Detecting Advanced Network Threats Using a Similarity Search

Milan Čermák, Pavel Čeleda

Detecting Advanced Network Threats Using a Similarity Search

Milan Čermák and Pavel Čeleda

Institute of Computer Science, Masaryk University, Brno
Czech Republic, {cermak|celeda}@ics.muni.cz

Abstract. In this paper, we propose a novel approach for the detection of advanced network threats. We combine knowledge-based detections with similarity search techniques commonly utilized for automated image annotation. This unique combination could provide effective detection of common network anomalies together with their unknown variants. In addition, it offers a similar approach to network data analysis as a security analyst does. Our research is focused on understanding the similarity of anomalies in network traffic and their representation within complex behaviour patterns. This will lead to a proposal of a system for the real-time analysis of network data based on similarity. This goal should be achieved within a period of three years as a part of a PhD thesis.

Keywords: similarity search, network data, classification, network threats.

1 Introduction

A large number of attacks threaten computer networks. Although the basis of these attacks is similar, a lot of variants exist differing in the protocol used, behaviour, or methods used to avoid their detection. Every day a lot of new attack variants emerge, which represents a challenge for current Intrusion Detection Systems (IDS) [13]. Almost all of these variants require a fast reaction by the authors of these systems to suppress it in time to defend their systems.

Currently, several approaches exists for network anomaly detection [3]. These approaches utilize various techniques for network data analysis from statistical to machine learning methods. Each of these has its own advantages and disadvantages that affect its success in detecting network attacks. Nevertheless, most of the currently used detection tools are based on only statistical analysis and the exact matching of pre-known behavioural patterns, due to their simplicity and lower false positives rate. This approach, however, has a drawback in the inability to detect advanced network threats such as hidden or obfuscated attacks which try to hide their specific characteristics within normal network traffic.

Our research aims to overcome the constraints of these network data analysis approaches through similarity search techniques [15]. This will allow us to combine the advantages of both knowledge-based detection (capable of easily identifying network anomalies without high false positive rates) and cluster-based detection (which allows us to recognize unknown attack variants). We will

focus on network traffic classification based on the same principle as automated images annotation using a similarity search [4]. We plan to classify the analysed network traffic by comparing its similarity with a collection of annotated patterns, reflecting pre-known characteristics of common network attacks. The network traffic will be classified as well as patterns that are closest to it. This approach allows us to take advantage of our knowledge of network anomalies' characteristics, while also revealing their unknown variants.

2 Research Questions

The aim of our research is ***to use similarity search techniques for detecting advanced network threats based on similarity of traffic behaviour patterns***. Our intention is to analyse these behaviour patterns in a similar way as a security analyst would usually do, which is an investigation of known attack patterns and searching for their variants. We have identified the following three research questions that reflect the main topics of our research:

1. *How can we characterize similarity in network traffic?*

The majority of current methods for measuring network traffic collect only specific traffic information, which is not comprehensive enough to describe complex behaviour [3]. Thus, we need to explore methods of transforming these simple data into more complex data, and so providing a reasonable amount of information for comparing their similarity. This research question covers the research of a suitable representation of complex behaviour patterns and the selection of appropriate distance functions to measure their similarity.

2. *How can similarity search techniques be utilized for detecting network anomalies?*

Our second research question is focused on research into the transformation possibilities of fundamental methods for detecting network anomalies into the similarity search concept. We plan to utilize network traffic classification based on a similarity search of defined behaviour patterns and propose a proof-of-concept system for detecting anomalies. This research question, besides other factors, also includes the creation of a collection of anomalous behaviour patterns. Based on their similarity, analysed network traffic should also be classified.

3. *What possibilities do the similarity search techniques have for detecting advanced network threats?*

The main goal of the third research question is to explore the benefits of the proposed method for network anomaly detection in detecting advanced network threats, such as hidden or obfuscated network attacks. We will pay attention to the verification of different characteristics of similarity searches and to the various representations of network behaviour patterns. Specifically, we will focus on different functions to measure the patterns' distances and the identification of combinations of smaller behaviour patterns based on general models of network attacks.

3 Proposed Approach

We will focus on the interconnection of similarity search techniques and methods for detecting anomalies in network traffic. To achieve this aim, we will progressively work through the research questions.

3.1 Characterization of a Similarity in Network Traffic

To achieve the aim of our research, it is necessary to understand network traffic characteristics and their similarities, which should be used within anomaly detection. Weller-Fahy et al. [14] demonstrate that almost every method for detecting network anomalies utilizes some kind of a similarity measure. Thus, we plan to study publications focused on this area and extract all the presented network traffic characteristics and their similarity properties. These findings will be evaluated on publicly available datasets [1,5] and on live network traffic to verify their suitability for specifying network behaviour patterns.

Based on the identified traffic characteristics and their similarities, we will specify behaviour patterns which reflect all important events in network traffic. For their specification, we plan to utilize the Bro Network Security Monitor [11] and IP flow monitoring systems [7], which are suitable for collecting data quickly and analysing in large networks. We will consider two forms of network behaviour patterns: *aggregated*, represented by the aggregation of specific traffic features per unit of time and *sequential*, consisting of sequences of traffic features ordered in time. Each of these forms has its advantages and disadvantages, which we want to understand. Our goal is to associate the appropriate form of specific anomalies to correctly represents their characteristics and avoid behavioural aliasing [9], which occurs when anomaly behaviour looks like normal traffic.

During the definition of network traffic behaviour patterns, it is important to consider their characteristics' similarity. The selection of suitable distance functions and methods for their utilization plays a crucial role in our research since it directly affects the success of detecting network anomalies. For this purpose, we plan to analyse publications focused on the similarity measure, discuss our observations with specialists in the similarity search topic and verify our findings using simulated and live network traffic. For this verification, we plan to utilize the Metric Similarity Search Implementation Framework (MESSIF) [2], which provides a suitable environment for verifying the selected similarity measure techniques.

3.2 Utilizing Similarity Search Techniques for Detecting Network Traffic Anomalies

As we are focused on knowledge-based anomaly detection, the next necessary part of our research is the preparation of annotated behaviour patterns corresponding to network anomalies. To define these patterns we plan to analyse current network attacks and anomalies observed within live network traffic. These patterns will form the basis of a proof-of-concept framework. This framework will

be based on the kNN-classification [15] of ongoing traffic using a similarity comparison of analysed traffic, with the annotated collection of anomaly behaviour patterns.

The anomaly detection capabilities of the proof-of-concept framework will be verified within real and simulated network traffic using simulated network attacks. To validate within simulated traffic, we plan to utilize the KYPO platform [10] that enables network attacks to be securely testing and provides extensive options for monitoring them. Thanks to this environment we will be able to properly compare our anomaly detection approach with other approaches specified in specialized publications or used within common anomaly detection tools, such as Snort [12], Bro [11] or Flowmon ADS [8].

3.3 Detection of Advanced Network Threats Based on a Similarity Search

After verifying our approach, we will focus on the optimization of similarity search attributes and test different distance functions. The aim of this effort is a complex study of impacts, the possibilities of similarity search techniques and advanced network threat detection. These attacks are characterized by their effort to bypass known detection techniques and by hiding their specific characteristics within normal network traffic, which makes them difficult to detect by current anomaly detection methods.

Apart from the complex behaviour patterns of network traffic anomalies, we will also focus on recognising their individual phases. We plan to extend Drašar's [6] research into network attack models and utilize these models as a basis for defining our anomaly behaviour patterns. The updated form of patterns will correspond to the attack phases instead of the whole attack. We believe that such behaviour patterns will make it possible to identify multiple forms of attacks, even those that are as yet unknown.

4 Conclusion

Since this is the first work of its kind, it is necessary to deal with several complications that come with the application of similarity search into the field of network data analysis. The most important part of such an application is the proper understanding of the similarity of network traffic and the specification of complex behaviour patterns reflecting this similarity. The correct specification of these patterns is crucial for achieving adequate results in the detection of common network threats and their unknown variants.

Acknowledgement

This research was supported by the Security Research Programme of the Czech Republic 2015 - 2020 (BV III / 1 VS) granted by the Ministry of the Interior of the Czech Republic under No. VI20162019029 *The Sharing and analysis of security events in the Czech Republic.*

References

1. Barbosa, R.R.R., Sadre, R., Pras, A., Meent, R.v.d.: Simpleweb/University of Twente Traffic Traces Data Repository. Technical Report TR-CTIT-10-19, Centre for Telematics and Information Technology, University of Twente (April 2010), http://eprints.eemcs.utwente.nl/17829/
2. Batko, M., Novak, D., Zezula, P.: MESSIF: Metric Similarity Search Implementation Framework. In: Thanos, C., Borri, F., Candela, L. (eds.) Digital Libraries: Research and Development, Lecture Notes in Computer Science, vol. 4877. Springer Berlin Heidelberg (2007)
3. Bhuyan, M.H., Bhattacharyya, D.K., K., K.J.: Network Anomaly Detection: Methods, Systems and Tools. IEEE Communications Surveys Tutorials 16(1), 303–336 (2014)
4. Budikova, P., Batko, M., Botorek, J., Zezula, P.: Search-Based Image Annotation: Extracting Semantics from Similar Images. In: Mothe, J., Savoy, J., Kamps, J., Pinel-Sauvagnat, K., Jones, G.J., SanJuan, E., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction, Lecture Notes in Computer Science, vol. 9283, pp. 327–339. Springer International Publishing (2015)
5. CAIDA: The CAIDA UCSD Anonymized Internet Traces 2015 - 20150219-130000 (2015), http://www.caida.org/data/passive/passive_2015_dataset.xml
6. Drašar, M.: Behavioral Detection of Distributed Dictionary Attacks. Doctoral theses, dissertations, Masaryk University, Faculty of Informatics, Brno (2015)
7. Hofstede, R., Čeleda, P., Trammell, B., Drago, I., Sadre, R., Sperotto, A., Pras, A.: Flow Monitoring Explained: From Packet Capture to Data Analysis with NetFlow and IPFIX. Communications Surveys Tutorials, IEEE PP(99), 2037–2064 (2014)
8. INVEA-TECH a.s.: Flowmon ads. Web page (2015), https://www.invea.com/cs/produkty-sluzby/flowmon/flowmon-ads, [Accessed: 2016-01-06]
9. Kompella, R.R., Singh, S., Varghese, G.: On Scalable Attack Detection in the Network. IEEE/ACM Transactions on Networking 15(1), 14–25 (February 2007)
10. Kouřil, D., Rebok, T., Jirsík, T., Čegan, J., Drašar, M., Vizváry, M., Vykopal, J.: Cloud-based testbed for simulation of cyber attacks. In: 2014 IEEE Network Operations and Management Symposium (NOMS) (May 2014)
11. Paxson, V.: Bro: a System for Detecting Network Intruders in Real-Time. Computer Networks 31(23-24), 2435–2463 (1999), http://www.icir.org/vern/papers/bro-CN99.pdf
12. Roesch, M.: Snort - Lightweight Intrusion Detection for Networks. In: Proceedings of the 13th USENIX Conference on System Administration. pp. 229–238. LISA '99, USENIX Association, Berkeley, CA, USA (1999)
13. Symantec Corporation: 2015 Internet Security Threat Report. Tech. Rep. 20, Symantec Corporation (Apr 2015), http://www.symantec.com/security_response/publications/threatreport.jsp
14. Weller-Fahy, D.J., Borghetti, B.J., Sodemann, A.A.: A Survey of Distance and Similarity Measures Used Within Network Intrusion Anomaly Detection. Communications Surveys Tutorials, IEEE 17(1), 70–91 (July 2015)
15. Zezula, P., Amato, G., Dohnal, V., Batko, M.: Similarity Search: The Metric Space Approach, Advances in Database Systems, vol. 32. Springer Publishing Company, New York, NY 10013, USA (2006)