Network Monitoring for Todays Internet

Felix Erlacher
Institute of Computer Science, University of Innsbruck, Austria
erlacher@ccs-labs.org

***Abstract*—The task of network monitoring gives operators the ability to analyze their traffic in order to keep their communication network up and running. Monitoring appliances have to adapt continuously to the changes in how the Internet is used. The increase in functionality, usability and secrecy necessitates revisions in monitoring mechanisms. The challenges are manifold: The shift of transport services to the application layer or the rising usage of encryption makes monitoring evermore complex. The solutions currently available are addressing largely individual issues only and are mostly closed source. This PhD thesis will tackle monitoring issues related to high bandwidth links and encryption as well as problems that came up due to the changed usage of HTTP.**

I. Introduction

The observation and analysis of network communication is indispensable for every operator. It provides valuable information about the traffic and the overall state of the network. Hence network authorities have to rely on powerful network monitoring tools. Due to the tight coupling of every communication network to the Internet, network monitoring appliances have to constantly adapt to the evolution of the Internet.

The rather static, technical basis of the Internet is the same as ever. What has changed is the way we use it. To benefit from the capabilities of modern devices, changes in the communication had to be made. Because of the flexibility of the higher layers these changes took place there.

There are two major changes that this PhD thesis focuses on:

One is the growth of powerful web applications. Made possible through effective meta- and scripting languages, it changed the way web developers design web pages. Previous websites were designed strictly according to a pattern, where the server provided static content that was fetched and displayed on the clients' browser. Then came scripting languages such as JavaScript, which gave developers the possibilities to interact with the client side, followed by web applications that are rendered individually and dynamically on the client side and communicate asynchronously with multiple servers. Today's web applications finally are highly interactive using multiple communication channels and benefit greatly from user generated content.

All these innovations entailed changes in the way the Hypertext Transport Protocol (HTTP) is used. Instead of using it for the transfer of hypertext, web applications nowadays use HTTP as a transport protocol requesting resources from a multitude of servers resulting in multiple long lived TCP streams. Many desktop applications even encapsulate their communication traffic in HTTP when firewalls block every other protocol (e.g. Skype). This is also underlined by the increasing usage of HTTP which can be seen in Internet traffic statistics [1], [2].

The other major change is the increasing use of encryption in Internet communication. This as well, was made possible through powerful computing devices, but the bigger influence is the awareness of people for privacy and secrecy which skyrocketed with the publication of the massive surveillance practices of intelligence agencies by Edward Snowden. Encryption, if used properly, can raise the level of privacy and secrecy, but on the other hand makes monitoring much more difficult because a lot of information is not visible for the monitoring appliance.

All these changes let to a different Internet for the users, to a better usability, more functionality, more secrecy but it poses many challenges in the field of network monitoring. The following are the most pressing ones:

- Raise in complexity because of high bandwidth links and thus higher throughput and packet rates
- Degradation of HTTP to a transport protocol and thus pushing the application payload one layer up
- Deep Packet Inspection (DPI) being ineffective because of encryption
- Increasing distribution of communication on multiple channels due to cloud services

II. State of the Art

It is common knowledge that Internet traffic cannot be distinguished anymore by looking solely at the protocol or at the port numbers. As a remedy, manufacturers came up with mechanisms to pinpoint the application based on the network flow. Nowadays most monitoring devices offer traffic categorization mechanisms. Some analyze the IP address the flow is directed to, some use header information and others do expensive DPI operations. Using header information resulted in not being very reliable while DPI over the whole payload is too computational expensive for high bandwidth links.

In the field of intrusion detection Limmer *et al.* [3] improved the DPI scheme by only using the first n bytes of TCP payload of every dialog element and exporting this payload together with the corresponding IPFIX [4] flow to an Intrusion Detection System (IDS). They called this method Dialog based Payload Aggregation (DPA). This resulted in capturing only a very low portion of the payload while still having a high detection rate. We successfully used DPA to detect web applications [5] based on the monitoring tool Vermont [6]. The

remaining challenge is to apply this method to the application layer as well.

The problem of expensive operations during the packet capturing process was tackled by Fusco and Deri in [7]. They propose new driver architectures handing packets directly from the Network Interface Card (NIC) to the application while exploiting queues present on modern NICs and multi core CPUs. Now the challenge is to develop applications that can handle this massive stream of packets in a smart way.

The privacy of users and organizations generating the monitored traffic is affected most when this traffic is made visible to people and appliances other than the operators of the network. For these scenarios the scientific community came up with anonymization techniques. These techniques try to change attributes of the network traffic that might reveal sensible details about the generator of the traffic. It is a thin line between anonymizing enough information to preserve the privacy and still being able to use the gathered traffic for monitoring purposes.

Especially the anonymization of packet header information is a major challenge. Xu *et al.* [8] proposed a well thought out scheme for IP address anonymization. But there is a lot more in network traffic concerning the privacy then just the IP addresses as Zalewski points out [9]. Pang *et al.* [10] proposed an anonymization tool and a guideline for securing a sites' permission to publish the network traces, anonymizing the trace and validating its correctness.

But these approaches do not guarantee proper handling during the monitoring process itself. This problem was tackled in the PRISM project [11]. They propose a network monitoring architecture with a front end collecting network traffic, encrypting it, exporting it to the back end and if necessary anonymizing it for export. Although very thought trough, to the best of our knowledge this approach has never been implemented.

III. Contributions

The main research questions that are going to be solved with this PhD thesis are the following:

- **How to categorize traffic?**
  Because of the simple *port equals application* assignment not working anymore, improved mechanisms for traffic categorization are needed. The currently proposed solutions are mostly closed source and/or not satisfying. What makes traffic categorization even more demanding is the nesting of protocols and the widespread usage of overlay protocols. For example HTTP is used by most web applications as transport protocol and web applications use multiple communication channels This introduces many complications and makes analysis and categorization more challenging.
- **How to cope with very high transmission speeds?**
  Monitoring is typically done at the edge of a network where all the traffic has to pass a gateway or a similar networking device. Transmission speeds of $10\,\mathrm{Gbit/s}$ are more than common on these appliances. This results in very high data and packet rates, making monitoring processes very complex.
- **How does the increasing usage of encryption affect monitoring, and what can we deduce from encrypted traffic?**
  Due to the advancing computing capability and the rising awareness of people the trend goes steadily in the direction of more encryption. In the near future every bit that is not essential for a network to function is likely to be encrypted. This results in methods relying on traditional DPI not being effective anymore. Services like IDSs or virus scanners that need packet payload information will have to look for other means to fulfill their purpose.
- **What are the privacy issues with traffic monitoring and how can they be solved?**
  Privacy concerns are becoming evermore important, this is also reflected in recent discussions and laws. Already an IP address can be uniquely associated to a person. Common monitoring tools until now do only have very little awareness of privacy.

The answers to these questions should also cover the challenges mentioned earlier.

The goal is to implement the resulting solutions in a single, Free and Open Source Software (FOSS) appliance using open standards and interfaces. All evaluation efforts should be made as transparent as possible using freely accessible and recognized traffic traces as well as making own traces and results public again.

References

[1] S. Gebert, R. Pries, D. Schlosser, and K. Heck, "Internet Access Traffic Measurement and Analysis," in *TMA 2012*, Vienna, Austria, March 2012, pp. 29–42.

[2] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, and K. Cho, "Seven years and one day: Sketching the evolution of internet traffic," in *IEEE INFOCOM 2009*, Rio de Janeiro, Brazil, April 2009, pp. 711–719.

[3] T. Limmer and F. Dressler, "Improving the Performance of Intrusion Detection using Dialog-based Payload Aggregation," in *IEEE INFOCOM 2011, IEEE GI 2011*, Shanghai, China, April 2011, pp. 833–838.

[4] B. Claise, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information," IETF, RFC 5101, January 2008.

[5] F. Erlacher, "Monitoring Support for Efficient Web 2.0 and overlay Network Detection," Master's Thesis, University of Innsbruck, October 2012.

[6] R. T. Lampert, C. Sommer, G. Münz, and F. Dressler, "Vermont - A Versatile Monitoring Toolkit for IPFIX and PSAMP," in *IEEE/IST MonAM 2006*, Tübingen, Germany, September 2006, pp. 62–65.

[7] F. Fusco and L. Deri, "High Speed Network Traffic Analysis with Commodity Multi-core Systems," in *ACM IMC 2010*, Melbourne, Australia, November 2010, pp. 218–224.

[8] J. Fan, J. Xu, M. H. Ammar, and S. B. Moon, "Prefix-Preserving IP Address Anonymization: Measurement-based Security Evaluation and a New Cryptography-based Scheme," *Elsevier Computer Networks*, vol. 46, no. 2, pp. 253–272, October 2004.

[9] M. Zalewski, *Silence on the wire: a field guide to passive reconnaissance and indirect attacks*. No Starch Press, 2005.

[10] R. Pang, M. Allman, V. Paxson, and J. Lee, "The devil and packet trace anonymization," *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 1, pp. 29–38, 2006.

[11] G. Bianchi, E. Boschi, D. I. Kaklamani, E. Koutsoloukas, G. V. Lioudakis, F. Oppedisano, M. Petraschek, F. Ricciato, and C. Schmoll, "Towards privacy-preserving network monitoring: Issues and challenges," in *IEEE PIMRC 2007*, Athens, Greece, September 2007.