DITL: Collecting Global Empirical Data to Support Internet Research

kc claffy, Min Zhang, Wolfgang John

I. Introduction

This poster[1] aims to promote the DITL project, which challenges the research community to develop the means to capture a complete *Day In The Life* of the Internet to gain a deeper understanding of Internet properties. Our tactical goal is to establish a tradition of periodic, synchronized measurements, and to provide supporting tools, analysis, visualization, and data indexes [1], [2]. Our strategic goal is to significantly increase the quantity, quality, and accessibility of empirical data supporting Internet research, helping to answer many open research questions regarding Internet workload, topology, routing, and performance [3].

Commercial pressures make it next to impossible to get many types of Internet measurement data to the research community, but empirical network science is not possible without such data. The proposed project involves building a global cooperative community to support the simultaneous capture of measurements from and across many strategic links around the globe for further analysis by researchers. Participating partners provide a variety of trace data with privacy-sensitive techniques for anonymization, aggregation, or analysis [4]. If you have access to or influence over Internet measurement infrastructure and can contribute datasets (anonymized according to your needs [5]), please email *ditl-info@caida.org* for details regarding already planned measurement dates, times, locations, and types of data. Questions about the Internet are also welcome.

II. Status of DITL

Several projects at CAIDA provided the impetus for our first attempts to coordinate large-scale, distributed measurement activities in 2006. To our knowledge, these annual events stand as the largest scale simultaneous collections from critical components of the global Internet infrastructure made available to academic researchers. We consider these events prototypes for eventual regular *Day in the Life of the Internet* measurement events. Since the first DITL attempt in 2006, more participants [6] from around the world have joined each year, extending the scope of this community collection experiment from its initial focus on DNS root servers, to include data from other authoritative name servers, as well as commercial and academic networks. Data types have included pcap files (packet traces), netflow files, topology data, web cache (squid) logs, and BGP data.

III. Case study: DNS root measurements

While our previous DITL measurements can be considered prototypes of a *Day in the Life of the Internet*, our goal was narrower: to collect as complete a dataset as possible about the DNS root servers operations and evolution, particularly as they deployed new technologies, e.g., anycast. Currently, the DNS root zone is facing further dramatic changes: cryptographically signing the root zone with DNSSEC, deploying Internationalized Top-Level Domain (TLD) Names (IDNs), and addition of other new global Top Level Domains (TLDs) are all planned for the next 12-24 months. As a result, there is growing interest in measurement, testing, and provisioning for foreseen (or unforeseen) complications.

We have undertaken four annual large-scale data collection events. In 2006, three DNS root servers participated in the event; five in 2007, eight in 2008 and 2009, and hopefully twelve in 2010. Since 2008, TLD servers and open/caching resolvers have also started to contribute data. The results [7] serve as a baseline for future measurements of this critical infrastructure. For example, DITL measurements revealed disquieting news for the impending transition to DNSSEC, by showing a decreasing trend in the fraction of DNSSEC-capable clients (Fig. 1). We hope that annual measurement activities to support scientific study of macroscopic questions about the Internet will gather increasing momentum, including expanded partnerships with public and private sector stakeholders, as well as legal and policy expertise to advise and review privacy-protecting data disclosure control mechanisms [4].

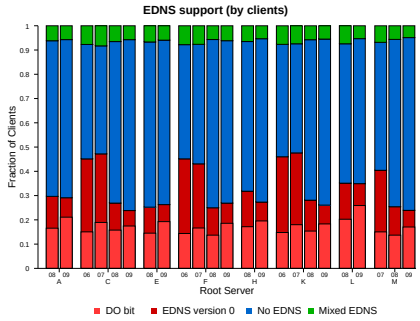

Fig. 1. Decrease in EDNS support measured by clients.

References

[1] C. Shannon, D. Moore, K. Keys, M. Fomenkov, B. Huffaker, and k claffy, "The Internet Measurement Data Catalog," *SIGCOMM CCR*, 2005.

[2] "PREDICT: Protected Repository for the Defense of Infrastructure against Cyber Threats," https://www.predict.org/ (accessed 2010.02.01).

[3] "What Researchers Would Like to Learn from the DITL Project," http://www.caida.org/projects/ditl/questions/ (accessed 2010.02.01).

[4] E. Kenneally and K. Claffy, "Dialing Privacy and Utility: A Proposed Data-sharing Framework to Advance Internet Research," *IEEE Security and Privacy (S&P) special issue on Privacy-Preserving Sharing of Sensitive Information, http://www.caida.org/publications/papers/2009/dialing_privacy_utility/.*, 2009.

[5] "Data Anonymization," http://www.caida.org/data/anonymization/ (accessed 2010.02.01).

[6] "A Day in the Life of the Internet - DITL," http://www.caida.org/projects/ditl/ (accessed 2010.02.01).

[7] "DNS Root Traffic: Interactive Graphs," http://www.caida.org/research/dns/roottraffic/evolution/interactive-graphs/ (accessed 2010.02.01).

[1]We plan to present a poster without demo. Min Zhang and Wolfgang John are PhD students.