Anomaly Detection Based on Traffic Records

Leonard Bradatsch
Institute of Distributed Systems
Ulm University
Ulm, Germany
leonard.bradatsch@uni-ulm.de

***Abstract*—Anomaly detection is a popular approach to detect unusual behavior by subscribers in a network. Different network monitoring approaches, which can indicate such anomalies, with pros and cons each are available. My thesis is developed as part of the bwNetFlow project that aims at realizing a working toolchain for live data analysis and anomaly detection based on NetFlow records. In the course of this project, I analyze different monitoring approaches and weight them up against each other. Common technologies show unresolved problems in high-speed environments. Further research is necessary to meet the requirements of modern high-speed networks.**

***Index Terms*—Networks; Security; Anomaly Detection; NetFlow**

I. Motivation

Nowadays, most computers in companies, institutes or universities are connected with other nodes over a network. Thus, they can be a potential target of a network-based attack. DDoS attacks, port scans or spam e-mails are only a few examples of such attacks. Anomaly detection based on live-analysis of recorded network traffic can help to either prevent networks and computers from such attacks or mitigate their impact.

Network traffic based anomaly detection follows two popular approaches: (1) signature based and (2) and non-signature based anomaly detection. The first approach checks recorded traffic in real time against signatures of known anomalies. In this way, popular attacks can be detected and subsequently mitigated. The drawback of this approach is that only known attacks can be detected. For this approach, the system needs to know beforehand what is a suspicious behavior. In contrast, non-signature based detection tries to identify unusual traffic behavior on the basis of statistical information about the traffic in the monitored network. Thus, anomalies that are not known beforehand can be detected but the system needs to learn normal network behavior before it can work. [1], [2]

Anomaly detection systems can work on the basis of information provided by different network monitoring technologies (e.g., sFlow and NetFlow) or software-defined management solutions (e.g., OpenFlow, P4) that allow access to forwarding devices's control planes. All these approaches can be used to collect data about the traffic that passes the monitored router/switch interfaces. Network packets are combined to flows according to specific matching criteria such as destination and source IP address, destination and source port numbers and physical inbound/outbound router interfaces. All the information can be periodically exported in flow records. Such records can contain information about these flows such as MAC and IP addresses, VLAN identifiers, flow direction or the layer-7 protocol according to the flow's port numbers. Exported records can then be processed and analyzed in real time by an analyzer to extract the desired information. Subsequently, the flow information can be used for congestion detection, ISP billing techniques or as mentioned before for anomaly detection. Changes in the recorded flows can indicate unusual behavior in the monitored network.

Each technology shows pros and cons that must be traded off against each other for the respective use case. sFlow is an open standard and widely available. However, it only samples packets that pass the monitored interfaces und thus it is possible to miss entire flows which can be critical in threat detection. In contrast, NetFlow provides 100% accuracy but is not supported by all router or switch vendors and is a proprietary solution. In high-speed environments, it falls back on sampling packets on flow level. OpenFlow can also provide 100% accuracy but shows scalability problems in high speed networks where it can not ensure this accuracy. The programming language P4 is not widely deployed yet and needs further analysis.

As part of this thesis, all available technologies are investigated and weighted up against each other. The common issues of the mentioned technologies in high-speed networks motivate further research to meet the requirements of modern networks.

II. Related Work

Wagner et al. [5] uses collected NetFlow data from the Swiss Academic and Research Network for entropy based worm and anomaly detection. The authors have focused on real-time detection of worm outbreaks in fast IP networks on the basis of changing entropy contents of traffic parameters derived from the NetFlow data.

Parades-Oliva et al. [3] introduced an anomaly detection technique based on classifying frequent traffic patterns. The anomaly detection process is structured in two phases. During the first offline phase specific traffic characteristics, which are derived from recorded NetFlow records, that occur frequently together during specific attacks are used to build a model that can be used to classify anomalies. In the online phase, the incoming traffic can be matched against this model to identify unusual behavior.

Risto Vaarandi [4] proposed two novel algorithms that can be used for anomaly detection in organizational private

networks. Both algorithms try to discover the actually used layer-7 protocol by using NetFlow data as input. On this basis, the normal service usage for each client can be determined and thus possible anomalies in the traffic can be found.

III. Problem Statement

The thesis is realized as part of the bwNetFlow project under the supervision of Prof. Dr. Frank Kargl. In the course of this project, a toolchain for live data analysis is to be implemented. The first step of this toolchain is to record raw network traffic data. Source of these data is the production traffic in the BelWü network[1], a research network that connects all research facilities in Baden-Württemberg over 100G connections. All data is recorded at the Cisco peer routers that span the network by mirroring the router ports and processing the forwarded packets by a dedicated server. The records conform to the NetFlow protocol standard. The second and third steps in the chain are to process and analyze these stored NetFlows and consequently to detect anomalies. The flows as well as possible anomalies must be analyzed and detected in real time.

The thesis focuses on the live-analysis of captured NetFlows in order to detect security-relevant anomalies among the flows and , subsequently, to inform responsible or interested parties.

A. Approach

The task can be divided into four sub-tasks: (1) The expression of monitoring rules and the evaluation of relevant monitoring tools, (2) the implementation of a prototypical toolchain, (3) the evaluation and optimization of the prototype and (4) the transfer of the prototype into normal operation.

1) The NetFlow protocol provides information about the recorded flows, e.g., source and destination IP-addresses and port-numbers, Type-Of-Service information or the used layer-7 protocol type. On the basis of that information, suitable monitoring rules must be defined that can be used to detect suspicious behavior of network participants. For instance, tremendous e-mail traffic to different destinations could indicate an infected computer that is abused for spaming e-mails. These defined rules must be realized with the aid of suitable monitoring tools, e.g., *tflow2*[2], a NetFlow analyzer released by Google, or the open-source tool *NfSen*[3]. Pros and cons of such tools must be evaluated.
2) The next step is the deployment of the first prototype, including the identified rules and tools, into the operating network to test their applicability.
3) After the deployment, all identified rules and used tools are evaluated regarding their efficiency, scalability, correctness and potential for optimization. When necessary, the existing rules must be expanded by more sophisticated ones.
4) The last step is the transfer of the tested software tools into normal operation and the release of the implemented solution.

[1] BelWü - das Landeshochschulnetz, https://www.belwue.de/

[2] tflow2, https://github.com/bio-routing/tflow2

[3] NfSen, http://nfsen.sourceforge.net/

IV. Research Questions

Resulting from the outlined approach, the following research questions in the area of the introduced bwNetFlow project arise:

- What are modern approaches for anomaly detection in high-speed networks? What approaches are freely available and what are the pros and cons of these solutions?
- How can monitoring rules be expressed and what are suitable applicability criteria?
- Do modern monitoring tools cover all needs of efficient anomaly detection and where is potential for optimization?
- Where are starting points for new research topics that can improve the existing work in the area of anomaly detection?

In this context, additional research questions in a broader scope can be discussed, including:

- How are recorded network data to be handled? What are the necessary steps to make the recorded data anonymous and what are the criteria for preparing the data in a way that it is allowed to grant third parties access to these data?
- What are possible counter approaches against deanonymization attacks? What is the influence of sampling techniques on the applicability of such kind of attacks?
- Is it possible to recognize the used layer-7 application type (e.g., video streams or file downloads) in a recorded network flow according to the application type's specific flow characteristics? Is such an approach applicable for security devices?
- Can the available data from the BelWü network used to provide up-to-date datasets that can be used to test network or security devices?

References

[1] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, *Network Traffic Anomaly Detection Techniques and Systems*. Cham: Springer International Publishing, 2017, pp. 115–169. [Online]. Available: https://doi.org/10.1007/978-3-319-65188-0_4

[2] H. Huang, H. Al-Azzawi, and H. Brani, "Network traffic anomaly detection," 2014.

[3] I. Paredes-Oliva, I. Castell-Uroz, P. Barlet-Ros, X. Dimitropoulos, and J. Solé-Pareta, "Practical anomaly detection based on classifying frequent traffic patterns," in *2012 Proceedings IEEE INFOCOM Workshops*, March 2012, pp. 49–54.

[4] R. Vaarandi, "Detecting anomalous network traffic in organizational private networks," in *2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, Feb 2013, pp. 285–292.

[5] A. Wagner and B. Plattner, "Entropy based worm and anomaly detection in fast ip networks," in *14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise (WETICE'05)*, June 2005, pp. 172–177.