CROSSMobile:
A Cross-Layer Architecture for Next-Generation Wireless Systems

Bob Iannucci, Patrick Tague, Ole Mengshoel, and Jason Lohn

{bob,patrick.tague,ole.mengshoel,jason.lohn}@sv.cmu.edu

CMU-SV-14-001
Abstract—Commercial cellular networks emerged from the wired public switched telephone networks (PSTN) in an evolutionary manner. Market pressures fueled upgrades in bandwidth and functionality. Four decades later, these networks maintain historical artifacts from PSTN networks, and the artifacts work against the fundamental needs of today’s wireless systems. Non-cellular wireless networks are used beneficially within geographically limited domains (enterprise, at home), but such networks lack the architecture to scale geographically.

In this paper, we step back and re-evaluate existing wireless network architectures, identifying inherent limitations and offering a new set of architectural principles that, we contend, will lead to significantly improved overall system performance and scalability. Based on these principles, we propose the CROSSMobileSM architecture that is enabled by controlled cross-layer information exchange between radio, network, and application layers (both on-device and in-cloud), coupled with information-owner-based privacy and security controls. We discuss how this architecture can provide increased value to equipment and device manufacturers, application and network service providers, and end users. We close by outlining a number of open research questions.

I. INTRODUCTION

Mobile computing is rapidly overtaking traditional computing for many, if not most, applications. The power of having access to one’s world of information at any place and time is transforming lives, businesses, and economies. The devices that have fueled this transition are remarkable and have themselves undergone a dramatic transformation.

Initially, and constrained by the technology of the day, mobile phones were used to make and deliver voice calls with spectrum efficiency, protocol agility, and good intelligibility while on the move. Pioneers recognized the potential value of the programmability of mobile devices [25], anticipating the inevitable advance of enabling technologies [20]. The embedded computing model, with voice as the “killer app,” gave way to the platform computing model. By bringing desktop-class operating system thinking to increasingly capable mobile phones, programmers were enabled by the same tools, libraries and techniques that they knew from the desktop world. This, in turn, led to the broad spectrum of apps that has transformed our lives. However, as apps have become increasingly dependent on cloud services, issues related to the underlying network have arisen.

The cellular networks behind these devices have similarly undergone transformation, but unlike phones with life cycles measured in months, cellular network life cycles are measured in significant fractions of decades. Capital-intensive build-outs, the need to preserve backward compatibility, and the natural pace of standards activities have all contributed to substantial evolution within network layers and elements. But evolution of the architecture overall has been slower.

If the phone plus the network are to be the next-generation computing platform, the network should be able to evolve as rapidly as phones. Software-defined networking (SDN) and network functions virtualization (NFV) are steps in the right direction but may not go far enough. In addition, within the scope of existing cellular network architecture, significant state information that could be beneficially shared across layers is hidden (imagine apps informing the network of their anticipated future needs and the network informing apps about impending changes in the condition of the wireless channel). This results in sub-optimal network performance and user experience.

Further broadening the scope of the problem, non-cellular wireless networks (e.g., WiFi) have also matured in bandwidth but, due to lack of an architecture for federation, remain as isolated islands. Efforts to bring cellular and non-cellular wireless networks together have had only modest success to date.

In this paper, we consider the challenges and opportunities
central to commercial wireless networks (CWNs) and argue that these can be addressed by

- creating an overall architecture that supports compositional construction, relying on the forming of negotiated “agreements” between network elements rather than on slowly-evolving standards to assure network integrity,
- defining mechanisms by which information available at various layers in the network can, under policy-based governance, be beneficially shared across layers, and
- establishing the means by which processing and communication resources for both network functions and “app” functions can be deployed efficiently and in real time based on supply and demand to create a platform architecture for wireless computing.

In Section II, we go deeper into the standing challenges faced by cellular networks from the point of view of several different constituencies. In Section III, we highlight relevant prior work. In Section IV, we lay out a new architectural approach for CWNs and explore how it addresses the identified challenges. In Section V, we outline some relevant research questions.

II. CHALLENGES AND PLAYERS

A. Background

A look back at the evolution of cellular networks reveals strong ties to the wired public switched telephone network (PSTN). While commercial mobile phone service of sorts dates back to the 1920’s, we focus on the last four decades of cellular mobile telephony (see Figure 1). In their beginning, cellular networks were designed for voice signals carried over analog radio channels. Despite early efforts with low-data-rate modems, the networks had essentially no significant data-handling capability.

GSM- and CDMA-based digital networks changed this. These networks offered the raw potential for data transport,
But the transformation from analog was motivated more by a desire for increased channel capacity than by a vision to ignite a mobile computing revolution. Nevertheless, end-to-end digital capabilities did indeed change our thinking, and the bandwidth race was on. Gateways were created to bridge the Internet to these bandwidth-limited networks.

Since that time, channel coding efficiencies have improved, spectrum allocations have been increased, and large cells have been split into small ones—resulting in significantly more available capacity. But the basic architecture for data transport has remained relatively unaltered since its inception. Until the end of 2009, voice traffic dominated data traffic, and the underlying network architecture was largely circuit switched—an artifact from the pre-mobile-network era. With the appearance of smartphones and the development of cloud-connected mobile applications, this began to change. Today, data traffic dominates voice traffic (Figure 2) [1], and the core network architecture is evolving toward packet switching.

In parallel with the evolution of cellular networks, non-cellular networking using unlicensed spectra has matured. WiFi networking based on the IEEE 802.11 family of standards has become an important component of the mobile landscape. The premise of building this next generation of mobile apps on top of unlicensed networks is alluring when one considers the cost (essentially free), the typical network speeds, and the possibility of actually building a high-coverage network through crowdsourcing.

Attempts to federate home and enterprise unlicensed networks have mostly failed to gain widespread acceptance, at least when one compares the footprint of federated networks to the coverage of cellular networks. Recently, network service providers have made a push toward integration of WiFi into cellular networks, offering subscribers unmetered broadband connectivity while simultaneously offloading traffic from heavily loaded cellular networks. We observe that in most cases, the use of WiFi in such networks is tangential, not central, to the network’s architecture. Rather than accepting WiFi as just another radio bearer—completely interchangeable with the cellular radio for both data and voice services—with few exceptions, WiFi is relegated to data services only. Voice calls continue to be routed over and terminated on the cellular network. We see this as an artifact of the clash between the evolutionary paths of voice and data that will remain until both systems are jointly reconsidered.

In summary, cellular networks are far behind what we might be possible today if we had the luxury of starting over.

B. Challenges

As applications grow in significance and become more distributed (cloud-dependent, cloud-based, peer-to-peer, M2M), the traditional layered approach to network architecture exposes its limitations, and the ways in which networks have been built out resist change. The traditional monolithic model for commercial cellular networks imposes severe constraints on both the analysis of network performance and the rapid introduction of new capabilities that may go beyond established standards.

In particular, we observe the following challenges of cellular networks and their clients (wireless devices and the apps that use the network):

- **Network Impermeability**: From the perspective of today’s cloud-connected mobile apps on cellular networks, a round-trip involves passing across the network operator’s radio access network (RAN), across the operator’s core network, across a core-to-Internet gateway, across the Internet to the cloud service and back again. The RAN and the centrality of core-to-Internet gateways are sources of latency—in some networks, significant latency. But, perhaps more importantly, cellular networks are essentially monolithic entities, presenting an impenetrable barrier into which cloud computing cannot be readily migrated.

- **Hidden RAN State**: The mobility of clients, and the implications of that mobility, may be known by elements of the RAN, but they are essentially invisible to mobile apps and the cloud services behind them. Knowing both a client’s radio history and its largely-predictable radio future could be used, for instance, by media streaming applications to adapt intelligently. As an example, we observe that traditional cellular networks, dating back to the days of analog, have taken advantage of sophisticated, but only relatively static, antenna design. Sectorized receive arrays are commonplace as are electronically-steerable sector antennas. But antenna reconfiguration was initially envisioned as a relatively static activity to mitigate seasonal variations in RF propagation. Handoff techniques were created to apply intelligence to RF signal management as mobile devices moved from one base station’s coverage to another. Beamforming and MIMO techniques have further enhanced the capabilities of cellular systems to improve available bandwidth between

---

3The Generic Access Network (GAN) system [24] offers the potential to extend a carrier’s network over alternate IP networks. Its use requires a mobile device that is GAN capable. The system is more commonly referred to in the marketplace as “Unlicensed Mobile Access” (UMA). GAN is also used as a mechanism to extend a carrier’s network to femtocells.
the network and devices. However, the information used to manage and adapt the RF portion of cellular networks remains mostly limited to the physical layer and is mostly stateless. Measuring and modeling the trajectory of mobile devices, over time, could provide better predictive input to the process of managing the RF channels between the network and all of the devices seeking to use it.

- **Hidden Network State:** In early telephone networks, congestion manifested itself in very simple ways—a call either went through or it didn’t (the caller received *user-busy* signal or an *all-circuits-busy* signal). Allocation of network resources (to a connection) was static and, in essence, guaranteed by the fundamental nature of *circuit switching*. But nearly every aspect of telecommunications networks has abandoned this model in favor of *packet switching* in which connection quality is only statistical and subject to variability throughout its lifetime. The state of the IP path from client to server is largely invisible to mobile apps and their cloud services but can have a significant impact on the app’s performance [17]. *Bufferbloat* [8] has been identified as one significant factor in network performance. More generally, the ability of an app or its cloud-side counterpart to adapt to varying network characteristics is limited. We acknowledge that visibility to network state is valuable in both wireless and wired networks. But in cellular networks where the core is, presumably, under a single administrative point of control, there is an untapped opportunity for optimization.

- **Hidden App State:** Just as the RAN and network states impact cloud-connected mobile apps, the state of apps impacts the RAN and network. While the developer may possess significant knowledge about an app’s network-impacting characteristics—such as its bandwidth needs over time, its latency tolerance, and its ability to benefit from enhanced quality-of-network-service—he or she has essentially no means to express this. The user may also have similar knowledge (*e.g.*, the value of being able to download a large file now) and likewise has no means to make this known to the network.

As such, the structure of current networks precludes potentially significant opportunities for system-level optimization. Moreover, their closed nature resists the very studies that would make such opportunities more apparent.

**C. Players**

The wireless ecosystem (encompassing present-day cellular plus non-cellular networks) is made up of many players, each with issues and opportunities unique to their role. A rational approach to the above problems must make sense in the context of each of these players.

- Telecommunications equipment manufacturers (TEMs) seek differentiation, yet they operate within the bounds of slowly-evolving standards.
- Mobile Device Manufacturers (Phone OEMs) are similarly constrained with respect to standards, but the growing importance of on-phone computing and apps indicates an opportunity to better leverage a network with dynamic properties. They straddle the desire of mobile service providers to maintain closed, carefully-governed networks and the desire of the market to take advantage of “free” bandwidth available via WiFi networks that offer few service guarantees.
- Commercial Mobile Service Providers (Carriers) face the challenge of generating more revenue per unit of investment, offering a differentiated service in a market that views bandwidth as a commodity. Further, closed, fixed-functionality network elements from TEMs offer little flexibility to compose differentiated services, and network refresh cycles are measured in years while changes in user behavior occur on the timescale of days.
- Operators of unlicensed wireless networks—enterprises, small businesses, and homeowners—lack the means to pool their individual investments in WiFi (and other) networks even when it makes economic sense to do so.
- Application Service Providers (ASPs) and App Developers face a nearly completely opaque network that stands between them and their customers, yet the network’s time-varying characteristics (especially at the RAN level) have a first-order effect on service quality. App developers in particular have access to a wealth of information about the application’s intent (network access patterns, QoS needs) that would be valuable to the network in making resource allocation decisions.
- End Users regularly experience variable network performance, manifested as call drops and apps that “hang” on high-latency or failed network transactions. This experience is the product of the contributions of TEMs, carriers, and ASPs with the user left as the system integrator. But users lack tools and expertise to diagnose problems. User frustration often results in the “churn” that carriers experience. But the solution users seek can rarely be found by simply switching to another carrier.

A better systems architecture would provide the means to diagnose and manage problems despite the multi-party nature of its structure. It could also offer the possibility of dynamically adapting to change up to, and including, dynamic additions and deletions of new capabilities to and from the network. All that we have experienced in the last 40 years of cellular networks (and the further scaling-up we anticipate with the coming Internet of Things) points to the conclusion that demand for service will continue to outstrip supply—consequently, a better systems architecture will allow the participants to treat the network as a dynamic marketplace in which the resolution of supply and demand happens more organically.

**III. Enabling Technologies and Related Work**

While we believe strongly that a new perspective on CWN architecture is needed to satisfy these needs, much prior work has been done on technologies and subsystems that will undergird most any proposed solution.

- **LTE and 4G Networks:** LTE [2] and subsequent cellular architectures promise increased wireless bandwidth and reduced latency for data services, but these performance gains
come at the cost of quality of service (QoS), as circuit switching is eliminated. Core latency and the potential for bufferbloat are eliminated. Moreover, LTE networks do not provide efficient or cost-effective data services for the embedding of computing and storage into the network. This presents a barrier to low-latency mobile-to-mobile apps (e.g., games) but also to the IoT. The fraction of IoT data using cellular networks remains near zero. While there are several reasons for this (cost being primary), we also believe that this pattern will change, and substantial traffic over CWNs will originate from and terminate on IoT devices. One barrier is sensor-to-control-to-actuator latency which is both high and variable in today’s cellular networks. Solutions that fundamentally reduce latency, improve predictability, and support embedding of control functionality (built on general-purpose computing) in the network are enablers.

**Cross-layer Design:** Modeling wireless networks with the same layering concepts as wired networks does not adequately address the significant differences between mostly-reliable wired links and mostly-variable wireless links. The concept of cross-layer design to mitigate the impacts of wireless links has been studied, and numerous past proposals have been reported in the literature [19], [21]. We believe these and related techniques are essential in the design of next-generation wireless networks. We note that a significant body of this work took place prior to the dramatic rise in importance of mobile applications and, therefore, we see opportunity to extend these concepts to further explore physical-to-application cross-layer design.

**Broadband Wireless for Unlicensed Spectrum:** In a parallel line of evolution, unlicensed wireless networks were designed to handle high-rate data transfer without restriction from service providers or network operators. The IEEE 802.11 family of standards has gradually evolved over recent decades to the point that WiFi implementations can reliably support hundreds of megabits-per-second of data throughput. More recently, WiFi has been incorporated into mobile devices, partly to provide users with the higher data rates that they desire and partly to offload traffic from the constrained cellular networks. While there have been efforts to allow cellular and non-cellular networks to converge in terms of services and capabilities, the disparity of service models and opportunities for financial gain have more-or-less prevented real convergence into a heterogeneous system.

**Fog Computing:** Fog computing [4] argues that the cloud should penetrate the network, instead of existing on the “other side” of the core. Architecturally, integrating the cloud with the network can improve performance substantially but in a transparent way. While fog computing was originally conceived for WiFi networks in the form of Cloudlets [18], suggesting that previously “WAN-limited” computational services can be offered at or near WiFi access points, we are motivated to explore application of the concept (a) with full integration into CWNs and (b) with the flexibility of migrating computation transparently as users move and as utilization of resources elsewhere in the network may dictate.

**Software-Defined Networking:** Software-defined networking (SDN) [12] offers architectural degrees of freedom that are most welcome in CWNs. The OpenFlow SDN framework [13] enables deeper control over IP network resource allocation and management, while providing transparency between control and data plane operations. While SDN drastically improves network and resource management, it alone cannot solve many of the previously described challenges in mobile systems.

**Software-Defined Radio:** Software-defined radio (SDR) demonstrates that computational cost has declined to the point that radio functionality previously dependent on expensive and inflexible custom hardware can now be implemented with inexpensive general-purpose computing elements. High-performance radios built using commercial, off-the-shelf (COTS) computers, augmented with frequency-agile radio-frequency (RF) “front end” hardware (e.g., the Universal Software Radio Peripheral (USRP) from Ettus Research [7] and similar devices) are used in commercial deployments today, and the trend toward SDR is increasing. Open source software packages such as e.g., GNURadio [3] make SDR accessible to many. Implementations of cellular standards such as OpenBTS [6] and emerging open implementations of LTE go further to demonstrate the degree of sophistication possible with commodity equipment. SDR itself offers the potential to open heretofore closed network elements and to provide significant freedom in use of available spectrum.

**Coding Theory:** Coding has reached the point where we are now approaching the Shannon limit. For example, Spinal codes [16] build high quality real-time channel models intended to inform encoding, thereby providing high rate communication over time-varying wireless channels. We believe that the channel modeling can inform and can be informed by other layers in the network, especially in the realm of how user histories and physical network models may be predictors of variations in channel characteristics.

**Application Development Environments:** A key enabler for CWNs was the opening of computing and communications capabilities in mobile phones to third-party application developers. Two aspects of this well-known trend are especially important. First, apps are increasingly a combination of on-phone code and in-cloud services—a single app developer controls both the logic on the phone (at least for the app itself) and the logic of the cloud service. This presents an opportunity to capture key characteristics of both ends of the CWN connection via a single locus of control at the same time. Second, they represent an important point of policy injection. Apps offer the opportunity to capture the user’s view of the instantaneous value of network resources (e.g., willingness to pay extra for temporary increases in bandwidth allocation in order to play a game, download a file, stream a video, and the like). They also represent the mechanism to use this knowledge for negotiating with the network itself—assuming that a suitable interface can be created.

IV. A NEW WIRELESS SYSTEMS ARCHITECTURE

The substantial, accumulated investment in legacy cellular networks and related standards have both created the demand for a true mobile computing platform and established a complex set of constraints that inhibit its rapid evolution. Barriers
such as opening closed networks and introducing heretofore-unknown interfaces to existing telecommunications equipment border on the unthinkable. At the same time, we observe that enabling technologies such as those outlined in Section III make possible, in concept, the creation of a functionally-equivalent, fully open network that could serve as a research platform for studying the challenges and concepts we have outlined.

To that end, we are pursuing the creation of an open research testbed for future wireless systems (similar in spirit to the Arpanet), operated among research institutions with the intent of enabling a broad array of research studies related to creating a next-generation wireless computing platform. We call this the CROSSMobile testbed and corresponding CROSSMobile architecture. It is our hope that the creation of this testbed and architecture will accelerate innovation and ultimately yield results that can also be “back-ported” to legacy networks. We have begun the creation of such a network at Carnegie Mellon University’s Silicon Valley campus. CROSSMobile architecture, at the conceptual level, is illustrated in Figure 3 and is based on the following four principles:

**Cross-layer design revisited:** While transparent and layered networking allows for simplified and modular design, it prevents service integration and precise resource allocation and control. Cross-layer design enables these. Proprietary hardware components that manage a single network layer should give way to open sub-systems that expose internal state information to partner sub-systems at different layers.

As an example, the RAN must be able to expose operational and performance statistics and to take input about configuration and operational requirements from higher protocol layers, applications, and services. Instead of relying on each layer to individually optimize based on inferred operations, the system can operate as a whole to optimize system performance based on informed cross-layer exchanges of information.

In addition to sharing of information across existing elements, we also imagine the creation of a distributed infrastructure for gathering real-time information about the network and its usage, building dynamic models of overall system behavior, and applying insights gained to control and optimize network behavior at various levels.

CROSSMobile architecture exposes an application programming interface (API) at each layer in the system together with mechanisms to allow information to traverse this API in both directions (reporting and control) subject to policies, expressed as sets of rules.

**Open, compositional structure:** With policy-governed APIs at each layer, the need to assure system integrity by keeping the network closed can give way to a more open approach, even when the network is composed of Mutually-Suspicious Elements (MSEs). MSEs can join and leave the network in much the same manner that merchants join and leave an open marketplace. Dynamically negotiated trust relationships based on mutual benefit can be created between pairs of network elements. A cluster of WiFi access points can join a nationwide network, bartering local bandwidth for global access. An enterprise can offer IP transport across its private core to a commercial cellular provider in exchange for call-termination services. Self-Organizing Networks (SON, defined within 3GPP Release 8 and subsequent specifications [23]) are a significant step in the right direction for radio access networks. For true open, compositional structure within the CROSSMobile architecture, we imagine a related capability that spans the entire network up to and including agents in apps and in cloud services that negotiate trust relationships under which state information can be beneficially and securely shared.

**Application- and service-specific resource allocation:** A
system-level approach to resource allocation and management affords tighter control and more optimal use of scarce resources. At a minimum, we imagine a systems-level mechanism to record, correlate (across layers), analyze and control low-level, layer-specific events for the purpose of making more informed resource allocation decisions based on sophisticated machine learning algorithms. Richly-instrumented network elements, contributing real-time state information about themselves and their environments (radio, IP connectivity, computational load) offer a substantial opportunity to optimize resource usage at the system level rather than just at the element level.

**Integrated computing:** The classic separation between networking and computation limits performance in today’s cellular networks, similar in spirit to layered network design. Pushing computing in CWNs as close as possible to clients can significantly enhance performance. We refer to the capability of flexibly running user application (and other) code at the edge of CWNs as the FogBank.

Looking at the current evolution in mobile apps, we observe a growing trend toward the use of cloud services presented as libraries built into apps (e.g., Dropbox, AdMob, Flurry Analytics, game engines). We imagine the app API for CROSSMobile architecture will follow this design pattern—specifically, a purpose-created CROSSMobile library can be compiled into apps as the mechanism by which apps conveniently share state information with the rest of the CROSS-Mobile network.

Providing the means by which computation can move from the cloud into the CWN itself has the potential to improve the user experience and reduce the load on both the core network and centralized cloud services. Providing the on-phone app and the corresponding cloud service with visibility into the state of the wireless medium brings the potential for cooperation across layers in managing the fundamental characteristics of the CWN links.

Certain types of applications can be greatly enhanced by this integrated networking and computing paradigm. One obvious example is real-time gaming. Players who are geographically nearby can enjoy very low player-to-player and player-to-engine latency, providing a high performance gaming experience. Another example is machine-to-machine (M2M) communication in which real-world sensing and actuation are controlled by FogBank-based computation. Once again, having access to low-latency computation and storage in the network broadens the applicability of CWNs to the realm of cyber-physical systems.

The concept of integrating computing and storage directly into the CWN makes the CWN itself an integral element in the mobile computing platform. As in the case of past computing generations, the platform (e.g., IBM System/360 in the mainframe generation of computing, the DEC VAX in the minicomputer generation, and the IBM PC in the desktop computing generation) becomes the focal point for software reuse across generations of compatible hardware and is, itself, a driver of innovation. This is a substantial change-in-role for networks which have largely been, to this point, filling the role of bit pipes. Further development of the role of the network as part of the mobile platform will create new sources of network value.

V. **Research Opportunities**

CROSSMobile architecture opens (and, in some cases, re-opens) interesting avenues of investigation. The list here is by no means exhaustive and merely serves to illustrate the rich set of topics that relate to this proposed next-generation approach.

A. **Network Resource Allocation**

One of the key characteristics of CWNs is that devices and users in the network are mobile. Constantly moving users, mobile devices, connected vehicles, UA Vs, and mobile sensors pose challenges to the network to allocate resources such as bandwidth, data storage and computing cycles to accommodate such dynamic needs. For example, a CWN deployed in the field for a disaster recovery mission needs to proactively reconfigure itself when vehicles, personnel and UAVs move at the scene. The ability to predict traffic patterns to reduce overhead [22], proactively reallocate network resources based on predicted mobility patterns of mobile devices, and dynamically prioritize traffic flow are essential.

Existing research in mobility pattern modeling [10], prediction [11], and location trace modeling [15] demonstrate the feasibility of a dynamically reconfigurable CWN based on network behavior prediction. However, most prior work is predicated on offline analysis and modeling due to the complexity of integrating with current CWNs. Incorporating these proven mobile behavior modeling and prediction techniques in next generation CWN designs will significantly improve the ability of networks to allocate resources.

B. **Autonomic Service Orchestration**

CROSSMobile architecture opens the door for mechanisms that seek to optimize end-to-end user-visible services and, in particular, the performance of cloud-connected mobile applications. Such applications are judged both at the level of how well an individual user is served (one-to-one) as well as how the aggregate service (one-to-many) performs. This raises the question of how to optimally allocate (virtual) compute and network resources and set their parameters. It is an open research challenge to frame this as a multi-objective optimization problem, where objectives may include elements such as latency, bandwidth, and power consumption. A proper formulation will enable investigation of stochastic optimization algorithms (such as genetic algorithms) for this purpose—as a starting point. However, the inherently dynamic nature of the problem necessitates automation. Autonomic and feedback computing techniques [5], [9], [14], which integrate feedback control and artificial intelligence, are worthy of exploration.

---

4The FogBank can be implemented using virtualization techniques at any of a number of levels—from full-fledged containers to software as a service. The choice among these implementation alternatives depends on the nature of the computation—in particular, the degree to which the service depends on sharing (or not) of rapidly-changing information.
Another important research direction is attempting to understand, in an empirical way, CROSSMobile network performance using statistical and machine learning techniques. This will provide the foundation for multi-objective optimization as well as autonomic and feedback computing.

C. A Service Marketplace

Possibly the greatest feature enabled by CROSSMobile architecture is the ability for users, applications, services, and providers to negotiate for resource allocations and operations in an open economic marketplace. App developers and end-users gain better overall experience from adaptive and optimal networking and computing, but service providers can gain an economic incentive. With networking and computing resources being dynamically shifted based on need comes the opportunity for a real-time marketplace. The aggregated demands, as expressed in real time by all the collected apps running on the network and the fixed supply of network resources (both computing and communication) present the possibility for real-time auctioning and other creative means for monetization. In particular, CROSSMobile architecture exposes security and privacy risks related to sharing of information to support data and network analytics; heterogeneity of computing resources and capabilities across network, edge, and mobile devices; stringent requirements of latency-sensitive applications; distribution of network elements and users over a wide geographic area; and coexistence of multiple management domains. Hence, there exists an opportunity to expand the recently designed cloud computing primitives addressing aspects of trusted computing, isolated execution, and privacy-preserving analytics to apply to the unique features of mobile fog computing.

D. Security and Privacy in Mobile Fog Computing

By incorporating mobile edge computing and distributed information management/sharing into a CWN consisting of mutually-suspicious elements, we must also consider the security and privacy risks that are exposed in regard to distributed computation, user service, and data management, including but not limited to availability, secrecy, and information leakage. In particular, CROSSMobile architecture exposes security and privacy risks related to sharing of information to support data and network analytics; heterogeneity of computing resources and capabilities across network, edge, and mobile devices; stringent requirements of latency-sensitive applications; distribution of network elements and users over a wide geographic area; and coexistence of multiple management domains. Hence, there exists an opportunity to expand the recently designed cloud computing primitives addressing aspects of trusted computing, isolated execution, and privacy-preserving analytics to apply to the unique features of mobile fog computing.

E. Anticipatory Antenna Systems

Gains in spectrum efficiency at the system level are enabled by combining detailed knowledge of the real-time channel characteristics between each radio (base station, access point) in the network and all of its “visible” clients. Phased array and other antenna subsystem approaches can provide a CROSSMobile network’s machine learning infrastructure with elevation angle, azimuth angle, and range for each client. Combining this with knowledge of each client’s anticipated network demand (bandwidth and position) and adaptive array smart antennas and/or switched beam smart antennas create the possibility of antenna systems that adapt on the fly, anticipating the behavior of mobile clients.

Adaptive beamforming at a system level would allow the collective set of network antennas to steer beams toward clients while simultaneously nulling interfering signals.

F. Policy Enforcement

While open interfaces create a wealth of opportunity for improved user experience, increased availability of network
resources, and services that cannot currently be provided, these gains do not come without risks. If the interfaces are uniformly open, with no access control policies, monitoring, or management, then users or applications can potentially cheat the system, for example by requesting an unnecessarily large resource block and then not using it, which effectively denies availability of the resource block for others and wastes it. Hence, one of the side effects of our proposed open architecture is a significant challenge to integrate policy and management controls for each interface. While this is not an infeasible or unreasonable requirement, it does present several problems that need to be addressed before systems employing CROSSMobile architecture should be deployed.

Of course, there may exist conflict between the policy desires of the different parties involved in the management problem. We envision the creation of a collaborative policy agreement mechanism that allows the different parties to provide a list of their required and desired features, then the parties can agree on a policy based on “least common requirement” and “greatest common desirable.”

VI. CROSSMOBILE ARCHITECTURE AND TESTBED AT CMU–SILICON VALLEY

While modeling and simulation may bring some insights, we believe that the characteristics of CROSSMobile architecture, and, in particular, its advantages, can best be studied with real traffic from real users. As such, we have begun construction of an on-campus testbed that implements all levels of CROSSMobile architecture and opens its software-defined elements for research and exploration. Figure 4 shows our first node (as of this writing, two nodes are operational). Our intent is to integrate the CROSSMobile testbed into most of the on-campus wireless-systems-related research projects and to afford them degrees of freedom not available on commercial networks. In so doing, their use of the network will generate valuable low-level status trace information that we will use for cross-layer inferencing studies.

With success, we hope to expand the CROSSMobile testbed from a simple on-campus network to a multi-research-institution mobile virtual network operator (MVNO)—offering a much larger community the capabilities of this flexible testbed and scaling up the number of real users for the purpose of broader and more insightful inferencing studies.

VII. CONCLUSION

In this paper, we have highlighted four key challenges of today’s cellular networks: network impermeability, hidden RAN state, hidden network state, and hidden app state. We have argued for opening today’s closed networks to permit the embedding of cloud-side computing and adding mechanisms for exchanging state information in a controlled way across layers. This creates possibilities for system-level optimization and a win-win-win between mobile clients, network operators, and app developers.

To validate and characterize our proposed CROSSMobile architecture, we are developing an on-campus CWN testbed at CMU–Silicon Valley. Recognizing that real network behavior can only be properly characterized with substantial, real traffic, we invite others to join us in the creation of a large-scale, federated research network.

VIII. ACKNOWLEDGEMENTS

Development of CROSSMobile “Node 0” is due to Ervin Teng, Harry Chan-Maestas, Derek Kozel, Bruce DeBruhl, Phoorichet Thepdusith, and Mridula Chappalli Srinivasa. Special thanks go to the team at Range Networks, Kurtis Heimerl, and the other contributors to OpenBTS on which our initial RAN was based. Thanks also go to Wendy Fong for proofreading and numerous suggestions for improving the manuscript.

REFERENCES


[24] J. Vikberg. 3GPP TS 43.318 Generic Accesss Network (GAN); Stage 2. 3GPP Specification Detail.