The 6th Workshop on Active Internet Measurements (AIMS-6) Report

kc claffy
CAIDA/UCSD
kc@caida.org

David D. Clark
CSAIL/MIT
ddc@csail.mit.edu

Mike Wittie
Montana State U.
mwittie@cs.montana.edu

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 It has NOT been peer reviewed.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

ABSTRACT

On 26-27 March 2014, CAIDA hosted the sixth Workshop on Active Internet Measurements (AIMS-6) as part of our series of Internet Statistics and Metrics Analysis (ISMA) workshops. As with previous AIMS workshops, the goals were to further our understanding of the potential and limitations of active measurement research and infrastructure in the wide-area Internet, and to promote cooperative solutions and coordinated strategies between academics, industry, policymakers, and funding agencies in the area of active Internet measurement. This year, we explored capabilities and opportunities for network measurement in the wireless domain, and research infrastructure to support it. Participants found the workshop content challengingly diverse, with substantial knowledge exchange regarding the wireless research infrastructure landscape(s) and existing measurement capabilities. But attendees agreed that the conversation was only beginning, and that some challenges merit further discussion, such as finding consensus on standard metrics to measure, and constructing a road map for wireless measurement research infrastructure and activities for the next decade. This report describes topics discussed at the workshop, and summarizes participants' views of priorities for future funding as well as follow-on workshops in this area. Materials related to the workshop are available at http://www.caida.org/workshops/aims/1403/.

Categories and Subject Descriptors

C.2.3 [**Network operations**]: Network monitoring; C.2.5 [**Local and Wide-Area Networks**]: Internet; C.4.2 [**Performance of Systems**]: Measurement techniques—Active

Keywords

wireless, cellular, active Internet measurement, validation

1. MOTIVATION

For six years, the AIMS workshops have helped stakeholders in Internet active measurement projects to communicate their interests and concerns, and explore cooperative approaches to maximizing the collective benefit of deployed infrastructure and gathered measurements. Each year we take on a theme based on feedback from previous workshops. This year, we explored capabilities and opportunities for network measurement in the wireless (including cellular) domain, and research infrastructure to support it. One motivation was the increasing attention to measuring the character of broadband access, including the U.S. FCC Measuring Broadband America effort (for both wireline and mobile connectivity). The more complex structure of cellular access, with its signaling protocols and potential to shape different sorts of application traffic, makes it harder to decide what to measure in order to characterize or compare wireless service. Our goals for this two-day workshop were to: (1) understand the wireless research infrastructure landscape(s), and measurement capabilities that support (or should support) it; (2) articulate important questions that measurement can help answer, and who would be the customer for such measurements; and (3) propose a vision/roadmap for wireless measurement research infrastructure and activities for the next decade. This report reviews highlights of the discussions at each session, reflects on what was learned at the workshop, and lists collaborations that resulted from the workshop.

2. OBSERVED MOBILE BROADBAND PERFORMANCE

Our first session included three presentations on projects trying to gather empirical data on mobile broadband performance: an individual measuring mobile performance in Santa Cruz, the Federal Communications Commission's Measure Broadband America program, and a Norwegian government-supported research project to measure mobile reliability in Norway.

Jim Warner (UC Santa Cruz) rode his bicycle around the western part of Santa Cruz with a Verizon data tablet executing Ookla speedtests at intersections, to gather performance measurements of the cellular data infrastructure. His "warbiking" revealed significantly different (lower) coverage and quality than that reported by cellular providers. Maps provided by cellular carriers are based on signal propagation models, rather than actual measurements, and overstatement of coverage on such carrier-provided maps can prevent broadband assistance and deflect public investment and involvement in infrastructure deployment.

Walter Johnston (FCC) described FCC's Measuring Broadband America (MBA) Program, which started measuring wireline broadband residential connections four years ago, and more recently released mobile performance measurement apps for Android (Nov 2013) and Apple iPhone (Feb 2014) devices to enable crowd-sourced measurements of cellular infrastructure. The app tests speed, latency, and packet loss (metrics agreed to by the project's participating cellular providers), cell tower identification (standardized to a network provider, but not consistently named across providers). Level3 PoPs host the test destination servers (unlike the wireline broadband tests which used MLab servers). The app differentiates indoor from outdoor measurements, relying on median statistics to reflect performance, which thus far has correlated well with the carriers' own reported street-level measurements. The eventual goal is a real-time map of network performance, including the impact of natural disasters.

Given the sensitivity surrounding privacy aspects of measurements on personal mobile devices, the FCC went through an extensive process to ensure there would be no collection of any personally identifiable information from phones – the app only gathered data related to performance. Walter expressed strong interest in developing common privacy policies for collection of mobile data, at least as collected by government agencies. This goal triggered discussion of what performance-related information current phones provide to users or developers. Android reveals a richer set of device data information than Apple's iOS, and allows background process execution unlike iOS. But there is no way to obtain consistent signal strength measurement over wireless networks, e.g., the bars shown on the phone display are an approximation that may include history and averaging. In particular, iOS does not provide signal strength information to apps, because there is no common definition of signal strength, or more to the point, RF signal strength is not a reliable predictor of download speeds. Android provides a public signal strength API but there is little information about how the measurement is calculated.

Ahmed Elmokashfi (Simula Research) presented an update on a government-sponsored project to measure Norway's mobile broadband reliability. Reliability is a more complex notion to try to measure than raw speed metrics. Operational since July 2013, their infrastructure includes 300 dedicated stationary nodes in schools and government buildings. They tested five commercial carriers, four UMTS operators and one CDMA 2000 operator. They separated measurements into connection reliability, data plane reliability, and application reliability. They define connection reliability failure as the loss of an IP address and the requirement to reestablish context. They found the mean time between failures varied widely across operators, but operators that failed more frequently recovered more quickly. They managed to find failures that carriers were not finding; carriers are generally unaware of user-perceived performance problems unless they result in disconnection. They have not yet correlated user-based measurement with these infrastructure-based measurements. They are currently working on a system that monitors not just dropped connections but also performance, such as the effects of congestion in or beyond the provider's network.

3. NON-CELLULAR WIRELESS MEASUREMENT

We shifted our attention to wireless measurement in non-cellular infrastructure. Craig Partridge (Raytheon BBN Technologies) presented on the evolution of wireless network measurement in a world of software-defined radios (SDRs). The big gain from SDRs is their ability to reduce radio protocol development time from months to days, ideal for graduate student exploration. SDRs can also be much more responsive to application needs, e.g., selecting specific bands or protocols based on requirements, allowing one to avoid writing more capable (and thus expensive) code by using more capable filters. An open question is how to use radios to monitor spectrum, including inferring transmitters and protocols. For example, maybe a cellular radio can receive but not transmit signals. One limitation of SDRs for measurement is that an external observer cannot see the internal state of an SDR, or what is driving its logic. Some architectures provide a control channel that could potentially reveal some of this information. One speculation was that WiFi technology was undercutting SDR, another was that LTE radios would replace WiFi everywhere as hardware became cheaper.

Aaron Schulman (Stanford) talked about how to enable innovative measurements of wireless networks. Wireless measurements are essential to fault and interference diagnosis, and to adapting protocol behavior and spectrum utilization to different observable classes of users or conditions. Today's ASICs are the heart of current wireless networks, but are generally proprietary, limiting the ability to use ASIC-provided failure counters and signal strength reports. Perhaps 10-15 parameters are used to quantify signal strength, and the chips are opaque about how they compute it. Future wireless networks will use programmable hardware (digital signal processors and FPGAS) rather than ASICs. Although protocols will likely remain proprietary, emerging open DSP development frameworks will enable, e.g., programming DSPs in C to build open source wireless protocols that can be a foundation for experiments and measurements. It is not trivial; for example, programming DSPs requires knowing the timing of each part and deriving timing of interdependent parts hooked together, where changing code for one part may affect the timing of other parts of the system. Also, radios may have a huge amount of data at their head end, which generally requires highly reducing it in the DSP before sending it on.[1]

Nick Feamster (Georgia Tech), presented recent work on locating performance problems in home wireless networks using his Bismark measurement infrastructure as well as the FCC's SamKnows infrastructure that supports the Measure Broadband America (MBA) project. Like the FCC's MBA project, Nick first developed and deployed the Bismark infrastructure to study broadband access performance, but has expanded its scope to study other phenomena including whether the residential performance bottleneck tends to be the broadband access link or the home wireless network. Bismark nodes are custom OpenWrt firmware capable of active and passive measurements, including collection of packet traces. They discovered that a high coefficient of variation of packet delays of user traffic was correlated with the access link being the bottleneck, which suggests a simple method to detect the bottleneck link. They found wireless bottlenecks were common, especially at higher access bandwidth rates. In particular, if access bandwidth exceeded 35 Mbits/s, the bottleneck was not the access link. Nick is now working with the FCC to deploy similar measurements on the SamKnows boxes, which have slower processors, no continuous packet capture capability, and are no longer on-path, which reduces user support requirements but also the quality of the data gathered.

There was discussion of how to understand better what is really happening in home networks, or even what statistics to gather to demonstrate whether wireless is the problem. This challenge will only become more important as access link bandwidths increase, since we still do not understand what prevents gigabit LAN performance in the home. Information is needed from clients on whether they are even trying to send, and cross-layer interactions sometimes require Layer 1 knowledge. The standard for cable modem measurement may eventually incorporate functionality that would enable the customer to self-diagnose problems. This discussion led to an obligatory digression of the home router as one of the most severe vulnerabilities in the global Internet infrastructure, since they are such cheap elements riddled with software problems and an inability to upgrade in the field, ever. Sometimes these vulnerabilities can impact performance, e.g., if home routers are compromised they may be launching DoS attacks that might consume a large fraction of upstream access link bandwidth. All agreed that this topic was for another workshop.

[1] Aaron's work is part of the OpenRadio project (http://snsg.stanford.edu/projects/openradio/), which envisions a programmable wireless network with a consistent control plane, analagous to what SDN is providing for wired networks today.

4. COLLECTING MEASUREMENTS ON USER SMARTPHONES

Over lunch we cataloged all mobile performance app projects represented by workshop attendees, and encouraged participants to install and try the available mobile performance apps on their phone, and discuss their experience with the app creators:

FCC Speed Test	Android, iOS
UMich MobiPerf	Android
ICSI Netalyzr	Android
MySpeedTest	Android
Microsoft Network Speed Test	Windows [|Phone] 8
PortoLAN	Android
Justin Cappos' Seattle	Android
SciWiNet's CyberTiger	Android

After lunch, Ethan Katz Bassett (USC) and David Choffnes (Northeastern) led a discussion on the difficulties of measuring the mobile Internet, starting with the fundamental epistemological problem that what you measure is not always what you think. For example, an obvious low-cost way to estimate page load time (estimate ping time plus DNS lookup time) does not work accurately with the split TCP used by many mobile devices. Ethan described the Google-sponsored Speedometer data set, which has broad but not necessarily representative coverage.[2] David also summarized a recent study of circuitous routes from mobile users to Internet content as a contributor to suboptimal performance [9]. He emphasized that progress in mobile measurement will require: a larger set of participating devices, real-time monitoring of performance, and on-demand adaptive measurement. He also expressed concern about the risks of the proliferation of performance measurement apps, which motivated the development of a single common measurement library, Mobilyzer, that many apps could use for specific mobile performance measurements. David and Morley Mao (U. Michigan) designed Mobilyzer with researcher incentives in mind, so it simplifies app development, uses a PlanetLab-like participation model, provides dynamic server-side control of measurement, and has a published privacy policy.[3]

Valerio Luconi (U. Pisa) introduced Portolan, a smartphone-based crowd-sourced project to measure wired and wireless infrastructure. The architecture includes a server at the university, proxy devices that act as front ends to handle teams of smartphones, and 300 Android devices running the Portolan app, mostly in the United States and Europe. The authors did not know about the Mobilyzer library so created their own measurement types, both manual and background, both of which send data to the server. One background measurement they focused on was AS-level topology discovery using Paris traceroute. They analyzed a regional traceroute campaign, with the source and target in the same country (Italy). They used UDP traceroute because ICMP yielded much lower coverage. They mapped IP addresses to ASes using Isolario BGP data. The battery consumption was low and traffic was about 2MB/day for AS-level topology discovery measurements. The manual measurements included signal coverage mapping (RSSI samples), maximum throughput estimates, and discovery of traffic differentiation (by simulating BitTorrent). In the future they would like to explore the number and locations of cellular packet gateways, and topology surrounding them. They also hope to extend traceroute coverage using RIPE Atlas data, and to integrate IP address geolocation in collaboration with ELTE. Right now the incentive to download the app is the ability to run manual tests, but they are working on developing a diagnostic tool for operators to show how a given network is reached from the outside. They are not yet sharing the data.

Narseo Vallina Rodriguez (ICSI, UC Berkeley) provided a brief overview of Netalyzr (an active measurement software client for diagnosing host and network configuration limitations) for Android, a huge challenge given the opaqueness of the cellular infrastructure. They wanted to be able to identify configuration problems and behavioral anomalies, including metrics related to performance of DNS resolvers and caches, location and impact of proxies and caches, network topology, throttling behavior (e.g., after volume caps are reached), the openness (traffic discrimination properties) of and relationships between MNO and MVNO infrastructure. They redirect measurements to Amazon EC2 server infrastructure, and provide a JSON report similar to the original Netalyzr. They have not ported to Apple iOS due to lack of raw socket access (needed for traceroute) and lack of support for background measurements without rooting the device. One challenge has been the lack of available radio level control plane information, which required some driver hacking [8]. As of the March 2014, there were 15K Android installs of Netalyzr, across 290 operators, 90 countries, and having executed 25K Netalyzr sessions. There is a clear geek bias in the deployment: 60% are rooted.

All of the mobile measurement projects found that the challenge of acquiring and keeping users was essential to measurement coverage and inference accuracy, and app developers were interested in other approaches to expand their user base, including advertising and publicity, internationalization support, clear explanations of results, gamification and sharing capabilities, improved GUI and reduced testing time, and privacy and app permissions to foster user trust. The Netalyzr team found it difficult to maintain an application on Google Play, due to the huge diversity in hardware and Android OS software. Nick Feamster experienced a similar challenge in trying to maintain his team's MySpeedTest mobile performance measurement app for Andoird, especially since negative reviews on the application can result in uninstalls. They considered the best path to wide deployment of measurement tools was press coverage, or having app developers include measurement software as part of their code base.

Mike Wittie (Montana State U.) presented MITATE, a new scalable Mobile Internet Testbed platform for researchers to prototype mobile apps against live (custom) mobile traffic. MITATE offers an alternative to the challenge of finding volunteers to deploy untested code in real environments, and includes a generic scripting capability to construct transactions of what the mobile device should do. Separating traffic generation from application logic, MITATE allows users to define a traffic pattern, when and where it will be sent, and which mobile device should execute the experiment. They have used MITATE to learn about differences in traffic shaping across carriers [7], select the fastest CDN through a given mobile carrier's peering points, and estimate how large of an update message can be reliably delivered within a given time frame.

Sharad Agarwal (Microsoft Research) discussed why understanding mobile app performance is so difficult, including the communication gap between app developers and network researchers. Because the interfaces to both network and UI elements are highly asynchronous, the critical path can be difficult to determine. Most mobile network measurement is divorced from what apps do or need, so is of limited use to app developers. And yet, the cellular domain is sufficiently complex, constrained, and dynamic that app design requires support for performance adaptation to current network conditions. For example, app designers tend not to think

[2] Speedometer was an internal Google tool that employees use for active mobile network measurement; U. Michigan's Mobiperf now uses an extended version of the Speedometer codebase, which Google made public.

[3] http://www.mobiperf.com/privacy

about usage caps (which are becoming more prevalent), instead optimizing only for performance, and retrofitting an existing app to conserve data consumption is challenging. There are generally many dependencies to unwind to ascertain what network activity can be deferred.

5. SPECTRUM UTILIZATION

Ranveer Chandra (Microsoft Research) described how Microsoft's Spectrum Observatory[4], one of several spectrum databases being established to support emerging spectrum sharing paradigms. Spectrum databases help improve public understanding of spectrum utilization and availability, of increasing popular interest to a variety of stakeholders. Use cases include: allowing dynamic determination of which bands are best to use at a given time; detecting rogue transmissions or locations of transmitters; and modeling of realistic spectrum conditions. This presentation inspired a discussion of the current status of white space, what databases exist or are emerging to help manage white spaces, and what chips are in the pipeline to support whitespace standards.

Walter Johnston (FCC) reviewed current approaches to increasing spectrum availability, both for licensed and unlicensed spectrum. Most spectrum usable for mobile device traffic is already allocated (6000 MHz to 5 Ghz today), and allocated spectrum is extremely difficult to reallocate due to issues of property rights and federal government rights. (By a recent estimate, it will take $18B and 10 years to move government off of 1755-1850 MHz spectrum). In the near term, there are efforts to reallocate spectrum where possible (H Block, AWS-e, 600 MHz incentive auction), but spectrum sharing is more likely achievable than reclamation. The biggest challenge with spectrum sharing is that incumbents (for a given piece of spectrum) are always protected from interference, which imposes constraints, and incumbents tend to use worst case analyses to protect themselves, which delays resolution. Walter also reviewed the FCC/NTIA's reexamination of the 3.5 GHz band given the emergence of small cell deployment, including WiMAX. Unfortunately, small cells may not be deployed within 80km of shore, where 55% of the U.S. population lives. A current focus is whether small cells can co-exist with radar without harmful interference. The FCC's Technical Advisory Committee (a multi-stakeholder group) has drafted a recommendation on "harm claim thresholds", i.e., what level of interference should constitute a harm to the incumbent.[5] Additionally, there is increasing research interest in using the large amount of spectrum available in higher bands ($>$ 20 GHz) for communication, since complex antenna arrays can now be built using small antennas, allowing beam forming to mitigate attenuation over short distances.

Hiroyuki Ishii (Docomo Innovations) briefly articulated a vision of the 5G world, with data rates 100X higher than today, but with grand challenges, including how to deploy small and MIMO cells in combination at reasonable operational cost. Hiroyuki thought that 5G environments will also require new measurement approaches, e.g., how to measure in the presence of beam-forming.

We finished the day with a poster session that included five posters (links to poster summaries are on the workshop web page):

1. Sanae Rosen (U. Mich.), Measuring performance impacts of RRC state transitions in cellular networks with user devices
2. Xing Xu (USC), Investigating Performance Enhancing Proxies in Cellular Networks
3. Abbas Razaghpanah (Stony Brook U.), Identifying Traffic Differentiation on Cellular Networks
4. Ashkan Nikravesh (U. Mich.), Toward Meaningful Mobile Network Performance Measurements
5. Hongyi Yao (U. Mich.), A Network Measurement Library for Android Platform

6. FEDERAL AGENCIES COLLABORATION AND SUPPORT FOR MEASUREMENT

Vijayarangam Subramanian (NTIA) presented an overview of research needs and support for national wireless communications policy, especially with regard to usage (and interference) measurement and analytics. While understanding the performance of an existing system might have been sufficient to inform traditional wireless communications policies, the future will be more complex. Relentless growth in demand will require new approaches to routing, security, trust, and measurement, all of which will require innovations in technology, business models and accompanying policy innovation. Several national initiatives to support innovative use and management of spectrum are either advancing, including those that are trying to improve the integrity of spectrum usage measurements, most of which today are ad hoc, static, and reactive. Spectrum measurement to support policy must contend with many inter-related issues, including interference, receiver standards, certification, enforcement capabilities, and security.

This talk led to discussion about how to best generate input into the research agency funding process, since neither NTIA nor FCC fund research nor provide any structured guidance to research funding agencies. There was also debate on whether authority over spectrum should be delegated to regional levels, such as with public safety spectrum. The PCAST-recommended Test Cities program [6] is one proposal to experiment with providing regional testbeds designed to support policy and technology development in the context of dynamic spectrum sharing. There was some hope that the 3.5GHz spectrum transition would take less time since lessons from previous efforts could be applied. Moving people *out* of spectrum is what takes so long, which is why paths to peaceful co-existence are more popular recently. Vijayarangam noted that the NTIA was hosting a workshop the following week on using data and monitoring to improve spectrum utilization.[6]

Ann Cox (DHS), a program manager in the Cybersecurity Division of the Department of Homeland Security (Science and Technology Directorate) talked about her view of the role of federal government in mobile measurement R&D. DHS's perspective on measurement is about improving security more than performance, and their focus is mostly on the development ("D") end of R&D, rather than research. But much of security is enabled by understanding more about the network, from tracking growth of the attack surface and security-related performance issues, to accurate measurement and modeling of network structure and behavior, all of which require monitoring normal as well as anomalous activity.[7]

Michael Piatek (Google) gave his view of what researchers should study in the field of mobile measurement, based on his experience in Google's mobile web performance group, where he worked to optimize Chrome on Android platforms. He explained a few reasons that web performance is so poor on mobile devices, including cross-layer protocol incongruities (e.g., across HTTP, TCP,

[4]http://observatory.microsoftspectrum.com

[5]http://transition.fcc.govoettac/tacdocs/reports/TACInterferenceLimitsIntrov1.0.pdf

[6]http://www.nitrd.gov/nitrdgroups/index.php?title=Wireless_Spectrum_Research_and_Development_(WSRD)

[7]DHS subsequently released a solicitation that included mobile technology security as a technical topic area, available at https://baa2.st.dhs.gov/portal/BAA/ (HSHQDC-14-R-B0005).

3/4G) and poor application designs, which might themselves be side-effects of the lack of good development tools or experienced web software developers. He described the measurement infrastructure that drives this web optimization, by gathering timing information about nearly everything, including enough to drive design of improvements like SPDY. He emphasized that Google had plenty of data, but needed more useful questions to ask of the data. Many question among participants related to energy usage, e.g., can the OS enable individual app designers to assess energy usage implications on performance needs?

James Martin (Clemson) led the session about testbed infrastructures, where we tried to take an inventory of currently operational wireless research infrastructures, understand their utility and purpose, and speculate on directions that could benefit the research community. James introduced SciWiNet,[8] a large scale wireless testbed research infrastructure for the academic community, but an actual MNVO, i.e., it sells data services to users. SciWiNet represents a community of users that collectively provide a support system for academics whose research requires cellular data services. He distinguished research wireless systems infrastructures such as Rutger's Orbit and GENI's WiMAX testbed, from commercial wireless system infrastructures that researchers use, such as Phonelab and MITATE. He envisions SciWeNet as bridging these two ecosystems.

Justin Cappos (NYU) talked about and demonstrated how to use Seattle[9] and the Sensibility testbed, which allows researchers to deploy code on end user devices. The Seattle testbed is providing open peer-to-peer application hosting (i.e., a type of cloud computing), established by users donating 10% of their (isolated) resources. It is used like a P2P PlanetLab, for deploying experimental services (e.g., intelligent distributed storage, dynamic DNS remapping, transparent network optimization, censorship measurement) as well as supporting curriculum activities in classrooms. The testbed software has been ported to Android, OpenWrt, Raspberry Pi, and iPad, and runs on PlanetLab, Emulab, GENI and DOME. Justin also described a newer platform, the Sensibility testbed, which provides access to sensors on user devices.

Jacobus Van der Merwe (U. Utah) described PhantomNet: a programmable end-to-end mobile network testbed with an Emulab control framework. The testbed supports creation and exploration of standard mobile network architectures, e.g., 4GL LTE + EPC. The physical devices are LTE small cell-based, SDN-capable, and come with two radios: one can be used for a normal commercial service, and the other is hooked to the testbed. The current status of the testbed is that they have integrated OpenEPC with Emulab; topology is specified using an NS (network simulator) file format, and it was made publicly available shortly after the workshop.[10] Jacobus encouraged feedback and suggestions on how PhantomNet could better serve researcher needs.

7. ECONOMIC AND POLICY CHALLENGES

William Lehr (MIT) led the final session, on interdisciplinary economic and policy challenges of deploying, managing, and using distributed mobile/wireless measurement platforms for market and policy decision-making. He tied together some previously discussed threads: privacy, advantages of licensed vs. unlicensed spectrum, and how to promote policy/economics-aware measurement and research infrastructure. Markets need better data and metrics to function effectively, but data interpretation is contentious, as will be design of the measurement ecosystem.

Broadband and spectrum policy should be disentangled because broadband concerns more than spectrum availability, and mobile broadband is not the only use of spectrum. But from a measurement perspective, integrating wireless and mobile broadband efforts is desirable. Both efforts confront similar challenges related to location, device/service heterogeneity, and cross-layer interactions; combining efforts can lower costs and take advantage of scale and scope economies. But integration involves challenges. Whereas mobile broadband measurement focuses on questions of universal service, user-experience performance, and security issues, spectrum/wireless measurements are directed at identifying underutilized spectrum to inform sharing and allocation decisions. Spectrum measurement is also more politicized since spectrum ownership and governance are hotly contested topics. The multidisciplinary challenge confronted by policymakers in jointly addressing economic, legal and technical concerns is further complicated by the need to bridge the divide between radio and Internet engineering expertise. This workshop provided a valuable first step in bridging these communities.

David Reed (CU Boulder) spoke on the implications of WiFi as a commercial service, i.e., a WiFi service platform with extended roaming as a platform [2]. Earlier providers did not try to monetize WiFi, but circumstances have changed. WiFi devices have proliferated with 2B shipped per year, 7 per WiFi-enabled household, and industry efforts to harmonize standards activities by the WiFi alliance and 802.11 have improved prospects for coordinated interference management which is critical to support commercial WiFi services. Finally, the need for carriers to offload mobile traffic to WiFi networks to meet growing mobile broadband demand has been a strong driver. Two additional facts will further promote WiFi as a commercial service: public WiFi access points that are blanketing populated areas are interfering with private users; and a new QoS feature in 802.11n (Enhanced Distributed Channel Access or EDCA) can prioritize traffic (e.g. voice), enabling carriers to offer differentiated services.

Bendert Zevenbergen (Oxford Internet Institute) gave the last presentation, on ethical privacy guidelines for network researchers that are collecting, measuring, analyzing, and sharing mobile connectivity data. To support Google's interest in guidance for Measurement Lab users. Ben drew on recent work, e.g., [4], as well as on international legal frameworks. Given that the international landscape is fragmented, and perceptions are changing around the word, he suggested an approach that would give researchers and IRBs tools (and time) to talk, with basic changes to current processes such as new forms with assistive questions and links to informative background related to a given guideline. Subsequent discussion explored the risk of disaster-driven progress, and the future benefit the community may gain by being prepared for overreactive policy positions; such preparation could include developing guideline document drafts, e.g., [1, 5, 3]. There was extensive discussion across multiple sessions of privacy models in mobile data, including whether it was possible to transfer some knowledge or processes from the Census Bureau to control use of privacy-sensitive data. The group debated the value of gathering raw data and imposing privacy protections on the analysis as opposed to limiting data collection to that which can be used freely.

8. FINAL THOUGHTS

The workshop yielded significant knowledge exchange across a diverse community, resulting in a better understanding of the wireless research infrastructure landscape(s) and measurement capabil-

[8]http://www.sciwinet.org
[9]https://seattle.poly.edu
[10]http://www.phantomnet.org

ities that (should) support it. We learned about redundant capabilities across available measurement tools, and early attempts to develop open source libraries to support community development (Mobilyzer) and avoid redundant development effort.

There was also consensus that the conversation was only beginning, and that some challenges merit further discussion, including articulating the most important measurement questions and who will value their answers, finding consensus on standard metrics to use, and constructing a real roadmap for wireless measurement research infrastructure and activities for the next decade. The latter exercise would require more discussion on the scope of measurement infrastructure, how to share it, and whom it should support: the network research community, the broader academic research community, or both, plus an outreach component to benefit the country? Fortunately, recently NSF-funded projects indicate that NSF is interested in investing in wireless research infrastructure and measurement capabilities, although there was interest in a better understanding of what government funding should prioritize. Most participants were interested in attending a future workshop to make progress on these challenges.

Standard privacy policies to facilitate mobile measurement and protect resulting data was a recurring theme of discussion at the workshop. Several attendees felt a big problem was the vast amount of data being collected without limiting data retention to specific window (e.g., 6 months) to minimize its potential threat to privacy. Other issues related to mobile measurement data sharing could (and have) called for a dedicated workshop: What responsibilities should exist in terms of privacy protection and informed consumer consent in the collection and retention of mobile broadband data? What is current practice among researchers? Are existing privacy impact assessments useful? How should implementable policy be developed in this area? How can we motivate wireless carrier cooperation? How can we crowd-source mobile network measurements while preserving privacy and utility? How can we make data easy to digest for users and policymakers?

The group created a wish list of research areas, many of which were discussed at the workshop:

- how cellular networks peer with the rest of the Internet, and the efficiency of the resulting overall topology
- correlating claimed coverage with reported coverage
- investigating and optimizing application layer performance, including understanding the implications of observed network performance on quality of user experience with mobile applications, and how applications could directly use network measurement data in real time
- studying the impact of traffic shaping, content caching, and other middlebox behavior on user-perceived network performance (for different cellular carriers)?
- measurements to support the next level of innovation (IoT, smart grids and vehicles, etc.);
- service observatories, to gain information on service characteristics to complement spectrum observatories
- architectures to support full-stack, end-to-end service measurements (as opposed to black box measurements)
- cooperative initiatives with industry to build user-visible measurement capabilities into equipment

9. RESULTING COLLABORATIONS

Several current collaborations continued at the workshop, including those that originated at previous AIMS workshops. Continuing and newly initiated collaborations included:

1. Nick Feamster, Morley Mao, and David Choffnes continue to discuss how to further integrate Mobilyzer into Feamster's MySpeedTest app. Since the workshop they have released a version of the app that uses Mobilyzer in a limited way.
2. Justin Cappos developed a collaboration with Jim Martin about integrating the Sensibility Testbed code into SciWiNet.
3. Justin continued discussions with the FCC and is exploring how the Seattle software can support multiple experimenters sharing a SamKnows environment.
4. Jacobus Van Der Merwe will include SciWiNet as a third-party network in PhantomNet's hetnet system.
5. David Reed and Jim Martin are exploring shared adaptive HTTP-based video broadcasting tools for educational use.
6. Jim is trying to integrate SciWiNet support for support users that want to build custom applications that use the Mobilzyer measurement library, and in general provide a focal point for community-based initiatives.

ACKNOWLEDGMENTS. The workshop was supported by the National Science Foundation Division of Advanced Cyberinfrastructure (ACI-1127500) and U.S. Department of Homeland Security (DHS) Science and Technology (S&T) Directorate (FA8750-12-2-0326). We thank all participants for their insights and feedback at the workshop and on this final report.

10. REFERENCES

[1] Coull, Scott E. and Kenneally, Erin. A Qualitative Risk Assessment Framework for Sharing Computer Network Data. In *TPRC 40: The 41st Research Conference on Communication, Information and Internet Policy*, 2012.

[2] David P. Reed and Jim Lansford. Wi-Fi as a Commercial Service: New Technology and Policy Implications. In *TPRC 41: The 41st Research Conference on Communication, Information and Internet Policy*, 2013.

[3] Dittrich, David and Kenneally, Erin and Bailey, Michael. Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report. 2013. `http://ssrn.com/abstract=2342036`.

[4] Erin Kenneally and Kimberly Claffy. Dialing Privacy and Utility: A Proposed Data-sharing Framework to Advance Internet Research. *IEEE Security and Privacy*, July 2010.

[5] Kenneally, Erin and Dittrich, David. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research, 2012. `http://ssrn.com/abstract=2445102`.

[6] Presidents Council of Advisors on Science and Technology (PCAST). Realizing the Full Potential of Government-held Spectrum to Spur Economic Growth, July 2012.

[7] U. Goel and A. Miyyapuram and M. P. Wittie and Q. Yang. MITATE: Mobile Internet Testbed for Application Traffic Experimentation. In *MOBIQUITOUS*, 2013.

[8] Narseo Vallina-Rodriguez, Andrius Auçinas, Mario Almeida, Yan Grunenberger, Konstantina Papagiannaki, and Jon Crowcroft. Rilanalyzer: a comprehensive 3G monitor on your phone. In *ACM SIGCOMM Internet Measurement Conference (IMC)*, 2013.

[9] Kyriakos Zarifis, Tobias Flach, Srikanth Nori, David Choffnes, Ramesh Govindan, Ethan Katz-Bassett, Z. Morley Mao, and Matt Welsh. Diagnosing Path Inflation of Mobile Client Traffic. In *Passive and Active Measurement Conference (PAM '14)*, March 2014.