It’s Time for an Internet-wide Recommitment to Measurement, and Here’s How We Should Do It

Dr. Paul Andrew Vixie
CEO, Farsight Security, Inc.
Woodside, CA, USA

Abstract

There has never been a greater need for comprehensive Internet metrics than now. Even basic security-critical facts about the Internet, such as “How many systems are botted?”[1] or “What networks still don’t do Source Address Validation?” [2] remain murky and poorly quantified. Likewise, traffic characterization and summary inter-AS flow data typically remain closely-held proprietary information,[3] rather than routinely-shared basic operational data. Without trustworthy Internet measurements of this sort, we’re “driving blind” and will routinely make suboptimal choices about critical technical policies, [4] including issues as fundamental as Network Neutrality.

From the beginning, system and network measurements were once an integral part of Internet practice,[5] something hardly surprising given the Internet’s roots in the university community. Scientists naturally make observations, record data, and analyze that data to document phenomena and advance the state-of-the-art.

More recently, however, a variety of factors have created an online environment that’s hostile to legitimate academic Internet measurement and monitoring efforts. Major drivers contributing to that public hostility include overly-aggressive marketing analytics[6] and domestic pervasive monitoring by the intelligence community.[7] It all feels like eavesdropping to the public, even though important real differences exist and reforms[8] have taken place. Bottom line, the public is having none of any of it.

ISPs are also increasingly reluctant to share data. ISPs worry about potential regulatory actions which may be taken based on any data they voluntarily proffer, so they defensively resort to sharing little by default. From the regulators’ perspective, that closed-mouth ISP posture spurs calls for compulsory reporting (and those threats reinforce ISP worries). It’s a vicious circle.

What of data collected by commercial cyber security companies? Data that might once have been freely shared is now all-too-often a critical company asset, hard won at great cost, and a key differentiator in a highly competitive market. Reluctance to give that data away is both rational and understandable, yet researcher frustration over the existence of academically-inaccessible data is equally easy to empathize with.

Other factors are purely technical. Deployment of encryption continues to increase,[9] rendering a growing volume of network traffic opaque to end-to-end observation. NAT/PAT is hiding the edge of the network. Core network speeds are escalating making it more difficult to conveniently collect un-sampled data with high fidelity. Automated protective systems conflate active scans by Internet researchers with pre-attack reconnaissance efforts by cyber criminals, and active throughput tests with volumetric DDoS attacks, squelching it all alike. What’s a researcher to do?

Some sources of technical “blindness” may be inherent in evolving networks. Mobile devices relying on “closed” cellular and 4G networks are playing an increasingly significant role online, but wireless and wireline phone companies have a genetically-inherited reluctance to engage in non-mandated network monitoring, a trait traceable to ECPA and Bell-head traditions. Expansion of the Internet in the southern hemisphere and in former totalitarian countries continues to occur, and that’s terrific -- except that there may be a limited tradition of academic network instrumentation in those regions.

All these phenomena and more have largely served to damp forward progress on Internet measurement. The result is that the Internet is increasingly becoming as little-measured as the most remote galaxies of our universe. We must fix that. We must recommit to an Internet-wide comprehensive program of measurement activities. There are many unmet needs when it comes to Internet measurement at this time, including:

- We need a comprehensive review of what’s currently being collected, with the identification of gaps and opportunities for synergistic collaboration
- We should consider creation of a volunteer citizen Internet measurement corps, kin to the Citizen Weather

WTMC’16, May 30, 2016, Xi’an, China.

ACM 978-1-4503-4284-1/16/05.

DOI: http://dx.doi.org/10.1145/2903185.2903186

Observer Program (CWOP),[10] or the SETI@home effort at UCB.[11]

- If we want the public to embrace Internet measurement activities, they will need to be made aware of its importance, and the potential role that the public can play in collecting and reporting data using standardized tools.
- There will need to be a "Chinese wall" between academic network measurement efforts and any national security or law enforcement-related data users, reassuring potential measurement contributors as to what's being collected, by whom, and for what purposes.
- Federal funding for global collection of Internet telemetry is essential and must be stable to allow for long term longitudinal data collection efforts. [12]
- The community needs to move from batch-oriented data collection and analysis approaches (worthy of the dark ages of computing!) to a real-time continuous processing paradigm. These days, data from hours ago may as well be data from years ago.
- ISPs and others, such as mobile carriers and operating system vendors, are critical to our efforts. They need statutory protection so that they can share measurement data without worrying that what they share will be used against their business interests, or result in penalties for "privacy violations."[13]
- ISPs and others who agree to contribute data should receive tax relief, in recognition of the fact that measurement data is economically valuable, and the process of collecting that data isn't cost-free.
- To have a holistic picture of all parts of the Internet, we need systematic outreach to academic institutions and network operators, particularly in under-represented and developing regions of the Internet, such as much of the Southern Hemisphere.

Time will be reserved for Q&A at the end of the talk.

Short Bio

Dr. Paul Vixie is the CEO of Farsight Security. He previously served as President, Chairman and Founder of Internet Systems Consortium (ISC), as President of MAPS, PAIX and MIBH, as CTO of Abovenet/MFN, and on the boards of several for-profit and non-profit companies. He served on the ARIN Board of Trustees from 2005 to 2013, as ARIN Chairman in 2008 and 2009, and was a founding member of ICANN Root Server System Advisory Committee (RSSAC) and ICANN Security and Stability Advisory Committee (SSAC). He operated the ISC's F-Root name server for many years, and is a member of Cogent's C-Root team. He is a sysadmin for Op-Sec-Trust.

Vixie has been contributing to Internet protocols and UNIX systems as a protocol designer and software architect since 1980. He wrote Cron (for BSD and Linux), and is considered the primary author and technical architect of BIND 4.9 and BIND 8, and he hired many of the people who wrote BIND 9. He has authored or co-authored a dozen or so RFCs, mostly on DNS and related topics, and of Sendmail: Theory and Practice (Digital Press, 1994). His technical contributions include DNS Response Rate Limiting (RRL), DNS Response Policy Zones (RPZ), and Network Telemetry Capture (NCAP). He earned his Ph.D. from Keio University for work related to DNS and DNSSEC, and was named to the Internet Hall of Fame in 2014.

[1] "[...] another key barrier to effective Code participation is the current inability to uniformly measure the bot population and the results of activities to reduce bots. Without consistent objective agreed upon industry based measurements, ISPs may find it difficult or impossible to tell the extent of the bot problem on their network, and if so, whether efforts to correct it will have, or have had, any material effect." https://transition.fcc.gov/bureaus/pshs/advisory/csric3/CSRIC_III_WG7_Report_March_%202013.pdf at section 3.2 (see also Appendix 4).

[2] http://www.redbarn.org/internet/save

[3] "Many network operators do not collect or store data of the type required for such a study, and many more regard it as proprietary or covered by privacy legislation with provisions such that no researcher is ever likely to see it. So we can see that study of the inter-domain matrix is likely to be a long-term, and rather challenging project," *Internet Traffic Matrices: A Primer*, Paul Tune and Mathew Roughan, http://sigcomm.org/education/ebook/SIGCOMMeBook2013v1_chapter3.pdf at page 46.

[4] "The lack of consistent, meaningful, and widely-applicable baseline metrics to assess the current national cybersecurity posture is a major challenge to improving cybersecurity practices...", http://csrc.nist.gov/cyberframework/rfi_comments/
041213_fcc_pshsb.pdf at PDF page 6.

[5] See for example Kleinrock's Network Measurement Center in *The ARPANET: The First Decade*, http://www.dtic.mil/get-tr-doc/pdf?Location=U2&doc=GetTRDoc.pdf&AD=ADA115440 at pp. III-39.

[6] A telling indicator: the most-downloaded add-on for Firefox is AdBlock Plus (see https://addons.mozilla.org/en-US/firefox/extensions/?sort=users)

[7] *U.S. Voters Say Snowden Is Whistle-Blower, Not Traitor, Quinnipiac University National Poll Finds; Big Shift On Civil Liberties vs. Counter-Terrorism,* http://www.quinnipiac.edu/news-and-events/quinnipiac-university-poll/national/release-detail?ReleaseID=1919

[8] *Statement by the President on the USA Freedom Act,* https://www.whitehouse.gov/the-press-office/2015/06/02/statement-president-usa-freedom-act

[9] At the time this abstract was prepared, just to provide one sample metric, 83% of Gmail's outbound email traffic was encrypted in transit. See https://www.google.com/transparencyreport/saferemail/

[10] http://www.srh.noaa.gov/epz/?n=cwopepz

[11] http://setiathome.ssl.berkeley.edu/

[12] *The Federal Networking and Information Technology Research and Development Program: Background, Funding, and Activities*, https://www.fas.org/sgp/crs/misc/RL33586.pdf , outlining Federal networking funding, shows inadequate support for focused work on Internet metrics.

[13] See for example *Chairman Wheeler's Proposal to Give Broadband Consumers Increased Choice, Transparency & Security With Respect to Their Data*, http://transition.fcc.gov/Daily_Releases/Daily_Business/2016/db0310/DOC-338159A1.pdf