ISSN:0975-9646

Pratiksha R. Naik et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3104 - 3106

Anomaly Extraction In Backbone Network Using Association Rules

Pratiksha R. Naik, Shruti S. Kedari, Snehal G. Pawale

Computer Engineering Department
ISB&M School Of Technology, Nande Pune-412115.

***Abstract*— Anomaly extraction refers to automatically finding in a large set of flows observed during an anomalous time interval, the flows associated with the anomalous event(s). It is important for several applications ranging from root cause analysis, to attack mitigation, and testing anomaly detectors. In this work, we use meta-data provided by several histogram-based detectors to identify suspicious flows and then apply association rule mining to find anomalous flow, and summarize the flow.**

***Keywords*— put your keywords here, keywords are separated by comma.**

I. INTRODUCTION

A. Motivation

Anomaly detection techniques are the last line of defence when other approaches fail to detect security threats or other problems. They have been extensively studied since they pose a number of interesting research problems, involving statistics, modelling, and efficient data structures. Nevertheless, they have not yet gained widespread adaptation, as a number of challenges, like reducing the number of false positives or simplifying training and calibration, remain to be solved.

In this work we are interested in the problem of identifying the traffic flows associated with an anomaly during a time interval with an alarm. We call finding these flows the anomalous flow extraction problem or simply anomaly extraction. At the high-level, anomaly extraction reflects the goal of gaining more information about an anomaly alarm, which without additional meta-data is often meaningless for the network operator. Identified anomalous flows can be used for a number of applications, like root-cause analysis of the event causing an anomaly, improving anomaly detection accuracy, and modelling anomalies.[1]

B. Anomaly Extraction

In Figure 3.1 we present the high-level goal of anomaly extraction. In the bottom of the figure, events with a network level footprint, like attacks or failures, trigger event flows, which after analysis by an anomaly detector may raise an alarm. Ideally we would like to extract exactly all triggered event flows; however knowing or quantifying if this goal is realized is practically very hard due to inherent limitations in finding the precise ground truth of event flows in real-world Traffic traces. The goal of anomaly extraction is to find a set of anomalous flows coinciding with the event flows.

An anomaly detection system may provide meta-data relevant to an alarm that help to narrow down the set of candidate anomalous flows. For example, anomaly detection systems analysing histograms may indicate the histogram bins an anomaly affected, e.g., a range of IP addresses or port numbers. Such meta-data can be used to restrict the candidate anomalous flows to these that have IP addresses or port numbers within the affected range. In Table I we outline useful meta-data provided by various well-known anomaly detectors.

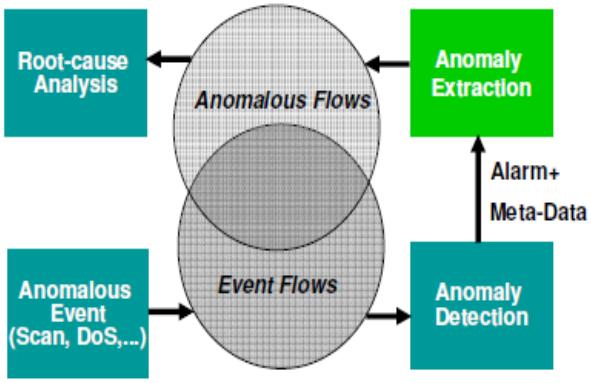

Fig.1.1. the high-level goal of anomaly extraction is to filter and summarize the set of anomalous flows that coincide with the flows caused by a network event such as Denial of Service attacks or scans.

To extract anomalous flows, one could build a model describing normal flow characteristics and use the model to identify deviating flows. However, building such a microscopic model is very challenging due to the wide variability of flow characteristics. Similarly, one could compare flows during an interval with flows from normal or past intervals and search for changes, like new flows that were not previously observed or flows with significant increase/decrease in their volume[6], [8]. Such approaches essentially perform anomaly detection at the level of individual flows and could be used to identify anomalous flows.

II. EXISTING SYSTEM

Identifying network anomalies is critical for the timely mitigation of events, like attacks or failures that can affect the security and performance of network. Traditional approaches to anomaly detection use attack signatures built in an Intrusion Detection System (IDS) that can identify attacks with known patterns. Significant research efforts have focused on building IDS's and, therefore, related production systems are presently employed in many networks. Although signature-based detection finds most known attacks, it fails to identify new attacks and other problems that have not appeared before and do not have known signatures.

III.PROPOSED SYSTEM

Our system contains three different phases. One is histogram detector that will observe the network traffic and alert the system upon anomaly detection. Second phase consists of histogram cloning which assures the anomaly detection and finds the suspicious flows from network traffic. Finally third phase is to apply association rule mining algorithm i.e. Apriori to find the frequent item sets.

Process Summary

1] Form network between computers or laptops.
2] Histogram detector will observe network for certain interval.
3] On anomaly detection form clones of histogram and find suspicious flows in Network.
4] Apply Apriori algorithm to this suspicious flows.
5] Find frequent item sets from the set of suspicious flows.

IV.METHODOLOGY

A. Overview

An overview of our approach to the anomaly extraction problem is given in Figure 2. It contains two sub figures that illustrate the individual steps of our approach. The upper sub figure depicts the anomaly detection and meta-data generation steps. These steps are applied for each traffic feature.The lower sub figure shows how association rule mining is applied to suspicious flows. A subtle point of our approach is filtering flows matching any meta-data (in other wordswe take the union of the flows matching meta-data) instead of flows matching all meta-data, i.e., the intersection of the flows matching meta-data. Assume for example the Sasserworm that propagated in multiple stages: initially a large number of SYN flows scanned target hosts, then additionalflows attempted connections to a backdoor on port 9996 ofthe vulnerable hosts, and finally a third set of frequent flows resulted from downloading the 16-Kbyte worm executable.Anomalies often result in such distinct sets of frequent flowswith similar characteristics. In addition, different meta-datacan relate to different phases of an anomaly. In our example,the anomaly could be annotated with meta-data about the SYNflag, port 9996, and the specific flow size. The intersection ofthe flows matching the meta-data would be empty, whereasthe union would include the anomalous flows.Our approach consists of four main functional blocks.

Histogram cloning:

To obtain additional traffic views the distribution of a traffic feature is tracked by multiple histogram clones. Each clone randomizes the distributionusing one of k independent hash functions. Upon detection of a disruption in the distribution each clone compiles a list Vk of traffic feature values that are associated with the disruption.

Voting:

Meta-data is compiled from the individual featurevalue lists Vk by voting. Specifically, if a certain featurevalue is selected by at least l out of k clones, it is includedin the final meta-data. We analyze the impact of differentparameter settings for l and k on the accuracy of ourapproach.

Flow pre-filtering:

We use the union set of meta-dataprovided by n different traffic features to pre-filter a setof suspicious flows. This pre-filtering is necessary sinceit typically eliminates a large part of the normal flows.

Association rule mining:

A summary report of the most frequent item-sets in the set of suspicious flows is generated by applying association rule mining algorithms. The basic assumption behind this approach is that the most frequent item-sets in the pre-filtered data are often related to the anomalous event. A large part of our evaluation results are devoted to the verification of this assumption.

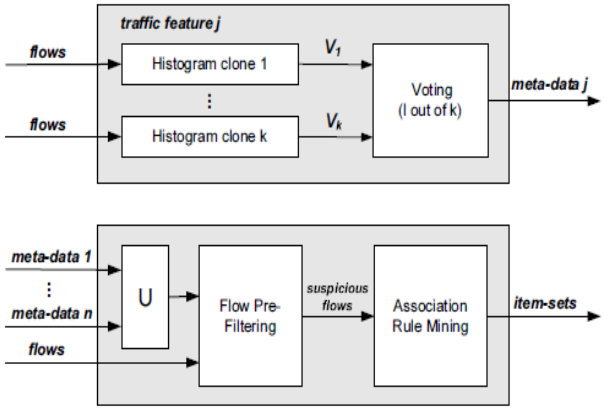

Fig. 2. Overview of our approach to the anomaly extraction problem. The upper figure illustrates how meta-data for a single traffic feature j is generated by voting from k histogram clones. The lower figure illustrates how the metadata for filtering flows is consolidated from n traffic features by taking the union, and how suspicious flows are pre-filtered and anomalous flows are summarized in item-sets by association rule mining.

V. RELATED WORK

Substantial work has focused on dimensionality reduction for anomaly detection in backbone networks [2], [23], [25], [16], [11], [4], [13]. These papers investigate techniques and appropriate metrics for detecting traffic anomalies, but do not focus on the anomaly extraction problem we address in this paper.

VI. CONCLUSIONS

We have studied the problem of anomaly extraction that is of uttermost importance to several applications such as root-cause analysis, anomaly mitigation, and detector testing. We presented a histogram-based detector that provides fine-grained meta-data for filtering suspect flows.

Further, we introduced a method for extracting and summarizing anomalous flows. Our method models flows as item sets and mines frequent subsets. It finds large sets of flows with identical values in one or more features. Using datasets from a backbone network we showed that rule mining is very effective, extracting in all studied cases the involved event flows and triggering a low number of false positives in certain cases that could be trivially sorted out.

ACKNOWLEDGMENT

We acknowledge our sincere thanks to those who have contributed significantly to this project. We thank each and every one's efforts who helped us in some or the other way for small and significant things.

REFERENCES

1. D. Brauckhoff, X. Dimitropoulos, A. Wagner, and K. Salamatian, "Anomaly extraction in backbone networks using association rules," in *IEEE*, 2012.
2. A. Kind, M. P. Stoecklin, and X. Dimitropoulos, "Histogram-based traffic anomaly detection," *IEEE Transactions on Network and Service Management*, vol. to appear, 2009.
3. K. H. Ramah, K. Salamatian, and F. Kamoun, "Scan surveillance in internet networks," in *Networking*, 2009, pp. 614–625.
4. M. V. Mahoney and P. K. Chan, "Learning rules for anomaly detection of hostile network traffic," in *ICDM '03: Proceedings of the Third IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2003, pp. 601–604.
5. R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile*. Morgan Kaufmann, 1994, pp. 487–499.
6. C.Krishnamurthy, S. Sen, Y. Zhang, and Y. Chen, "Sketch-based change detection: methods, evaluation, and applications," in *IMC '03: Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM Press, 2003, pp. 234–247.
7. X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone, and A. Lakhina, "Detection and identification of network anomalies using sketch subspaces," in IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement. New York, NY, USA: ACM, 2006, pp. 147–152.
8. G. Cormode and S. Muthukrishnan, "What's new: finding significant differences in network data streams," *IEEE/ACM Trans. Netw.*, vol. 13, no. 6, pp. 1219–1232, 2005.
9. M. P. Stoecklin, J.-Y. L. Boudec, and A. Kind, "A two-layered anomaly detection technique based on multi-modal flow behavior models," in *PAM: Proceedings of 9th International Conference on Passive and Active Measurement*, ser. Lecture Notes in Computer Science. Springer, 2008, pp. 212–221.
10. M. P. Stoecklin, J.-Y. L. Boudec, and A. Kind, "A two-layered anomaly detection technique based on multi-modal flow behavior models," in *PAM: Proceedings of 9th International Conference on Passive and Active Measurement*, ser. Lecture Notes in Computer Science. Springer, 2008, pp. 212–221.
11. A. Wagner and B. Plattner, "Entropy based worm and anomaly detection in fast ip networks," in *WETICE '05: Proceedings of the 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise*. Washington, DC, USA: IEEE Computer Society, 2005.
12. Book: Balguruswami "JAVA Programming"
13. Website:http://nikhilvithlani.blogspot.in/2012/03/apriori-algorithm-for-data-mining-made.html.
14. IBM Research. Aurora – Network Traffic Analysis and Visualization.
15. http://www.zurich.ibm.com/aurora/.