AUTOMATIC VALIDATION AND EVIDENCE COLLECTION OF SECURITY RELATED NETWORK ANOMALIES *

Ignasi Paredes-Oliva[1,**], Pere Barlet-Ros[1], Maurizio Molina[2]

[1] {iparedes, pbarlet}@ac.upc.edu
Universitat Politècnica de Catalunya (UPC)
Computer Architecture Dept., Jordi Girona 1-3, Campus Nord D6, Barcelona 08034, Spain

[2] maurizio.molina@dante.net
DANTE Ltd.
City House, 126-130 Hills Rd, Cambridge, CB2 1PQ, United Kingdom

DANTE has recently benchmarked and deployed several commercial tools for anomaly detection based on Sampled NetFlow. According to this experience, the number of false positives (even in commercial tools) is still significant (in the order of 10-20% even for the best performing ones). Therefore, human mediation is still fundamental before taking actions to mitigate and prevent recurrence of security related anomalies, especially if this involves the cooperation of a neighbouring network. Currently, this anomaly validation process mainly relies on security engineers' skills.

This work presents an automatic methodology to collect and present, in a compact form, all the IP flows potentially associated to a security related anomaly. Our goal is to give as much precise information as possible to Level 1 NOC operators (possibly not in depth security experts), in order to reduce the cases where the problems need to be escalated to more skilled L2 and L3 security engineers. This proposal aims at creating a closer integration between a NOC and a CERT, when some of the security tools are operated by the NOC. Even when the problems are escalated, the automatic collection and presentation of information related to detected anomalies will help security engineers to speed up incident investigation and resolution.

Recently, Brauckhoff et al. proposed a method [1] to extract the flows associated to an anomaly. This approach is based on an algorithm called Apriori, which uses data mining to find abnormally large sets of flows sharing a combination of some features (each of these sets is called *itemset*). For instance, in case of a Network Scan, Apriori could reveal if the same source IP has scanned more ports than those reported by the anomaly detection tool. In case of a point to point DoS, Apriori would reveal if there are some other sources targeting the same destination, or if the source has heavy communication with somebody else at the same time (this is frequent in botnet-orchestrated DoS).

In this work, we improved the Apriori algorithm to make it work properly in the network operated by DANTE (GÉANT). These improvements are based on our experience gained by manually investigating thousands of anomalies during the tools benchmarking [2]. For example, we have observed that if an anomaly is not characterized by a significant volume of flows, Apriori would not be capable of spotting it. For instance, this occurs in the case of point to point UDP floods (involving a small number of flows but a large number of packets), which are quite frequent in our network. We implemented a variant of the algorithm that, depending on the anomaly type, looks for sets of flows with a huge number of packets instead of flows. We have also added to Apriori the capability of automatically self-adjusting its main parameter called *minimum_suport*, which indicates the minimum size of a bunch of flows to be considered as an *itemset*. This parameter is critical for evidencing parallel activity related to the main anomaly signalled by the anomaly detection tool. Finally, we incorporated as well a mechanism to filter out the false positives returned by Apriori. This filtering technique outputs just the top N of the most frequent itemsets (the ones with most packets or flows, depending on the anomaly type).

We implemented a frontend application based on the adapted algorithm that uses a NfDump [3] backend to retrieve the NetFlow data necessary for Apriori's operation, and we have started using it along with the anomaly detection tool deployed in our network (NetReflex [4]). This GUI will be presented as a demo and the source code of the GUI and the backend will be made publicly available.

---

* We would like to thank the COST Action IC0703 to make this work possible and Daniela Brauckhoff and Xenofontas Dimitropoulos (ETH, Zurich) for kindly sharing their implementation of Apriori.

** Ignasi Paredes Oliva is a PhD student at the Computer Architecture Dept. of the Universitat Politècnica de Catalunya (UPC).

[1] Brauckhoff, D. et al, Anomaly extraction in backbone networks using association rules. In Proceedings of the 9th ACM SIGCOMM Conference on internet Measurement Conference (Chicago, Illinois, USA, November 04 - 06, 2009)

[2] W. Routly, "A Quantitative Cross-Comparative Analysis of Tools for Anomaly Detection", TF-CSIRT/FIRST technical seminar, Riga, Jan 2009, http://www.terena.org/activities/tf-csirt/meeting26/routly-anomaly-detection.pdf

[3] http://nfdump.sourceforge.net/

[4] http://www.guavus.com/