THE NEED FOR HOLISTIC NETWORK DESIGN

BY ALBERTO LEON-GARCIA AND MARTHA STEENSTRUP

At the ComSoc annual retreat in January of 2021, we established an ad hoc committee to assess the future directions of communications and network research. It has been clear for a few years now that networks are undergoing rapid and disruptive changes. They are no longer utility pipes providing transport; now they offer services to those that are attached to them. With the addition of services, the control planes of networks need to be a lot smarter and more responsive and handle the task of providing security.

Vincent Chan, ComSoc President

The communication network is changing from a provider of connectivity to a provider of services that enable a gamut of applications that are distributed, mobile, and intelligent. At the same time, consumers of network services are expanding beyond humans to a myriad of things, machines, and automated digital processes. In addition, the network must now respond to new socio-economic priorities such as privacy and environmental concerns. These trends drive the network to become an entire, or in some cases, an integral part of an application platform. We use the term 'holistic network design' for the essential challenge of creating a network architecture that supports the diversity of applications, while meeting new priorities, within a cohesive yet flexible set of design elements.

The emergence of networks as massive-scale application platforms both builds on and reinforces cloud computing. Networks are essential to the creation of interconnected multi-clouds encompassing private, public, and network operator clouds that can support global-scale applications. Networks also enable the emergence of distributed multitier clouds that span devices, edge computing, and core computing that will support mobile and IoT applications. On the other hand, cloud computing principles, such as virtualization, microservices and orchestration, are revolutionizing how network services are realized largely by functions implemented in software running on this multitier cloud.

The emerging multitier network will play a central role in the creation of intelligent applications that leverage AI and machine learning to provide continuous situational awareness, event detection and identification, and automated response. Some applications will be used by the network/cloud operator to manage the complexity of their infrastructure and services, while other applications will provide intelligence-as-a-service to customers of the network. Sensor data for both types of applications will be the coin of this realm because it will be the raw material needed in both model development and in live inference, classification, and control. The volume of data has the potential to generate huge demands for computing and bandwidth. However, the multitier cloud will allow the optimized placement of functions that implement the data analytics pipelines to collect and analyze data and formulate actions and responses. The orchestration system that defines and deploys the sequence of functions in a pipeline can promote the efficient use of resources as well as meet the diverse quality and cost constraints of different applications. The centrality of data will also favor content-oriented and name-based architectures.

Network security is the overarching challenge in emerging networks. The ongoing introduction of 5G mobile networks will provide higher data speeds, denser coverage, improved network quality and reliability, and very low latency. The intent is that these capabilities will make 5G networks essential in sectors such as e-Health, smart manufacturing, connected vehicles, transport and logistics, smart retail, and drone delivery. These capabilities will also transform critical infrastructures and systems including intelligent multi-mode transportation systems, environmental monitoring, smart-energy networks, water supply, and agriculture. 5G networks incorporate innovations that bring major benefits that enable new services and applications. However, the redesign of a highly complex system, such as the redesign of the 5G control plane to a service-based architecture, inevitably creates new vulnerabilities. Despite enormous effort to define protocols to eliminate vulnerabilities, it is impossible to know all vulnerabilities ahead of time, and there is the potential for very high impact from intentional or accidental disruptive events. We emphasize that while 5G networks are a central component of future networks, they are not the only component and must coexist with the Internet and its evolutions.

The emerging network is being conceived during a period of changing socio-economic expectations that place new requirements on the network. Climate change demands that new networks must be sustainable and have low carbon footprint. Abuses in the unfettered use of personal information calls for effective and customizable privacy mechanisms. The pre-eminence of information and intelligence in society dictates that there be unbiased access to information technologies to promote equity and inclusiveness. The evolution of the network toward an application platform has the potential to address some of these requirements.

It is clear from the above trends that the emerging network must integrate elements that fit its architecture in a holistic fashion within a future applications environment. We hesitate to call the emerging network "6G" because it may well turn out to be a dramatic departure that is unrecognizable from the ongoing progression of network generations. In particular, the above trends underscore the importance of the logical definition of network architectures which focus on the "what" instead of the "how" in the operation of a network. Hence logical functions can be implemented in a variety of ways, from software running on various types of computation devices, to processing integrated into radio, optical or quantum systems. One can discern the new network architecture as consisting of two new overarching layers, a Logical Layer, encompassing all logical aspects, and a Physical Layer, encompassing physical enabling technologies. Reflecting back on the five decades since the introduction of the OSI Reference model, we appreciate both the foresight of the early network architects as well as the revolutionary advances in networking that have transpired

The principal challenges in designing such an emerging network arise in creating a physical realization, as devices and algorithms, of the logical structure, given the capabilities, constraints, and costs of potential physical constituents. We argue that adopting the holistic perspective is essential not only for the successful design of such a network, but also for the development of new communications technologies and network architectures.

A communications network is, from the holistic perspective, a dynamic distributed system of interacting components that collectively act as a medium for transporting information among users of and services accessible via the network, ideally anytime and anywhere with minimal delay and loss. Over time, applications and services have become more tightly coupled with communications networks to the point that they are usually considered to be components thereof, thus blurring the distinction between what is and what is not part of a network. Moreover, the behavior of a communications network is inextricably linked with that of its users and the environment in which it operates, and therefore the holistic view includes the network as well as these two external elements and their impact on it.

A network's effectiveness and efficiency depend not only on the careful design and analysis of its individual components, but also on a clear understanding of the functional interactions of all of the elements, both within and outside the network, so that synergistic interactions can be exploited and destructive interactions avoided. For example, the cost of network infrastructure, performance of applications, security and privacy of information transported, and energy consumed in network operation each depend on multiple diverse factors such as user needs, operating environment, transmission media, signal processing, device electronics, information encoding, and allocation and scheduling of network resources for users, to list only a few, together with the interplay among these factors.

As networks evolve and increase in complexity, the acquisition of a holistic view becomes more difficult but also arguably more important. The continued rapid growth in the number and type of users and devices connected to a network, applications and services available through the network, and communications devices and technologies forming the infrastructure of the network, have resulted in commensurate growth in the size and capabilities of networks and the heterogeneity and dynamics of their constituent parts. A clear understanding of the interaction of the various parts of a large and complicated communications network is key in designing the network to perform well and to allow rapid and accurate diagnosis and correction of problems, ideally without human intervention, when it does not.

No one person's mind can likely hold the entirety of the holistic view of such a network, but each person contributing to its design can formulate a partial view with varying degrees of granularity, and these partial views can be combined, according to the logical structure of the network, to form a full holistic view. The focus of each partial view is a portion of the physical realization of the logical structure, together with the interactions of those physical components. The remainder of a partial view is at the logical level, showing the position of this physical portion within the logical structure and its interactions with other parts of that structure. Formulating each partial view demands expertise in particular aspects of communications, computation, and control; assembling these partial views to obtain the complete holistic view demands collaboration among the holders of the partial views. IEEE Communications Technical Committees and, more generally, related IEEE Societies ranging from Antennas and Propagation to Systems, Man, and Cybernetics, are ideally positioned to provide the expertise and foster the collaboration necessary for formulating holistic views of emerging communications networks. Adopting the holistic perspective can lead not only to the discovery of modifications that improve the performance of existing networks, but also to the creation of entirely new network architectures and communications technologies that satisfy continually evolving needs for new services.

There are several aspects that will need to be fleshed out either by this Committee or some future follow-on Committees when the current ad hoc Committee expires at the end of 2021. Some questions that come to mind that need more detailed and concrete answers are:

1. The layered network architecture was created to simplify the research and design of a very complex system into weakly interacting subsystems so they can be designed with little coupling. In the future, when we optimize across layers and also include sensing, storage, computing and applications, how can we understand the complexity and build a sensible architecture?

2. It is dubious AI can design a new network architecture. Ultimately the human mind will have to perform this task. How can the human mind that strives for simplicity incorporate these complex components in its formulation of a new architecture?

3. How should and would government policies, regulations and legislation play a part in shaping the future network architecture?

4. Is there a way to assure "fair" access to services across the economic strata and prevent a widening of the "digital divide"?

5. How does the position on privacy and civil rights affect network architecture?

6. Where do we draw the boundary of privacy/civil rights and cyber/physical security monitoring, flagging and automated remedial actions?

7. Our privacy laws are historically based on human observing written and visual media. It has been hard to translate those rules to the cyberspace. Is data analytics on captured data an invasion of privacy? How can we detect maleficent intent and behavior?

8. What should be a good and productive partnership between technocrats, civil rights activists, legal scholars and legislators look like?

Vincent Chan