ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

TIK Institut für Technische Informatik und Kommunikationsnetze

Semester/Master Thesis Proposal: Visualization of HTTP Network Traffic for Digital Forensics

Motivation

Network traffic records are one of the most exhaustive data sources for digital forensics and particularly for data leakage investigations. Network traffic records can be used to show all of an offender's actions, like a videotape of a convenience store robbery [1]. However, already the network traffic of a single workstation can easily account for millions of packets per day. This makes identifying relevant events a really cumbersome and time consuming task, especially since an analyst has at the beginning of a data leakage investigation often only a vague idea about the source of a leak and how information could have been exfiltrated.

Task

The task of this thesis is to develop and implement an approach for aggregation and visualization of HTTP requests issued by a workstation[1]. The visualization should allow to quickly get an idea of the big picture, that is to answer questions like "Which Web sites did a malicious employee visit?" or "Are there HTTP request patterns that do not correspond to typical Web browsing and might be a sign for malware infection?". Typical Web sites nowadays no longer load content from a single domain, instead a user clicking on a hyperlink triggers up to several hundred Web requests being issued to content distribution networks, tracking and advertising sites or social networks. That is why even just compiling a list of Web sites, on which a malicious employee navigated, is more cumbersome than it might seem at first glance.

A possible approach for the task at hand would be to employ frequent pattern mining to aggregate HTTP requests to meta events (e.g. Web browsing on 20min.ch) and only displaying meta events instead of individual HTTP requests.

More Information

To get more information on this thesis, please contact David Gugelmann (gugelmann@tik.ee.ethz.ch).

References

[1] E. Casey. Network traffic as a source of evidence: tool strengths, weaknesses, and future needs. *Digital Investigation*, 1(1):28 – 43, 2004.

[1] We focus specifically on HTTP traffic as this is the protocol accounting for a large portion of the overall traffic in company networks.