# The 8th Workshop on Active Internet Measurements (AIMS-8) Report

kc claffy
CAIDA/UCSD
kc@caida.org



## ABSTRACT

On 10-12 February 2016, CAIDA hosted the eighth Workshop on Active Internet Measurements (AIMS-8) as part of our series of Internet Statistics and Metrics Analysis (ISMA) workshops. This workshop series provides a forum for 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. Discussion topics included: infrastructure development status and plans; experimental design, execution, and cross-validation; challenges to incentivize hosting, sharing, and using measurement infrastructure; data access, sharing, and analytics; and challenges of emerging high bandwidth network measurement infrastructure. Other recurrent topics included paths toward increased interoperability and cooperative use of infrastructures, and ethical frameworks to support active Internet measurement. Materials related to the workshop are at http://www.caida.org/workshops/aims/1602/.

## CCS Concepts

•**Networks** → **Network measurement; Public Internet;** *Network dynamics;*

## Keywords

active Internet measurement, validation
cectionActive Measurement
Infrastructure Updates

Many participants at the workshop operate or contribute to the operation of existing Internet measurement infrastructure, and provided updates and answered questions about the status of their projects. An ongoing challenge in the community is to understand the potential and limitations of different active measurement research and infrastructure, especially in the context of discussing coordinated strategies among academics, industry, policymakers, and funding agencies. Each project aims to enable novel analyses, development of new tools, educational opportunities, as well as feedback and contributions to improve the platform. But each platform is subject to different resource, policy, and technical constraints. Active measurement infrastructure projects represented this year included:

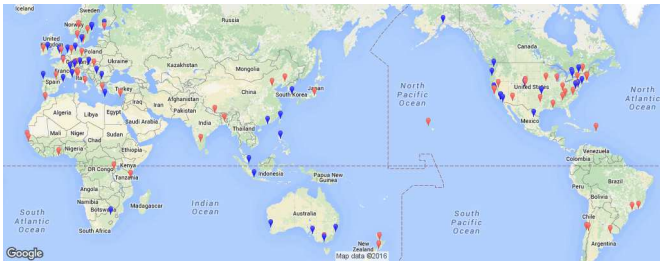


**Figure 1: Ark deployment (blue: IPv6+IPv4; red: IPv4 only)**

1. **Archipelago (Ark)** (CAIDA/UCSD). Approximately 160 monitors around the world (Figure 1), Ark supports vetted measurement experiments on a security-hardened distributed platform. The most important Ark infrastructure development this year is a new interface (both web-based and a more powerful command-line) and back end database to enable browsing, querying, and visualizing the data gathered, specifically to find traces with desired properties, and visualize topological properties of these traces. Sample queries include selecting all traceroutes that transit or reach a set of IP addresses, prefixes, ASes, or countries.
2. **Atlas** (RIPE NCC). About 9300 probes around the world, RIPE Atlas supports measurement of Internet connectivity and reachability in real time, and shares the resulting data via visual maps or an API. The most important development for Atlas this year is software modules to stream data in close to real-time.
3. **PerfSonar** (ESnet, Internet2, Indiana University, GEANT). About 1700 servers on R&E network infrastructure around the world, mostly 10Gb-connected hosts, PerfSonar supports a variety of measurements (e.g., iperf) to enable scientific users to establish empirically grounded expectations about network performance, and to diagnose soft failures, i.e., those that reduce performance but not to zero. PerfSonar was not intended to sup-

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

port network research itself, but to help fix network problems in a multi-domain high bandwidth (>= 1Gb links) R&E network environment.

4. **WAND active measurement project** (U. Waikato). A spin-off of NLANR's AMP project, WAND AMP has deployed monitors across New Zealand ISPs to support operational monitoring, including event detection, and to share resulting data. With New Zealand government support, WAND has used NZ AMP to compare broadband performance across New Zealand ISPs. ISPs can also request tests for specific services: DNS/DNSSEC, HTTP, VOIP, streaming video.
5. **PEERING testbed** (USC, UFMG). The PEERING testbed allows experiments to establish BGP peering sessions, and exchange routes and traffic at locations around the world. Researchers can actively inject new routes and observe the effects on path computation, and use active probing in parallel to measure the effects, e.g., on path latency.
6. **OpenINTEL** (U.Twente). OpenINTEL is new infrastructure to support long-term, large-scale high-performance active DNS measurements, specifically sending a comprehensive set of DNS queries for every name in select TLDs, once per day.[1]. Since most Internet services rely on DNS in some way, this instrumentation can provide a lens into Internet conditions and trends. They now (expanded since the workshop) cover over 190M names (over 60% of the global DNS), including most new gTLDs, and are seeking more ccTLDs.

We discussed operational issues and architectural directions for these infrastructures. One theme was compatibility issues that hindered coordinated use of multiple infrastructures, e.g., different methods for probing, sampling, file management. Measurement infrastructures are created for different reasons, e.g., some are intended to support scientific research, others are operationally driven, and different cost and management models affect who can use them and how. Approaches that try to abstract away or unify these differences could cause more harm than benefit. An AIMS workshop topic from years ago was the idea of creating and sharing "do not probe" lists so that new measurement projects could avoid probing those who have previously complained about active measurements. But several participants argued that many network operators might find it acceptable to be probed for some experiments but not others, so this idea never got traction. Instead, active measurement infrastructures follow an established best practice of identifying themselves in packets, e.g., by an identifying DNS hostname, or including a user-agent string in the payload, either of which point to a URL of the project web site.

Another theme was how to improve ties to operators, in particular how to apply measurement experiences and visualization to questions of interest to operators. Many organizations are now hosting hackathons that focus on measurement and data analysis software development, e.g., RIPE NCC, CAIDA, NANOG, which is an opportunity to further expand ties to operators. WAND's group had been successful with New Zealand government support and regulatory interest in measurement of Internet infrastructure on behalf of consumers, and the success of SamKnows is in large part due to growing regulatory interest in broadband performance on behalf of consumers. But sustained funding and support for such infrastructure is an on-going challenge, and no country (to our knowledge) has stable sources of government funding specifically for measurement infrastructure to enable Internet research.

# 1. OTHER MEASUREMENT INFRASTRUCTURE PROJECTS

We also covered other related measurement infrastructure projects not specific to active measurement:

1. **BGPStream** (CAIDA). BGPstream is a new framework for historical analysis and real-time analysis of BGP data provided by other infrastructure operators and data providers, such as Route Views and BGP-Mon. BGPStream is a component of other CAIDA projects, including CAIDA's monitoring platform that combines diverse data sources (traffic, BGP, active probing) to support research on detecting and characterizing macroscopic Internet outages.
2. **scamper** (Waikato/CAIDA). The scamper software module performs efficient parallel probing of many paths on the Internet. The biggest development for scamper this year is the ability to control a scamper process via a remote node, which opens the possibility to collaborate with infrastructures composed of lower-resourced nodes than Ark, e.g., FCC's SamKnows platform or BISmark, both of which use OpenWRT platforms with limited CPU and memory.
3. **Sibyl** (UFMG Brazil, USC, UNapoli, UMich, CAIDA). The Sibyl system accepts rich queries expressed as regular expressions, and coordinates submission of queries to a diverse set of vantage points across different measurement infrastructures to return paths of interest to the user. The challenge is in navigating the constraints of different measurement platforms, while relying on historical measurements to estimate which vantage point is most likely to be able to capture the path queried but not yet observed. Sibyl combines paths it has measured with knowledge of routing in order to reason about which unobserved paths to measure.
4. **Reverse Traceroute** (USC, Northeastern). a revival of an older project that uses available vantage points to tries to infer a reverse path without control of the destination. The system now provides a public API that allows users to issue batches of measurements.[2]
5. **DNSViz** (Sandia, Verisign). DNSViz is a tool for visualizing the status of a DNS zone, originally designed for troubleshooting DNSSEC configurations. It provides a visual analysis of the DNSSEC authentication chain for a domain name and its resolution path, and detects and reports configuration errors. This year Casey Deccio (Verisign) extended DNSViz to create a *DNS Looking Glass*, to facilitate observation of a broad range of DNS-related behavior along paths between stub and authoritative resolvers, from diverse vantage points. Casey hopes that measurement infrastructure operators will integrate this new functionality, inspired by BGP and DNS looking glasses, into their platforms.
6. **Root Name Server research platform** (ISI/B-Root).

[1] http://www.openintel.nl/

[2] http://www.revtr.ccs.neu.edu

A collaboration with the B-root operator to enable experiments and testing in real traffic conditions. To broaden participation in this community, ISI is co-hosting a workshop at USC in November 2016.[3]

7. **Target list distribution**[4] (RIPE NCC): Emile Aben has customized a DNS server to support distribution of target lists for large-scale measurement experiments, which could facilitate parallel coordinated use of different platforms.
8. **Haystack** (ICSI). A platform for passive capture of traffic on mobile devices, to support large-scale measurement experiments that intercept mobile network traffic and app activity in user space, Haystack also runs (re)active measurements based on passive observations, e.g., if Haystack detects a new network it checks for vulnerability to TLS interception. Haystack enabled development of an interactive map of third-party tracking activity on mobile apps.[5]
9. **Mobilyzer**[6], a measurement library that supports QoE characterization from mobile devices, as well as coordination and scheduling of measurements across multiple devices.
10. **Spoofer** (CAIDA/UCSD and U.Waikato). CAIDA reported on its revival of the Spoofer project that Rob Beverly started in 2005 as a PhD student. Newly funded by DHS, CAIDA's Spoofer project will overhaul the software and hardware infrastructure (released in May 2016). The client software runs in the background, testing every new network it encounters, including behind NATs. The project will publish individual results by default, as well as publishing results aggregated by country, network type, provider, and auto-generate customer cone prefix information for interested ASes who want to bootstrap ingress access list configuration.

A lively theme of discussions of the infrastructures that tried to measure forward paths was whether one could assume ground truth in the origin AS of a prefix, since every routing speaker has relative view of universe. One strong view was that artifacts of destination-based routing, address exhaustion leading to suspicious use of addresses, and other idiosyncrasies of IP addresses and protocols that use them, creates an unavoidable, and unfortunately rapidly growing, gray area, which denies any objective reality of a prefix in space time, and thus sheds serious doubt on the validity of stitching pieces of captured traceroutes together. A contrasting view was that, empirically, most (e.g., non-anycast) prefixes do have a stable ground truth if you measure over time.

Another theme was the quest to improve active measurement coverage, especially in the context of making claims about the entire Internet, such as prevalence of networks that allow spoofing. Unlike DNSSEC, where one can measure compliance from anywhere else on the Internet, active measurement of the presence of network filtering requires a vantage within each network of interest, or potentially even more than that, to the extent that ASes have different filtering policies or configurations throughout their network infrastructure. Otherwise, the inherent sample bias – data is only available on networks who have downloaded the tool – limits the representativeness of the data. One obvious way to increase coverage is to run the software on much larger measurement infrastructures, such as RIPE Atlas. RIPE Atlas has an established policy of prohibiting any spoofed traffic from RIPE Atlas nodes, even to support research. We discussed the trade-offs involved in publishing such results; CAIDA's compromise with the Spoofer project was to warn users in advance that the new version is reporting publicly, and sending to remediation groups (CERTs, governments), unless the user explicitly opts out. One suggestion was to point to the Spoofer project in the RIPE Atlas FAQ as a compromise. Another challenge is how to delineate the boundaries of acceptable probing if spoofing is allowed. A few people observed that using the term *spoofing* does not help with perception issues, and phrases like BCP38-compliance testing, or network hygiene testing, might be more auspicious when soliciting cooperation. Providers themselves might object to (or even formally prohibit) the testing, even if their customer hosting the probe does not.

## 2. INFRASTRUCTURE USE CASES

For broadly distributed measurement infrastructure, obvious use cases of interest to both researchers and operators relate to stability and security of connectivity, and performance. Characterizing the stability of connectivity (or lack thereof, an outage) requires a distributed set of vantage points because a small number of nearby VPs will not distinguish local reachability issues from global ones. Ideally an outage detection system will combine information from multiple vantage points of active probing, passive traffic analysis, and BGP data to verify the occurrence of an outage. Other stability related measurement experiments have included comparing IPv4 and IPv6 stability, and middlebox behavior.

CAIDA's Ark infrastructure focuses on security-related measurement experiments, including vulnerability assessments of TCP, IPv6, and DNS implementations.[7] Most recently, Ark is supporting an NSF-funded project to detect a specific path compromise known as a BGP-based traffic hijack. Like outage detection, detecting a BGP-hijack also requires comparing prefix reachability information as observed from different vantage points. The most common hijacks manifest as two or more distinct ASes announcing exactly the same prefix, or a portion of the same address space, at the same time. These two use cases (outages and hijacks) motivated CAIDA's development of the BGPStream platform to provide real-time information from many BGP vantage points, that could then be confirmed with data from active measurement vantage points.

A more technically and politically complicated challenge that has emerged this year has been how to design, deploy, and use active measurement infrastructure to support assessment of user quality of experience. While, researchers, operators, and application developers could all benefit from increased awareness of network performance characteristics, there are tremendous technical challenges in QoE-related measurement, including detection of traffic differentiation

[3]https://ant.isi.edu/events/dinr2016/
[4]https://labs.ripe.net/Members/emileaben/measuring-more-internet-with-ripe-atlas
[5]https://haystack.mobi/panopticon/
[6]http://mobilyzer-project.mobi
[7]Ark experiments listed at: http://www.caida.org/projects/ark/.

by network operators and quantifying its impact on QoE. A recent NSF/FCC Workshop on Measurement of Quality of Experience in the Internet[8] considered the prospect of integrating threads of QoE research into a collaboration, in order to measure, analyze, and improve the state of QoE in the current Internet. Among the takeaways from that workshop was the recognition that there was substantial recent work in controlled (laboratory) measurement of QoE (mostly in Europe), but the only attempts at large scale field measurement of actual users of QoE were commercial, proprietary, and thus inaccessible to the research community.

The application-specific nature of QoE, and the range of factors and behaviors that affect it (home network, ISPs, IXPs, server, application adaptation) shed doubt on the viability of a general Internet infrastructure to support QoE measurement. Such platforms are likely to require aggregation and correlation of heterogeneous signals to reveal QoE-related properties, a capability even more complex and multi-dimensional than the capabilities needed for detection of outages and route hijacks. One inevitable tension associated with any of these use cases (outages, hijacking, QoE assessments) is the divergent interests of operators in supporting the kind of transparency into network infrastructure required to identify and localize their causes. But QoE assessment involves a layer of political tension since there is a recent history of finger-pointing with respect to who has responsibility for upgrading capacity to mitigate perceived QoE impairment. Ricky Mok (Hong Kong Polytechnic) described difficulties experienced with crowd-sourced measurement of video streaming QoE, concluding that collaboration from service providers or users is essential.

Adding to these obstacles is the fact that each of these capabilities currently constitutes a research as well as infrastructure challenge. Thus far, incentives to provide and maintain sustained funding for infrastructure capable of capture, aggregation, and analytics, in parallel with funding research on how to best perform all those tasks, have not emerged. But, acknowledging these over-arching challenges, participants highlighted various attempts to provide platforms and components to support QoE measurement research. Comcast is now leading an effort to develop a new distributed and automated platform for measuring achievable throughput, which could serve as the basis for a wider, open platform. Ashkan Nikravesh (U.Mich) gave an update on the Mobilyzer mobile measurement library support for QoE characterization from mobile devices. Mobilyzer supports coordination and scheduling of measurements across multiple devices, and could serve different QoE-measurement apps. There was some optimism that even without solving the deepest aspects of the QoE problem, the community could develop a framework for measuring simple parameters that reflect meaningful QoE characteristics but are still tractable, and could be monitored over a broad segment of the infrastructure. One obvious next step in the U.S. would be to explore a collaboration between the National Science Foundation, which funds community research infrastructure, and the Federal Communications Commission, who established and maintains the Measure Broadband America (MBA) program, which outsources to a private company a specific set of measurements thus far focused on broadband access link performance. Although the FCC would like to extend the MBA program to measure performance more broadly, i.e., to video streaming services, points of interconnection, and cellular networks, there is a rich research agenda associated with these objectives, and neither the FCC nor the MBA program is architected to support research.

[8]http://aqualab.cs.northwestern.edu/conference/276-nsf15-qoe-internet

# 3. NEW TECHNIQUES FOR ACTIVE MEASUREMENT

Several participants solicited early feedback on new ideas for active measurement methods.

1. **More efficient topology probing**. Inspired by Zmap's[9] success in probing every IPv4 address in less than an hour, Robert Beverly (NPS) wanted to create the analogous capability for it to path measurement, not just for probing the other end of the path. He proposed a technique that randomly permutes the <IP,TTL> space statelessly, achieving Internet-scale probing from a single vantage point. One can then stitch together topology computations after the probing completes. The challenge is to know when to stop probing. He used Ark traceroute archive data to construct a distribution of unique interfaces discovered and where and when they were discovered. The data revealed the probing does not discover much very near to or far away from the vantage point.
2. **Characterizing DHCP lifetimes**. Ramakrishna Padmanabhan (UMD) proposed to use RIPE Atlas probe device log files to characterize DHCP behavior, including root causes of renumbering (e.g., user or ISP action) and resulting distributions of address lifetimes.
3. **In-country path analysis**. Emile Aben (RIPE NCC) presented his work using RIPE Atlas for in-country traceroute probe meshes, to investigate which Internet paths stay inside a country.[10] He used RIPE's OpenIPMap[11] to geolocate intermediate hops in traceroutes, and published results of running this tool in 100 countries.[12]
4. **Detecting Carrier Grade NAT (CGN) in paths**. Amogh Dhamdhere (UCSD/CAIDA) is leading a collaboration to deploy NAT Revelio, a CGN detection tool, on the FCC's Measure Broadband America platform operated by SamKnows. This experiment will not only reveal CGN usage in U.S. broadband networks, but also be a valuable test case of using SamKnows for large-scale measurement experiments by the research community.
5. **Web-based network performance measurements**. Zubair Shafiq (U Iowa) proposed a high-level architecture for an open, web-based network performance measurement platform that would use JavaScript to crowdsource measurements, and leverage different platforms underneath (M-Lab, perfSONAR, Ark, RIPE, cloud instances) to expand coverage. He acknowledged the huge variability and noise in such measurements, but noted that client-side measurement precision is improving, including with methods to filter out the noise.

[9]https://zmap.io
[10]https://github.com/emileaben/ixp-country-jedi/
[11]http://marmot.ripe.net/openipmap/
[12]http://sg-pub.ripe.net/emile/ixp-country-jedi/history/2016-01-01/

6. **Measuring gigabit networks**. In a world with gigabit access networks but not gigabit paths to every destination, we need a clearer understanding of what to expect from performance measurements. Today's measurement programs cannot accurately capture gigabit end-to-end performance, and current performance expectations are not appropriate during the transition to a gigabit broadband world. If ISPs are penalized, even in public relation terms, e.g., a measurement-based report that suggests their gigabit access links are not providing gigabit access, it will likely delay or disrupt deployment of gigabit broadband. Steve Bauer (MIT) proposed five possible sets of reasonable expectations for gigabit broadband: (1) Gbps everywhere (not feasible today, yet the only one that is consistent with today's performance expectation); (2) Gbps island (can average this bandwidth consistently within the access network); (3) Gbps in aggregate (aggregation of flows to multiple destinations, not necessarily to a single destination); (4) Gbps to select services; and (5) observable growth toward Gbps paths, e.g., with regulator-visible interconnection agreements to keep capacity at pace with demand. A rational policy would recognize not everyone will have (or need) Gbps access soon, but over time the minimum need will increase. Policy focus will need to shift from peak speed to concerns about minimal access speed.
7. **Merging multiple platforms onto a single device**. Phillipa Gill (Stony Brook) posed the thought experiment of merging multiple measurement platforms onto a single Pi to optimize installation and hosting coverage.[13] Given the different user bases, capabilities, and management and scheduling models of different infrastructures, she was not greatly optimistic, but suggested small interoperability steps such as common measurement code modules. If that worked, the community could think about a common API. Robert Kisteleki (RIPE NCC, Atlas) considered it much harder to synchronize code bases than interfaces. Alberto Dainotti (CAIDA/UCSD) noted that this proposal sounded like the European *mPlane* project from few years ago, so revisiting lessons learned from that project might be useful. Steve Bauer noted that the robotics community accomplished an analogous goal in standardizing on an open source operating system for robotics (ROS), which might offer lessons. To many in the room it seemed an obvious question: given so many active measurement platforms, should we integrate them and have them work together via open standard protocols? A related exercise in scoping such a platform's capability would be to survey what minimal functionality would enable 80% of the IMC papers using active measurement in the past X years. (The caveat is that researchers do researach with data they can get, not necessarily what they would do if they had better data.)
8. **Crowd-sourcing measurements using Google AdWords** Geoff Huston (APNIC) reported on his progress with experiments for measuring end user characteristics. He started years ago trying to improve metrics for assessing IPv6 and DNSSEC deployment, because people tended to use metrics that were easy to measure, but did not accurately reflect usage, e.g., announced IPv6 routes, AAAAs in DNS zones, DNSSEC-signed zones. He leverages Google's advertising service to measure millions of users, by launching (originally flash, now HTML5) ads through Google with embedded code that attempts to retrieve specific content from servers he controls. The retrieval process, i.e., the server logs, reveal behavior and capabilities of end user host software. He has used this method to study the evolution of IPv6, DNS, and DNSSEC behavior. He publishes the results of these experiments on his blog (4-8M unique end points/day, 1B to date). The main limitation of this approach is that he can only have the user probe or request data from his own (server) vantage point. For example, measuring IPv6 using one or a few destination servers does not give a good sense of a given user's IPv6 connectivity, since users sometimes have incomplete IPv6 routing tables which probes to only a few servers may not capture.

This last thread triggered discussion of ethical considerations of doing reachability measurements via ads, unbeknownst to the users clicking on the ads. Geoff noted that if users click on the ads, he complies with Google's requirements to tell people how to opt out. He also noted that while he publishes aggregated statistics about individual ISPs, he does not release the end point IP addresses. There is an ethical quandary related to the study of censorship using any method that involves probing for a censored piece of content from a machine in an authoritarian country, without the knowledge of the machine's user(s). It was unclear whether university researchers could get such an experiment through a university IRB.

# 4. REFLECTIONS ON ETHICS

Recurring discussions of ethics and deep contrasting feelings about potential associated harm in various scenarios suggest the need for ongoing dialogue in this area. One possible goal (often discussed at networking research venues in the last few years) is to frame some consensus on boundaries where, with informed consent, the level of harm is low enough that people agree it is ethical. Another concern was the frustration academics feel that they cannot do things that researchers in the commercial world do all the time, some of which might be quite harmful. One participant noted that we might think of ourselves as the wild west in terms of ethical assessments, but we would not if we looked at the Internet advertising and brokering industry, which has no formal standards except to exploit information (i.e., track and target ads) to make money. Another participant pointed out that regulators are paying more attention to this industry now and may eventually limit some of these practices, but also cautioned awareness that plenty of people would choose "free" over "ad-free", given the choice, although admittedly users lack meaningful transparency over what is actually collected and how it is used to target. We quickly ended up in territory that merits its own workshop series: how to enable and incentivize third parties, including consumer advocacy groups, to provide transparency to users about how data aggregators are using information about

[13]This idea was also extensively discussed at a Dagstuhl workshop in January 2014, as reported in CCR's April issue: http://www.sigcomm.org/sites/default/files/ccr/papers/2016/April/

them. (The U.S. FTC has begun to have workshops related to this topic.)

Many thought highly of the current trend in the research community, where program committees require that authors provide an explicit discussion of ethical decisions and tradeoffs in their submitted papers, and if they are operating out of a clearly ethical area, to formally address the costs and benefits of their experimental method. An operator participant at the workshop indicated that when they share data with researchers, they require a similar written explanation of ethical issues, to make sure both the researcher and operator have considered them. For this operator, if there is any PII (personally identifiable information) being shared, it has to go through a formal ethical review anyway.

One example came up where an operator received illegally collected data that contained user names and passwords, including some accounts of their own customers that were now known to be compromised. Their legal counsel was unsure what was allowed with this data; they concluded that enabling each customer visibility into that customer's compromised user accounts so they could notify them was okay. But when outside researchers wanted the same data to study common passwords, the operator felt it unethical and shredded the disks instead. [14]

## 5. DATA SHARING FRAMEWORKS

We discussed recent developments of the DHS-funded IMPACT project[15] (previously called PREDICT) which supports operational costs related to acquiring and sharing cybersecurity-relevant data, including administrative and legal overhead (acceptable use policies, non-disclosure agreements). IMPACT data sets include those related to active measurement censuses, outages, topology, and malware. Erin Kenneally (DHS) is leading the next phase of this project, and one of her goals is to experiment with a prototype information marketplace of data and analysis exchange between researcher and data providers. She solicited other ways of bridging the gap between data providers (operators), academic and industrial researchers, and cyber security technology and policy development.

Several projects in the room mentioned they were relying on IMPACT products (the Menlo Report, and guidance from IMPACT PIs on development of AUPs) in their own data-sharing policy development. We also discussed what data we would like to see in such a shared repository. Although this topic also merits its own workshop, in the context of shared active measurement data sets the most popular idea was a continually updated repository of AS-level paths, which everyone agreed would support multiple existing research projects and enable new ones.

## 6. REFLECTIONS OVER LAST EIGHT YEARS OF AIMS WORKSHOPS

We took this opportunity to list common themes over the last eight years as we consider future directions.[16] We view these as potential topics for future workshops or other research attention.

1. **Data and infrastructure building blocks**
   (a) dedicated and scalable operational platforms to support active measurements, and for testing of experimental services such as detection of hijacking or outages
   (b) accessible wireless measurement testbed instrumented across all layers of the wireless stack
   (c) stable sources of funding to support research-enabling Internet measurement infrastructure
   (d) collaborative ways to store, retrieve (query), and share data, e.g., federating clusters to share backup resources
   (e) techniques and supporting platforms that facilitate data collection, aggregation, correlation, and analysis of heterogeneous types of data, e.g., active, passive, BGP
   (f) unifying interfaces across measurement and data aggregation platforms, while articulating desired functionality and reasonable expectations in doing so, e.g., how does one write an experiment for one infrastructure and port it to another
2. **Promoting synergies among industry, government, researchers, vendors, Internet service providers**
   (a) programs to incentivize researchers and industry to share data and otherwise cooperate
   (b) concerted effort to transfer measurement technologies to private and government sectors, e.g., cooperative initiatives to build user-visible measurement capabilities into equipment (routers, CPE, mobile devices)
   (c) identifying important measurement questions and who will value their answers
   (d) methods, tools, ethical guidelines, and policies to facilitate measurement research while protecting privacy and utility of resulting data
   (e) leverage scarce funding by coordinating projects among different agencies (e.g., FCC, NSF) needing common functionality from measurement infrastructure
3. **Path measurement research**
   (a) open-source path measurement tools that require use of ancillary knowledge bases, e.g., AS-level traceroute
   (b) navigating and quantifying the trustworthiness of traceroute output and associated inferences
   (c) understanding peering between cellular networks and the rest of the Internet, and efficiency of the resulting overall topology
4. **Quality of experience measurement research**
   (a) reliable and low-impact network performance measurement techniques, especially for mobile networks
   (b) metrics to capture the end-user quality of experience (not just network performance)
   (c) how applications could use network measurement

[14] This episode was described at the January 2016 Dagstuhl workshop; slides at http://boemund.dagstuhl.de/mat/Files/14/14052/14052.LeenesRonald.Slides.pdf. accessed 25 July 2016

[15] Information Marketplace for Policy and Analysis of Cyber-risk & Trust; http://www.impactcybertrust.org.

[16] Many of these are also highlighted in the report from the 2014 NSF Workshop on Mobile Community Measurement Infrastructure: http://www.ccs.neu.edu/home/choffnes/nsf-meas-wkshp/

data in real-time to optimize performance

5. **Network architecture research**
   (a) studying the impact of traffic shaping, content caching, and other middlebox behavior on user-perceived network performance of different cellular carriers)
   (b) measurement architectures to support engineering and management of emerging information-centric network architecture

*ACKNOWLEDGMENTS. The workshop was funded by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHS S&T/CSD) Broad Agency Announcement 11-02 and SPAWAR Systems Center Pacific via contract number N66001-12-C-0130, and by Defence Research and Development Canada (DRDC) pursuant to an Agreement between the U.S. and Canadian governments for Cooperation in Science and Technology for Critical Infrastructure Protection and Border Security. The work represents the position of the authors and not necessarily that of DHS or DRDC.*

# 7. WORKSHOP PARTICIPANTS

The main reason we continue this workshop is the enthusiastic participation it attracts from some of the brightest and most productive people in the community. We are grateful for their engagement and insights, many of which are reflected in this report.

- Adnan Ahmed (University of Iowa)
- Alexander Isavnin (the open Net)
- Andrei Robachevsky (Internet Society)
- Ann Cox (DHS S&T Cyber Security Division)
- Ashkan Nikravesh (U. Michigan)
- Brandon Schlinker (USC)
- Brian Tierney (Berkeley National Lab)
- Casey Deccio (Verisign)
- Christoph Dietzel (TU Berlin / DE-CIX)
- Danilo Cicalese (Telecom ParisTech)
- David Choffnes (Northeastern)
- David Clark (MIT)
- Drew Taylor (Comcast)
- Emile Aben (RIPE NCC)
- Erin Kenneally (DHS)
- Ethan Katz-Bassett (USC)
- Geoff Huston (APNIC)
- Hiroshi Yamamoto (NTT)
- Italo Cunha (UFMG, Brazil)
- John Heidemann (USC/ISI)
- Matthew Luckie (U. Waikato)
- Mattijs Jonker (Twente U.)
- Mobin Javed (UC Berkeley)
- Narseo Vallina Rodriguez (Int. Computer Science Institute)
- Nathan Owens (Comcast)
- Phil Roberts (Internet Society)
- Phillipa Gill (Stony Brook University)
- Rachee Singh (Stony Brook University)
- Ramakrishna Padmanabhan (University of Maryland)
- Ricky Mok (Hong Kong Polytechnic University)
- Robert Beverly (Naval Postgraduate School)
- Robert Kisteleki (RIPE NCC)
- Roderick Fanou (IMDEA Networks Institute / CAIDA)
- Roland van Rijswijk-Deij (Twente University / SURFnet bv)
- Shane Alcock (University of Waikato)
- Steven Bauer (MIT)
- Tanja Zseby (TU Wien)
- Yi-Ching Chiu (University of Southern California)
- Zubair Shafiq (University of Iowa)
- kc claffy (UCSD/CAIDA)
- Alberto Dainotti (UCSD/CAIDA)
- Amogh Dhamdhere (UCSD/CAIDA)
- Marina Fomenkov (UCSD/CAIDA)
- Bradley Huffaker (UCSD/CAIDA)
- Young Hyun (UCSD/CAIDA)
- Chiara Orsini (UCSD/CAIDA)
- Joshua Polterock (UCSD/CAIDA)