The 7th Workshop on Active Internet Measurements (AIMS-7) Report

kc claffy
CAIDA/UCSD
kc@caida.org

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 It has NOT been peer reviewed.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 31 March - 2 April 2015, CAIDA hosted the seventh Workshop on Active Internet Measurements (AIMS-7) 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. This report describes topics discussed at the workshop, including current state of Ark and related infrastructure, current and proposed experiments using these infrastructures, and participants' views of challenges and priorities. Materials related to the workshop are at http://www.caida.org/workshops/aims/1503/.

1. MOTIVATION

For seven 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. This year we went back to our roots, and focused on CAIDA's a active measurement infrastructure (Archipelago or Ark): its status, role, activities, research results, and relationships with other measurement infrastructures. This report describes topics discussed at the workshop, and summarizes participant views of challenges and priorities.

2. ARK INFRASTRUCTURE

Young Hyun (CAIDA) began the first session with an update on the status and future plans for the Archipelago active Internet measurement infrastructure. As of March 2015 the infrastructure included 107 monitors spread across 40 countries (Figure 1). Of these, 59 use Raspberry Pi hardware, 44 have IPv6 support, and 36 have RADclock support. Institutional hosting sites include academic (48), residential (24), business (23), and network infrastructure (10). The Raspberry Pi nodes include the first generation (700MHz ARMv6 CPUs with 0.5GB RAM) and second generation (900MHz quad-core ARMv7 CPUs with 1GB RAM).

Supporting software includes the Marinda distributed tuple space [11], which allows users to request execution of measurements using structural pattern matching. Ark's probing engine continues to be Matthew Luckie's *scamper* software [1]. Young used *scamper* as

Figure 1: Archipelago monitor deployment as of March 2015.

the basis for a lighter weight probing engine called *mper* [4] which sends and receives individual packets (ICMP, UDP, TCP) without high-level supporting functions such as traceroute. *mper* provides a measurement API, allowing users to write measurement scripts in Ruby.

Young described *Dolphin* [2], a new software tool he designed to conduct parallel reverse PTR DNS lookups (of hostnames) of IPv4 and IPv6 addresses as our traceroute measurements encounter them. Dolphin performs millions of lookups per day from a single host, and retries failed lookups once per day for up to 3 days; to reduce load on authoritative DNS servers, Dolphin looks up any given IP address at most once in any 7 days, regardless of TTL in the response. Dolphin is built on *libunbound* (part of Unbound by NLnet Labs), a validating, recursive, caching resolver. Dolphin is a single Python source file (845 lines). Young also created a related script *qr*, which performs only simple DNS lookups using the *ldns* [3] library for low-level structured access to raw DNS response packets. In August 2014 CAIDA used *qr* to perform a lookup of the entire IPv4 address space, in 8.5 days (317M queries/day) a rate chosen to minimize load on hosting networks [16].

Young described two other tools he developed to support on-demand measurements: (1) *tod-client* (topology-on-demand), a scriptable command-line interface for performing IPv4 and IPv6 traceroutes and pings [10], and (2) the *Vela* web interface (vela.caida.org), which allows remote execution of ping and traceroute (ICMP, TCP, UDP) measurements on specified Ark monitors. For experiments that need full control of the Ark nodes, CAIDA provide shell access to the Unix nodes themselves, where users can compile and

run their own software and/or write measurements in Ruby with Ark software. Experiments that have used Ark in this full access mode include methods for detecting packet modifications by middleboxes [6], IPv6 alias resolution [13], and Casey Deccio's DNS root key rollover preparedness study (Section 6).

In addition to supporting researcher experiments on demand, Ark runs other measurements in the background:

1. **IPv4 topology**: traceroute to a random address in each routed /24 (570M/month)
2. **IPv6 topology**: traceroute to a random address and ::1 in each routed prefix; pings to IPv6 addresses of Alexa 1M sites (16M/month)
3. **PTR DNS lookups** of observed IPv4 and IPv6 addresses
4. **MIDAR alias resolution** [5]
5. **congestion at inter-domain peering links**: topology mapping to identify AS borders, subsequent ICMP probing to identify evidence of congestion at those borders [14].

CAIDA's focus for Ark is shifting to improving data accessibility: creating an interface for browsing, querying, and visualizing the data gathered by the infrastructure command-line and web interfaces. For browsing, the goal is to support viewing of broad properties and summary statistics (e.g., response rates, path lengths, RTT distributions, inferred AS links) over multiple time scales and aggregation levels. For querying, the goal is for researchers to be able to find the most relevant historical data for one's research, which means either directly answering a question via the interface, or identifying which data to downloaded for further study. For example, one might want to find: all traceroutes through a given region and time period toward/across a particular prefix/AS; all router address aliases for a given IP address; all inferred links to a router identified by a given IP address; or all routers in a given city. Some of these queries are more challenging to support than others; we hope at next year's AIMS to have an interface to present to the community.

2.1 Supporting software: RADclock

Darryl Veitch gave an overview of his RADclock system [12], which he developed, and CAIDA deployed on Ark, to mitigate the clock drift that most hosts experience because they use hardware counter-based clocks. Because GPS instrumentation (to fix the problem) is so expensive, most hosts instead use NTP (Network Timing Protocol) to continually correct (i.e., compensate for the drift of) their clocks based on transactions with a network of dedicated time servers across the Internet. But this feedback approach inherently suffers from large and variable network (and host or server) delays. RADclock provides a mechanism to obtain more accurate absolute timestamps (on the order of hundreds of microseconds rather than milliseconds), and much higher robustness to network delays and disruptive events. Darryl has used RADclock to discover weaknesses in the extant NTP infrastructure, and believes that the combination of Ark+RADclock is an ideal platform for further testing the integrity of time servers, and could provide a service that enables Ark itself as well as others to better select their own stratum 1 servers. (This would require only that some Ark monitors to be stratum-1, his wish for this year).

Darryl posed a loftier challenge that Ark with RADclock support is in a unique position to support: research and development of an NTP-replacement Internet timing infrastructure. The two projects – monitoring the integrity of public timing infrastructure, and R&D of a new public timing architecture – complement each other as well as Ark's goal of providing the research community with high-precision timestamps at reasonable cost. He is working with the FreeBSD development community to achieve full inclusion of RADclock in FreeBSD 11, and Linux versions exist via patches up to 2.6.32. Darryl is looking for developers and partners to help with this project.

2.2 Supporting software: ArkQueue

Ark and its topo-on-demand (ToD) system does not expose monitor status, and given the widely distributed infrastructure operated by volunteers, head-of-line blocking can occur when using topo-on-demand for large experiments. Rob Beverly (NPS) talked about *ArkQueue*, a Python module he developed to navigate submission of probe requests to topo-on-demand (ToD), and handle many common failure scenarios. To deal with a changing subset of unavailable or slow Ark monitors, ArkQueue sorts and queues user probe requests by vantage point (VP), runs an instance of tod-client per vantage point, tracks VP response time and stops submitting to that VP if it is unresponsive, and reports unresponsive VPs for future reference. ArkQueue facilitates intelligent probing patterns, where future probes depend on feedback from earlier probes. Rob is sharing these and other Ark-related tools to facilitate wider use of Ark.[1]

Rob offered the following wish list for Ark: ("not in order"):

1. Expose list of available monitors (and tell user of new monitors put into production)
2. Warts output (ToD produces tab delimited partial output)
3. Ability to clear tuples in a timely manner
4. Full control over scamper options
5. Fix Marinda memory leak
6. Visibility into outstanding request tuples
7. Visibility into individual monitor queue/status

3. CURRENT EXPERIMENTS ON ARK

Julien Gilon (graduate student at University of Liège, visiting scholar at CAIDA) presented his preliminary analysis of BGP *more specific* prefix advertisement. His goal is to investigate whether the significant fraction of more specific BGP-announced prefixes (smaller segments of address space covered by another prefix that represents a larger segment) observed in a global routing table are primarily due to traffic engineering. He used an Ark monitor that had a co-located BGP view, tracerouted to more specific prefixes, inferred borders between ASes, and compare inferred paths with those observed for less specific prefixes. [2]

Amogh Dhamdhere, Matthew Luckie (UCSD/CAIDA) and Steve Bauer (MIT/CSAIL) presented various aspects of a new collaboration to use Ark to study Internet interdomain congestion, The technical goals are to develop methods and infrastructure to measure and monitor the location and extent of interdomain congestion, i.e., happening across two directly connected service providers. They currently use traceroute data from Ark monitors, combined with BGP and other meta-data to infer boundaries between ASes, and then uses TTL-limited time-series latency probing (TSLP) to characterize episodes of impaired performance on the far end of an AS boundary with no simultaneous impairment on the near end of an AS boundary. The method has shown surprising promise thus far [14], although involves challenges, the most prominent of which is the interdomain router-level topology inference.

We reviewed attempts to use Ark to improve the state of IP mapping and resource geolocation. Bradley Huffaker (UCSD/CAIDA) gave an overview of a current DHS-funded effort to improve the coverage of CAIDA's AS-level topology by integrating peering links

[1] http://www.cmand.org/direct.

[2] Julien wrote the results up for his Masters thesis later in the summer, and is working on submitting a version to a workshop.

observed via traceroute and IXP servers, relying on BGP community data to inform AS relationship inference [8]. Using 106 Ark monitors, CAIDA inferred adjacent peering links, limiting the inference to only one hop from the monitor to minimize false inferences due to traceroute artifacts. Bradley compared the coverage of different AS topology data sources: BGP-only, traceroute-derived peerings, IX-derived peerings, and all data combined. He compared topological metrics of the simple (BGP-derived) vs. combined graph, including metrics such as eccentricity, betweenness, degree, coreness, clustering, and customer and peer cone.

Bradley also reported on two related CAIDA systems to support use of DNS hostname information to improve geolocation inference coverage and accuracy of router infrastructure The first system (DDec – http://ddec.caida.org) extracts DNS hints from hostnames, and provides a public interface to resulting data for lookups and validation/correction by external parties. CAIDA's DNS-based Router Positioning (DRoP) system [9] uses active measurements from Ark and a large library of known geographic strings (including those gathered from DDec above) to *automatically* infer geographic hints in hostnames. RTT measurements from different Ark hosts constrain inference of IP addresses as candidate routers in the same geographic region. Future directions include using DRoP-geolocated routers to geolocate adjacent routers, inferring less common geographic hints, improving methods for validation and feedback, and increasing its visibility to operational communities.

Michael McCarrin (NPS) used DRoP to extend landmark-based geolocation methods to router interfaces. He performed traceroutes to a set of landmarks and to a target, determined the point at which the traceroutes diverged, estimated the delay between each landmark to a given target, and then approximated the target's location to be nearest the lowest-RTT landmark. The underlying assumption of the method is that routers are frequently co-located with other routers. He used DRoP's 6M interfaces and 8K unique locations as ground truth to validate his method, using half (4K) of the locations as landmarks and the other half (4K) as targets. In the process he helped CAIDA gain a better understanding of inconsistencies in DRoP inferences. NPS and CAIDA continue to collaborate to scale up error detection, strategically select landmarks to maximize inferential power, and investigate the value of historical traceroute data for improving inference coverage.

Erik Rye (NPS) presented his work developing an Emulated Router Inference Kit (ERIK), which generates Internet-like, flat, random network topologies, as well as individual router configurations (including IP addressing) based on the generated topology and policy. The system also configures a Dynamips hypervisor to run router images and interconnect virtual routers and switches. With a configured set of topologies, one can run automated topology inference exhaustively on the resulting graph, compare topology generation models, evaluate the effects of number and selection of VPs, evaluate resiliency of topologies under failure scenarios, and expose implementation-specific behaviors. Opportunities for others to extend ERIK included scaling up the number of emulated routers, combining intra-AS and inter-AS topologies, integrating JunOS topology emulation, validating topology inference methods, and supporting new probing/inference algorithms, e.g., for IPv6.

We ended the day with Alberto Dainotti's (UCSD/CAIDA) introduction to a new NSF-funded collaboration[3] with Phillipa Gill (Stonybrook) that will rely on the Ark infrastructure to detect and characterize BGP hijacking events. The goal of the project is to develop live monitoring methodologies to detect traffic interception, test and evaluate the system with test hijacks (using USC's PEERING infrastructure[4]), and quantify the impact of detected events. The project will combine and and correlate BGP data from RIPE RIS and RouteViews, continuous daily as well as triggered traceroutes data from Ark nodes, and external data to support geolocation of interception events. They hope to improve the number of Ark monitors with co-located BGP feeds (currently only 20), and develop new methods to infer AS paths from traceroutes.

Throughout the day there was a lively debate on the feasibility of a general protocol and system architecture to serve as an interface to request measurements from different infrastructures. There was recognition of the tradeoffs across different measurement needs that lead to architectural diversity of systems, including with methods for message queuing, e.g., rabbitmq vs Marinda. Steve Bauer presented some thoughts on tools and approaches that can help accelerate the (ideas $\rightarrow$ implementation $\rightarrow$ evaluation) chain when dealing with large scientific data sets, and how he has applied them to classification of congestion events.

4. RELATED PROJECTS

The second day of the workshop began with a review of other topology measurement infrastructure projects that present opportunities for collaboration. Guilherme Martins (Georgia Tech) gave a BISmark platform update and roadmap. As of March 2015 there were 120 reliable online routers (typically in homes) over 5 continents. Nodes are either Netgears running OpenWrt or Raspberry Pis. Each node runs automated active measurements (netperf, fping, paris-traceroute), and supports other performance monitoring tools, as well as historical charts for bandwidth and latency.[5]

Spiros Thanasoulas and Christos Papadopolous (Colorado State University) gave a status update on BGPmon, which collects and stores BGP data from hundreds of BGP peers, and responds to queries. The BGPmon project has been around for years, pre-dating other projects with the same name e.g., bgpmon.net. Christos is leading a complete rewrite of the system architecture and code base, from an XML-streaming model to a language-agnostic system that leverages advances in database technologies and industry best practices. They are using Golang, and protocol buffers as the internal data format. Christos' team is using BGPmon to track outages, by correlating ISI's active measurements with BGP messages observed before and after the outage.

Alistair King (CAIDA/UCSD) presented *BGPStream*, CAIDA's new software framework for BGP data analysis. RIPE NCC's *BGPdump* is the de-facto standard for BGP data analysis, performing low-level extraction of information from MRT data. Processing historical data requires (semi-)manual download and curation of data, and processing across time, collectors, and data types requires custom code. *BGPStream* is the first set of tools, libraries, and interfaces that can perform both historical analysis as well as real-time monitoring of BGP data. The framework provides transparent access to different MRT sources, including previously downloaded local files, RIBs and updates from RouteViews and RIPE RIS, and real-time streams for Colorado State's BGPmon (RouteViews collectors). The framework also includes *BGPCorsaro*, a fork of CAIDA's Corsaro tool that transforms a stream of BGP records into a set of structures and metrics representing specific time intervals, and then supports modular plugins to execute desired analyses. CAIDA supports a production deployment of BGPStream, which it currently uses for its outage and hijack detection projects. CAIDA plans to release an open source version of the code later this year. Beta access is available upon request.

[3] http://www.caida.org/funding/hijacks/

[4] http://peering.usc.edu

[5] http://networkdashboard.org

	BISmark	Ark	SamKnows	RIPE
Continuous active	*Y*	*Y*	*Y*	*Y*
Passive	*Y*	*N*	*Y/N*	*N*
Scope of experiments	*High*	*Higher (better CPU/storage)*	*Medium(resource constraints)*	*Low (only use tools compiled in)*
Heavy duty exp	*?*	*Y*	*N*	*Y/N*
Local storage	*N*	*Y*	*N*	*N*
Scale	*~*	*~*	*Y*	*Y*

Figure 2: Feature matrix of active measurement platforms (Srikanth's slides).

Robert Kisteleki (RIPE NCC) presented highlights of recent RIPE Atlas activities. New developments include better user interfaces and APIs, support for tagging probes, new measurement types, data streaming, and anchor nodes that are more powerful than regular nodes. The measurement API now supports functions to query/search, create, change, or stop measurements, as well as to download results and state checks. They are planning new measurement types including: querying NTP servers, wifi association and authentication (while being on a wired network), HTTP (against anchors), and TLS checks. They support streaming real-time access to data, with the ability to replay recent historical data.[6] He introduced the concept of a RIPE Atlas anchor, which is a rack-mounted PC (Soekris) that is more powerful and can receive measurements as well as source them. Atlas has about 120 anchors now, mostly in data centers. They are now looking into revamping the RIPE RIS BGP collection architecture to leverage all the lessons learned (and code developed) for Atlas. The RIPE Atlas team supported a successful hackathon just days before this workshop, where 25 hackers worked on projects that visualized RIPE Atlas (and related) data.[7]

Srikanth Sundaresan (ICSI) led an extended discussion on how researchers would like measurement platforms to work together, e.g., desired functionality, interfaces, and reasonable expectations for infrastructure operators and researchers in the context of a unified platform. He used a feature matrix (Figure 2) which framed discussion on the range of capabilities and limitations of different platforms, and shed insight on how a researcher might choose a platform to use. He discussed the challenge of writing an experiment for one infrastructure and porting it to another, reporting his learning experience in attempting to port his WTF (Where's the Fault?) tool from the Bismark to the SamKnows platform administered by the FCC for the Measure Broadband America program. After he extensively tested it on BISmark (65+ homes, 2 months), in June 2013 the FCC invited his group to try porting it to SamKnows. The experience ran into several hardware and software interoperability hurdles, but in late 2013 they finally launched WTF? on several thousand SamKnows nodes, crashing 30-40% of them within 36 hours. The experiment was pulled, although they did get some interesting data.

Srikanth posed the question to those running research infrastructures (SamKnows was not designed to support research): is it possible to agree to a set of basic constraints that would enable a unified platform for experiment development, allowing an expansion of each platform's visibility, and facilitating research that relies on measurements from the edge. The practical problems span memory, CPU, and bandwidth constraints, and infrastructure-specific quirks. The platform should provide an open, easy-to-use development tool chain, support for data synchronization, and transparent enforcement of computational and network constraints. He acknowledged this last task was the most difficult; he found it was nearly impossible to vet experiments on BISmark with confidence. But of the existing platforms, he thought that BISmark and Ark were probably easiest to integrate via an external interface. He urged as a practical first step to try to run basic experiments on each others platforms.

The room was split on the feasibility of a unified platform. Robert Kisteleki (RIPE) felt the risk of someone providing misbehaving code, either intentionally or not, would override any desire to support research. He thought others were underestimating the complexity of communicating with fundamentally different architectures. He was more optimistic that we could build an interface to enable systems to talk to each other, e.g., "Tell Ark to measure X." (Mplane has apparently tried to do something similar.[8])

Amogh (CAIDA) noted another obstacle: experiments on these platforms typically require babysitting, e.g,. restarting with revised code. We would need middleware, e.g,. Puppet[9], to support interaction with the remote platform. Matthew Luckie (CAIDA/U. Waikato) wrote functionality to have scamper issue commands over a control socket to a remote node. Ethan has been using this functionality on PlanetLab and MLab in the early steps towards rebuilding reverse traceroute. The main problem Matthew saw was (human) cycles available on both sides for infrastructure support.

5. PROPOSALS TO USE ARK

We shifted the discussion to researchers who would like to use the Ark infrastructure for their research. Phillipa Gill (Stony Brook) reported on her platform (ICLab) for network measurement of censorship. ICLab's approach is to try to fetch a Web page from a location with suspected censorship, and fetch the same Web page simultaneously from a location without censorship, then compare the results. She supports a baseline set of network measurements: HTTP request, traceroute, DNS queries, HTTP header, fingerprinting (Netalyzr test), customized IP TTL header to localize the censor in the network, etc. She proposed using Ark and Atlas for censorship measurements, using their basic traceroute and ping support, but she acknowledged the risk of putting hosting sites at risk. Robert K. emphasized that to protect hosting sites, RIPE Atlas has a policy of not supporting censorship measurement.

Danilo Cicalese (Tlcom ParisTech / UPMC), spoke on his recent work (presented at INFCOM2015) developing a protocol-agnostic methodology for detecting, enumerating and geolocating an anycast instance. They developed an architecture that can launch a periodic fast census to probe for anycast instances. This capability could enable monitoring and diagnosing problems of DNS root servers, detecting BGP hijacks, and monitoring anycast-reliant CDNs. The main challenges of an anycast census are the dependencies of the measurement platform, and efficiency of data collection and analysis. Distribution of vantage points greatly affects detection as well as accuracy of geolocation of the anycast instances. He hoped that Ark could complement the other platforms he is using.

Benoit Donnet (Universit de Lige), talked about how to measure the transit tunnel diversity of MPLS deployments. His motivation was that previous MPLS measurement studies have focused on its impact on topology discovery, rather than actual usage of MPLS by operators. Operators may build tunnels using the basic Label Distribution Protocol (LDP), which does not allow traffic engineering

[6]https://atlas.ripe.net/docs/
[7]https://atlas.ripe.net/hackathon/2015/
[8]http://ict-mplane.eu
[9]http://sourceforge.net/projects/puppet/

(TE), or with RSVP, which enables TE. They found that LDP was much more common, and use of RSVP-TE to manage distinct Forwarding Equivalent Classes (FEC) seemed minimal. (Work subsequently published in IMC2015 [15].)

Daniel Zappala (BYU) surveyed recent measurement work, including his own, on using measurements to understand the fragility of the TLS certificate ecosystem. There is abundant evidence that the CA system's weaknesses are getting worse. He reviewed five methods to measure its vulnerabilities: scans from a single vantage point; passive monitoring (e.g., Bro); mobile apps (e.g., Netalyzer); flash apps with millions of views (e.g., using Google Ads), and user surveys. Daniel proposed the use of measurement testbeds such as Ark (as well as Dasu/NameHelp and Atlas) to contribute to a comprehensive view of certs seen by clients from as many vantage points as possible, hopefully resulting in construction of a heatmap of TLS proxy location and behavior.

Neil Spring (U. Maryland) gave a fascinating look at round-trip times much higher than anything that seems reasonable on the Internet. The conventional wisdom based on extant active measurement systems is that a probe should allow 1-3 seconds for a response, before giving up. They used the ISI survey data set to explore the validity of this assumption. They sampled 2000 of the highest RTT IP addresses in the data and re-probed them using scamper, finding the RTT was still just as high. RTTs as high as 5 seconds occurred for 5% of his probes, and the highest RTT was 159 seconds. Participants were fascinated by these anomalies and encouraged Neil to continue his investigation. (They were inspired to write up the results which were accepted to IMC2015.)

6. USING PLATFORMS TOGETHER

We started Day 3 with short talks continuing the theme of using multiple measurement platforms together. Rocky Chang (Hong Kong Polytechnic) summarized his group's recent work on improving the accuracy of browser-based measurement and measurement with embedded systems, such as home routers and Pis. Vasileios Giotsas (UCSD/CAIDA) talked about his work developing a platform to query Looking Glasses (web interfaces to routers or hosts that allow execution of active measurements and/or routing table queries, e.g., traceroute, ping, or sho ip bgp.) Looking glasses sometimes offer what many active measurement platforms struggle with: traceroute and BGP vantage points at the same location. Challenges include the lack of a centralized repository of available looking glasses, lack of standardized querying or output formats, the fact that they are generally intended for low-frequency (manual) querying, attrition over time, and changes in supported commands. He reviewed his methodology for automatically discovering looking glasses, and reported results of his recent crawl, finding active looking glasses in 2,984 locations across 499 autonomous systems (ASes). The goal of the interface is to be able to use customer cone information to select traceroute vantage points that will maximize the probability of crossing a specific desired link. Future plans include setting up periodic measurements to popular locations, e.g., large CDNs, hybrid relationships, multilateral peering links; a public REST API for querying the looking glasses, and optimizing the distribution of queries to different platforms.

Ethan Katz-Bassett (USC) gave an introduction to his new project Sibyl, which is an attempt to provide a unified interface to traceroute platforms. He has proposed to architect and implement a system that provides routing information based on rich queries that researchers and operators can express naturally. Sibyl integrates diverse traceroute vantage points that provide complementary views of Internet routing, from high-rate, low-diversity vantage points (like Ark and PlanetLab) to low-rate, high-diversity vantage points (like Atlas and looking glasses), to enable queries for measurements from thousands of ASes. Because users may not know which measurements will traverse paths of interest, and because probing rate limits keep Sibyl from tracing to all destinations from all sources, Sibyl uses previous measurements to intelligently predict which measurements will most likely match a given query. Ethan was still developing the query language, soliciting feedback on types of queries users would like to make of such a system, e.g., "give me (at least) one path that matches", "give me as diverse a set of matching paths as possible". He gave an overview of how his system splices existing traceroute paths to find a path to probe that will likely match a given query.

Renata Teixeira (Inria) gave an update on her effort to develop networking technology that can guide network performance and diagnosis (where is the problem, and if in the home, what is the cause?), as well as infer user dissatisfaction with application performance. She announced the updated Fathom 2.0, browser-based programmable interface for writing and launching either passive or active measurements from web pages. The new version of Fathom is written on top of the add-on SDK, and supports Mobile Firefox (on Android), and common JS module support. Built-in capabilities include connection debugging, homenet discovery, and network performance monitoring. She hopes that Ark or RIPE nodes within homes can collaborate with Fathom, i.e., a Fathom-enabled browser could trigger a request for an Ark or RIPE node to perform a specific measurement, request historical data from an in-home Ark or RIPE node, or query Ark/RIPE data archives in real-time to locate potential WAN problems.

Steve Bauer briefly reviewed his "*net.info*" proposal for sharing network service information, e.g., contracted upload and download speeds. Customers generally do not know and cannot easily find this information, and inferring it from measurement data is difficult. He proposes that an http get of net.info redirects to an ISP-supported http-accessible page, e.g., http://net.info.csail.mit.edu/, with conventions subject to community consensus. The information returned would be specific to the client IP address, and would allow for integration of provider response data with the vast amount of measurement test data currently collected across projects. Providers could also use this channel to expose other service parameters, network traffic alerts, or operational conditions to users. He listed some next steps: developing knowledge representation formats, feedback from privacy experts, broadband ISPs, measurement projects, identifying holes to determine how challenging this would be, and build a demonstration prototype.

Ioana Livadariu (Simula/CAIDA) gave an update on her recently started work on comparing IPv4 and IPv6 routing stability using BGP data from RouteViews and data plane data from 9 Ark monitors probing dual-stacked targets. Thus far her measurements reveal more routing changes in IPv6 than IPv4, but most IPv6 routing dynamics are generated by a few unusually unstable prefixes.

Ramakrishna Padmanabhan (U. Maryland/CAIDA) summarized his PAM2015 paper on UAv6, a new alias resolution technique that uses partially used IPv6 prefixes to find aliases. UAv6 finds aliases in two phases. The first "harvest" phase gathers potential alias pairs, based on the empirical observation that addresses adjacent to router interface addresses are often unused. UAv6 probes these unused addresses (of /126 prefixes), eliciting ICMPv6 Address Unreachable responses. The assumption is that the source address of such a response belongs to a router directly connected to the prefix containing the unused and router interface addresses. The second "disambiguation" phase determines which interface address is an alias of the Address Unreachable's source address. UAv6 uses both new and established techniques to prove or disprove that two ad-

dresses are aliases. They confirmed the accuracy of UAv6 by running the Too-Big Trick test [13] on discovered aliases, and comparing them with limited ground truth from the Internet2 topology. They concluded that UAv6 and the Too-Big Trick are complementary to existing address-based techniques for resolving IPv6 aliases, finding alias pairs that other methods do not.

Casey Deccio (Verisign Labs) presented his new work using Ark to provide diagnostic measurements of authoritative root and top-level domain services. Instrumenting these measurements from diverse vantage points is fundamental, as middle boxes can induce incorrect or inconsistent response behavior from the perspective of the DNS resolver. Such misbehavior is often masked by DNS resolver implementations that work around path brokenness for the sake of functionality, leaving operators of both recursive and authoritative services potentially unaware of the underlying problems. With a change in version, configuration, or implementation of the resolver, or response content of the authoritative servers, the hidden problems might reveal themselves, yielding outages of sizable impact. This concern is relevant to ongoing discussions of rollover of the DNSSEC root key, which will increase the size of responses for the root zone's DNSKEY set during transition and possibly beyond the rollover, depending on algorithms, key sizes, and number of keys involved in the future root DNSKEY set. The dynamics and observations of the TLDs can yield some insights into the impact of future changes at the root. He uses Ark to instrument DNS measurement from diverse perspectives, in order to establish a baseline of response behavior and quantify current connectivity issues as well as those that might emerge with a root key rollover. He installed DNSViz code on 32 Ark nodes in 27 countries, and ran basic queries (NS/SOA/DNSKEY/DS, NXDOMAIN/NODATA) using multiple network and transport protocols (TCP, UDP, IPv4, IPv6), 4 times per day for 6 days. Results from these preliminary experiments showed the following results: root server communication is generally quick and stable from all instrumented locations; most ccTLD/gTLD servers have reasonable response rates and response times; some (ccTLD) servers are not available from any vantage point; response times from root are generally lower than those from gTLD/ccTLD servers; and median IPv6 response time from ccTLD servers is less than median IPv4 response time. As future work he plans to refine his measurement methods, analyze path similarity between clients and servers, identify EDNS/PMTU issues between clients and servers, and try to quantify the impact of response rate/response time.

Ethan reminded folks of the availability of his BGP PEERING ("Pairing Emulated Experiments with Real Interdomain Network Gateways") testbed, which he built to allow researchers to exchange routes and traffic with real ISPs for research purposes. The PEERING testbed (AS47065) has 9 universities as upstream providers, and peers at AMS-IX with 500 peers including 13 of the 50 largest ISPs (per CAIDA's AS Rank). He would like to provide more support for outside users and experiments, including RPC support to control announcements without BGP, software control of packet processing at routers, and automated deployment of experiments. Separately, he briefly described the effort to build a more industrial strength version of the Reverse Traceroute system he prototyped for his NSDI 2010 paper [7].

Ann Cox gave a brief review of DHS recent and upcoming activities, including the goals of the National Conversation on Homeland Security Technology that occurred in the summer of 2015, which will inform DHS's effort to update its 5-year roadmap and strategic plan for federal investment in cybersecurity R&D.

7. PLATFORM CHALLENGES

In addition to an intense impromptu breakout on potential unified interfaces for BGP measurement infrastructure, we had two deep dives, where a group leader interviewed selected participants to catalog infrastructure challenges. On the third day, the group leaders summarized and reported the results of these conversations.

7.1 Hardware coordination and software robustness

The first was led by Aaron Schulman (Stanford), who interviewed the operators of the main platforms present: Ark, Atlas, BISMark. He classified comments by topic: storage, power supplies, platform hardware.

- **Storage**
  "Minimize writes to SD Cards and flash storage, e.g., on USB nodes). These forms of storage are not durable; they die for unknown reasons."
  "Expect to lose the node in a year if you keep writing to it. It is easier to just upload the data."
  "Avoid using a file system if possible, to minimize writes."
  In the Raspberry Pi node, the SD card seem to be the most common point of failure. Failures are more correlated with manufacturer of SD card than with time in field. Also, unlike Linux, which can detect its file system has been corrupted and diagnose and fix it, there is no warning (e.g., dmsg) that an SD card is failing. Furthermore, SD card slots are not all created equal – in some cases the slots on the Pis make it hard to seat the card well. Storing all data externally is convenient (can yank the disk), but can weaken security if even authentication info is stored that way.
- **Power supplies**: *"Make them easy to replace, or (better) give backups with original device."* Platform operators found an average lifespan for power supply to be 1.5 years. The Ark operators found the custom power supplies for the Pis pretty reliable, although the Pis themselves have voltage issues. As with SD cards, not all power supplies are created equal. The Atlas operators learned that not all USB power supplies can supply enough power, even if rated to do so. Using USB for power avoids the issue entirely, and also avoids dealing with many different plug types in highly global deployments. Another observation was that poor quality phone chargers can lead to SD card corruption.
- **Components** All three platforms considered the fundamental components quite reliable, rarely if ever a need to send any back. Each platform acknowledged the difficulty of supporting older models, perhaps multiple older models, of hardware as newer versions came out. The strongest advice was to pick a platform that can be debugged remotely. There was consensus that timing was an issue as we move toward smaller hardware components for nodes.
- **Deploying software** All projects pre-loaded long-term software, and pulled ephemeral packages from the net and write them to RAM. BISmark used the external flash for common experiments. Pi supports the Puppet package[10], which makes software configuration much easier. Ark stores common measurement software on the drive.

7.2 Back end data processing infrastructure

Steve Bauer (MIT) led the working group on back end data processing infrastructure, using the following questions: What is or is

[10]https://puppetlabs.com/

not working well? What technologies are you considering adopting going forward? How do you capture and share lessons learned?

He also posted a design challenge question: "Imagine that we tried to work together on a simple infrastructure project that say tried to coordinate/automate/trigger a network test on different infrastructures. (The equivalent of astronomers asking each other to point their telescopes at interesting astronomical events.) How might we architect that? What technologies should we use?"

In the "Research ideas → Implementation → Evaluation loop", he observed that we talk to each other a lot about the second transition (evaluating results to trigger more research ideas), but we should find ways to talk to each other about the engineering details more. It is fun, educational, therapeutic.

His highest level takeaway was *"Any piece of technology can work well at small scales, the challenge is scaling up."* The responses reminded him of a distributed systems talk which he cited,[11] but he drilled down to networking-domain specific workloads and challenges in his interviews. Taking the list of common challenges from that talk, he noted the need to balance: simplicity, scalability, performance, reliability, generality, and features. A goal should be to try to anticipate how requirements will evolve, and try to design for scale expansion of 1-2 orders of magnitude. As a method to capture infrastructure engineering how-tos, he emphasized the value of easily searchable non-authoritative engineering notes. It is not necessary to have one wiki that is the right answer; in fact it would create barriers to participation. He also noted that papers from Google/Yahoo/Facebook tend to offer insights that are far from what the academic community will experience. Finally, he noted that many in the community are evaluating hybrid systems – a combination of building vs. buying measurement vantage points, e.g., from cloud platforms.

8. CONCLUSIONS

We review insights shared during our interactive sessions at the beginning of each day:

1. Measurement infrastructure operators were pleased to see the infrastructure used in unexpected ways, and hoped to better support those uses. Some researchers did not know the extent one could actually run experiments on Ark, and the range of experiments possible, so the workshop opened their eyes to new possibilities.
2. Many participants were surprised to learn that even after 20 years of measuring the network, and developing many tools to capture and model network structure, we still have wide open problems in areas we believed to be mostly finished, e.g., geolocation. There was an extended discussion on how to improve the state of IP geolocation in the core.
3. There was a feeling that different measurement systems not only experienced many of the same operational challenges, including hardware and software issues, but were converging on a common measurement set, which made the concept of a unified interface to measurement infrastructures more viable. Many people were eager to use prototypes of such interfaces immediately for their research. Ann (DHS) agreed that it was probably time to fund some level of system integration of data collection, rather than only investing in separate pockets of data collection itself.
4. Similarly, many saw the value of cooperating on data analysis systems as well as measurement systems, such as a collaborative way to store, retrieve (query), and share data, rather than having everyone rolling their own back end data processing infrastructure. The idea of formalizing a service like hijacking or outage detection – where people can see charts in real-time and connect them with the data they've been collecting on their own – appealed to everyone.
5. Because general purpose measurement infrastructure takes so much effort to build and deploy, there is a tendency to leverage other means of gathering data, e.g., Flash ads via Google, which allows massive deployment of measurement. But there was recognition of the need for both approaches to measurement. One potential gold mine was trying to convince the OpenWrt platform developers to integrate network measurements into their platform.
6. (Mentioned at previous AIMS workshops): To promote sustainable infrastructure, participants recognized that in addition to the need for funding infrastructure construction and maintenance, the community needs venues (and thus incentives) for publishing papers on measurement infrastructure experiences and results. Srikanth Sundaresan (ICSI) collated a list of venues that included experience tracks (see appendix).
7. People were stunned to learn they should not be really trusting Stratum 1 time servers. Darryl's proposal to use Ark for testing and evaluating a new Internet timing system was compelling, although daunting in terms of resources required, e.g., GPS clocks attached to as many Ark nodes as possible.

[11] Jeff Dean of Google was not at the workshop, but his slides are at: http://static.googleusercontent.com/media/research.google.com/en/us/people/jeff/stanford-295-talk.pdf

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. Matthew Luckie is now working on reverse traceroute with Ethan's group, as they transition to use Scamper, and possibly eventually integrating Ark.
2. A side discussion among the BGPmon/RIS/BGPStream teams led to planning for a BGP hackathon adjacent to the next AIMS.
3. Ethan first met Italo Cunha at AIMS years ago, and they are now repeat collaborators (Italo is on both PEERING and Sibyl).
4. Phillipa provided Ethan with a use case from her Tor work in which she needs a system like Sibyl.
5. Ethan began talking to the BGPMon and BGPStream folks about integrating with PEERING.
6. Ethan first met Italo Cunha at AIMS years ago, and they are now repeat collaborators (Italo is on both PEERING and Sibyl).

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.

APPENDIX

*Venues to Publish Measurement and Data Processing Infrastructure Research

Srikanth compiled an initial list of potential venues one can submit infrastructure papers or deployment experiences to:

1. NSDI (operations track
2. SIGCOMM started an operations track in 2015.
3. Usenix's Annual Technical Conference has started explicitly soliciting experience reports
4. Surprisingly, IMC does not have a call for deployment experience. In the past, PAM has filled that gap.
5. CCR could be a possible venue, although they do not say anything about infrastructure in their scope.
6. SIGCOMM's HotNets may be a fit for certain papers.
7. International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities (TRIDENT)
8. International Workshop on Computer and Networking Experimental Research Using Testbeds (CNERT)
9. ACM Workshop on Information Sharing and Collaborative Security (WISCS)
10. International Conference on Internet Monitoring and Protection (ICIMP)

A. REFERENCES

[1] Scamper. http://www.caida.org/tools/measurement/scamper/".

[2] *Dolphin: Bulk DNS Resolution Tool*. http://http://www.caida.org/publications/presentations/2014/dolphin_dhs/, 2014.

[3] *ldns*. http://www.nlnetlabs.nl/projects/ldns/, 2014.

[4] *mper*. http://www.caida.org/tools/measurement/mper/", 2015.

[5] CAIDA's Macroscopic Internet Topology Data Kit (ITDK). `http://www.caida.org/data/active/internet-topology-data-kit/`.

[6] Ryan Craven, Robert Beverly, and Mark Allman. Detecting packet header manipulations with HICCUPS. 2013.

[7] Ethan Katz-Bassett. Reverse Traceroute, 2010. `http://research.cs.washington.edu/networking/astronomy/reverse-traceroute.html`.

[8] Vasileios Giotsas, Shi Zhou, Matthew Luckie, and k claffy. Inferring multilateral peering. In *CoNEXT*, December 2013.

[9] B. Huffaker, M. Fomenkov, and k. claffy. DRoP:DNS-based Router Positioning. *ACM SIGCOMM Computer Communication Review (CCR)*, 44(3):6–13, Jul 2014.

[10] Young Hyun. Ark Topology on Demand: scriptable command-line interface for performing IPv4 and IPv6 traceroutes and pings, 2015. `http://www.caida.org/projects/ark/ark.xml`.

[11] Young Hyun. Marinda tuple space, 2015. `http://www.caida.org/tools/utilities/marinda/`.

[12] Julien Ridoux and Darryl Veitch. Principles of Robust Timing over the Internet. *ACM Queue*, 2010.

[13] Matthew Luckie, Rob Beverly, William Brinkmeyer, and kc claffy. Speedtrap: Internet-scale ipv6 alias resolution. Oct 2013.

[14] Matthew Luckie, Amogh Dhamdhere, David Clark, Bradley Huffaker, and K Claffy. Challenges in Measuring Internet Interdomain Congestion. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, 2014.

[15] Y. Vanaubel, P. Mèrindol, J.-J. Pansiot, and B. Donnet. MPLS Under the Microscope: Revealing Actual Transit Path Diversity. *ACM Internet Measurement Conference (IMC).*, 2015.

[16] Young Hyun and Amogh Dhamdhere. Reverse Lookups of IPv4 space, 2015. `http://www.caida.org/data/active/complete_dns_lookups_dataset.xml`.