Research Topics related to Real-Time Communications over 5G Networks

[CCR Editorial Note]

Carol Davids
Illnois Institute of Technology
davids@iit.edu

Vijay K. Gurbani
Bell Laboratories, Nokia Networks
vkg@bell-labs.com

Gaston Ormazabal
Columbia University
gso@cs.columbia.edu

Andrew Rollins
Nokia Networks
andy.rollins@nokia.com

Kundan Singh
Avaya
kundansingh_99@yahoo.com

Radu State
University of Luxembourg
radu.state@uni.lu

This article is an editorial note submitted to CCR. It has NOT been peer reviewed.
The authors take full responsibility for this article's technical content. Comments can be posted through CCR Online.

ABSTRACT

In this article we describe the discussion and conclusions of the "Roundtable on Real-Time Communications Research: 5G and Real-Time Communications — Topics for Research" held at the Illinois Institute of Technology's Real-Time Communications Conference and Expo, co-located with the IPT-Comm Conference, October 5-8, 2015.

CCS Concepts

•General and reference → Surveys and overviews; •Networks → Network design principles; Programming interfaces; Network manageability; Programmable networks; •Social and professional topics → Computing and business; Computing profession; •Computer systems organization → Architectures; Cellular architectures;

Keywords

5G networks; Real-time Communications; Blockchain; Virtualization

1. INTRODUCTION

Cellular carriers are preparing to roll out 5G networks in the first half of the 2020's. The goals, requirements and standards for 5G are still being debated in the industry and the technical community and the defining characteristics of 5G are still being identified [3, 8]. Aiding the technical efforts to define 5G there are various industry standard efforts as well. Prominent among these are the METIS 2020 project [1] [11], the 5G NORMA project [2] [5] and the Next Generation Mobile Alliance (NGMN [3]). Academic institutions have also been at the forefront of 5G; the UK government announced the establishment of a 5G Innovation Center at the University of Surrey in 2012 while in the US, NYU Wireless was established as a multidisciplinary 5G wireless research center funded by the US National Science Foundation (NSF) and a board of 10 major wireless companies.

While definition and standardization are still under way, any goals for 5G have already been proposed to distinguish 5G from earlier mobile generations. These include: orders of magnitude increases in capacity and speed and orders of magnitude lower latency such that 5G networks can support machine to machine communications and real-time operations both medical and industrial; programmability using virtualization, Software Defined Networks (SDN) and Network Function Virtualization (NFV); interoperability with 3G, 4G and fixed networks; and support for *enterprise class* use cases of many kinds [10].

A panel of scientists and researchers from industry and academia addressed these topics at a roundtable discussion in October 2016, as part of the Illinois Institute of Technology 11^{th} Annual Real-Time Communications Conference [1] and the 8^{th} IPTComm Conference [2], which were co-located at the Illinois Tech Main Campus in Chicago, Illinois.

The panelists were:

- Luis Lopez, Universidad Rey Juan Carlos;
- Henning Schulzrinne, Columbia University;
- Vijay K. Gurbani, Bell Labs, Nokia Networks;
- Kundan Singh, Avaya;
- Radu State, University of Luxembourg.

Gaston Ormazabal of Columbia University and University of Luxembourg moderated the panel. Andrew Rollins

[1]https://www.metis2020.com
[2]https://5gnorma.5g-ppp.eu
[3]https://www.ngmn.org

ACM ISBN 123-4567-24-567/08/06.
DOI: 10.475/123_4

of Nokia and Carol Davids of Illinois Institute of Technology created the minutes, based upon which this editorial note was written. The panel first developed a working definition of 5G both for the purpose of general research and specifically as it relates to real-time communications over 5G and the mobile Internet. A discussion of useful research topics followed with many interesting observations by the audience. Interest was so high that the panel extended on to a second day. We share the results of this discussion in the paragraphs below and hope it will encourage further research and reflection.

2. DEFINING 5G AND IMPLICATIONS FOR RESEARCH

The 5th Generation Wireless System (5G) is the next major release of mobile telecommunications standards beyond the current 4G standards. 1G to 2G represented a fundamental shift characterized by a move from analog to digital cellular systems, new air interfaces, new devices, better security and mobility. 3G and 4G networks offered increasingly more bandwidth that allowed faster access to the Internet, and have made the mobile endpoints (smartphones) and content providers primary players. Today's devices simultaneously support 2G, 2.5G, 3G and 4G network interfaces plus non-cellular access channels such as WiFi, Bluetooth and ZigBee. A pertinent question is whether the user will end up with an umbrella technology that provides a unifying architecture where heterogeneous networks and a plethora of devices (phones, tablets, computers) work in concert, or the user will end up with dedicated and targeted networks: A special network for cellular data access and another specialized network to connect low-energy, Internet of Things (IoT) devices, for examples.

One school of thought is that 5G will act as a unifying network. The Next Generation Mobile Network Alliance (NGMN) predicts that the first, and perhaps early, release of 5G will occur in 2020. 5G is defined differently by different engineering disciplines. To researchers in the physical (PHY) and media access control (MAC) layers, 5G represents spectral efficiency that will deliver data rates between 1Gbps and 100Gbps to mobile devices that support multiple access technologies. To researchers in the transport and network layers, 5G represents a programmable and agile network driven by technologies such as Network Function Virtualization (NFV), Virtual Network Functions VNF), Software Defined Networking (SDN), and network slicing (defined and explained below). To researchers in the network and session layer, 5G promises low latencies for applications, such as gaming, real-time communications and 3D telemedicine. To researchers in the presentation and application layers, 5G opens up new venues in Machine-to-Machine (M2M) and Internet of Things (IoT). Lastly, to end users, 5G promises an adaptable network that conforms to the expectations of the user instead of forcing the user to adapt to the network. Unarguably, at any layer 5G raises the prospect of what the network can deliver to the user.

There are interesting implications for novel research directions in 5G to deliver on its promise. The more important ones are listed below, with the understanding that this is far from being an inclusive collection of research initiatives in 5G:

- Exhaustive testing of 5G systems from the physical layer to the application layer.
- Managing interference intelligently in heterogeneous networks.
- Novel beamforming techniques and advanced channel acquisition methods.
- Impact of virtualization of the radio access network, edge network, core network and data center networks that host and deliver 5G services.
- Role of analytics in 5G networks.
- Understanding and harnessing the Mobile Cloud Computing (MCC) revolution driven by dramatic increased network capacities of 5G.
- Scaling up mechanisms for network intrusion detection and security event classification in 5G networks.
- New programming paradigms and APIs for 5G networks.
- The role of standardization bodies in 5G: what to standardize while leaving room for innovation and how fast should standardization move?
- The greening of 5G: how to make batteries last longer and minimize power consumption in 5G mobile and network devices.
- Ensuring reliability and availability in a network with virtualized functions and perhaps multiple SDN controllers.

3. VIRTUALIZATION OF NETWORK ELEMENTS AND FUNCTIONS

The virtualization of network function is inherently challenged by two major factors: scalability and security. One current approach to virtualize network functions consists of running multiple virtual machines over a hypervisor. In most cases, these virtual machines are scaled-down versions of complete operating systems such as Linux, where many additional unnecessary processes and applications run on these systems beyond what's strictly needed for the specific operation of a virtualized network function. These additional processes offer a larger attack surface and a significantly larger resource footprint.

While the Docker application[4] is one step towards a micro-service based virtualized network function, highly promising approaches based on unikernels [9] might represent the longer-term solution for implementing virtualized network functions. Unikernels are small footprint containers, running a specific dedicated functionality. They allow for better security such as memory randomization, no shell, and no additional services, and thus are ideal candidates for implementing networked virtualized functions. In terms of research objectives, we expect significant work to be performed in areas related to compiler generation, benchmarking and testing, as well as unikernels orchestration.

[4] http://www.docker.com

4. NETWORK SLICING

Network slicing is a new concept in 5G. A 5G network slice is analogous to a virtual local area network (VLAN) in traditional IP networking, although more complex. The intention of a 5G slice is to provide uniform traffic treatment to all applications that belong to, or are executing in, the slice. In other words, a 5G network slice supports a communication service that has certain expectations from the network - such as low latency and high bandwidth - and provides specific ways of handling the control- and user-planes for this type of service. A 5G slice is composed of a collection of specific network functions and specific radio access settings that are combined together for a specific use case. A 5G slice incorporates all layers and domains of the network: it includes software modules running in data center cloud nodes, specific configurations related to the radio access network, the wired network and even the configuration of the specific 5G device. For example, a 5G network slice that provides security, low latency and reliability could be used for real-time traffic while a slice that provides high bandwidth could be used for large-scale data transfers. A 5G network would support multiple slices concurrently, with the attributes of each slice being different from the others. The 5G network slice concept could be a key enabler for the service provider to expand existing businesses and create new business opportunities. The expectation is that trusted third parties will be provided with permissions to control aspects of the slice via suitably defined APIs.

4.1 Role of the controller in network slicing

SDN Controllers can be used to support network slicing in 5G. The SDN controllers could move from the SDN flow paradigm prevalent today to include the service paradigm. More specifically, instead of manipulating network elements based on flows, the 5G SDN controllers could be made service aware and could interface with the different elements in the 5G network that execute service logic and orchestration.

4.2 Programmability and P4 in support of network slicing

The proposed P4 programming language for programmable switches will likely have a disruptive impact on both research and technology as it extends the programmability of network devices towards the data plane [6]. This language was initially developed by an informal consortium comprised of Barefoot Networks, Google, Intel, Microsoft, and Princeton University and lead to the creation of a recent formal P4 Language Consortium[5].

A P4 program requires the specification of five elements. The first element contains the header definition. Headers are ordered lists of field names and their lengths. The second element provides the parser required to parse the headers. The parser describes the transitions in a state machine traversing the individual headers. Each extracted header and corresponding value is sent to the match+action tables. These tables compose the third element, where table entries specify the action to be undertaken, when a specific match occurs. Matches can be exact, wildcards, or ranges and are implemented by various hardware specialized circuit components (TCAM, SRAM). Once a match occurs, an action is executed. Actions represent the fourth component of a P4 program. Actions can be elementary/primitive, or parallel executions of several elementary/primitive actions. The final P4 program component is the flow of control among the different tables. This flow is specified via a mix of conditions, references and functions. The P4 language specifies this control flow between a series of logical match+action tables in an abstract, programmer-friendly way making it easy to add new protocols that can be executed efficiently. Because P4 is openly developed in cooperation between networking industry and academia, it provides academia with an opportunity to shape the specification and operational semantics.

[5]http://p4.org

5. BLOCKCHAIN

The technological innovation behind the virtual currency Bitcoin's unexpected takeoff is the underlying distributed public ledger. Publishing transactions which are packaged in chained blocks, with cryptographic hashes as safe-guards against very resource-powerful attackers, the blockchain and the incentives related to the *mining* performed by a large volume of working nodes solve an important problem in distributed computing: achieving distributed consensus without needing trust models or a trusted third party.

With blockchain technology as a support for Internet of Money (IoM) and an ever-increasing range of application domains ranging from asset management, IoT, insurance, and alternative coins, the underlying network protocols and network infrastructure need to address the timing requirements of blockchain-based applications. While current transaction rates in the Bitcoin network are modest, due to the design of Bitcoin itself, the extensions towards micro-payment schemes and proof-of-stake based consensus schemes require efficient peer-to-peer flooding protocols, such that blocks and transactions may efficiently reach all the corresponding peers. For the relatively simple existing protocols, with demonstrated performance only for peer sizes in the 4 digit ranges, scalability will become crucial in the future. Network assistance can be prototyped and designed in the short term, with existing approaches like Application-Layer Traffic Optimization (ALTO [4, 7]), and can guide the peer connection process thus improving the bootstrapping of the existing blockchain. Longer term strategic research is required for scaling up to millions of nodes that inter-operate through a blockchain.

Lightweight virtualization of network resources is another stringent requirement in this area. While simple Bitcoin nodes can run simplistic scripts, the architecture proposed by Ethereum[6] generalizes the virtual machine running on a Blockchain towards a Turing complete specification.

With nodes on the blockchain implementing smart contracts, or even allowing for unbounded participants (as is the case of decentralized applications, or Dapps) the nodes currently require their own virtual storage space and virtual CPU resources. In the near future, virtual network space may be the logical evolution proposed for such approaches. One potential enabler is network slicing on a per Dapp basis.

6. ANALYTICS AND SECURITY

Big Data and Analytics are expected to be key components of a 5G architecture. When we move beyond the technical magic of 5G into the realm of what it promises

[6]https://www.ethereum.org/

its users (an adaptive network that provides personalized and contextualized services, a *network of you*) we come to the realization that this promise largely rests on intelligently leveraging Big Data and Analytics in 5G. We need Big Data Analytics to allow the 5G network to create a new slice, to maintain end-to-end performance, to continue providing assured network services to customers, to combat security threats to the network in (near) real time, and to maintain the agility and elasticity of the network. The volume and velocity of data in 5G is expected to far surpass other well-studied global networks. Twitter and WhatsApp have had record-breaking days during 2015, and it is expected that they and other over-the-top (OTT) services will continue to drive network volume and velocity as 5G is defined and rolled out.

There is a significant fear that end users may reject analytics because of privacy issues. Privacy is not only a technical artifact. Technically, privacy could be afforded to any protocol. Whether or not the user understands the privacy implications of the protocol is a separate question. Privacy research should look beyond how technology can increase privacy, and beyond security vulnerabilities, to discover what are the plausible uses of that technology or application by end users who are not engineers.

7. APPLICATION DEVELOPERS

Application developers have the privileged opportunity to shape the user experience of 5G networks. This is because mobile applications are often the only means by which an end user touches the underlying wireless technology. Applications research can go in multiple ways to investigate and exploit the potential of 5G for end users, and at the same time protect operators and their networks.

5G's many goals include four in particular that application developers and researchers should seek to exploit and study. These are:

1. Very high bitrate compared to 4G,
2. High bitrate at high movement or speed,
3. Support for large number of nodes,
4. Reasonably high bitrate at low latency.

An application could, for example, process and analyze the real-time video from a car dash camera, or publish it in real-time to a cloud service. This example utilizes one or more of the factors mentioned above. The application developer and researcher should place particular emphasis on using 5G capabilities that are not available on 4G networks.

Learning from 4G experience, there is a need to separate those who develop 5G implementations and those who certify them as meeting the 5G technology requirements. It would be useful to develop publicly available tools and applications that an end user can quickly install, test and provide input regarding how well a particular 5G implementation works in real world and in daily use. Such tools and test beds would help create a realistic crowd-sourced comparison of various implementations. Such a test bed will allow consumers to learn whether an implementation adheres to the manufacturer's or service provider's claim under average conditions.

Today's mobile apps are still in their early stages of evolution. Restricting them to only a few constrained environments such as an iOS or Android ecosystem hinders their potential for explosive growth. There is a need to explore a wider range of wireless systems and interfaces beyond what is available on our favorite mobile phones. For example, low cost consumer grade 3D cameras could add a new dimension to immersive video collaboration experience, but add more challenges at the system- and application layers. Distributed applications built out of commodity devices require a distributed- components software architecture, e.g., to allow conference room participants to expose their mobile phones as microphone input to create spatial audio streams delivered to remote participants. There are two parts to componentization: considering mobile devices as part of a larger universe, versus using elements of your device as individual separate entities. Software systems and architectures for quickly developing such applications could fuel the innovation promise of 5G.

Applications and systems research often go hand-in-hand. Application developers use APIs to interact with the system. Creating elegant and simple APIs to expose the control knobs of the emerging technologies is not easy: To select an underlying network policy, for example, a simple socket option may suffice. However, depending on the application's requirements, fine-grained new APIs to control the feature from the application may be necessary. Moving such controls from the network or service to the application opens doors to many more types of use cases than what we see today. There is a need in the research community to focus on APIs: Are they appropriate and homogeneous? Are they at the right level of abstraction? We should be able to characterize APIs along cognitive dimensions of notations - abstraction or viscosity, for example. In short, research should consider mechanisms for evaluating APIs.

With high bitrate, it is also desirable to move complex processing outside battery- constrained devices: A mobile phone, for example, could move the video processing of a call to the corresponding phone if the latter is on a charger and the former is not. Collective CPU resources of one's devices can be exploited to create innovative distributed applications. Interconnecting, inter-operating and handing-off between devices for bandwidth-intensive applications are possible too. For example, it will be interesting to do application level multicast using only mobile phones to distribute a live video event to hundreds of viewers.

In the past, wireless technology controls have largely remained hidden from application developers. With recent growth in mobile applications and software-defined networks, some special network capabilities are moving to developers. This calls for research and development not only in underlying systems, but also in novel and specialized applications as well as the interfaces between the applications and the systems. We need software systems and frameworks to easily create new applications demonstrating a wide range of system and network capabilities, and under various types of user scenarios.

8. SOCIOLOGICAL IMPACTS OF NEXT GENERATION NETWORKS — THE CONTINUING SPECIALIZATION OF NETWORK PERSONNEL

Engineers and scientists continue to specialize as technology evolves. Each update to networks requires further specialization for development of that network, innovation of the network, and even maintenance. The list of research topics for 5G will undoubtedly result in specializations within the electrical engineering, computer engineering, and computer science disciplines.

In the 80s and 90s, the engineer or student would build the network to support their work. This would include updates to computing devices, and components, networking equipment, routing protocols, and any applications. Through the construction of the network and applications, the researcher was familiar with all aspects of the network, including functionality, troubleshooting, performance, etc. Over time, these advances were commercialized and standardized. Standardization has allowed for the commoditization of components, allowing the research to move onto more challenging topics. At the same time, IT personnel have assumed support for the base network, abstracting that from application(s) running on it. IT personnel have further fragmented to specialize in hardware, routing/switching, security, and device support. Thus, engineers and students have relinquished the network programming and support to others, focusing instead on the next set of topics.

The complexity of networks has increased, and networks are layered on each other: Software clients communicating via virtual networks to servers running on virtual machines communicating via software defined networks running on computers communicating via networking technologies via physical connectivity through physically cabled routers and switches. Support for the different network layers in this multi-layered scenario falls to different personnel. Further, troubleshooting issues in any given network initially assumes that the underlying network is intact and working. Issues that may have previously been solved by a single engineer now require a team of experts to dig through the networking layers to find the cause for the observed behavior.

Driving the specialization issue are the dynamics of the players in the industry. There are essentially four main actors in the mobile ecosystem: Equipment vendors who build the networking equipment; carriers who mostly perform customization and integration in today's networks; device vendors (distinct from equipment vendors) who provide new mobile devices (phones, tablets, etc.); end users who are the consumers of the networks and the devices running on them. There is still another actor, the content provider who creates content that the users consume, but the role of content provider is not germane to this specific discussion.

The issue of specialization of network personnel is made acute simply because operators have decided to invest in deploying and running networks at the expense of conducting novel research into related technologies. This trend is evident in Europe and the Americas where the operators assume the role of marketing and collecting revenues from subscribers while subcontracting the management of the network to equipment vendors. The side effect of this decision is the operator does not have personnel who understand the key technologies like SDN that will be deployed in their networks. The operators depend on equipment vendors for the research, deployment and management of similar technologies. In such an ecosystem, much of the value is added through verticals (IoT, for example) and revenues are generated from supporting the verticals. The carriers do not have network personnel to create custom applications and must depend on equipment vendors for new applications.

To some extent this behavior mirrors what has occurred in the enterprise IT space: the decreased funding for IT exhibited itself in deployment of Commercial Off The Shelf (COTS) hardware that did not require programmers who understood scripting, programming, or a deep knowledge systems administration.

As stated, analytics and big data will have a large impact on the operation and maintenance of these networks. The challenge will be driving collaboration between the engineers performing the analysis of the data and those working in the network to resolve the issue.

9. CONCLUSIONS

5G technologies encompass a variety of disciplines, and extend the network in new and innovative ways, allowing end users even more interaction with devices and other users. Virtualization and SDN will be key parts of this evolution, allowing users and operators to build environments that meet their needs, while coexisting with other networks in the same physical space.

In addition to expanding engineering education, leadership training and education should be added as part of robust engineering/technology/scientific curricula. This should include topics such as improving communication (written and verbal), team building, and general troubleshooting techniques and algorithms (especially tracing a problem through multiple systems and networks the researcher has little direct experience with or knowledge of). As network complexity increases (and more and more virtualized networks are layered on other networks), teams of engineers will need to be able to work together to quickly and efficiently build and troubleshoot networks.

10. ACKNOWLEDGMENTS

The authors of the paper thank Henning Schulzrinne and Luis Lopez for participating on the panel and imparting insights that are reflected in this paper.

11. REFERENCES

[1] Illinois Tech Real-time Communications Conference and Expo. http://www.rtc-conference.com/, 2016.

[2] IPTComm: Principles, Systems and Applications of IP Telecommunications. http://www.iptcomm.org/, 2016.

[3] P. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour. Design considerations for a 5G network architecture. *IEEE Communications Magazine*, 52(11):65–75, 2014.

[4] R. Alimi, R. Penno, Y. Yang, S. Kiesel, S. Previdi, W. Roome, S. Shalunov, and R. Woundy. Application-Layer Traffic Optimization (ALTO) Protocol. RFC 7285 (Proposed Standard), Sept. 2014.

[5] A. Banchs, M. Breitbach, X. Costa, U. Doetsch, S. Redana, C. Sartori, and H. Schotten. A novel radio multiservice adaptive network architecture for 5g networks. In *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, pages 1–5. IEEE, 2015.

[6] P. Bosshart, D. Daly, G. Gibb, M. Izzard, N. McKeown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat, G. Varghese, and D. Walker. P4:

Programming protocol-independent packet processors. *SIGCOMM Comput. Commun. Rev.*, 44(3):87–95, July 2014.

[7] V. K. Gurbani, V. Hilt, I. Rimac, M. Tomsu, and E. Marocco. A survey of research on the application-layer traffic optimization problem and the need for layer cooperation. *IEEE Communications Magazine*, 47(8):107–112, August 2009.

[8] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu. 5g network capacity: key elements and technologies. *Vehicular Technology Magazine, IEEE*, 9(1):71–78, 2014.

[9] A. Madhavapeddy, R. Mortier, C. Rotsos, D. Scott, B. Singh, T. Gazagnaire, S. Smith, S. Hand, and J. Crowcroft. Unikernels: Library operating systems for the cloud. In *Proceedings of the Eighteenth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '13, pages 461–472, New York, NY, USA, 2013. ACM.

[10] NGMN Alliance. NGMN 5G White Paper. https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf, 2015.

[11] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren. Scenarios for 5g mobile and wireless communications: the vision of the metis project. *IEEE Communications Magazine*, 52(5):26–35, May 2014.