Paul Schmitt —Research Statement

Ubiquitous, high-bandwidth Internet and cellular connectivity is no longer a goal, but an expectation for modern-day users. Such connectivity is heavily dependent on infrastructure that is both costly and slow to deploy. As data consumption and demand increase at rates akin to Moore's Law, traditional infrastructure models begin to break down. In my thesis work, I have studied problems surrounding wireless and mobile connectivity in a multitude of environments, and have been particularly interested in locations where infrastructure fails to meet demand, as we can learn from these situations in order to inform new, novel solutions.

My research interests span the areas of *wireless network systems design*, and *network measurement and performance analysis*. A common thread in my research is in-depth characterization of real-world connectivity in resource-limited environments, and the use of this analysis to subsequently inform system designs. Through comprehensive analysis across all tiers of real-world telecommunications infrastructure, my research produces solutions that solve fundamental connectivity problems, and in particular those that that manifest in challenging network environments. The contributions of my research to-date have in two primary dimensions:

1. **Characterization and analysis** of existing wireless systems to understand the current operating state of connectivity and inefficiencies therein [1, 2, 3, 4, 5, 6, 7].
2. **System design** of solutions for providing agile wireless connectivity in resource-limited settings [8, 9, 10, 11, 12, 13, 14].

1 Characterization and analysis

To design effective solutions for resource-limited environments, we must first understand the realities faced by users of infrastructure in these contexts. My research has led to field work in multiple locations in order to collect traffic traces of operational wireless networks and study real-world performance. Specifically, I have done field work in Zambia, Guatemala, and a Syrian refugee camp in Jordan to assess the current state of Internet and cellular connectivity in these disparate environments. My findings have led to critical new insights in the performance and usage of networks in these poorly understood and understudied settings.

Cellular: I have studied cellular networks across multiple tiers of the telecommunications hierarchy. At the last-mile, I have extensively analyzed infrastructure serving cellular users in resource-limited situations. I traveled to Guatemala, the United States, and the Za'atari refugee camp in Jordan and collected roughly 6.5 million cellular broadcast messages. I mined these traces with the goal of understanding of commercial cellular network performance at the access level and how infrastructures differ between the locations and contexts. I leveraged the broadcasts to infer radio resource congestion at the base station by creating a new *method and metric that enables third-party congestion detection*. Previously, such operational knowledge was unavailable to third-party users and only observable by operators themselves from within the cellular core network. The results of the measurement studies, which led to novel systems designed for alleviating the observed challenges, were published in top Computer Science venues [8, 9], as described in (§2).

I used my congestion detection techniques, along with physical layer measurements, to characterize the camp-serving infrastructure in Za'atari. One of my findings was that the camp-serving infrastructure for the most popular cellular provider was severely under-provisioned, leading to residents' repeated inability to access cellular services during peak times. Additionally, I found that the quality of cellular coverage and data speeds varied drastically between different locations as well as between the different cellular carriers in the area. The result was that, even within the confines of a single camp, there were substantial *community-level divides*. This study was published in the leading interdisciplinary conference on information and communication technologies for development (ICTD). The work, along with a technical report [15], was provided to the United Nations High Commissioner for Refugees, the agency that manages Za'atari, as we offered practical recommendations to assist aid organizations in refugee camp planning.

To better understand the challenges to improved connectivity in the network core, research is necessary to fully discover and characterize hidden inefficiencies present within networks. Whereas traditional data network topologies and routing policies are comparatively transparent, cellular core networks generally remain a 'black box' for networking researchers. To this end, I studied network behavioral differences between mobile virtual network operators (MVNOs) and traditional mobile network operators (MNOs). Through collection and analysis of over 290,000 cellular path measurements, I found that, *despite often sharing infrastructure with the underlying MNO, MVNO traffic frequently takes drastically more-inefficient geographical paths and ingresses onto the Internet in different, often suboptimal, locations*. For example, the median physical distance traveled for traffic on an MVNO network was roughly *300% longer* than traffic on the corresponding MNO network. This work was published in a leading measurement conference [3].

Broadband: In my work on broadband Internet characterization, I utilized network traffic traces collected from a small, rural network in Zambia and a rural tribal network in California to measure performance and identify *locality* in web multimedia production and consumption [14]. I found high locality-of-interest for Internet media in both networks. For example, locally-produced Facebook media files were downloaded an average of six times more than external files. The limited bandwidth at the Internet gateway led to extremely poor user experience for multimedia (e.g. videos frequently stalled due to buffering). Using these findings I created a system that allows for transparent local transformation and redistribution of media *uploaded* through an Internet cache. Critically, the system *caches and redistributes local content without requiring a single download from the Internet*, which is ideal for networks with limited bandwidth on the access link to the Internet. The system virtually eliminates video stalls, and time-shifts uploads to low-demand times of the day.

In another project, I studied the Zambian network trace to understand Internet adoption in resource-limited environments. Interestingly, the trace spans time before and after an eight-fold bandwidth upgrade for the Internet gateway. I studied the network performance and the end-user experience before and after the upgrade [4, 5]. I found that simply increasing bandwidth for the community link did not result in improved user experience. Rather, the capacity increase appears to have encouraged more multimedia Internet usage, compared with before the upgrade when video streaming and large file transfers were simply impossible. The change in user behavior ultimately led to a *deterioration of network performance and user experience*, as the relatively modest bandwidth increase was not sufficient for supporting heavy multimedia web consumption. This work identified unforeseen consequences of Internet adoption that provided new insight in the research literature. The findings were published in a highly-regarded interdisciplinary journal [4], as well as a Computer Science conference dedicated to ICTD challenges [5].

2 System design

The second major dimension of my dissertation research is the design and development of robust, reliable systems motivated by the real-world connectivity challenges observed through my network measurement and analysis work.

Cellular: Motivated by the cellular network congestion discovered in the Za'atari refugee camp analysis, I designed a system that leverages community cellular technology, software-defined radios, and an Android application to provide seamless *hybrid, heterogeneous cellular coexistence* between commercial cellular and community cellular networks [8]. Community cellular networks are rapidly-deployable, low-investment infrastructure that can provide basic cellular services using open-source software and software-defined radios. Previously, community cellular networks have been confined to areas with no commercial cellular coverage. In my work, I created a hybrid network that offers two unique features: (1) although both the commercial and community networks are utilized, they appear as one to users; and (2) the hybrid community network *is cognitively aware of the commercial networks' 'health' in real-time, and dynamically reconfigures its behavior in response*. During times of high congestion, the system shifts usage to the community network to alleviate commercial congestion as well as provide means for users to obtain cellular service [8]. In another, related project, I created a standalone Android application that utilizes my cellular congestion detection techniques to alert users of deteriorated commercial service [9]. The app informs and empowers users in near real-time so that they are able to seek out improved mobile connectivity by using alternative network SIM cards or physically changing location. Without information about the quality of the commercial cellular network provided by the system, users are not aware that a better connectivity option may be available, and are likely to continue to suffer with poor performance.

I designed a system for unique environments where users are located near, but just outside of the coverage area of their cellular providers [13]. This happens commonly in refugee camps, which are typically located just across borders; or, in disaster response scenarios, just outside of cities. The system combines community cellular networks and long-distance, point-to-point wireless links to *virtually extend* nearby commercial coverage. I created a novel *mechanism for solving the rendezvous problem* encountered when users switch between multiple, non-cooperative networks. The system is particularly useful for situations where physically traveling into the existing coverage area is impossible or dangerous for users. This work was inspired by my field work in Za'atari, where residents live near the Syrian border in Jordan, just beyond the range of their 'home' Syrian cellular providers.

Dynamic spectrum: A fundamental problem facing wireless networking is the scarcity of frequency spectrum that is available to use as a medium for connectivity. Currently, spectrum is managed and allocated in a top-down manner, typically at the national scale. Carriers are given exclusive use of spectrum across large geographic areas. This paradigm results in *artificial spectrum scarcity* that leads to high cost for services and limits wireless connectivity options. Recently, academics, policy makers, and industry have recognized the problem and have begun exploring solutions for the next generation of spectrum management. By allowing multi-tiered, dynamic occupancy of radio spectrum, we can more efficiently utilize this limited resource. A key requirement for realization of dynamic spectrum access is deep understanding of spectrum use in the time, frequency, and space domains. In my work, I have proposed a *single-pass system that achieves rapid transmitter detection* and is robust to noise [10]. I use wavelet decomposition to amplify and detect transmissions at a given time instant, and use time-decay smoothing to reconcile transmission duration. The system successfully identifies transmissions, even those with power levels *near the noise floor*, and proves to be more accurate than power thresholding and other wavelet-based alternatives.

3 Future work

Looking forward, I will continue to investigate wireless connectivity usage and performance in both traditional and non-traditional contexts, and I will continue to push the envelope of network systems design to offer higher quality connectivity within these environments. Research I expect to pursue in the near-term includes the following:

Protection of cellular user privacy and anonymity: Many of the user populations I have studied suffer not only with poor network performance, but are additionally oppressed for other reasons. Surveillance and censorship of network usage is a critical problem facing users in politically-charged environments. Many commercial cellular providers, as well as governments, force traffic through middleboxes, surveil, and / or censor user data sessions. Likewise, small-scale cellular technology has increasingly been leveraged as a means for monitoring user locations and activities using weaknesses inherent in older cellular technologies such as GSM. Such censorship and surveillance must be thwarted if we are to trust networks with our private communications. I am interested in designing systems that *protect cellular client privacy, location, and anonymity*, and *automatically detect and report to end-users the presence of external entities* in the end-to-end network path.

Heterogeneous cellular data and agile cellular networks: My prior work has focused on offloading basic cellular services onto community cellular networks in areas where infrastructure is failing or non-existent. The field-wide push toward 5G includes extensive use of heterogeneous networks consisting of many, small-footprint cells. To facilitate the coexistence between these two technologies, I am interested in enabling high-speed data (e.g. 4G LTE) offloading onto community cellular networks. Whereas systems I have designed in my thesis work have appeared seamless to the user, phones in this context currently must physically switch between community and commercial networks. I intend to explore the use of nascent soft-SIM technologies to create a hardware device that acts as multiple user SIMs in real-time. Given this ability, I will design what is essentially a *cellular network address translation (CNAT)* system that will enable multiplexing of users through a single point of access. Further, I intend to explore *software-defined MVNOs*, where software-defined radios can be harnessed to build cellular base stations that are able to programmatically reconfigure themselves to enable dynamic client offloading for non-cooperative failing infrastructure.

Alternative network infrastructures: There remains millions of people that, to this day, live beyond the reaches of telecommunications infrastructure. As we build networks to connect the Internet's frontier, we have the opportunity to fundamentally rethink connectivity models and explore new network architectures and topologies. The Internet was originally envisioned as a decentralized network of networks. While it can be argued that it remains such today, many critical services are relatively centralized and user connectivity relies on underlying assumptions of global connectivity. I will explore protocols and technologies that *enable network self-organization, self-configuration and self-recovery*, ultimately allowing for *decentralized, bottom-up, user-extensible infrastructure*. Wireless Internet service providers (WISPs) are a common solution for providing broadband in a cost-efficient manner, using commodity wireless equipment that has become affordable in the past few years. However, WISP networks are often tree-like to simplify management and are brittle due to their topologies and reliance on unlicensed wireless bands. I want to design solutions to enable new architectures that increase reliability and capacity of WISP networks. One architectural alternative is the addition of multiple, parallel links in the network to create more highly-connected topologies. To achieve such a solution and avoid problems associated with network loops, a multipath protocol must be used. Prior multipath solutions that have been designed for datacenters often assume homogenous hardware with equal-capacity links, while WISP network link capacities and qualities can be dynamic over time. I want to explore new *network protocols that enable link-agnostic multipath across wireless networks with arbitrary topologies and heterogeneous hardware*. Additionally, I will investigate protocols that allow for *flexible tuning to choose between reliability and performance* in order account for user traffic asymmetry and the unreliable nature of wireless networks (e.g. to improve reliability of the ACK path or performance to improve user download speeds).

New cellular network architectures: Cellular networks are increasingly relied upon to provide high-bandwidth, low-latency data services, yet most cellular core networks are deeply hierarchical and their topologies are reflective of carriers' long-standing priorities toward voice services. As it stands, most cellular core network topologies are highly inefficient for data traffic. As we trend toward 5G data speeds, it is increasingly important to remedy latencies and throughput limitations attributable to inefficient topologies, as well as to explore alternative architectures. Additionally, modern cellular standards transition user traffic to routable IP nearer the network edge than did previous generations. I am interested in taking advantage of the increased edge computing capabilities that now exist (i.e. smartphones) to explore lines of research related to *'core-less' cellular networks*, shifting from the current, hierarchical client/server architectures toward distributed networks.

References

[1] P. Schmitt, D. Iland, E. Belding, B. Tomaszewski, Y. Xu, and C. Maitland. Community-level access divides: A refugee camp case study. In *ICTD '16*, Ann Arbor, MI, Jun 2016.

[2] P. Schmitt, D. Iland, E. Belding, and M. Zheleva. Cellular and Internet connectivity for displaced populations. In *ICTs for Refugees and Displaced Persons*. MIT Press, forthcoming.

[3] P. Schmitt, M. Vigil, and E. Belding. A study of MVNO data paths and performance. In *PAM 2016*, Heraklion, Crete, Greece, Mar 2016.

[4] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. Internet bandwidth upgrade: implications on performance and usage in rural zambia. *Information Technologies & International Development*, 11(2):1–17, 2015.

[5] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. The Increased Bandwidth Fallacy: Performance and Usage in Rural Zambia. In *ACM DEV'13*, Cape Town, South Africa, Dec, 2013.

[6] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. Bringing Visibility to Rural Users in Côte D'Ivoire. In *ICTD'13*, Cape Town, South Africa, Dec, 2013.

[7] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. Community Detection in Cellular Network Traces. In *ICTD'13*, Cape Town, South Africa, Dec, 2013.

[8] P. Schmitt, D. Iland, M. Zheleva, and E. Belding. HybridCell: Cellular connectivity on the fringes with demand-driven local cells. In *IEEE INFOCOM 2016*, San Francisco, CA, Apr 2016.

[9] P. Schmitt, D. Iland, and E. Belding. Smartcell: small-scale mobile congestion awareness. *IEEE Communications Magazine*, 54(7):44–50, Jul 2016.

[10] T. LaRock, P. Schmitt, E. Belding, P. Bogdanov, and M. Zheleva. AirPress: Mining Transmitters for Next Generation Spectrum Management. In *in submission*.

[11] T. Pötsch, P. Schmitt, J. Chen, and B. Raghavan. Helping the lone operator in the vast frontier. In *HotNets 2016*, Atlanta, GA, Nov 2016.

[12] P. Schmitt and E. Belding. Navigating connectivity in reduced infrastructure environments. In *LIMITS '16*, Irvine, CA, Jun 2016.

[13] P. Schmitt, D. Iland, E. Belding, and M. Zheleva. Phonehome: Robust extension of cellular coverage. In *ICCCN 2016*, Waikoloa, HI, Aug 2016.

[14] P. Schmitt, R. Raghavendra, and E. Belding. Internet media upload caching for poorly-connected regions. In *DEV '15*, London, United Kingdom, Nov 2015.

[15] C. Maitland, B. Tomaszewski, E. Belding, K. Fisher, Y. Xu, D. Iland, P. Schmitt, and A. Majid. Youth mobile phone and Internet use, January 2015, Za'atari Camp, Mafraq, Jordan. Technical report, Penn State College of Information Sciences and Technology, 2015.