Transitioning to 6G: Part 2-Systems and Network Technology Areas

Sharad Sambhwani, Arjun Bharadwaj, Christian Drewes, Kenza Hamidouche, Ayman Naguib, Dirk Nickisch, Sabine Roessel, Matthias Sauer, Yannis Schoinas, Tarik Tabet, Sree Vallath, and Yu-Ting Yu

Apple, Inc.

Part 1 of this two-part column presented potential radio technologies for 6G. In Part 2, we focus on potential systems and networking technology areas for 6G. Paradigm shifts in the areas of protocol stack support for extreme applications, subnetworks of devices, distributed computing, AI/ML in networks and devices, semantic communications, seamless mobility, and enhanced application frameworks are discussed.

Systems & Network Technology Areas

6G for XR

Extended reality (XR) features in 5G NR have been focused on ways to manage scheduling within the existing framework of the radio access network (RAN). To unlock the potential of next generation applications, the challenges posed by the end-to-end (E2E) system need to be addressed.

For the most part, cellular traffic enhancements have been enhancing best-effort traffic with a couple of notable exceptions. VoIP over LTE/NR necessitated the development of an entire back-end framework across multiple layers to enable audio calls between users. Industrial and Internet of Things (IoT) applications were enabled by ultra-reliable low-latency communications (URLLC), which reduced latency for smaller data rates. The key difference for XR applications is that it involves interactive experiences where the entire perspective of the user may be communicated via the network. If multiple users are involved, the E2E experience i.e., user-to-user latency) is the most relevant key performance indicator (KPI) to consider. Several companies and domain experts have identified that the necessary user-to-user latency for a seamless user experience is < 50 ms as shown in Fig. 1.

User-to-user latency includes the access link, gateway, backhauls, Internet service provider (ISP) links, and finally downlink to the other users. Variations in one or more dimensions can be used to maintain a consistent end-user experience. It is time to think of cross-technology optimizations across multiple nodes in the network. It is expected that XR applications will be compute-intensive. However, form factors of XR devices continue to reduce in size and weight. This poses significant thermal dissipation and power consumption challenges. The conventional model of the user equipment (UE) transmitting to the network cell radius of 500 m to 1.5 km may be completely infeasible for some of the applications. A different paradigm to manage compute is required.

One solution to manage power consumption and on-device computing is to simply offload the necessary compute to the cloud or to a server located at the cell edge (for low-latency applications). This requires the development of an entire infrastructure- and platform-independent framework of operation. Such proposals are being considered in the 3rd Generation Partnership Project (3GPP) SA and European Telecommunications Standards Institute (ETSI) forums; however, a comprehensive design including the 6G paradigms needs to be undertaken. It should be noted that solutions that involve deployments of network nodes at the edge would be expensive, and if the gains obtained are mediocre, mobile edge computing as such would not be justified from a return on investment (ROI) perspective.

Solutions under consideration should also include application-aware enhancements that probe deeply into the application data flows, and optimize performance based on individual app data units (ADUs). As such cross-layer schemes are developed, protection of user privacy is critical. The user has the right to expect complete privacy with regard to their Internet activities, and it is our mandate to ensure that this is not violated.

There are many challenges involved in enabling the next-gen XR applications. However, if the challenges are faced squarely and addressed, this may well herald the next generation of computing.

Lean Protocol Stack

One of the goals of 6G is to offload some of the computations to the edge, reducing the processing requirements on the device, allowing for more compute-intensive applications while protecting user privacy and security. This would require an improved communication between the UE, the edge nodes, and the base stations. At the same time, new and enhanced use cases such as XR, vehicle-to-everything (V2X), and sensing will demand a diverse set of requirements and device characteristics. For all these use cases and their associated target KPIs, the protocol stack needs to be redesigned for tackling issues such as jitter, packet loss, data rate fluctuations, E2E latency, and seamless mobility. Moreover, optimizations should be per-

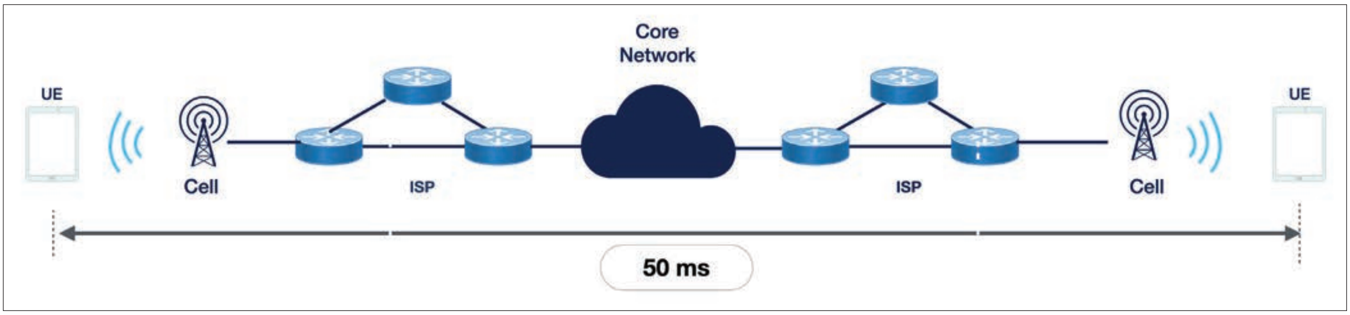

FIGURE 1. End-to-end user latency for XR.

formed across the different protocol layers on both the NW and the UE side, allowing for dynamically adapting application quality of service (QoS) requirements to radio conditions, improving as a result the end-user experience and keeping the UE complexity low. To realize diverse applications and diverse product types with their different requirements/KPIs, the protocol stack should be as lean as possible and adopt a flat control/data plane architecture. Functionalities across different layers as well as inter-layer interactions should be consolidated and optimized. This may lead to reduced memory, processing, and power consumption requirements, hence reduced UE complexity and, as a result, a reduced required cost for implementation.

AI/ML in Networks and Devices

The vision for next generation network design is to leverage artificial intelligence/machine learnging (AI/ML) capabilities for both applications and networks. On the network side, these techniques can be used to enhance the air interface performance in the network, for traffic prediction and RF planning, and other non-real time network operations. A vast number of applications and services use AI/ML to optimize their service and enhance the experience of users. To enable full functionality for AI/ML over the network, a new design paradigm should be devised in which AI/ML capabilities are native to the network architecture and woven into the network fabric.

Having said that, the distributed architecture of cellular networks presents several challenges that need to be addressed to enable AI-driven networks including: (1) resiliency of the learning mechanisms to nodes' failure and dynamic network topology, (2) the trade-off between computation, communication, and storage resource limitations and the performance of the learning models, (3) trustworthiness that depends on ensuring privacy for users, as well as the explainability of the AI decisions, and (4) handling the heterogeneity and scarcity of the data at the learning nodes to achieve highly accurate models. These factors will impact the choice of models and algorithms depending on the network conditions and requirements of the application.

Distributed Computing

The era of 6G will see ever-increasing density and heterogeneity of connected devices as well as the emergence of computationally heavy applications (e.g., immersive media, XR, V2X) that can go beyond the computation capabilities and power budget of a single device. Offloading to the cloud and to edge servers, and compute distribution to proximal devices in local networks will allow devices to perform tasks regardless of their processing resources and to benefit from more flexibility in managing computational and communication resources. Furthermore, the rise of cooperation-based AI/ML applications (e.g., federated learning) and collaborative sensing (e.g., for positioning, localization, mapping) will require multiple devices to effectively share their compute and communication resources. Enabling distributed computing is key to meeting these upcoming requirements. System nodes (devices, network nodes, cloud servers) will be able to offload and distribute tasks to and retrieve the related results from other nodes. Of course, distribution as well as collaboration involves sharing sensitive and private data, which calls for corresponding protection mechanisms. The 6G system will ultimately evolve toward a device-network-cloud fabric where each system node offers compute, content/data, and communication services with built-in security and privacy. Achieving this vision will require the technical mechanisms necessary for the system nodes to establish mutual trustworthiness such as cryptographically strong unique hardware identities and remote software attestation as well as a security assurance and certification regime that spans multiple regulatory jurisdictions. Alternatively, innovations in homomorphic computing and information coding techniques may lead to new approaches for offloading computation without a mutual trust requirement, at least for some workloads. Finally, enabling distributed computing will make it possible to have leaner devices with less costly computational components, thus achieving better sustainability.

Mesh and Sidelink Enhancements

In 4G/5G, 3GPP features such as sidelink, mainly target V2X and mission-critical applications, were developed to offer a framework for UE-to-UE connectivity. However, mobile device-based mesh network architecture has not been fully exploited commercially so far. Such frameworks may potentially be extended to support multihop/multi-path mesh networking. Development of device-based mesh networking in 6G has many benefits. From a system point of view, 6G can continue to improve the network scalability. With the inclusion of device-based mesh networking and advanced devide-to-device (D2D) techniques, lower latency, further power saving, and improved spectrum efficiency can be achieved. Moreover, new paradigms of mesh networking may enable new business opportunities, such as supporting personal device-based non-public networks to enable flexible spectrum utilization and new business opportunities for operators, while also providing opportunities to deliver enhanced QoE with privacy. Hybrid, heterogeneous mesh networks can provide extreme coverage with less infrastructure investment in extreme conditions, including for emergency operations (e.g., disaster response) or commercial applications operating outside of traditional network coverage.

UE Cooperative Communications

Operating in high frequency spectrum (mmWave, THz) results in a limited communication range. This motivates the use of local D2D communication to improve the coverage range, reliability, capacity, and spectral efficiency of communication with adequate protection mechanisms to ensure user privacy and security. It is expected that such a framework, referred to as UE cooperative communications, will become common in the 6G timeframe. This would represent a paradigm shift relative to previous generations and can include: a) efficient offloading of data/control transmission and reception to other devices; b) carrier aggregation in a network using multiple devices; c) carrier aggregation across networks at a single device or multiple devices; d) data sharing between devices; e) coverage improvement via relay (single/multi-RAT); f) multiple-input multiple-output (MIMO) across multiple devices; and g) distributed communication related processing across devices. In such a scheme, multiple devices form a group of cooperative devices for transferring user data as well as control information, as shown in Fig. 2. The group of devices relies on one or multiple access points to transfer the data and uses a high-speed D2D connection to exchange subsets or the full set of data between each other using licensed as well as unlicensed spectrum. While sidelink and D2D features have been specified from 4G onward, what is missing is an efficient E2E protocol to realize such a system.

Semantic Communication

Semantic communication (i.e., the communication of the intended meaning) is a technology vector being considered for 6G with the promise to significantly improve communication efficiency by reducing transmissions to semantically relevant content. 6G networks can translate these gains into gains in quality and range of service. AI/ML techniques like advanced generative adversarial networks or media transformer neural networks are expected to deliver on this promise.

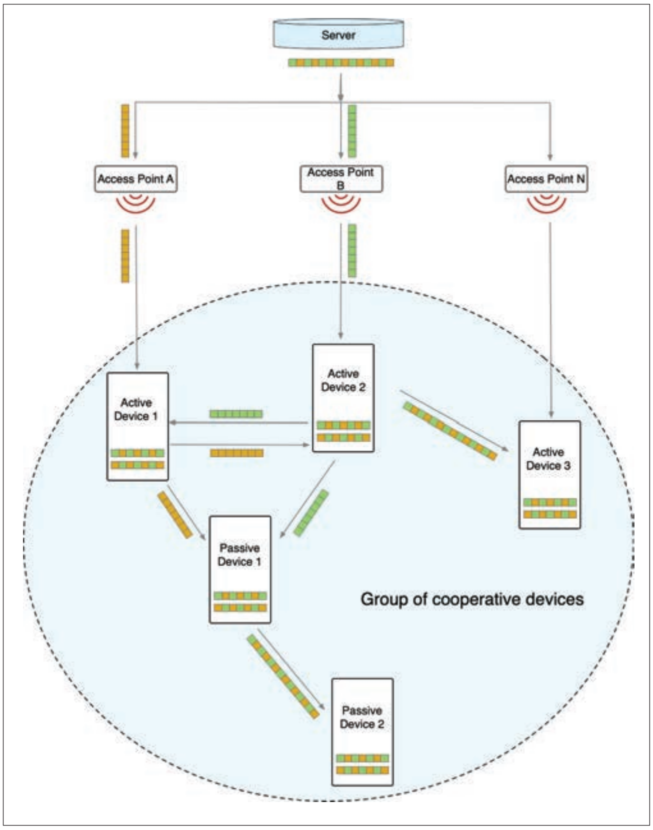

FIGURE 2. UE cooperative communications.

6G system designs need to consider properties of semantic communication. Similarly, it can be expected that properties of the wireless network will interact with the semantic coders. Thus, in contrast to Shannon [1], the semantic problem also becomes an engineering problem.

Research in semantic coding will include aspects ranging from fundamentals in information theory to application-specific AI/ML-based joint source channel coding. Candidates for the extensions to the communication model comprise introduction of semantic source coding that achieves the expected compression gains at acceptable quality, and of semantic channel coding that protects against semantic distortion. Appropriate semantic fidelity metrics need to be defined and formalized to enable control within the network. A third entity in the engineering problem is the required semantic knowledge and its provisioning at sender and receiver.

Seamless Mobility

For future wireless networks, further improvements to mobility performance and procedures will be needed to meet user experience and service reliability targets. This includes establishment of robust mechanisms for mobility procedures over the whole RF spectrum enabling highly reliable cell changes. One approach is to connect the device to multiple cells simultaneously for seamless and reliable transition of data transfer in addition to implementing procedures for devices in cooperation and D2D assistance to improve performance of mobility-related scenarios.

Furthermore, simple methods could be considered for service-interruption-free handover between cells of the same technology and across technologies including minimization of signaling and processing efforts and associated delays. Another important aspect is to investigate methods for power consumption reduction resulting in power optimized mobility procedures.

Enhanced Application Frameworks/Interfaces and dynamic QoE Provisioning

Each generation of cellular technology raises the performance and service quality thresholds for applications and user experience, and 6G definitely promises to push the envelope in both regards. Connected device use cases will generate unprecedented quantities of data. Multi-player gaming, group education, and industrial operations are examples of use cases that will involve distributed coordination and collaboration among multiple participants. As such, the 6G system is expected to be more open, distributed, and data- centric than 5G, with increased in-network application resources (compute, storage, and value-added capabilities such as AI/ML) that require coordination (e.g., scheduling, deployment, optimization) with network resources.

For these use cases to become reality in the 6G timeframe, the 6G system will require a rich set of 6G application services integrating with the 6G core network and 6G RAN. Therefore, these 6G open service frameworks will need to provide devices and applications with capabilities to interact with, coordinate, and complement the functions and capabilities of the underlying 6G core network and 6G RAN. Providing privacy is a key challenge in designing the open service framework. Moreover, dynamic orchestration of network capabilities can also be foreseen to ensure that E2E quality of experience is always met while the service is rendered. A flexible framework integrating applications, networks, and policy components will be required to make this a reality

Conclusion

Transitioning to 6G calls for further development of multiple technology areas to enable a versatile communication network that can accommodate a wide span of new applications and devices, some of which have yet to be discovered. These applications in some cases simply require better service coverage and in other cases require a fundamentally new framework, in terms of both the air interface and the network/system design, allowing for a convergence of wireless communication systems with related technologies such as distributed computing, sensing, AI/ML, and trustworthiness. Special emphasis must be placed in the design on device complexity and efficiency. In this two-part column, we have addressed what we view as the most relevant technologies areas that, if developed and adopted, will enhance the network flexibility, coverage, reliability, and efficiency. While the theoretical foundations of many of these topics have been long established in the literature, there are many implementation challenges yet to be overcome before the eventual adoption.

References

[1] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949.