She is currently an Associated Professor at Peng Cheng Laboratory, Shenzhen, China. She has published 17 research papers in IEEE journals and conferences and applied 67 patents. She has broad research interests including PHY layer modelling, cross-layer system design as well as network architecture.

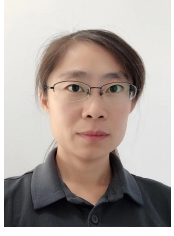
Meng Qin received the B.S. degree in Communication Engineering from the Taiyuan University of Technology, China, in 2012, and the M.S. and Ph.D. degrees in Information and Communication Systems from Xidian University, Xi'an, China, in 2015 and 2018, respectively. He worked as a Postdoctoral Fellow with School of Electronic and Computer Engineering, Peking University and Peng Cheng Laboratory, Shenzhen, China. He is currently an Assistant Researcher with Department of Strategic and Advanced Interdisciplinary Research, Peng Cheng Laboratory, Shenzhen, China. His research interests include Cybertwin-based cloud native network, green cloud storage, AI-aided self-organized wireless networks, edge intelligence in wireless networks.

Jiacheng Chen received the Ph.D. degree in Information and Communications Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2018. From Dec. 2015 to Dec. 2016, he was a Visiting Scholar at BBCR group, University of Waterloo, Waterloo, Canada. Currently, he is an Assistant Researcher in Peng Cheng Laboratory, Shenzhen, China. His research interests include future network design, 5G/6G network, and resource management.

Haibo Zhou (M'14-SM'18) received the Ph.D. degree in Information and Communication Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he was a Postdoctoral Fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. He is currently a Full Professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. He was a recipient of the 2019 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award, 2023-2024 IEEE ComSoc Distinguished Lecturer, and 2023-2025 IEEE VTS Distinguished Lecturer. He served as Track/Symposium Co-Chair for IEEE/CIC ICCC 2019, IEEE VTC-Fall 2020, IEEE VTC-Fall 2021, WCSP 2022, IEEE GLOBECOM 2022, IEEE ICC 2024. He is currently an Associate Editor of the IEEE Transactions on Wireless Communications, IEEE Internet of Things Journal, IEEE Network Magazine, and Journal of Communications and Information Networks. His research interests include resource management and protocol design in B5G/6G networks, vehicular ad hoc networks, and space-air-ground integrated networks.



Jing Ren (Member, IEEE) received the B.E. and Ph.D. degrees in Communication Engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2007 and 2015, respectively. Currently, she is an Assistant Researcher at UESTC. Her research interests include network architecture, protocol design, network modeling, optimization, and network security.



Ying Li received the B.E. and Ph.D. degrees in Communication Engineering from the National University of Defense Technology, Changsha, China, in 2001 and 2006, respectively. Currently, she is a Researcher at Peng Cheng Laboratory, Shenzhen, China. Her research interests include network architecture design, MIMO, and cognitive radio.

Jun Wu (M'05-SM'15) received the B.S. degree in Information Engineering and the M.S. degree in Communication and Electronic System from Xidian University in 1993 and 1996, respectively, and the Ph.D. degree in Signal and Information Processing from the Beijing University of Posts and Telecommunications, Beijing, China, in 1999. He is a Full Professor in School of Computer Science, Fudan University, Shanghai, China. He was a Professor in Department of Computer Science and Technology, Tongji University, and he served as a principal scientist in Broadcom before he joined Tongji University. His research interests include wireless network, machine learning and signal processing.

Yue Gao (S'03-M'07-SM'13) received the Ph.D. degree from the Queen Mary University of London (QMUL), U.K., in 2007. He is a Chair Professor at the School of Computer Science, Director of the Intelligent Networking and Computing Research Centre at Fudan University, China and a Visiting Professor at the University of Surrey, UK. He has worked as a Lecturer, Senior Lecturer, Reader and Chair Professor at QMUL and the University of Surrey, respectively. His research interests include smart antennas, sparse signal processing and cognitive networks for mobile and satellite systems. He has published over 200 peer-reviewed journal and conference papers. He was a corecipient of the EU Horizon Prize Award on Collaborative Spectrum Sharing in 2016 and elected an Engineering and Physical Sciences Research Council Fellow in 2017. He is a member of the Board of Governors and Distinguished Lecturer of the IEEE Vehicular Technology Society (VTS), Chair of the IEEE ComSoc Wireless Communication Technical Committee, and past Chair of the IEEE ComSoc Technical Committee on Cognitive Networks. He has been an Editor of several IEEE Transactions and Journals, and Symposia Chair, Track Chair, and other roles in the organising committee of several IEEE ComSoC, VTS and other conferences.

Wei Zhang [corresponding author] (S'01-M'06-SM'11-F'15) received the Ph.D. degree from the Chinese University of Hong Kong in 2005. Currently, he is a Professor at the School of Electrical Engineering and Telecommunications, the University of New South Wales, Sydney, Australia. His current research interests include 5G and beyond. He received 6 best paper awards from IEEE conferences and ComSoc technical committees. He was elevated to Fellow of the IEEE in 2015 and was an IEEE ComSoc Distinguished Lecturer in 2016-2017. He is Vice President of IEEE Communications Society.

Within the IEEE ComSoc, he has taken many leadership positions including Member-at-Large on the Board of Governors (2018-2020), Chair of Wireless Communications Technical Committee (2019-2020), Vice Director of Asia Pacific Board (2016-2021), Editor-in-Chief of IEEE Wireless Communications Letters (2016-2019), Technical Program Committee Chair of APCC 2017 and ICCC 2019, Award Committee Chair of Asia Pacific Board and Award Committee Chair of Technical Committee on Cognitive Networks.

In addition, he has served as a member in various ComSoc boards/standing committees, including Journals Board, Technical Committee Recertification Committee, Finance Standing Committee, Information Technology Committee, Steering Committee of IEEE Transactions on Green Communications and Networking and Steering Committee of IEEE Networking Letters. Currently, he serves as an Area Editor of the IEEE Transactions on Wireless Communications and the Editor-in-Chief of Journal of Communications and Information Networks. Previously, he served as Editor of IEEE Transactions on Communications, IEEE Transactions on Wireless Communications, IEEE Transactions on Cognitive Communications and Networking, and IEEE Journal on Selected Areas in Communications Cognitive Radio Series.