Network Security Incident Detection Based on Network Topology Patterns

Juris Viksna, Karlis Freivalds,
Mikus Grasmanis and Peteris Rucevskis

Institute of Mathematics and
Computer Science
Riga, Latvia
Email: {juris.viksna,karlis.freivalds,
mikus.grasmanis,peteris.rucevskis}@lumii.lv

Baiba Kaskina and Varis Teivans

CERT.LV, Institute of Mathematics
and Computer Science
Riga, Latvia
Email: {baiba.kaskina,
varis.teivans}@cert.lv

***Abstract*—In this work, we explore the option of using graph topology patterns for security incident detection in NetFlow data. NetFlow data sets in which data flows related to attacks are specially marked are analyzed using graph visualization techniques in combination with manual methods to identify prospective network topology patterns related to attacks. These patterns are subsequently validated and their merit for incident detection assessed. The current research shows that while such pattern based approach is unlikely to provide a highly reliable incident detection method on its own, it can well complement other methods and can detect attacks that remain unnoticed by statistical analysis of network traffic.**

***Keywords*—Network security; Data visualization; Graph topology patterns.**

I. INTRODUCTION

Network traffic data that is widely available for security incident detection in real time mostly is limited to information contained in NetFlow data format. There are several types of attacks (DoS, port scanning) that often can be detected by simple statistical analysis of NetFlow traffic and its changes over time, however such statistical analysis alone has limited capabilities for security incident detection. The possibilities to extract more information from NetFlow data have been extensively studied and one of the most widely used approaches involves use of different techniques of data visualization in combination with pattern identification in visualized data (a comprehensive survey of such approaches is presented for example in [1]). The usefulness of such methods is additionally demonstrated by commercial success of a number of proposed approaches of such type, e.g., *NFlowVis* system [2].

At the same time from the published use-cases, it is often not very clear what the capabilities and limitations of such methods are. In particular, few attempts seem to be devoted to formalization of patterns in visualized data that might indicate the presence of security incidents. Formal definition of such patterns is also unlikely to be achieved without formalization of what is exactly meant by data visualization. In this aspect, the most promising for formal treatment appear to be graph-based traffic visualization methods, not least because graphs themselves can be described by simple and well known mathematical structures, providing potential for formal definition of patterns of graph topology changes that might be indicative for specific types of attacks. For comparatively small graph topology patterns there is also a good prospect for development of efficient algorithms for detection of patterns in real time.

In graph based-representations of network traffic most often vertices represent traffic sources and edges network connections between them, however more complicated assignments, in particular for edges, are also possible. Probably one of the most formalized treatment of graph-based network monitoring is presented in [3], where a number of different graph patterns and security incidents associated with them have been identified. The analysis however is mainly done in terms of statistical attributes of such patterns (graph connectivity, average vertex degrees, etc.) and not their topology. Another comparatively formal treatment of graph patterns is presented in [4], but here the authors are focusing on the problem of network load monitoring and not on incident detection.

In our work, we use graph-based network traffic visualization with the aim to try to formally define patterns of graph *topology* that can be strongly associated with security incidents. For the study we use labeled data sets of NetFlow data integrated with data from application log files or data obtained by Deep Packet inspection (similar types of labeled data sets have been used and analyzed in [5]). Such labeled data sets for specific types of attacks allow to mark with high certainty the part of traffic involved in these attacks. Then a number of prospective topology patterns for specific attack type is selected and subsequently validated, including validation on sets of NetFlow data alone. The current results suggest that while certain types of attacks often have quite specific graph topology patterns, the same topology patterns are very likely to occur also in 'normal' traffic, and using them alone for security incident detection will give too many false positives. However, the situation significantly improves when these topology patterns are complemented by edge or vertex labels, derived from attributes of NetFlow records (port numbers, flags, number of flows, etc.) and it turns out to be possible to use such labeled topology patterns to detect several types of attacks with high certainty (i.e., with low number of false positives). Our approach of using labeled topology patterns somewhat resembles the one used in the study of

 ISBN: 978-1-61208-427-5

social networks [6]. However, we define patterns in a more formal way; also topological structures of social and network traffic graphs are very different.

In Section 2 of this paper we briefly describe the mathematical formalism used and some experimental results. In Section 3 the plans for future research are briefly outlined.

II. Formalization of Topology Patterns and Experimental Results

For the study, we use two types of data sets. The first data set is obtained by collecting NetFlow traffic from a number of dedicated servers which additionally provides labeling of 'bad traffic' on the basis of information from log files, together with traffic from the whole sub-network of these servers (together approximately 50 traffic nodes). Additionally, since the our institution is also one of the main internet service providers in Latvia, we have access to large amount of NetFlow data. However, this is largely unlabeled data, with only small part being manually analyzed by CERT.LV. The integrated analysis of these two types of data sets gets somewhat more complicated by the fact that the characteristic traffic patterns for internet service providers and end-user networks are different. For visualization and analysis purposes we use *Diagram Editor Engine Kit* – a powerful in-house developed graph visualization and clustering software suite.

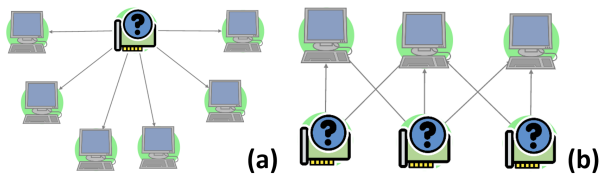

Figure 1. Two simple classes of graph topology patterns: (a) star-like patterns; (b) bipartite patterns.

The graph patterns are defined as (usually small) graphs with vertices and edges labeled by discrete and/or continuous attributes. Also each vertex and each edge is labeled by a Boolean expression having arguments in form $A \sim X$, where A is the value of attribute assigned to the particular vertex or edge, X is variable for the corresponding vertex or edge attribute in traffic graph and relation $\sim$ stands either for equality or inequality (for attributes with continuous set of values). The pattern is matched by network traffic graph, if it is found as its subgraph (or, if specified by pattern, as induced its subgraph) and all the Boolean expressions are true when their variables are substituted with attribute values from the mapped vertices or edges. Potentially perspective patterns are detected by visual analysis of network traffic graphs, and, if good candidates are found, mathematically formalized versions of these patterns are developed and then validated.

The topology of network traffic graphs being not too complex, it is not surprising that there are few patterns that are indicative of attacks by their topology alone. However, it turns out that the predictive power of such patterns considerably improves if they are complemented by edge or vertex labels derived from attributes available in NetFlow records, in which case even 'star-like' patterns that are ubiquitous in network traffic can be successfully used for detection of port scanning and DoS attacks. Examples of some simpler patterns are shown in Figure 1: star-like patterns with topological features being little specific to attacks, however with appropriate labeling added such patterns can become quite informative for attack detection; and bipartite patterns, the topology of these is already much more indicative of attacks, additional labeling is used to distinguish between different attack types.

III. Conclusions and Future Work

Our preliminary work shows that it is possible to define formal (and thus automatically detectable by a program) graph topology patterns that allow to detect security attacks with high certainty. Not all types of attacks could be linked with topology patterns however, and also the number of false negatives is comparatively high. Nevertheless such pattern-based method can detect attacks that remain unnoticed by statistical analysis of behavior of individual network nodes. Thus, they can provide a good complement to traffic statistical analysis and other widely used incident detection methods. The current aim of our research is to develop an annotated library of labeled graph topology patterns that have proved useful in incident detection together with efficient algorithms for detection of these patterns in NetFlow data.

A longer term challenge would be inclusion in pattern definitions the changes of network traffic over time. The problem of characterization of dynamic of networks is well known, however also very challenging, with very few formal results obtained. One of the most promising methods that may have some potential also for analysis of network traffic graphs is based on construction of ordered graphs of topology patterns describing their evolution with time [7].

IV. Acknowledgments

The research was supported by the project ERAF 2013/003/2DP/2.1.1.1/13/APIA/VIAA/027.

References

[1] H. Shiravi, A. Shiravi, and A. Ghorbani, "A Survey of Visualization Systems for Network Security," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, pp. 1313–1329, 2012.

[2] F. Fischer, F. Mansmann, D. Keim, S. Pietzko, and M. Waldvogel, "Large-scale Network Monitoring for Analysis of Attacks," in *Proc. of VizSec 2008*, ser. LNCS, vol. 5210, 2008, pp. 111–118.

[3] M. Iliofotou, M. Mitzenmacher, P. Pappu, S. Singh, M. Faloutsus, and G. Varghese, "Network Monitoring using Traffic Dispersion Graphs (TDGs)," in *Proc. of IMC'07*, 2007, pp. 111–118.

[4] E. Glatz, S. Mavromatidis, B. Ager, and X. Dimitropoulos, "Visualizing Big Network Traffic Data using Frequent Pattern Mining and Hypergraphs," *Computing*, vol. 96, pp. 27–38, 2014.

[5] A. Sperotto, R. Sadre, F. Vliet, and A. Pras, "A Labeled Data Set for Flow-Based Intrusion Detection," in *Proc. of IPOM 2007*, ser. LNCS, vol. 5843, 2009, pp. 39–50.

[6] R. Rossi and N. Ahmed, "Role Discovery in Networks," *IEEE Trans. Knowl. Data Eng.*, vol. 27, pp. 1112–1131, 2014.

[7] C. Vehlow, F. Beck, P. Auwarter, and D. Weiskopf, "Visualizing the Evolution of Communities in Dynamic Graphs," *Computer Graphics Forum*, vol. 34, pp. 277–288, 2015.

 ISBN: 978-1-61208-427-5