Network Traffic Analysis Using Weak Indicators

Ing. Richard Plný
supervisor Ing. Karel Hynek, Ph.D.
Faculty of Information Technology, Czech Technical University in Prague

Introduction

The majority of Internet traffic is encrypted nowadays meaning that insight into computer networks is very limited. Many methods already exist for both (encrypted) network traffic analysis and threat detection, mostly based on Machine Learning. However, such methods usually create large amounts of false positives in high-speed networks. We propose an approach based on "weak indicators" for developing heterogeneous detectors. Such detectors are more robust and precise, lowering the number of false positives. Moreover, the methods used in this approach are more explainable than the ones based solely on machine learning.

Heterogeneous Detectors and Weak Indicators

A weak indicator is a not necessary precise indication of an event or a hint about the contents of traffic. However, multiple weak indicators and their combination into a final decision can create a robust and accurate classifier. Moreover, output alerts of such methods can be easily understood because the alerts themselves contain explanations about their triggers.

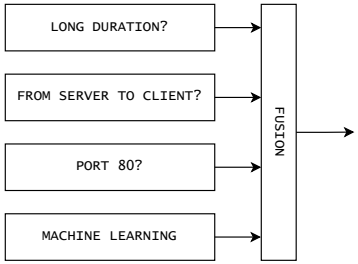

Figure 1. Heterogeneous method for HTTP detection

Figure 1 depicts an example of a heterogeneous method for the detection of the HTTP protocol. Machine Learning can serve as a possible indicator. However, methods without any ML can be designed as well. Nevertheless, we need a tailored module for the detection of the desired event in most cases, which is time-consuming.

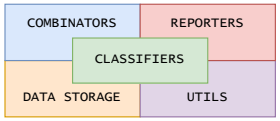

Figure 2. Parts of the Weak Indication Framework

Weak Indication Framework

Heterogeneous modules typically use similar methods for processing weak indicators. Therefore, the development time can be lowered by **re-using** the weak detectors. Weak Indication Framework (WIF) is a **C++ library** developed for this purpose. It contains the commonly used methods for network traffic classification and threat detection. Modules based on the WIF can be developed **much faster** because the methods are already implemented and ready to use for this approach. The WIF's structure is shown in Figure 2.

Weak Indication Framework also contains methods for **data fusion**, for example, it provides a combinator based on Dempster-Shafer Theory, a special theory of probability. Furthermore, reporters can be used for **increased explainability** of detectors based on the Weak Indication Framework. It is an easy-to-use and **highly efficient** library. Moreover, it can be used not only for network traffic classification but other use cases as well.

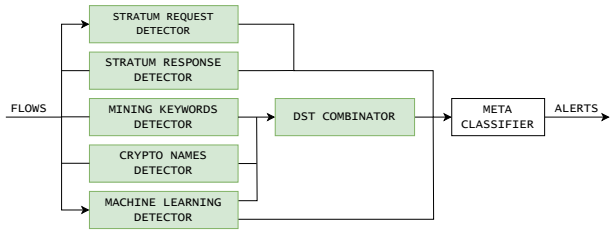

Figure 3. Cryptomining detector

WIF-based Detectors

Several detectors were designed and developed based on the Weak Indication Framework.

Cryptomining detector reveals cryptomalware. It uses three weak indicators. The level of this detector is depicted in the Figure 3: green color identifies components provided by the WIF, used in the implementation. We can see that only the Meta classifier performing the final decision process had to be implemented.

Tor detector looks for Tor communication. Tor detector enhances flows with the detection result which is further used by two consecutive detectors.

Tunnel detector looks for covert communication channels such as VPNs, typically used by malware for Command and Control communication and data exfiltration. It also uses aggregation over time.

Malware detector is a very complex detection method for discovering IoT malware, such as Mirai. It uses eight indicators, such as DHT or anomaly detection. Furthermore, it also uses time aggregation.

Evaluation and Deployment

All detectors were thoroughly tested on specially crafted data and their throughput was evaluated via performance tests. Finally, the detectors were deployed to the national network CESNET3 operated by the Czech NREN called CESNET. The network has more than half a million active users and uses 400Gbps on the backbone lines. Figure 4 shows the network where the detectors were deployed. Figure 5 shows number of detected events by the Cryptomining detector.

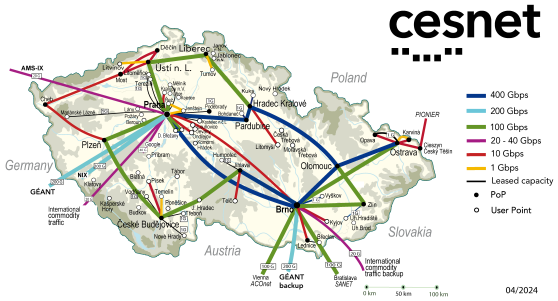

Figure 4. National CESNET3 network

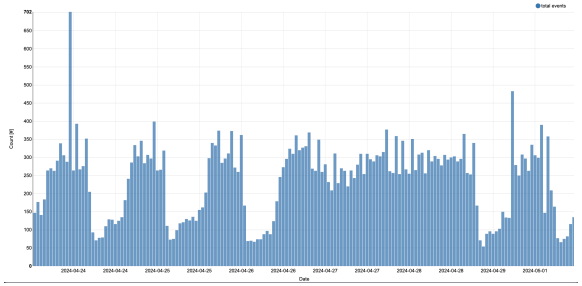

Figure 5. Number of alerts produced by Cryptomining detector

Conclusion

Heterogeneous approach to network traffic analysis provides accurate and explainable detectors capable of operation on national ISP-level networks. Weak Indication Framework can be used for fast development of such methods minimizing time needed for deployment of a detector for newly discovered threats.