NEXT G ALLIANCE
An ATIS Initiative

6G

Next G Alliance Report:

6G Distributed Cloud and Communications System

TABLE OF CONTENTS

FOREWORD

As a leading technology and solutions development organization, the Alliance for Telecommunications Industry Solutions (ATIS) brings together the top global ICT companies to advance the industry's business priorities. ATIS' 150 member companies are currently working to address network reliability, 5G, robocall mitigation, smart cities, artificial intelligence (AI)-enabled networks, distributed ledger/blockchain technology, cybersecurity, IoT, emergency services, quality of service, billing support, operations, and much more. These priorities follow a fast-track development lifecycle from design and innovation through standards, specifications, requirements, business use cases, software toolkits, open-source solutions, and interoperability testing.

ATIS is accredited by the American National Standards Institute (ANSI). ATIS is the North American Organizational Partner for the 3rd Generation Partnership Project (3GPP), a founding Partner of the oneM2M global initiative, a member of the International Telecommunication Union (ITU), as well as a member of the Inter-American Telecommunication Commission (CITEL).

For more information, visit www.atis.org. Follow ATIS on Twitter and on LinkedIn.

The ATIS Next G Alliance is an initiative to advance North American wireless technology leadership over the next decade through private sector-led efforts. With a strong emphasis on technology commercialization, the work will encompass the full lifecycle of research and development, manufacturing, standardization, and market readiness.

EXECUTIVE SUMMARY

The distributed cloud and communications system is one of the six audacious goals NGA developed for North American 6G vision. It aims at making 6G systems a wide-area cloud with ubiquitous computing and workload distribution across devices, network nodes and data centers. The 6G distributed cloud and communications system is a step towards converged communication and computing system capabilities beyond what has been achieved in 4G and 5G on network cloudification and edge computing.

This paper provides an overview on the distributed cloud and communications system from the perspective of design goals, driving forces, challenges, research directions, and the path to realization. The paper is intended to trigger research and design discussions across communication and computing communities to fully realize the distributed cloud and communications system audacious goal for 6G.

1 INTRODUCTION

One of the major drivers for network evolution is the convergence of mobile communications and cloud computing. 4G started the path of network function virtualization. 5G technologies, with the adoption of a service-based architecture, accelerated the enablement of network cloudification, edge computing, and network as a service. In 6G, we expect a further evolution of distributed cloud and communications systems, where 6G systems will provide a wide-area cloud with ubiquitous computing across and among devices, network nodes and data centers. The scaling out of mobile communication systems will be driven by the parallel scaling out of computing. Mechanisms for workload distribution among the computing nodes in the 6G wide-area cloud will enable a continuum of services and satisfy related Quality of Service (QoS) needs. A cloud-native philosophy will be a fundamental part of the design and deployment of 6G functions [1].

The 6G distributed cloud and communications system is expected to provide communication, computing, and data services. This contrasts with prior generations, which primarily provide communication services. The communication, computing, and data services can be provided in forms of infrastructure service (e.g., containerized communication-computing-data infrastructure), platform service (e.g., platform services for scaling out computing across mobile devices and network compute), and software services (e.g., data analytics services). The 6G system needs to be designed with the capability of providing those various forms of services (e.g., Everything as a Service (XaaS)). A computing service plane and a data service plane are expected to be introduced in 6G systems in addition to the communication service plane.

Development of new service subscription models are also foreseen. Subscribers to 6G XaaS are anticipated to include mobile users, mobile device vendors, application providers, cloud-service providers, etc. Figure 1 shows an example of the envisioned 6G distributed cloud and communications system, which includes:

> **User equipment** with a unique cloud and communications stack.

> **Cell sites** with far-edge capabilities and a facility cloud and communications stack.

> **Semi-private radio sites** (e.g., homes, offices, factories, other enterprises) with far-edge capabilities on a facility stack that may home to core sites (possibly more than one), and can potentially operate without connection to the core sites as needed.

> **Edge sites** with facility stacks.

> **Core sites** with facility stacks.

> **Central data centers** with facility stacks extending communication services to the data center end points, thus permitting end-to-end QoS and service feature guarantees.

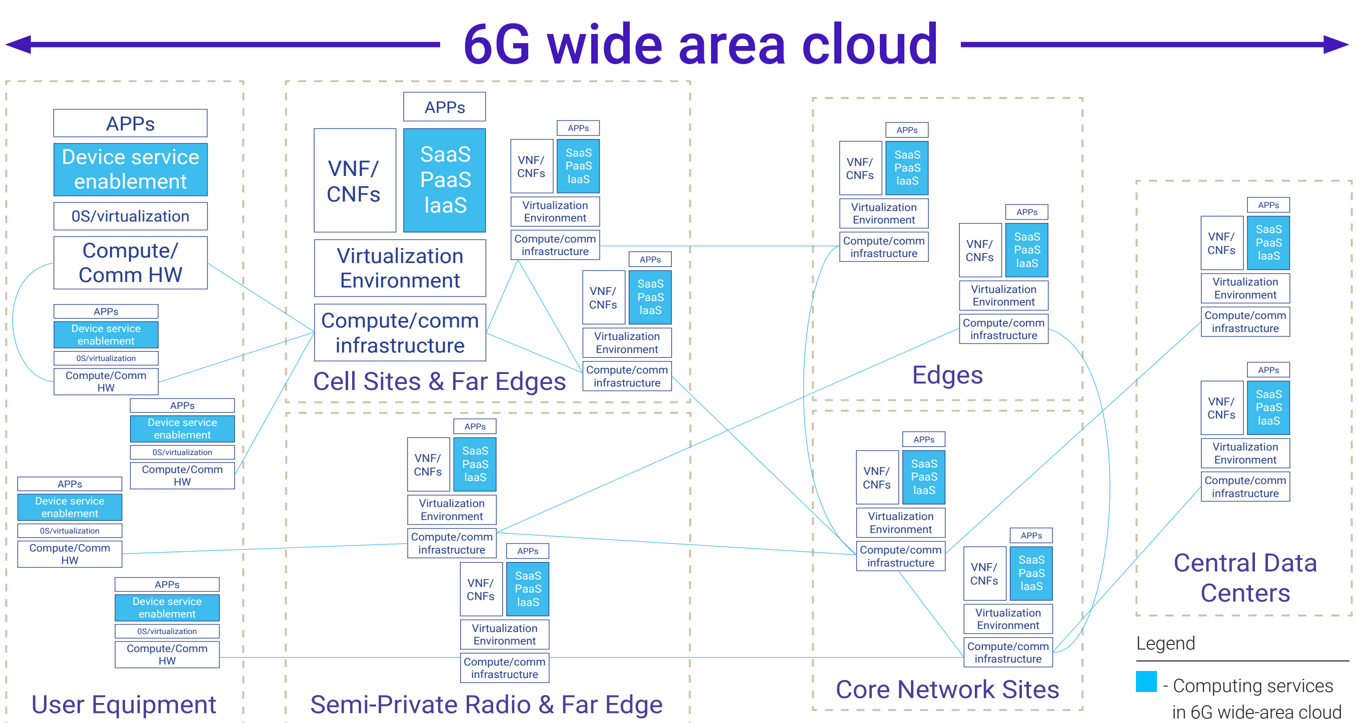


Figure 1: Distributed Cloud and Communications System Architecture

In prior generations, mobile communications systems and cloud computing systems were separately designed. The former focused exclusively on communications, while the latter operated “over the top” (e.g., 5G edge computing services at geographically localized break-out sites are fundamentally based on independent implementations of cloud and mobile services, except for the exchange of location and desired QoS information).

In the 6G distributed cloud and communications system, a higher level of integration between the mobile system and cloud computing is needed to attain better design characteristics. Specifically, a 6G system will provide balancing between localized computation and distributed computation to optimize resource usage and power consumption while meeting the task’s Key Performance Indicators (KPIs) requirements (e.g., in terms of latency and data rate). A 6G system should be flexible to allow computations to occur at different points in the network, depending on device capabilities and application requirements.

Therefore, the 6G system design needs to find the right level of coupling between communication and computing for optimized scalability, complexity, and performance.

Security and data privacy shall also be ensured, both for traditional mobile communications and computing requirements, and for new 6G radio features that include sensing data acquired by mobile elements, including human characteristics afforded by higher frequencies such as sub-THz.

The 6G distributed cloud and communications system will set the technology foundation for achieving the digital world experience. Applications and use cases intrinsic to the digital world experience are envisioned thriving in an ecosystem enabled and nurtured by the flexibility in terms of performance and scale offered by services inherited by the 6G distributed cloud and communications system.

Besides providing computing and data services to external subscribers, the 6G distributed cloud and communications system will supply the computing and data services and resources required to empower a distributed AI-native infrastructure expanded though wireless networks. This implies enabling artificial intelligence (AI) and machine learning (ML) distributed and collaborative learning/inference tasks throughout all network nodes and devices.

Sustainability is supported partly by energy savings achieved by bringing computing closer to data sources, thus reducing the amount of data requiring transport to central computing locations. Furthermore, specialized cloud computing resources for complex computing such as AI functions and signal processing workloads can provide both flexibility and energy efficiency.

Distributed cloud and communications systems are expected to enable more cost-efficient solutions through common reusable infrastructure, cloud-native function deployment, and DevOps software practice and automation that all reduce capital expenditure (CAPEX) and operational expenditure (OPEX). Moreover, cloud-based functions and services will enhance experiences using mobile devices, enabling lower device cost despite advanced capabilities due to computing workload offloading.

Finally, distributed computing and communications will be central in facilitating trust, security, and resiliency services, as well as securing the infrastructure and providing protection from new threat surfaces. With the addition of computing services and data services in the 6G system, new security and data privacy challenges are foreseen. Security and data privacy shall be key requirements in the 6G distributed cloud and communications system design.

2 DRIVING FORCES AND NORTH AMERICAN IMPERATIVES

The need for the 6G distributed cloud and communications system is driven by multiple factors, including North American national imperatives, application and market requirements, and technology evolution.

From the North American perspective, there is a national imperative to collectively progress mobile communications and cloud computing as the infrastructure foundation for empowering future inventions and growth. North America is home to several of the world's largest cloud providers, some of the most sophisticated and largest communications service providers (CSPS), and to a university-industry-government innovation engine that will be at the forefront of 6G technologies and services. To ensure a successful rollout of 6G solutions and services starting in 2030, North American government organizations, academia, and industry will need to move in unison to adopt and expand the distributed cloud and communications platform model. The emerging 6G ecosystem — encouraged by underlying technologies for computing, networking, and devices delivered by network and cloud infrastructure and service providers, CSPs, device vendors, service, and application providers — will be fundamental to our future society.

The proposed 6G system is intended to enable a wide range of applications. One application type is stringent, ultra-low-latency, and high-bandwidth wireless services for augmented reality (AR)/virtual reality (VR)/immersive applications, industrial and medical robotics, and vision/control systems. Other application types include low-bandwidth massive IoT and public sector applications such as smart cities, and novel applications such as holographic communication. The system architecture's flexibility is paramount in delivering elements of an application to different places of the network to optimize performance. From the device to the central cloud, the computing and communications stacks will serve to automatically provide the capacity and function needed for the future world of applications.

The technology evolution to enable this new vision will encompass:

- A new realization of the mobile device as a cloud stack.
- Underlying computing technologies to support high-performance and energy-efficient implementations of signal processing, augmented reality (AR)/virtual reality (VR) algorithms, and various application workloads.
- Integrated communication-computing scalable system stacks supporting a wide variety of distributed computing locations (e.g., from devices to homes to cell towers to major edge locations, to main data centers).
- Intelligent and highly dynamic automation support for a wide range of workloads, permitting orchestration and assurance functions to be efficiently enabled to deliver application features.

RESEARCH AND TECHNOLOGY DIRECTIONS

3.1 Challenges and Requirements

Compared to today's data center-based cloud systems and communication-centric mobile communications systems, the 6G cloud and communication system is envisioned with unique characteristics and requirements that pose several challenges:

> Computing resources will be widely distributed, spanning multiple security and ownership domains across mobile devices, network nodes, and data centers. One of the top system design requirements is scalability. Standardized interfaces are needed for interoperability. Security and authentication mechanisms need to be in place for trustworthiness.

> Mobile devices and network computing nodes have limited computing resources. Dynamic workload distribution needs to be enabled to efficiently use ubiquitous computing resources and meet service requirements.

> The 6G system will encompass both mobile computing devices and mobile wireless communications. This contrasts with today's data center cloud systems, where servers are interconnected via pre-established wireline connections so the system design can focus on computing service discovery and establishment. In the 6G distributed cloud and communications system, both wireless communication service discovery, establishment, continuity and computing service discovery, establishment, and continuity need to be addressed.

> The system needs to hide detailed communication, computing, and data service enablement from applications. This will free application developers to focus on realizing application features instead of being distracted by the underlaying communication, computing, and data platform. Standardized Application Programming Interfaces (APIs) need to be defined for applications and subscribers to use 6G system communication, computing and data capabilities, and services.

> Result consistency and computation ownership are central to the 6G system. Distributed task results must be consistent with on-device execution. Models and algorithms can be owned by an entity while being offloaded to another entity. By carefully considering the role of entities (e.g., in providing storage and compute resources), a system can be built offering consistent results.

> The protocols for communication and offloading need to be designed to meet requirements such as power consumption and workload distribution. The design should strive for leanness in terms of managing complexity and power consumption.

> Entities with a 6G system should be able to enforce privacy policies to control what data can be distributed. Guarantees must be available to ensure that the entities to which computation is offloaded do not also have access to the results of the computation. Instead, they should be provided with only the resources necessary to run those workloads.

3.2 Research and Technology Areas

Figure 2 illustrates the vision of the distributed cloud and communications system. The 6G cloud-native system design requires cross-discipline joint efforts from both the mobile communication and cloud computing communities. The following research directions are outlined.

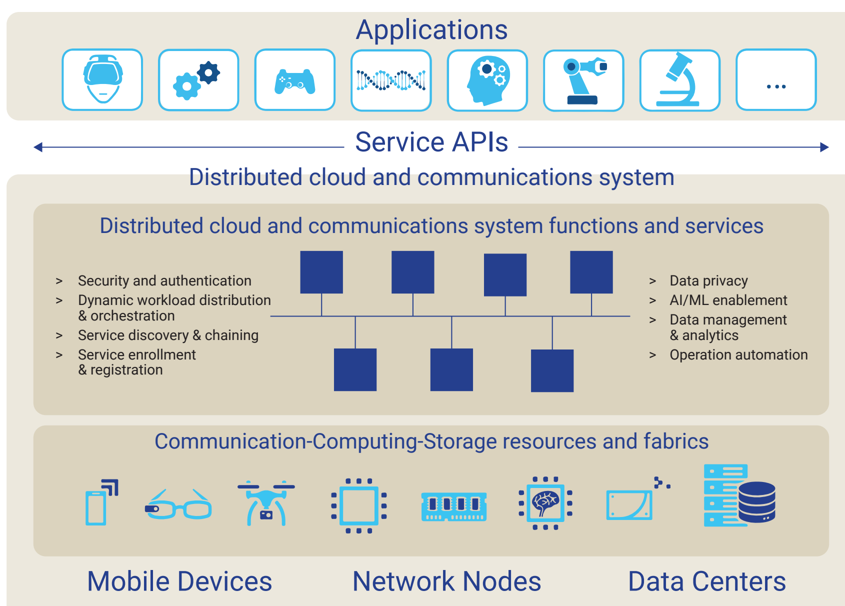


Figure 2: Technology Vision

System architecture, functions, and services

Mobile systems traditionally have been decoupled from cloud computing systems, including in 5G, where edge computing is designed as an application layer feature beyond the mobile core network [2][3][4]. 6G requires a mindset shift so that 6G is a system for both communication and distributed cloud computing. Integration between communication and computing domains is required so that 6G is a native part of the cloud. Research is needed to address questions such as:

> How can computing services and data services be introduced into 6G systems alongside communication services?

> What is the right level of coupling between communication and computing?

> How can seamless computing and storage access be ensured when using distributed nodes in the 6G system?

> How can the 6G wide-area cloud incorporate both mobile device compute and network compute?

> How can ubiquitous computing be achieved at scale?

> How can computing services and data services be abstracted and represented?

> How can communication, computing, and data services be jointly orchestrated and optimized?

> How can the system leverage cloud computing technologies such as orchestration, micro-services, and service mesh?

The 6G distributed cloud and communications system includes the distributed communication and computing infrastructure and the communication, computing, and data service functions instantiated on top (e.g., for providing data analytics, AI/ML computing services, vertical-specific services, etc.). From a logical system architecture perspective, the 6G distributed cloud and communications system can be built on the success of the 5G service-based architecture (SBA) [2]. The SBA can be further evolved to incorporate Radio Access Network (RAN) and devices. Computing service functions and data service functions can be modularized and added to the SBA. The functional partition among system functions can be further optimized based on factors such as the optimal coupling between communication and computing to maximize service reuse, system scalability, and performance.

Device-network-data center computing scaling and service continuum

In today's data center cloud computing, applications follow a client-server model with a predetermined work-load partition between client-side and server-side applications at the application development stage. Client-side applications often handle light computing tasks, while heavy computing tasks go to the server-side application hosted in data centers. The fixed client-server workload partition simplifies the computing scaling issue, but custom application development is needed for specific realizations of the service continuum. Moving toward the envisioned 6G distributed computing, the dynamic workload distribution across computing locations needs to consider factors such as computing resource availability, communication channel condition, processing time budget, and energy consumption budget. To avoid adding complexity to applications, 6G systems need to enable distributed computing at scale as a platform feature and hide details of computing workload distribution from applications. Standardized APIs can be developed for applications to deploy their workloads and set service requirements.

Distribution of the cloud fabrics and workload

In a 6G distributed cloud and communications system, services might be located or "distributed" to specific locations to reduce latency [5]. This distribution of services enables the system to meet very specific response time and performance requirements, regulatory or governance compliance mandates, or other criteria requiring cloud infrastructure to be located anywhere. A distributed cloud distributes both the application and the entire computing stack to the locations where it is needed. This requires novel distributed design solutions for data centers that focus on the benefits of a distributed cloud — such as increased compliance, scalability, flexibility, faster processing, and performance — and on environmental aspects such as energy efficiency.

Distributed computing workload modeling and evaluation methodology

Workload modeling and evaluation methodology development are critical in setting design requirements, identifying gaps, and validating performance. Mobile systems have developed effective traffic models (e.g., for full-buffer, File Transfer Protocol (FTP), Extended reality (XR), etc.) and evaluation methodologies (e.g., for link-level, system-level evaluations) for communications. 6G's tight integration of computing, data, and mobile communication creates a need to further develop computing workload models and evaluation methodologies. Major computing workloads — such as AI training/inference, video transcoding/feature extraction, and key-value-store data operations — can be analyzed and abstracted into a few typical workload models. Performance metrics and evaluation methodologies considering interactions between communication and computing including security and privacy need to be defined. In addition to workload, methods for evaluating and representing computing resources in devices, network nodes, and data centers need to be defined.

Air interface features

The air interface is the mobile device access link, so it plays a critical role in enabling the 6G distributed cloud and communications system. With computing and storage becoming new dimensions of resources provided by 6G systems, the air interface needs to be integrated in a system with features that can handle computing/memory resources, as well as radio resources in both mobile devices and the RAN. Work needs to be done on areas including:

> Methods of sharing and using remote computing/memory resources among mobile devices and RAN.

> Joint radio resource control and computing/memory resource control mechanisms.

> Joint communication and computing performance optimization.

> D2D and mesh network design for computing workload distribution among mobile devices.

Additional optimizations may also be needed for major distributed computing workloads — such as distributed learning or inference, video processing, and feature extraction — that may need specialized computing facilities.

AI/ML distributed computing – model vs. air interface optimization

Collaboration between multiple nodes requires exchanging large deep neural network models or datasets [6]. Given that communication resources might be limited, and the channel is time varying, the size of the neural networks might need to be jointly optimized with the allocation of the network resources to improve performance (e.g., in terms of latency). The impact of the wireless channel/spectrum on distributed learning/inference performance will require approaches to jointly balancing communication and compute resources to optimize task execution. Furthermore, 6G protocols will need to ensure that applicable security and privacy requirements are fulfilled while exchanging AI/ML data and workloads.

Service-aware transport

Mobile systems in prior generations used GTP tunnel-based network transport, which mainly serves as a data pipe between the RAN and cloud [2]. This approach works well in centralized computing. In scenarios that require multiple stages of processing in a chain of computing locations (e.g., multiple model aggregation stages in federated learning), the service-unaware transport is ineffective. It is desirable to have a service-aware transport that can be configured with service chain information and be able to steer packets to corresponding intermediate processing points along the service chain. To introduce service-aware transport into 6G system, a structural upgrade on the transport fabric and corresponding control plane functions is needed.

Service exposure and APIs

Efficient service-exposure mechanisms are important for conveying 6G system services to subscribers and achieving the XaaS goal. Conventional mobile system service subscribers are mobile device users, utilizing a mostly uniform communication service over the air interface. 5G expanded system services to industrial and enterprise subscribers. As a result, the network service exposure function was defined for external developers and enterprise users to customize network services. With expanded service dimensions beyond communication services, 6G systems shall further develop the service-exposure functions and APIs to allow easy use of 6G services by a larger subscriber base, enabling new use cases and innovations.

QoS/QoE definition and enablement

With the integration of communication, computing, and data services in the 6G system, QoS and Quality of Experience (QoE) definition and enablement will need to be revisited. Performance metrics such as throughput, latency, and energy efficiency need to be extended to reflect communication services, as well as computing and data services. New kinds of measurement and telemetry need to be defined. Compared to conventional decoupled communication-computing systems, the 6G system has a chance to take a holistic view of QoS/QoE enablement that considers both the communication and computing domains.

Computing resource and energy efficiency

Distributed computing allows local data processing and enables energy savings on data transport to remote data centers. Highly distributed 6G nodes may also be better positioned to utilize sustainable energy sources than data centers. However, the distribution of computing resources is conventionally less effective in achieving resource pooling gain. The low resource pooling gain may further lead to inefficient resource utilization and higher energy consumption if unloaded computing resources are not power gated. Research is therefore needed to optimize distributed computing resource and energy efficiency, such as close monitoring on workload condition and resource utilization, dynamic computing resource allocation and workload scheduling, and proactive power management based on traffic prediction.

Security, privacy, and trustworthiness

Distributed computing means that network and user data will be shared for compute purposes. Classes of data sensitivity and privacy-preserving methods should be researched to identify which data can be shared at which protection level. Security mechanisms and methods that preserve data integrity also are needed.

Billing/charging methods for new dimensions of services

Mobile systems have well-established charging methods for communication services. With the envisioned addition of computing and data services in 6G systems, new types of billing and charging methods shall be developed accordingly. The corresponding charging functionalities shall be added to responsible network functions as well.

4 PATH TO REALIZATION

To realize the goals of the 6G distributed cloud and communications systems described in this white paper, several developments across the industries involved will be needed, as Figure 3 illustrates.

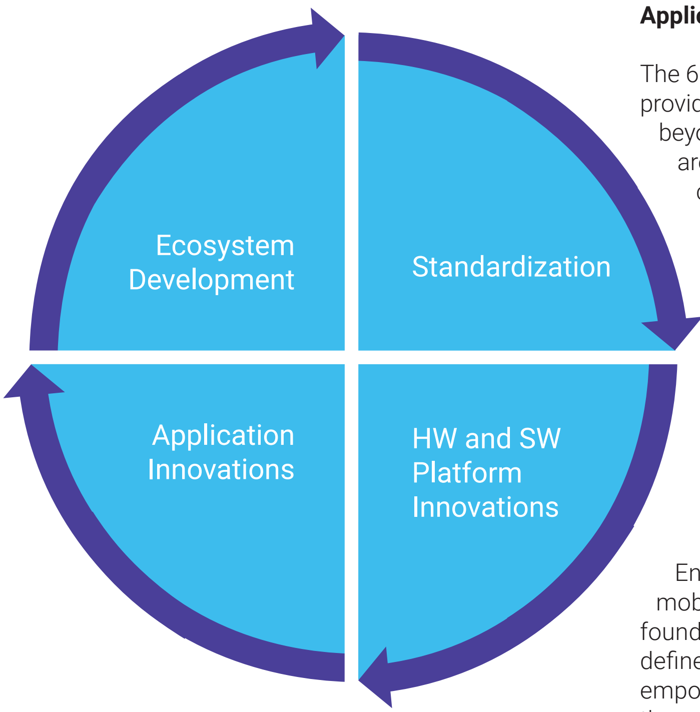


Figure 3: Paths to Realization

Standardization

Standardization is crucial for interoperability and achieving global scale. 3GPP is expected to continue to be the leading organization on 6G system specification development. With 6G heading toward distributed cloud and communications systems, 6G standardization would also evolve to accommodate computing and data components. An increased collaboration between 3GPP and other implementation communities/bodies — such as nternet Engineering Task Force (IETF), pen Radio Access Network (O-RAN, Open Networking Foundation (ONF), The Cloud Native Computing Foundation (CNCF), Linux Foundation (LF) Edge, and Distributed Management Task Force (DMTF) — is expected to pave the way for landing specifications into implementations.

Network and devices hardware and software platform innovations

Network and mobile device compute fabrics are foreseen becoming part of the 6G cloud. As a result, innovations in hardware and software are needed in areas such as secure and agile compute resource sharing among network functions and user compute workload, system programmability and configurability, and service enablement and exposure.

Application innovations

The 6G distributed cloud and communications system provides new system capabilities and services that go beyond communication services. Application innovations are needed to exploit the full potential of new system capabilities and services, all of which benefits end users. Application innovation is a crucial step to prepare the market and achieve 6G's full business potential.

Ecosystem development

The distributed cloud and communications system brings expanded dimensions to 6G. A tighter collaboration between the communications and computing communities, as well as with new players that have different backgrounds, will create a thriving 6G ecosystem end to end.

Enabling technologies in compute, communication, mobile device, and storage are envisioned to form a foundation upon which standards organizations will define specifications that tie these technologies together empowering a thriving 6G ecosystem and marketplace for the next generation of applications and services. A fine balance in terms of coupling between different technologies is needed to encourage innovation while arriving at industry standards for a marketplace that can deliver on the 6G expectations for QoE and mission-critical services. The advancement toward 6G will leverage the strength across industries, academia, open-source communities, standards organization, and government, where each stakeholder will play an important role.

5 CONCLUSION

The 6G distributed cloud and communications system is expected to fundamentally expand mobile system capabilities and services from being communication-centric to becoming communication-computing-data centric. A tighter integration between communication and computing is expected to achieve a wide-area distributed cloud across mobile device compute, network compute, and data centers. New applications and market opportunities will be empowered by the expanded capabilities and services of 6G systems. Realization of the 6G distributed cloud and communication vision requires joint effort from various players across industries, academia, open-source communities, and standards organizations.

Research and innovation are necessary in areas of computing-communication workload modeling, system architecture and functions, protocol stack, service enablement and exposure, security and trustworthiness, network and device hardware and software platform, applications, resource and energy efficiency, etc. The 6G distributed cloud and communications system is expected to enable revolutionary technologies, business models, and applications, bringing simplifications and conveniences to users, as well as new, mind-boggling experiences.

6 REFERENCES

1. Next G Alliance Report: Roadmap to 6G, https://nextgalliance.org/wp-content/uploads/2022/02/NextGA-Roadmap.pdf, Feb. 2022

2. 3GPP TS 23.501, Technical specification group services and system aspects; System architecture for the 5G system, stage 2, release 17, Mar. 2022

3. 3GPP TS 23.548, Technical specification group services and system aspects; 5G system enhancements for edge computing, stage 2, release 17, Sep. 2021

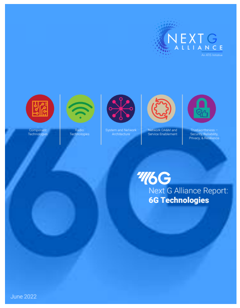
4. 3GPP TS 23.558, Technical specification group services and system aspects; Architecture for enabling edge applications, release 17, Mar. 2022

5. Costello, K. "The CIO's Guide to Distributed Cloud". Gartner. https://www.gartner.com/smarterwithgartner/the-cios-guide-to-distributed-cloud. Aug. 12, 2020.

6. J. Verbraeken, M. Wolting, J. Katzy, J. Kloppenburg, T. Verbelen, "A survey on distributed machine learning", ACM computing surveys, vol. 53, no. 2, Mar. 2020

COPYRIGHT AND DISCLAIMER

NEXT G ALLIANCE REPORTS



6G Technologies

6G Applications and Use Cases

Roadmap to 6G

Green G: The Path Toward Sustainable 6G

6G Distributed Cloud and Communications System

Trust, Security, and Resilience for 6G Systems

6G Market Development: A North American Perspective

Building the foundation for North American leadership in 6G and beyond.

nextgalliance.org

NEXT G ALLIANCE
An ATIS Initiative