Self-Evolving and Transformative (SET) Protocol Architecture for 6G

Lin Cai, Jianping Pan, Wenjun Yang, Xiangyu Ren, Xuemin Shen

***Abstract*—The fusion of digital and real worlds in all dimensions will be the driving force for future sixth-generation (6G) wireless systems. Ubiquitous in-time and on-time communication services between humans, machines, robots, and their virtual counterparts are essential, and they expand from the ground to air, space, underground, and deep sea. 6G systems are not only data pipelines but also large-scale distributed computing systems with integrated sensing, processing, storage, communication and computing capabilities. It is challenging to build ubiquitous and intelligent 6G systems, handling stringent quality-of-service (QoS) requirements, providing a rich set of communication modes, including unicast, multicast, broadcast, in-cast, and group-cast, and supporting user-centric mobile applications. In this article, we propose a new protocol architecture, Self-Evolving and Transformative (SET) architecture that can provide a wide range of control functions, and be intelligently configured for different types of 6G applications and networking environments. Its design principles, potentials, and open issues are discussed.**

I. Introduction

Wireless Internet, connecting more than half of the world population and tens of billions of devices, is an enabling technology for the coming Fourth Industrial Revolution, where the physical and digital worlds are connected for collaboration across departments, partners, products, and people. The recent wireless cellular standard, fifth-generation (5G) new radio (NR) has been developed and countries worldwide are launching 5G networks since 2019. Now is the pivotal time for fundamental research preparing for the next generation wireless systems, sixth-generation (6G), from network architecture to protocols supporting new 6G applications.

6G is anticipated to have the following new capabilities, compared to the current 5G standards and the Internet architecture developed in the 1970s. First, 6G will support massive machine-type communications with ubiquitous air, space, above/under-ground, above/under-water coverage. Second, 6G will support distributed applications through various service modes, including multicast, broadcast, incast, and group-cast. Third, 6G will support in-time and on-time services with ultra-high reliability and extremely-low latency requirements. Fourth, 6G will support ubiquitous intelligence services with integrated sensing, processing, storage, communications, and computing functions. In other words, 6G systems are not only data pipelines but also large-scale distributed computing systems, so communications, computing, and storage resources need to be managed together.

The Internet architecture was designed for supporting mainly point-to-point links with limited user mobility and best-effort services, not feasible to address the above 6G challenges. In the wireless domain, 5G takes the Software-Defined Networking (SDN) and Network Function Virtualization (NFV) approach, where the control plane functions are virtualized and placed in the cloud. However, the all-cloud approach cannot deal with small time-scale (sub-second level) control functions, insufficient to support many 6G applications.

To address the 6G challenges, this article presents a new protocol architecture. We apply the in-network flow/packet control in the control plane, which complements the in-cloud session control and user control. In each network entity, a protocol control agent is deployed to handle the flow/packet level control, which can assemble, configure, and switch protocol functions, making the protocols self-evolving and transformative (SET). With the distributed and autonomous in-network intelligence embedded in the new architecture, protocol control agents can coordinate with each other to handle new 6G services and scenarios.

In the following, we first investigate the 6G challenges and opportunities in Sec. II, and brief the recent network architectures and solutions in Sec. III. 6G system architecture and the SET protocol architecture are presented in Sec. IV and Sec. V, respectively. Several 6G use cases are given in Sec. VI, followed by conclusions and further research issues in Sec. VII.

II. 6G CHALLENGES AND OPPORTUNITIES

A. 3D coverage

5G and the previous generations aim to provide connectivity mainly in a two-dimensional space, i.e., network access points provide connectivity to ground equipment. In the 6G era, both human beings and things need to stay online anytime, anywhere, from ground to air, space, underground, and deep sea. As shown in Fig. 1, 6G will provide three-dimensional (3D) coverage, supplementing ground infrastructures with High-Altitude Platform Systems (HAPS) (drones, balloons, etc.) and satellites. These mobile elements can be deployed quickly to ensure continuity and reliability of seamless services, such as in rural areas or during events triggering a high volume of data traffic, thereby avoiding the operation and management costs of an always-on, fixed infrastructure.

The 3D coverage brings new challenges, including air-to-ground channel selection and link maintenance, topology and trajectory monitoring and optimization, resource management, and energy efficiency. In addition, since some 6G network backbones (e.g., satellite networks) and access networks (e.g., HAPS) can also be mobile, network protocols need to support not only user mobility, but also network mobility.

The existing protocols designed for terrestrial networks are difficult to deal with these challenges. For instance, data integrity and Internet stability rely on the end-to-end Transmission Control Protocol (TCP) protocol, which has poor performance for the non-terrestrial environment with long propagation delay, high jitter, and high uncertainty. In the network layer, a fixed-length IP address associated with the access network is inefficient for devices with limited resources and energy, incapable to support a massive number of connections with heterogeneous service requirements, and prohibitively costly to handle network mobility.

B. Stringent QoS requirements

In 6G, extended reality (XR) (the umbrella term for virtual reality, augmented reality, and mixed reality) using wearable devices, high definition images and holograms for remote diagnosis and surgery, and Metaverse communications between humans, things, and cyber systems will become realistic and rich. These emerging applications require stringent Quality of Services (QoS), such as extremely low delay and jitter, high reliability, and high data rate, as shown in Table I.

The dominant transport protocol, TCP, can ensure data reliability, but with no throughput, delay, or jitter guarantee. The IP layer is designed based on the best-effort concept, incapable of QoS guarantee. Overall, the current network architecture cannot support multi-dimensional QoS requirements.

C. *-cast communication modes

6G applications demand a rich set of communication modes among distributed entities, such as multicast, broadcast, incast, group-cast, and geocast, called *-cast, as shown in Fig. 2.

A representative *-cast scenario is for Digital Twin (DT), where users can monitor states, detect problems, and make changes remotely through the virtual twin of a physical object. For DT, how to reliably exchange sensing data and control messages between the physical twin, its virtual sibling, and the DT database with high fidelity, low energy cost, and ultra-low latency is critical. Other use cases of *-cast include the space-terrestrial integrated network, Industry 4.0, and unmanned mobility.

*-cast is challenging to networks mainly designed for unicast. When multiple end hosts request the same information, applying unicast to support them leads to lower efficiency and a higher chance of congestion. The problem is exaggerated in wireless networks where wireless channels are broadcast in nature and the spectrum is at a premium. For 6G, new *-cast communication modes should intelligently replicate data and distribute them to the receivers efficiently with a QoS guarantee.

D. Ubiquitous intelligence

Artificial intelligence (AI) is an indispensable part of 6G. 6G should be flexible to provide customized services to support large-dimension, heterogeneous, high-complex space-air-ground-water environments [1], [2], where AI is of vital importance [3], [4]. There will be a variety of AI applications supporting 6G networks, from resource/mobility management to network planning/optimization.

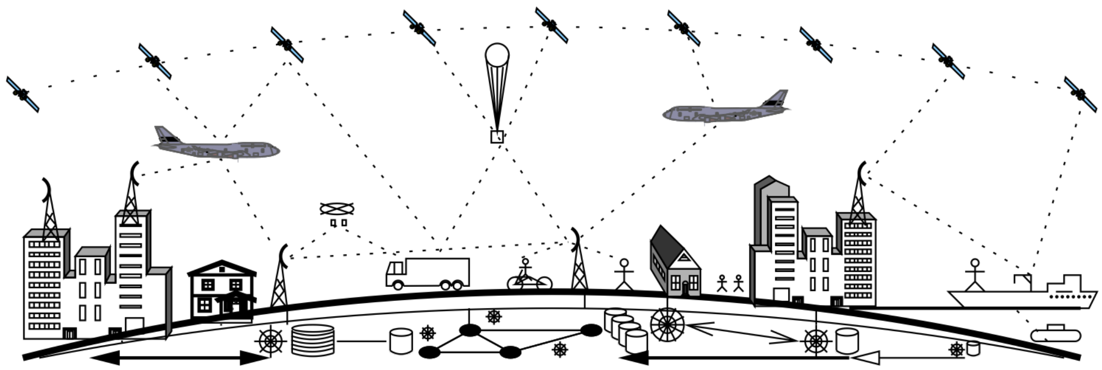

Fig. 1. 6G with ubiquitous space-air-ground-water coverage.

TABLE I
QoS Requirements for 6G Applications

Applications	Data rate	Delay	Loss	Jitter
Industrial control/Digital Twin	1 Gbps - 1 Tbps	< 1 ms - 10 ms	10^{-8} - 10^{-6}	100 μs
Immersive XR	1 Gbps - 100 Gbps	< 1 ms - 10 ms	$\leq 10^{-6}$	100 μs
Mobile hologram/Metaverse	1 Gbps - 1 Tbps	< 1 ms - 10 ms	$\leq 10^{-6}$	100 μs
Machine Learning	100 Mbps-10 Gbps	< 100μs - 10 ms	$\leq 10^{-5}$	$\leq$ 10 ms

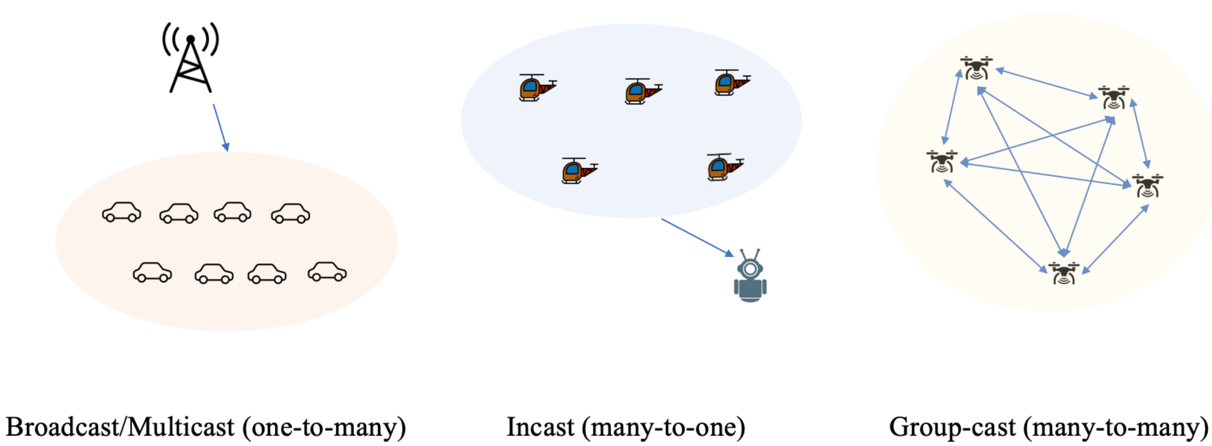

Fig. 2. *-cast communication modes.

In addition to applying AI for 6G, 6G can also enable and enhance future AI techniques such as distributed, large-scale machine learning (ML). ML has demonstrated its incredible capability to solve complex tasks. The state-of-the-art ML models require a large amount of data and a high volume of computing power for the training process. Training ML models with a large number of computing devices in a distributed way can improve training efficiency, ensure data privacy, and satisfy user-specific requirements. For example, with limited local data and energy and computing resources, mobile devices can first train a simple ML model independently and locally, and then aggregate to produce a more powerful ML model by servers at the edge or in the cloud.

Such a training hierarchy with mobile devices, edge servers and cloud servers can ensure a timely response and high accuracy, leveraging the data from others without compromising data privacy. With the increasing scale of the ML models, the communication requirements between training mobile nodes become higher. Therefore, 6G, with higher throughput and lower delay services, will be a critical infrastructure for training large-scale machine learning models, leveraging heterogeneous mobile end devices and orchestrating computing resources of end devices, edge servers, and cloud.

Furthermore, a new generation of mobile networks often relies on new communication technologies that provide unprecedented performance and functionality. For example, massive multiple-input multiple-output (MIMO) and millimeter-wave (mmWave) communications are key drivers of 5G networks. For 6G, in addition to the conventional spectrum (i.e., sub-6 GHz and mmWave), wireless networks may rely on new frequency bands such as Terahertz band and Visible Light Communications (VLC). From mmWave, Terahertz, to VLC, 6G wireless signals not only carry data information, but also have inherent sensing capabilities for location, position, imaging, material-detection, etc. With information sensing, processing, storage, communication and computing functions, 6G will become a large-scale intelligent system, bringing profound impacts to future society.

III. Existing approaches

There have been many attempts to enhance the current protocol stack or develop clean-slate architectures. We select a few approaches highly relevant to the above challenges. The literature scan here is neither comprehensive nor exhaustive.

A. New transport layer protocols

New transport layer protocols have been developed for important applications. QUIC [5] was developed to better support web applications by reducing TCP handshake time and using pluggable congestion control, using delay measurement for fast recovery. It also supports multiple streams to mitigate head-of-line blocking and uses connection ID to support connection migration. Xpress Transport Protocol (XTP) [6] can support a range of applications with additional services such as transport multicast, multicast group management, transport layer priorities, and traffic descriptions for QoS negotiation. With a fixed packet header, XTP has limited capacity for new service parameters and negotiations. Space Communications Protocol Specification Transport Protocol (SCPS-TP) [7] was designed for satellite networks, using explicit congestion notification, error code identification, and link interruption messages to identify different error sources in satellite networks and handle them accordingly.

B. Network layer enhancement

Virtual circuit (VC) technology relies on admission control with statistical reservation to provide statistical service guarantee. However, reservation-based solutions cannot explore all paths available in the mesh-topology network, and they are slow in response to network dynamics.

Deterministic Networking (DetNet) can control and reduce end-to-end latency in IP networks by determining transport paths with the worst-case bounds for latency, loss, and jitter [8]. With Segment Routing (SR), the source node specifies the forwarding path and converts the path into an ordered segment list in the message header, and routers forward packets accordingly [9].

Service function chaining (SFC) is an architecture to interconnect and maintain virtualized network function (VNF) services in a (partial) order applied to each flow. As many service functions are implemented at the hardware level which is complex to optimize, frequent changes in services lead to high computation costs and long delays.

Virtual routing and forwarding (VRF) allows multiple instances of a routing table to co-exist within the same router simultaneously. Since the routing instances are independent, the same or overlapping IP addresses can be used without conflicts. VRF as a form of network function virtualization (NFV) reduces hardware costs by converting hardware-based network appliances to software.

In path-aware networking, path information is explicit to the network and transport layer protocols, so endpoints can select a path for a packet or flow to meet specific QoS requirements. A few path-aware techniques have been proposed, such as Stream Transport, Integrated Services, Quick-Start TCP, SHIM6, and IPv6 Flow Label. They are not widely adopted yet, due to implementation costs.

C. Software defined protocol stack

Software-defined protocol stack (SDP) [10] aims to provide a set of services based on application requirements under the control of an SDP controller. Although it can select suitable protocols based on the QoS or quality of experience (QoE) requirements, the SDP controller does not allow configuring each protocol, which limits the scope of services.

D. Clean-slate Internet architectures

Several clean-slate Internet architectures have been proposed to address the IP network weakness. In Named Data Networking (NDN), the namespace is used for packet forwarding instead of IP, with the advantages of built-in multicast, in-network caching, and multipath forwarding capability. It encounters challenges in name server management and name resolution efficiency.

Recursive InterNetwork Architecture (RINA) builds upon the perspective that networking is not a layered set of different functions but rather a single layer of distributed Inter-Process Communication (IPC) that repeats over different scopes [11]. Communicating hosts only need to know each other's names and use the IPC interfaces to request communication. The upper IPC layer may recursively request services from its underlying IPC layers, thus forming a multi-layer structure repeated until an IPC facility can fit well with the physical medium. Thanks to the recursion feature, RINA can be more adaptive for finer-grained tasks.

The deployment of clean-slate architectures has been a big challenge. Upgrading the whole Internet is difficult if not possible, given its scale. It is more feasible to focus on domain-specific needs and develop a new architecture for some domains. The advances in these domains may influence others and eventually change the Internet architecture. Here, we focus on the new architecture for 6G for a better adoption opportunity.

E. 5G architecture

Since 4G, cellular systems have moved to all-IP networks to support various Internet applications. 5G further takes an open, software-defined model to decouple the control plane and data plane (named Control and User Plane Separation in the 3rd Generation Partnership Project (3GPP)). By shifting to an intelligent software layer running on commodity hardware, many user- and session-level control functions, such as Core Access and Mobility Management Function, Session Management Function, Policy Control Function, Unified Data Management, and Authentication Server Function, can be virtualized and deployed in 5G Mobile Core cloud as a service chain.

However, the 5G architecture still lacks the QoS support for many 6G applications, such as guaranteed delay and throughput per flow or per packet. The virtualized control functions in the Mobile Core are insufficient or too slow to manage flow/packet-level control. How to manage computing resources is not fully addressed in 5G either.

IV. 6G System Overview and Design Objective

A. 6G system overview

For 6G, in addition to the decoupling of data plane and control plane, we further divide the control plane based on control time-scale. As shown in Fig. 3(b), there are three levels of control in the control plane, the flow/packet level, session level, and user level. The flow/packet control is closely coupled with the data plane to handle individual flow/packet transmission related decisions. The session control and user control handle the session management and user accounting related functions, respectively.

Session and user control functions can be virtualized and deployed in the mobile core or at edge servers to handle the requests received by the base stations in Radio Access Networks (RANs), as their control action time can be longer. The session control can also manage the computing and caching functions in both 6G RANs (with edge servers) and mobile core (with cloud servers), as shown in Fig. 3 (b). However, flow/packet control needs to make decisions in sub-second time, which should be close to the data plane in the network, so it should be deployed in each network entity, as shown in Fig. 3 (a).

To handle flow/packet control, a protocol control agent (PCA) (a running process) is deployed in each network entity, which can take the input from applications and network/link/channel measurements, negotiate with other PCAs, and apply advanced learning technologies to assemble, configure, and switch functions of the data plane protocols. Then, the corresponding protocols can handle flow/packet-level service requirements timely and intelligently, as discussed in the next section.

B. Design objective

The existing TCP/IP protocol architecture has an hourglass model, i.e., a few transport-layer and network-layer protocols support all applications over all types of communication systems.

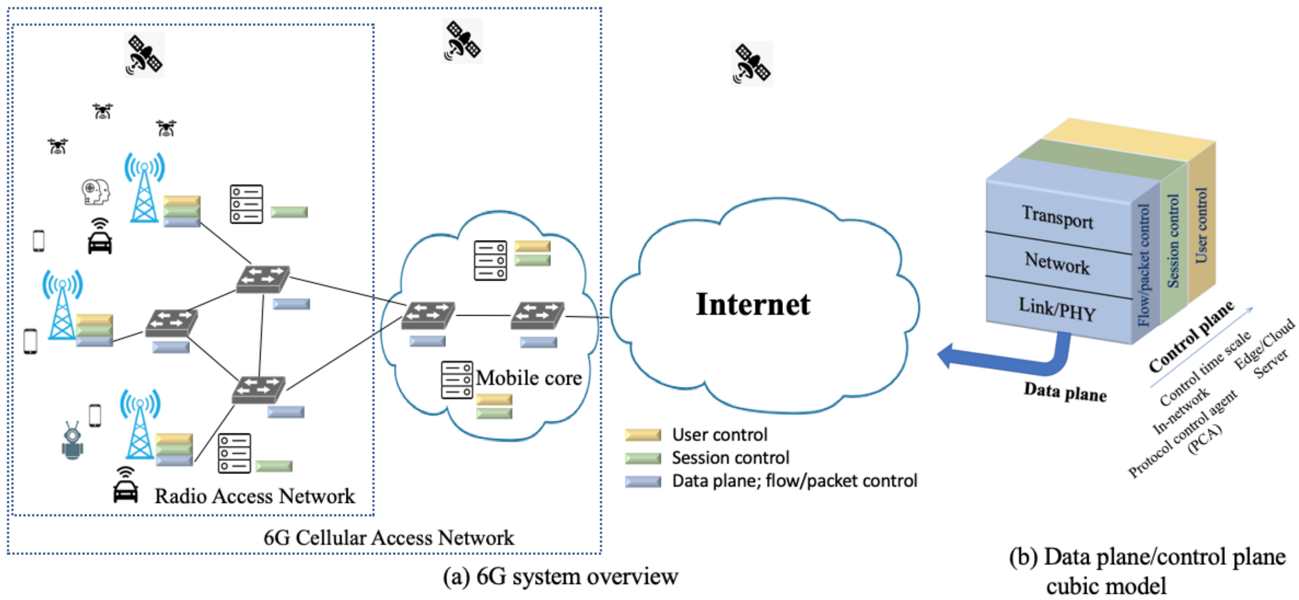

Fig. 3. 6G system overview and data/control-plane protocol model.

Applications may choose the all-in-one TCP with all the control blocks (error/flow/congestion control and connection management) or the User Datagram Protocol (UDP) with none. Some applications can tolerate a certain degree of packet losses but prefer timely and throughput-smooth services, so neither TCP nor UDP works well. In the network layer, the IP protocols are the only choices, where the interactions between the end-points and the intermediate systems are very limited, and not desirable for QoS provisioning.

Developing a new protocol for each new application in the transport and network layer will be too costly, given the fast advances of applications and the high heterogeneity of 6G communication systems. Therefore, based on the 6G system architecture in Fig. 3, the objective is to develop a new protocol architecture that can assemble a protocol to apply a wide range of control functions and be flexibly configured for different applications, supporting QoS, *-cast, 3D coverage, and ubiquitous intelligence.

V. SET PROTOCOL ARCHITECTURE

A. Developing SET architecture

Protocol architecture includes the data plane protocols and the control plane. In the control plane, we mainly focus on the flow/packet control which manages the data plane protocols in a small time-scale.

Learning from domain-specific computer architecture, object-oriented programming languages, and modular operating systems, we apply the domain-specific, object-oriented, and modular design principles to network protocols. With intelligent PCAs to manage the protocol modules, a protocol can be self-evolving and transformative (SET) to provide a wide range of services and be adaptive to the 3D environment. To make it feasible, we take a three-step approach.

The first step is protocol function decomposition. We first identify basic protocol functional blocks, and each block is independent and self-contained. Such decomposition allows the flexible assembly of selected blocks into complete, customized protocols. Protocol decomposition can help remove unnecessary redundancy and repetition in the protocol stack when PCAs assemble protocols.

Note that the existing layered architecture can be viewed as a type of "decomposition", but at a much coarse level. It requires to develop a new protocol for a new application if the existing protocols cannot meet the service requirements, and the protocol cannot be changed during the lifetime of a flow, which is a major limitation for many 6G applications in the highly dynamic 3D environment.

The second step is protocol assembly by PCAs based on the service requirements and network measurements. It includes details about how the protocol functional blocks are stitched together to form a

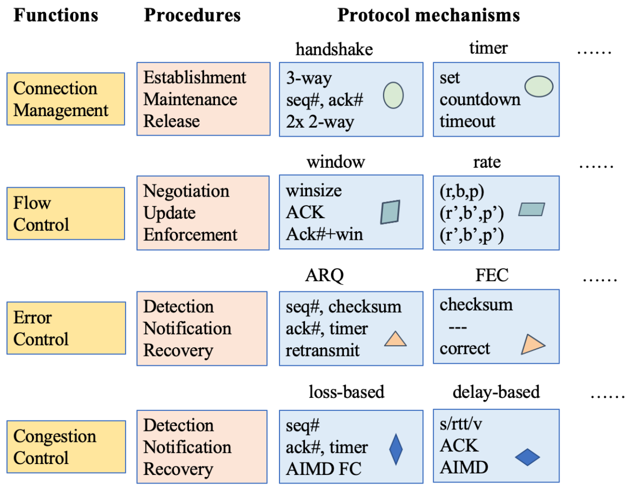

Fig. 4. Decomposition of the transport layer control functions.

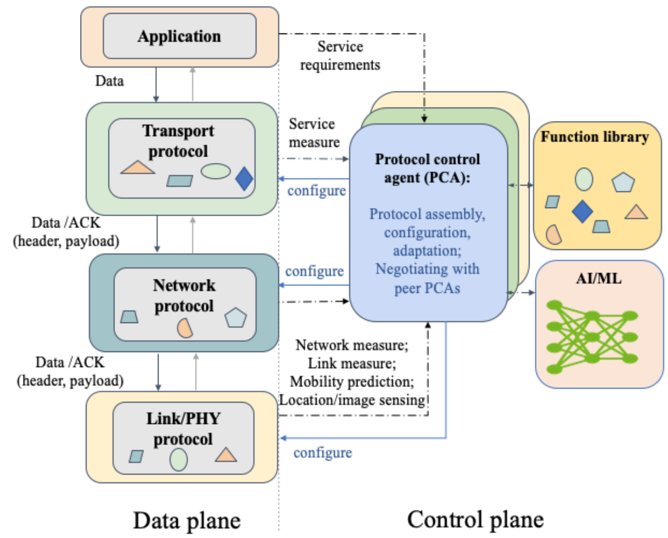

Fig. 5. SET protocol architecture.

complete protocol, how to define packet header with protocol control and configuration information, and how to interpret it during protocol processing. If multiple network elements are involved, control signaling between the corresponding PCAs is defined.

The last step is to develop intelligent control algorithms for PCAs to optimize protocol performance, by learning the network environment, measuring and estimating the service quality, and switching or re-configuring the protocol control functions.

B. Protocol control function decomposition

We use transport layer protocols as an example to demonstrate control function decomposition. In Fig. 4, the decomposed control functions include connection management, error control, flow control, and congestion control. For each control function, there are several possible policies and algorithms. For instance, connection management can be handshake- or timer-based; flow control can be window- or rate-based; congestion control can be loss-based, delay-based, or both; error control can use Automatic Repeat reQuest (ARQ), forward error correction (FEC), or a hybrid ARQ/FEC approach. These independent control functions and algorithms construct the control function and algorithm library. To deal with new service requirements, such as throughput-guarantee or delay-guarantee, more control blocks can be developed.

When we decompose the protocol into M control functions and each with N algorithms, the design space of the protocol can be enlarged to N^M, while the implementation space cost increases linearly $(N \times M)$. Therefore, by selecting and configuring control functions by the PCA with AI/ML support, a single SET protocol can provide a wide range of services. Since the design space is large, to avoid the long delay in searching the optimal configurations, PCA can prepare a lookup table with default settings for widely used services, and it can apply the online reinforcement learning technologies such as Multi-Armed Bandits (MAB) algorithms to switch or reconfigure the control functions when needed.

C. Protocol configuration and adaptation

Given the application service requirements, the controller (i.e., PCA) selects suitable control algorithms from the control function library, based on its knowledge, as illustrated in Fig. 5. The PCA (a running process in each entity), assisted by advanced learning and control mechanisms, configures the corresponding protocols in different layers according to the service requirements and the network, link, and physical layer (PHY) measurements collected locally or from the network.

To make it simple, the PCA can prepare a few default settings for different types of services, e.g., reliable and streaming service with throughput guarantee, unreliable and streaming service with elastic throughput, ultra-reliable and low-latency service, etc. The settings can be refined or changed during the session.

Since both the application traffic and the network environment may have high uncertainties, the default settings may not be optimal, or even infeasible to satisfy service requirements. In these cases, not only the protocol parameters (such as the congestion window increase/decrease rate/ratio [12]) can be adapted, but also the protocol control algorithms can be switched.

We give an example of how the agent adapts the protocol functions. A user application can "pay" for a high-priority service with the data rate, delay, and loss guarantee. The protocol agent starts with the default policies and algorithms at a low cost, e.g., window-based flow control and ARQ error control for the transport layer. Then the protocol agent observes QoS and the network performance. If QoS is no longer satisfied, it switches to higher-cost policies and algorithms, e.g., adding end-to-end FEC when some communication link has a high error rate or applying multi-path/multi-copy transmissions to achieve higher reliability. Advanced learning algorithms can be developed and applied to assist the PCA.

To ensure fast control for short flows or highly dynamic environments, PCA can apply online reinforcement learning such as the simple but effective multi-armed bandits techniques without prior training and with sequential learning and decision making.

D. SET features

1) In-network, distributed intelligence. The new architecture relies on the in-network intelligence of PCA deployed in each network entity (both end-systems and intermediate nodes). A PCA can take information not limited to that in traditional layered architecture and collaborate with others to make coordinated decisions for end-to-end QoS support. The distributed, in-network intelligence can react to channel, network, and application dynamics faster, more robust, and more efficiently.

Note that, with the SET architecture, the data plane protocol agents (PAs) and control plane protocol control agents (PCAs) are separated. In other words, once the PCA creates and configures the protocols, the corresponding PAs can handle the transmissions independently and efficiently, without waiting for further instructions from the PCA. The PCA is an independent running process to learn the network and service conditions, and make switching decisions on protocol control functions when necessary. The switching cost and delay will also be considered by the PCA to ensure that such switching is necessary and overall beneficial. Furthermore, the SET protocol architecture can be backward compatible with the traditional TCP/IP, as the existing protocols are among the options for PCA to choose.

2) Self-evolving and transformative. Note that decomposition of the existing solutions can prepare us with a library of control blocks to solve a wide range of problems. If a new problem is not solvable with the existing functions, new control functions can be developed. The growing function library makes SET protocols *self-evolving*. As a PCA can switch functions for a single flow when the service requirements or environments change, the SET protocols can be *transformative* in dynamic environments.

3) Backward compatibility and incremental deployment. PCAs can configure protocols following the existing TCP/IP layered protocol stack, so it is backward compatible and can be incrementally deployed. When a new node supporting SET works with other nodes using the existing architecture, its PCA can select the existing or compatible protocols to work with the others. Meanwhile, the PCA can still make some local optimization decisions to improve the system performance. For instance, the PCA of an end system (sender) can use a hybrid ARQ/FEC approach to deal with error control. If the other end system (receiver) also supports the option, the pair can adapt the control based on the delay/loss requirements and network conditions accordingly. Otherwise, the sender can still use the traditional TCP error control (ARQ-type) to ensure reliability at the cost of high delay/jitter.

On the other hand, PCAs can also negotiate with each other to support non-IP protocols and even clean-slate protocol stacks when needed, desirable for many *-cast applications.

4) Supporting virtualized link/PHY. By decomposing the link/PHY layer protocols, SET can configure virtualized link/PHY layers, which are decoupled from dedicated hardware. Devices can be equipped with heterogeneous 6G radio technologies, and PCAs can configure the link/PHY protocols to efficiently utilize multiple access links according to the service requirements and mobility scenarios.

VI. 6G USE CASES

We give a few 6G use cases based on the SET protocol architecture.

A. Distributed coordination for QoS guarantee

In the TCP/IP protocol stack, QoS support relies on the control blocks of multiple layers, including the physical layer (transmission power, modulation, coding, bit-level error control, etc.), link layer (access control, queue and resource management, link error control, etc.), network layer (routing, fragmentation, reassembly, etc.), and transport layer (flow/congestion/end-to-end error control, etc.). Reliability can be achieved given the error controls in different layers with limited or no inter-layer interactions. However, without careful inter-layer coordination, IP networks cannot ensure end-to-end throughput and delay/jitter.

For reliability with SET, the PCA can process both the application service requirements and the transport, network, link-layer measurements and adapt the configurations of multiple layers' protocols jointly, e.g., choosing link-layer retransmission and/or multi-path duplication for ensuring higher reliability.

By enabling distributed, autonomous in-network intelligence, SET is more powerful to break the limit of the existing layered architecture and can provide new services otherwise not possible, such as delay-guaranteed service for each packet or each flow.

For instance, SET protocols can configure the protocol headers to include per-packet delay requirements. Different packets from the same flow can carry different delay requirements and the intermediate systems can react to the per-packet QoS requirements to choose suitable control functions (e.g., priority queue) and parameters (e.g., using delay-based routing metrics) accordingly. Furthermore, with distributed, in-network intelligence by PCAs, intermediate nodes can jointly explore the explosive number of paths in mesh networks for higher reliability, throughput, and lower delay [13], much more efficient and robust than reserving a limited number of paths by a centralized/cloud-based SDN controller.

B. User-centric spectrum and mobility management

When a user equipment (UE) moves out of its current access point's coverage area, it leads to a handover event or connection interruption. To guarantee continuity during a handoff, methods at different layers were introduced. In the network layer, mobile IP (MIP) uses an agent to hide the mobility from the correspondent node, Locator/ID separation protocol (LISP) takes a mapping-based approach, and Network Driving IP Mobility (NDM) can support network mobility in Low-Earth-Orbit (LEO) networks. They all use single-link communications, suffering from disconnection due to hard handoff. To avoid disconnection, the transport layer in the UE can maintain multiple paths, using the multiple access links available during a handoff or in a dense network, e.g., MP-TCP and MM-QUIC [14], [15].

SET can support user-centric spectrum and mobility management, leveraging the above multi-path benefits. The PCA at a UE can manage multiple paths by using multiple interfaces to connect multiple APs using corresponding (virtualized) MAC and PHY components. For heterogeneous networks with both satellite and ground access links, UEs can access two or more networks with the same or different service providers in parallel.

PCAs of the UE and the RAN can collaborate to optimize link selection and combination. The PCA in the RAN can predict handover and inform the UE proactively. Meanwhile, the UE can explore multiple paths to achieve seamless handover and better network/link selection and QoS guarantee.

In addition, SET can also leverage the high data rate radio technologies of 6G (e.g., terahertz) to support both access and backhaul simultaneously, through negotiations among the relevant PCAs of UEs and RANs.

C. Adaptive *-cast

6G will support many new applications where multiple end-users (humans or things) need to exchange information and make decisions. End-users involved in the process may change over time.

For example, UAV swarms in a region need to exchange their locations, surveillance videos, and/or action requests with each other. When a new UAV arrives, it may need to join the group to share the information; when a UAV moves away from the swarm, it may or may not leave the group, depending on application scenarios. The traditional way to maintain multicast trees for message dissemination is difficult and costly, as all the nodes can be sources and the fast-changing topology may lead to intermittent links and broken multicast trees. New protocols are needed to support such high-dynamic *-cast applications.

Another application is the multi-streaming services: a user may demand the same streaming content in different devices or move from one device to another seamlessly, which has been a salient feature of distributed OS systems such as HarmonyOS. However, how to stream the data (probably in different resolutions, refresh rates, and quality levels) to different devices more efficiently than creating multiple unicast streams is an open issue.

Following the SET design principle, we can decompose the *-cast protocol functions and configure the protocol based on the application scenarios with distributed coordination. For the UAV swarms case, the address and group membership maintenance can be configured depending on physical locations, and packet routing for UAV swarms can be geographic-based to handle high mobility scenarios. For the multi-streaming case, a unicast session can be self-evolved into a multicast session when more receivers join, a unicast/multicast session can be evolved into a group-cast session when more senders need to disseminate messages within the group, and vice versa.

The SET protocol architecture can support advanced *-cast services with evolvable service mode, QoS guarantee, and mobility support in different application scenarios, e.g., robot swarms, drone swarms, and unmanned vehicle networks.

D. Ubiquitous intelligence

Ubiquitous intelligence applications typically rely on timely and reliable information exchange among a group of intelligent entities who can process and compute information to make real-time intelligent decisions. The proposed SET protocol architecture can handle the corresponding delay and reliability guarantee service requirements based on the networking conditions, and configure the protocols accordingly, e.g., using strong error coding to deal with channel impairment and exploring multi-link and multi-path diversity to ensure high reliability and low latency.

In addition to reliable and timely data transmissions, to support ubiquitous intelligence, 6G networks could acquire and process wireless sensing data and manage computing and storage resources for various AI applications. Wireless sensing data can be handled locally with the assistance of the PCA, while the resource orchestration is beyond the capability of the local PCA controller.

Instead, the session control in the cubic model shown in Fig. 3 can handle these applications by orchestrating the computing and storage resources of the end devices, edge servers, and cloud servers, based on the latency, privacy, and resource requirements of the applications, and loads of the servers and network communication links. The session controller deployed in the edge and the core can coordinate with each other, and they assist the applications to direct the data to suitable locations with sufficient computing and storage resources to satisfy the service requirements of ubiquitous intelligence applications.

VII. Summary and future work

We have proposed a novel SET protocol architecture for 6G. SET lets a protocol be configured to support a wide range of services. It can also be transformative during the lifetime of a flow, as the controller can switch to different control functions and settings based on its observations. SET enables distributed and autonomous in-network intelligence, so both end systems and network intermediate systems can be more intelligent for space-air-ground-water spectrum and mobility management, *-cast, QoS support, and ubiquitous intelligence with integrated sensing, communication, storage, computing, and control.

With the SET protocol architecture, many open questions beckon further research. First, how to decompose the whole protocol stacks and build new control functions. Decomposing the protocols in the existing layered architecture can be the starting point. Given the large number of existing protocols, it requires joint efforts from the research community to build the control function library by decomposing the existing protocols. In addition, how to categorize the decomposed functions and build general control functions with minimum overlapping with each other also requires extensive further research.

Second, for PCAs, the control process brings many open research issues. For instance, to effectively discover the network environment, it requires defining what to monitor, how to measure, how frequently to explore, and what are the affordable costs. There is no one-size-fits-all answer to these questions, and the solutions should be customized based on the network operation needs. In addition, given the historical data collected, PCAs may need to predict network dynamics using advanced signal processing and learning tools. Finally, PCAs need to intelligently map the QoS requirements to control policies accordingly, involving multi-objective optimization to minimize costs without compromising service quality in dynamic environments with advanced learning technologies. All these are critical further research issues for the success of the proposed SET protocol architecture.

Finally, in this article, the proposed SET protocol architecture mainly addresses the networking performance issues, while the security and privacy issues remained open for future research. The distributed intelligent agents, PCAs, can leverage advanced AI/ML AI technologies to enhance network security and data privacy, a promising direction to further explore.

REFERENCES

[1] S. Zhang and D. Zhu, "Towards artificial intelligence enabled 6G: State of the art, challenges, and opportunities," *Computer Networks*, p. 107556, 2020.

[2] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[3] J. Pan, L. Cai, S. Yan, and X. Shen, "Network for AI and AI for network: Challenges and opportunities for learning-oriented networks," *IEEE Network*, pp. 1–8, 2021.

[4] X. Shen, J. Gao, W. Wu, M. Li, C. Zhou, and W. Zhuang, "Holistic network virtualization and pervasive network intelligence for 6G," *IEEE Communications Surveys Tutorials*, to appear.

[5] A. Langley and et al., "The QUIC transport protocol: Design and Internet-scale deployment," in *Proceedings of SIGCOMM '17*, Los Angeles, CA, USA, 2017, p. 183–196.

[6] A. Weaver, "Xpress transport protocol version 4," in *Proceedings 1995 IEEE International Workshop on Factory Communication Systems. WFCS'95*, 1995, pp. 165–174.

[7] C. Secretariat—NASA, "Space communications protocol specification (SCPS) transport protocol (SCPSTP)—blue book," *Washington, DC, May*, 1999.

[8] N. Finn, P. Thubert, B. Varga, and J. Farkas, "Deterministic networking architecture," *draft-ietf-detnet-architecture-03 (work in progress)*, 2017.

[9] C. Filsfils, N. K. Nainar, C. Pignataro, J. C. Cardona, and P. Francois, "The segment routing architecture," in *2015 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2015, pp. 1–6.

[10] P. Djukic and H. Zhang, "System and method for providing a software defined protocol stack," *United State Patent WO2014166402A1*, May 2020.

[11] J. Day, *Patterns in Network Architecture: A Return to Fundamentals*. Prentice Hall, 2008.

[12] L. Cai, X. Shen, J. Pan, and J. Mark, "Performance analysis of TCP-friendly AIMD algorithms for multimedia applications," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp. 339–355, 2005.

[13] J. Hu, L. Cai, and J. Pan, "Mesh network reliability analysis for ultra-reliable low-latency services," in *2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, 2021, pp. 198–206.

[14] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure, "RFC 6824: TCP extensions for multipath operation with multiple addresses," *Internet Engineering Task Force*, 2013.

[15] W. Yang, S. Shu, L. Cai, and J. Pan, "MM-QUIC: A mobility-aware multipath QUIC for satellite networks," in *17th IEEE MSN'21*, Dec. 2021.

PLACE PHOTO HERE

Lin Cai [S'00, M'06, SM'10, F'20] is a Professor with the Department of E&CE at the University of Victoria. She is an NSERC E.W.R. Steacie Memorial Fellow, an Engineering Institute of Canada (EIC) Fellow, an IEEE Fellow, and a College Member of the Royal Society of Canada. Her research focuses on network protocol and architecture design supporting multimedia traffic and Internet-of-Things.

PLACE PHOTO HERE

Jianping Pan [S'96, M'98, SM'08] is a professor of computer science at the University of Victoria. His research interests include protocols for advanced networking, performance analysis of networked systems, and applied network security. He received the Telecommunications Advancement Foundation's Telesys Award (2010), the JSPS Invitation Fellowship (2012), and the NSERC DAS Award (2016).

PLACE PHOTO HERE

Wenjun Yang [S'19] received his M.S. degree in Information Science and Engineering from the Hunan Normal University, Changsha, China, in 2019. He is currently a Ph.D. student in the Department of Electrical & Computer Engineering at the University of Victoria, Canada. His research is on the next generation network architecture and protocols, including multi-path transport layer protocol and mobility management.

PLACE PHOTO HERE

Xiangyu Ren [S'19] received his B.Sc. degree from the Department of Automation Engineering, University of Electronic Science and Technology of China in 2019. He is currently pursuing his Ph.D. degree at the Department of Electrical & Computer Engineering, University of Victoria, Canada. His research interests include deterministic networks, vehicular networks, machine learning and optimization with applications in networking, and software defined network.

PLACE PHOTO HERE

Xuemin (Sherman) Shen [M'97, SM'02, F'09] is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on network resource management, wireless network security, AI for networks, 5G and beyond, and vehicular ad hoc networks. He is a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Chinese Academy of Engineering Foreign Fellow. He received the R.A. Fessenden Award in 2019 from IEEE, Canada; the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society; and the Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society.