"A Day in the Life of the Internet": Proposed community-wide experiment

kc claffy
www.caida.org

In 2001 the U.S. National Academy of Sciences convened a workshop to assess the state of networking research, and, in pursuit of objectivity and fresh insights, arranged for more than half of the attendees to be from other fields, in this case computer science. Among the most memorable conclusions:

.. the outsiders expressed the view that the network research community should not devote all or even the majority of its time to fixing current Internet problems.

Instead, networking research should more aggressively seek to develop new ideas and approaches. A program that does this would be centered on the three M's - measurement of the Internet, modeling of the Internet, and making disruptive prototypes. These elements can be summarized as follows:

Measuring - *The Internet lacks the means to perform comprehensive measurement on activity in the network. Better information on the network would provide the basis for uncovering trends, as a baseline for understanding the implications of introducing new ideas into the network, and would help drive simulations that could be used for designing new architectures and protocols. This report challenges the research community to develop the means to capture a day in the life of the Internet to provide such information.*

Modeling - *The community lacks an adequate theoretical basis for understanding many pressing problems such as network robustness and manageability. A more fundamental understanding of these important problems requires new theoretical foundations - ways of reasoning about these problems that are rooted in realistic assumptions. Also, advances are needed if we are to successfully model the full range of behaviors displayed in real-life, large-scale networks.*

Making disruptive prototypes - *To encourage thinking that is unconstrained by the current Internet, Plan B approaches should be pursued that begin with a clean slate and only later (if warranted) consider migration from current technology. A number of disruptive design ideas and an implementation strategy for testing them are described in Chapter 4.*

National Academies Press, *"Looking over the Fence at Networks: A Neighbor's View of Networking Research (2001)"* [1]

Per the above - "This report challenges the research community to develop the means to capture a day in the life of the Internet" - we admit that the research community has not come anywhere near this goal, nor does it seem a priority. We seek to open a discussion on what it would mean, require, and cost to capture a day in the life of the Internet with as much scientifically grounded methodology as possible, and with resulting data as widely accessible as possible. We recognize that the proposed project will involve building a cooperative community to support the simultaneous capture of a variety of measurements from and across many strategic links around the globe for further analysis by research scientists. But by establishing a periodic tradition of synchronized measurements, and supporting tools, analysis, visualization, and data catalog (DatCat, http://www.datcat.org [2], Internet Traffic Archive [3], CRAWDAD [4], MOME [5], Datapository [6], PREDICT [7]) in which to index collected traces, we hope to significantly increase the quantity as well as quality of empirical data supporting Internet research.

Several complementary projects at CAIDA provide the impetus for our first attempt to coordinate a distributed measurement activity in late 2006. As part of an NSF-sponsored DNS measurement project (http://www.caida.org/funding/dns-itr/ [8]), CAIDA and ISC plan to perform a 48-hour simultaneous measurement event on dozens of root server anycast nodes. Specifically, ISC will collect packet header traces from multiple (hopefully all) anycast instances of at least three root nameservers, based on feedback from the previous such measurement experiment. (http://www.caida.org/research/dns/dns-itr/resources/dnsroot_measurement_recommendations.xml [9]) Since to our knowledge this event will be the largest scale simultaneous collection from a core component of the global Internet infrastructure, we consider it an ideal time to prototype a "Day in the Life of the Internet" measurement event. Specifically, if you have access to or influence over Internet measurement infrastructure and can contribute datasets (anonymized according to your needs [10, 11]), please email ditl-info@caida.org for details regarding already planned measurement dates, times, locations, and types of data. (There will be an informal vetting process to avoid manipulation of the experiment.)

We also seek input from others interested in gathering specific complementary measurements on the same days, to help us maximize the return on investment of participation in the experiment.

Commercial pressures make it next to impossible to get Internet measurement data to the research community, but empirical network science is not possible without such data. We hope that over time, annual measurement activities to support "day in the life of the Internet" (DITL) data sets will gather increasing momentum.

Ideally, participating partners would provide simultaneous capture of a variety of trace data: workload, topology, routing, and performance, from a large number of strategic locations around the globe, anonymized appropriately according to local restrictions [12].

We recognize this project will involve global efforts to overcome logistic, technical, economic, and legal obstacles to measurement and data sharing. But through this activity we seek to determine whether, given enough interest in the community, it is possible to gather sufficient data not only to capture salient characteristics of 'a day in the life of the Internet,' but also to provide sufficient empirical grounding for the development of reliable predictive models of Internet traffic, topology, routing, and evolution.

Related References

[1] "Looking over the Fence at Networks: A Neighbor's View of Networking Research (2001)," National Academies Press, http://www.nap.edu/books/0309076137/html/

[2] Internet Measurement Data Catalog (DatCat), http://www.datcat.org/

[3] "The Internet Traffic Archive", http://ita.ee.lbl.gov/index.html

[4] "Community Resource for Archiving Wireless Data at Dartmouth", http://crawdad.cs.dartmouth.edu/

[5] "Cluster of European Projects aimed at Monitoring and Measurement -- MoMe Database", http://www.ist-mome.org/database/

[6] "The Datapository: A collaborative network data analysis and storage facility", http://www.datapository.net/

[7] "Protected Repository of Data for Internet CyberThreats" http://www.predict.org/

[8] "Improving the Integrity of Domain Name System (DNS) Monitoring and Protection" (NSF grant SCI-0427144)", http://www.caida.org/funding/dns-itr/

[9] "Recommendations for future large scale simultaneous DNS data collections", http://www.caida.org/research/dns/dns-itr/resources/dnsroot_measurement_recommendations.xml

[10] "Crypto-PAn: Cryptography-based Prefix-preserving ANonymization", http://www.static.cc.gatech.edu/computing/Telecomm/cryptopan/

[11] "The Devil and Packet Trace Anonymization", http://www.icir.org/enterprise-tracing/papers.html

[12] "Toward community-oriented network measurement infrastructure", http://www.caida.org/funding/cri/