Guest Editorial Measuring and Troubleshooting the Internet: Algorithms, Tools and Applications

Maurizio Dusi, Alessandro Finamore, Kimberly Claffy, Nevil Brownlee, and Darryl Veitch, *Fellow, IEEE*

THE ubiquity of Internet access and the wide variety of Internet-enabled devices and applications have made the Internet a principal pillar of the Information Society. Decentralized and diverse, the Internet is resilient and universal. However, its distributed nature leads to operational brittleness and difficulty in identifying and tracking the root causes of performance and availability issues.

We started this special issue with the idea of fostering original contributions which offer methods and tools for measuring and troubleshooting the current Internet at any level: the problem of collecting information from a variety of probes; the correlation of measurements from the application and network layers; and the need to gain insight on both network state and application performance.

The challenges include: the definition and deployment of distributed infrastructure able to monitor the health of the network in real-time and the implementation of scalable algorithms and mechanisms to address that issue; and methodologies for correlating the behavior of the network jointly with the behavior of the applications being exchanged on top of it, so that anomalies and network malfunctions can be identified or predicted, and properly addressed.

The editorial board has been designed to include leading academic and industry researchers in the area. The academics have industry experience and ongoing collaborations. This is important given the intense industry interest and activity on the topics. The board also includes broad geographic diversity and over a decade of experience in the area.

We believe the response of the research community has been very intriguing, with many original ideas that seriously contribute to advancing the current state of the art. We received a total of forty-nine submissions, out of which we selected nine works.

The first three contributions propose mechanisms for enabling scalable and real-time network monitoring on high-speed networks, and for orchestrating the collection of data from large population of probes.

In particular, the authors of the paper "Selective capping of packet payloads at multi-Gb/s rates" discuss an original idea to the practical problem of how to selectively capture packet payloads in high-speed networks, and retain the traffic which is most important for traffic analysis, while reducing write speed and storage requirements.

The authors of "Network packet processing with PFQ" tackle the problem of traffic collection by identifying a multi-core assisted framework for processing raw packets on the wire. They choose an approach based on functional programming, and effectively address the interesting problem of defining high-level primitives in the packet processing domain.

In "AMON: An open source architecture for online monitoring, statistical analysis and forensics of multi-gigabit streams" the authors introduce a system to compute traffic summaries at high-speed links in real-time, and discuss a series of statistical analysis for detecting anomalies that are common in networks, such as heavy hitters and DDoS attacks. The system is able to capture packets at 1Gbps+ speed and generate traffic summaries that are of common interest to network operators. Moreover, the authors make their implementation available as open source to the research community.

The next four selected contributions are a step forward towards traffic classification and topology discovery tools for monitoring the status of the network and the availability of the resources of the Internet at a scale.

The authors of "HEAP: Reliable assessment of BGP hijacking attacks" first discuss a model for the Internet routing and hijacking attacks, and then propose a framework for the validation of route hijacks alerts raised by third party systems. They do this by correlating inputs from Internet Route Registries, information about Internet topology, and measurements of SSL/TLS keys from a geographically distributed set of vantage points.

In "Lost in space: Improving inference of IPv4 address space utilization" the authors offer a study on the usage of the IPv4 address space, and investigate on how well the allocated addresses are actually used. The interesting aspect of this study is the correlation of measurements taken from different locations and types, different probing tools (for active measurements) and using different approaches, e.g., active and passive.

The authors of "A high-performance, scalable infrastructure for large-scale active DNS measurements" describe a measurement infrastructure aimed at measuring the DNS TLD

M. Dusi is with Copan Diagnostics Inc., Murrieta, CA 92562 USA (e-mail: maurizio.dusi@copangroup.com).

A. Finamore is with Telefonica Research, Distrito Telefonica, Madrid 28050, Spain (e-mail: alessandro.finamore@telefonica.com).

K. Claffy is with the Center for Applied Internet Data Analysis, San Diego Supercomputer Center, La Jolla, CA 92093-0505 USA (e-mail: kc@caida.org).

N. Brownlee is with the Department of Computer Science, The University of Auckland, Auckland 1010, New Zealand (e-mail: n.brownlee@auckland.ac.nz).

D. Veitch is with the School of Computing and Communications, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: darryl.veitch@uts.edu.au).

Digital Object Identifier 10.1109/JSAC.2016.2559239

domains in a scalable way. The authors effectively describe the related issues in terms of sampling the TLDs, the scalability of measurements, and the approach the authors used for the storage and analysis of the data. They do so by focusing on the applicability of their method to two interesting use cases.

An interesting implementation of a large-scale system is described in the paper "Latency-based anycast geolocalization: algorithms, software and datasets". In this work the authors point out the need to find locations for anycast instances, and introduce iGreedy, a lightweight, protocol-agnostic technique for doing that using "only a handful of latency measurements". The method is validated in the context of DNS servers and CDN servers, and advances the state-of-the-art in delay-based geolocation. The paper also offers a great survey on the subject.

The final part of this special issue is dedicated to the key topic of optimizing network resources for improving the QoS and QoE of users, as well as implementing network optimization and management functions.

In particular, the authors of "The power of SDN to enforce the traffic matrix estimation of an ISP" address the problem of how to use the measurement capabilities of available SDN nodes in a resource-aware way to improve the measurement of traffic matrices. The main novelty with this work is the use of the worse-case per-flow spread bounds as a means to prioritise the flows to be targeted for measurements.

Finally, the authors of "iRate: Initial video bitrate selection system for HTTP streaming" address the problem of predicting the best initial video bit-rate to select for HTTP video streaming, based on active traffic measurements. Through the system they proposed, the video stream can start at the video quality level which best fits the current network conditions. This work addresses a timely problem and industries could benefit from the solution offered by the authors.

The contributions included in this special issue offer original angles on different topics related to network measurement and troubleshooting at a scale, and result in always having practical implications on the industry. However, several challenges have yet to be investigated, such as identifying policy for different entities and domains (e.g., telecom operators, content providers) to co-operate and exchange measurements; identifying metrics which allow for effective troubleshooting while avoiding revealing sensitive information; conducting network planning and operating adaptive network management strategies to address both episodic malfunctioning and disaster recovery.

In particular, areas where the editors of this special issue see potential for novel contributions include: privacy-preserving schema for regulating and securing the exchange of information among different Internet entities and domains; novel visualization approaches to detect and diagnose network performance issues; novel testbeds for distributed performance measurements; and system architecture and protocol design for exchanging measurements and enhancing the collaboration among domains.

We hope that through this selection of works we can stimulate new discussions and new contributions to the interesting problem of monitoring and troubleshooting the Internet.

We hope you enjoy this special issue and we look forward to seeing your contribution to these topics!

Maurizio Dusi received the B.Sc. degree in information and communications technology and the M.Sc. and Ph.D. degrees in telecommunications from the University of Brescia, Italy. He was a Research Scientist with NEC Laboratories Europe, where he was involved in mining data from the Internet. He is currently part of the Research and Development Team with Copan Diagnostics Inc., where he is involved in scalable algorithms and mechanisms for the collection and the analysis of microbiological data. His research interests include real-time communications monitoring, data mining, pattern recognition, and modeling for data analysis and data characterization.

Alessandro Finamore received the Ph.D. degree in electronics and telecommunication engineering from the Politecnico di Torino, in 2012. He is currently a Researcher at Telefonica Research. He has co-authored over 25 publications (including a Best Short Paper Award at ACM CoNEXT'13). His current research interests include Internet measurements, big data technologies, video streaming, and CDN services. He served as a Reviewer for major journals (e.g., the IEEE TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, and *Computer Networks*) and conferences/workshops (e.g., IEEE TMA and ACM SAC).

Kimberly Claffy received the Ph.D. degree in computer science from the University of California at San Diego (UCSD), in 1994. She is currently a Principal Investigator for the Center for Applied Internet Data Analysis, and a Resident Research Scientist with the Supercomputer Center, UCSD. Her research interests include Internet workload/performance data collection, analysis, and visualization, particularly with respect to commercial ISP collaboration/cooperation and sharing of analysis resources.

Nevil Brownlee has been the Co-Chair of the IETF's IPFIX Working Group since it began in 2000. Since 2010, he has been the RFC Editor for Independent Submissions. He is an Associate Professor of Computer Science with The University of Auckland. His research interests include measurement and analysis of Internet traffic flows, the domain name system, unsolicited traffic, and understanding patterns in traceroute data.

Darryl Veitch (M'02–SM'02–F'10) received the Ph.D. degree in dynamical systems from the University of Cambridge, U.K. He is currently a Professor at the School of Computing and Communications, University of Sydney, Australia. His research interests include statistical aspects of network modeling and analysis, in particular network tomography and data sampling and sketching. His other interests include clock synchronization over networks and information theoretic security.