Towards an Observatory for Network Transparency Research

Stephan Neuhaus, Roman Müntener
Zurich University of Applied Sciences
School of Engineering
Switzerland
first.last@zhaw.ch

Korian Edeline, Benoit Donnet
Université de Liège
Montefiore Institute
Belgium
first.last@ulg.ac.be

Elio Gubser
ETH Zürich
Networked Systems Group
Switzerland
egubser@ee.ethz.ch

ABSTRACT

The Internet is full of middleboxes that change packets and flows. In fact, there is probably no IP or TCP header that is not affected by at least one middlebox. Obviously, middleboxes impede path transparency, i.e., the idea that an exchange of messages results in more or less the same packets, no matter what path the packets takes. But no one seems to have a truly global view of what middleboxes do to packets on what Internet paths, which would however be an essential knowledge for new transport protocols to be successfully deployed.

We address these concerns in the MAMI project by building an observatory of path transparency measurements. The project hosts an extensive set of path transparency measurements — we believe it to be the first dataset to deal specifically with middlebox involvement.

In this paper, we describe that Observatory and a number of questions that we want to address with the data in that Observatory. Eventually, the project will provide public access to that Observatory so that researchers and the interested public can ask their own questions about path transparency issues and middlebox involvement.

CCS Concepts

•**Networks** → **Middle boxes / network appliances; Network measurement;** *Network protocol design;*

Keywords

Network measurement, public observatory, middleboxes, tracebox

1. INTRODUCTION

The public Internet is very different from what its inventors had intended it to be. Instead of having intelligent end nodes and dumb packet-forwarding routers, the situation today is almost exactly reversed: from a networking point of view, end nodes are quite simple whereas middleboxes are very sophisticated, forwarding packets, doing Network Address Translation, balancing server load, transcoding videos depending on current throughput, and so on. This destroys a condition called *path transparency*, which means that an exchange of messages results in the same packets, no matter what path the packets take [14, 9].

Since these middleboxes need at least some knowledge of the underlying protocols in order to function, there has been transport protocol ossification: if it's not TCP or UDP, middleboxes cannot look into the packets and such packets might in the worst case be dropped on the floor.

However, today, no one seems to have a truly global view of what middleboxes do to packets on what Internet paths, which would however be essential knowledge for new transport protocols to be successfully deployed.

The MAMI project, and its Observatory, can be seen as an attempt to ask "What will break on the Internet if I do this?", where "this" is something that *ought* to work, but might not, due to middlebox involvement.

Path transparency has been discussed for some time now in the Internet measurement community in general and in the Internet Architecture Board in particular, and MAMI is an outgrowth of these discussions. For more information, see especially the work by Trammell, Kühlewind and others [16, 17, 18].

2. RELATED WORK

Of course, we are not the first network measurement infrastructure that can be queried. Prominent examples include infrastructures like RIPE Atlas [12] and RIPE TTM [8] (discontinued in 2014) for connectivity and reachability research, TopHat (OneLab/PlanetLab) [3] and Archipelago [4] for topology and iPlane [10] for path performance prediction. In addition, some infrastructures allow querying specifically for large-scale data analysis. One prominent example of this is mPlane [15]. MONROE is a project about (among other things) "large-scale monitoring and assessment of performance of [Mobile Broadband] (MBB) networks in heterogeneous environments". MONROE builds a "dedicated infrastructure for measuring and experimenting in MBB and WiFi networks, comprising both fixed and mobile nodes distributed over Norway, Sweden, Spain and Italy." [1] MAMI uses the MONROE infrastructure to provide it with vantage

ANRW '16, July 16, 2016, Berlin, Germany

ISBN 978-1-4503-4443-2/16/07…$15.00

DOI: http://dx.doi.org/10.1145/2959424.2959425

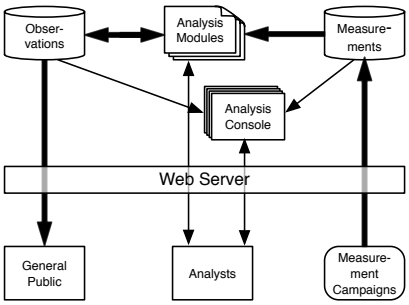

Figure 1: MAMI Observatory Architecture.

points for running measurements.

As far as we know, the observatory we build in MAMI is the very first one entirely dedicated to middleboxes data collection.

3. OBSERVATORY

As outlined above, MAMI is a project about path transparency, and thus the purpose of the MAMI observatory is to enable research on boolean statements about paths, called path conditions. Path transparency now means that the same path conditions hold on every path that leads to the same endpoint.

The observatory is designed to store a large number of raw measurements from different measurement campaigns. Analysis modules then extract observations from this data. An observation represents a statement about a given path condition at the time of measurement. Transforming measurements into observations allows us to separate low-level data handling and high-level analysis.

The observatory architecture is depicted in Fig. 1; the thickness of the arrows represent the volume of the data flow. All access to the data management infrastructure is mediated through a web server (nginx). Means of uploading data might change in the future to allow for bulk file import. Measurements are stored in an HDFS node, observations are stored in a NoSQL database (MongoDB).

Access to this data falls into three categories: (*i*) **Measurement Campaigns.** These may be people or machines. They are equipped with tokens that allow them to upload raw measurements to the measurement infrastructure. Measurements are stored in an HDFS node, measurement metadata is stored in a MongoDB database. (*ii*) **Analysts.** These are people who have access to a Jupyterhub[11] running on the infrastructure. This is responsible for spawning appropriate Jupyter instances. Jupyter is a web application providing an interactive python shell which allows users to run code on the observatory infrastructure as well as saving code for later editing or reruns. Analysts use Jupyter for exploratory analysis of both measurements and observations. Analysts write analysis modules that turn measurements and other observations into more observations. (*iii*) **General Public.** Members of the general public can access some observational data through specially vetted analysis modules. (The reason why the general public cannot access the raw measurement data is that this data potentially contains personally identifiable information.)

The part holding the measurements is currently implemented using HDFS, but this may well change in the future, since HDFS may not afford the guarantees that we would like for measurements.

The part holding the measurement metadata and the observations is currently in a MongoDB NoSQL database. We do not foresee this changing in the foreseeable future, as we are quite happy with MongoDB.

The analysis modules can be written in any language that supports reading from HDFS and writing to MongoDB. No specific analysis tool is prescribed for processing the data but we provide access to Apache Spark™ [2] infrastructure and may add more distributed computing tools later. Whenever new measurements arrive in the measurement database, appropriate analysis modules will be triggered that transform measurements into observations. This may trigger other analysis modules that perform deeper analysis and create new observations, and so on.

Analysis modules will be invoked and coordinated by a supervisor. This provides the modules the required information to access the infrastructure, and sets up temporary collections in the database for the modules' output. The analysis consoles are implemented using Jupyterhub, a multi-user Jupyter server.

4. USE CASES

4.1 Internet-Over-UDP

One issue with the prevalence of middleboxes on the Internet is that, in order to perform their function well, they have to do fairly deep packet inspection. For example, even a NAT box will have to alter IP and TCP or UDP packet headers, as does a load balancer or SSL termination proxy. A middlebox that transcodes videos on the fly would have to look even deeper into packets.

This creates a powerful incentive *not* to invent and deploy new transport protocols, since the middleboxes would not understand them, would not know what to do with them, and would probably simply have to drop them. This is known as "Internet transport ossification". There is in fact a need for new transport protocols, the most prominent example being QUIC [13] as a replacement for HTTP [7].

As can be seen in the case of QUIC, the most promising protocol to base one's new transport on is UDP. But will that work, or will middleboxes, being unaware what transport is routed on top of UDP, drop or otherwise mangle packets? Or will overly restrictive firewalls perhaps disallow UDP except for a few standard ports like 53 or 123? We are trying to answer the question "What will break if we try to run the Internet on UDP" in MAMI with the help of a large dataset obtained with `copycat` [6], a tool for comparing TCP loss, latency, and throughput with UDP by generating TCP-shaped traffic with UDP headers.

4.2 Presence of Middleboxes over Time

Transport protocols might reasonably make assumptions about the composition of middleboxes on the path from a source to a destination. For example, a transport protocol might assume that a middlebox that was on the path five minutes ago will still be on that path now. But there seems to be no empirical evidence either way. The question is thus: how dynamic is today's Internet? Do the same middleboxes appear on the same paths all the time, or do middleboxes routinely appear and disappear? We are using `tracebox` (a `traceroute` extension revealing middleboxes along a path) data [5] to answer that question.

5. ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688421, and was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0268. The opinions expressed and arguments employed reflect only the authors' views. The European Commission is not responsible for any use that may be made of that information. Further, the opinions expressed and arguments employed herein do not necessarily reflect the official views of the Swiss Government.

6. REFERENCES

[1] Ö. Alay, A. Lutu, D. Ros, R. Garcia, V. Mancuso, A. F. Hansen, A. Brunstrom, M. A. Marsan, and H. Lonsethagen. MONROE: Measuring mobile broadband networks in Europe. In *Proceesings of the IRTF & ISOC Workshop on Research and Applications of Internet Measurements (RAIM) 2015*. Internet Society, Oct. 2015.

[2] Apache Software Foundation. Apache spark™, 2016. See http://spark.apache.org/.

[3] T. Bourgeau, J. Augé, and T. Friedman. TopHat: Supporting experiments through measurement infrastructure federation. In *Proc. International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (TridentCom)*, May 2010.

[4] k. claffy, Y. Hyun, K. Keys, M. Fomenkov, and D. Krioukov. Internet mapping: from art to science. In *Proc. IEEE Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH)*, Mar. 2009.

[5] G. Detal, B. Hesmans, O. Bonaventure, Y. Vanaubel, and B. Donnet. Revealing middlebox interference with tracebox. In *Proc. ACM Internet Measurement Conference (IMC)*, Oct. 2013.

[6] K. Edeline. `copycat`, May 2016. See http://queen.run.montefiore.ulg.ac.be/~edeline/copycat/.

[7] R. T. Fielding and J. F. Reschke. Hypertext transfer protocol (HTTP/1.1): Message syntax and routing. RFC 7230, RFC Editor, June 2014.

[8] F. Georgatos, F. Gruber, D. Karrenberg, M. Santcroos, A. Susanj, H. Uijterwaal, and R. Wilhelm. Providing active measurements as a regular service for ISPs. In *Proc. Passive and Active Measurement Workshop (PAM)*, Apr. 2001.

[9] M. Honda, Y. Nishida, C. Raiciu, A. Greenhalgh, M. Handley, and H. Tokuda. Is it still possible to extend TCP. In *Proc. ACM Internet Measurement Conference (IMC)*, November 2011.

[10] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iPlane: An information plane for distributed services. In *Proc. USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, Nov. 2006.

[11] Project Jupyter. Jupyterhub, 2016. See http://jupyter.org/index.html.

[12] RIPE NCC. Atlas, 2010. See https://atlas.ripe.net/.

[13] J. Roskind. Multiplexed stream transport over UDP. https://docs.google.com/document/d/1RNHkx_VvKWyWg6Lr8SZ-saqsQx7rFV-ev2jRFUoVD34/preview?sle=true#! Warning! Only retrievable with JavaScript and cookies enabled.

[14] J. Sherry, S. Hasan, C. Scott, A. Krishnamurthy, S. Ratnasamy, and V. Sekar. Making middleboxes someone else's problem: Network processing as a cloud service. In *Proc. ACM SIGCOMM*, August 2012.

[15] B. Trammell, P. Casas, D. Rossi, A. Bär, Z. B. Houidi, I. Leontiadis, T. Szemethy, and M. Mellia. mPlane: an intelligent measurement plane for the Internet. *IEEE Communications Magazine*, 52(5):148–156, May 2014.

[16] B. Trammell and M. Kühlewind. Observing Internet path transparency to support protocol engineering. In *Proceedings of the first IRTF/ISOC Workshop on Research and Applications of Internet Measurements (RAIM)*, Yokohama, Japan, Oct 2015.

[17] B. Trammell, M. Kühlewind, D. Boppart, I. Learmonth, G. Fairhurst, and R. Scheffenegger. Enabling Internet-wide deployment of explicit congestion notification. In *Proceedings of the 2015 Passive and Active Measurement Conference*, New York, Mar 2015.

[18] B. Trammell, M. Kühlewind, E. Gubser, and J. Hildebrand. A new transport encapsulation for middlebox cooperation. In *Proceedings of the 2015 IEEE Conference on Standards for Communications and Networking*, Tokyo, Japan, Oct 2015.