VAST 2013 Mini-Challenge 3: AnNetTe - Collaboration oriented visualization of network data

Siming Chen * Fabian Merkle† Hanna Schaefer† Cong Guo * Hongwei Ai * Xiaoru Yuan *
Thomas Ertl‡

Peking University and Universitaet Stuttgart

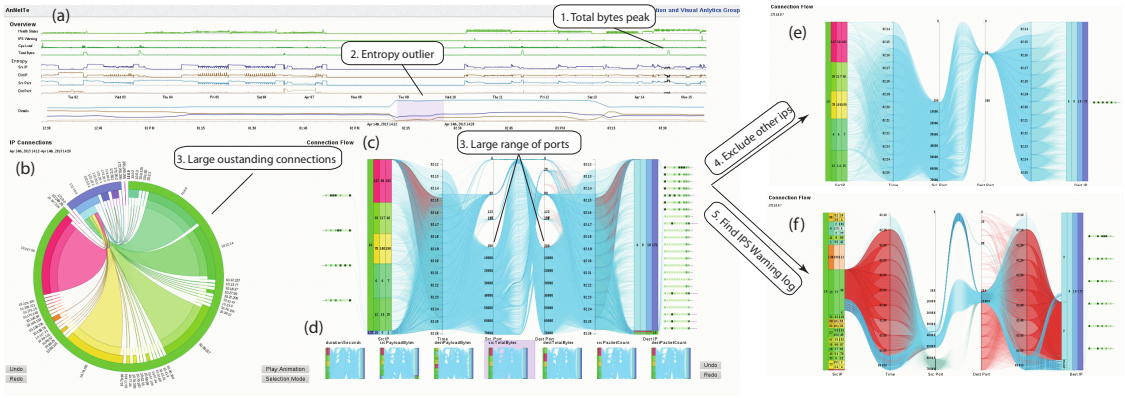

Figure 1: Summary of events of DOS attack and network scan using all visualizations of AnNetTe

ABSTRACT

In this paper we briefly describe the tool submitted to VAST 2013 Mini-Challenge 3 [1]. It is concepted to monitor the network of the Big Marketing Company and find different attacks or security issues. The tool provides traditional (time line and graph) as well as newly generated (connection river) visualizations to show patterns and anomalies in the dataset. The aim of our tool was to present the complex informations inside the given dataset obvious and simple in order to make it usable for collaboration.

Index Terms: H.1.2 [Human information processing]: Collaboration—Event comparison; K.4.3 [Automation]: Anomaly Detection—Visual Hints

1 INTRODUCTION

The VAST 2013 Mini Challenge 3 deals with a network security dataset of the fictitious Big Marketing Company. The companies network has around 1200 user PCs and 24 servers divided into 3 company parts.They provide three different datasets over a time of two weeks. The main dataset is a network connection protocol with IPs, ports, protocol information and transferred bytes, packages and payloads. Another dataset is a health protocol with overall status value of each IP and health measures like CPU load, disk usage and memory usage. For week two an IPS log is provided which records teardown, built up or denial activities of the prevention system. The task is to find critical behaviour in these datasets. In AnNetTe (Analyzing Network Technology) we adress this challenge by using a three level representation of the data with a clear interaction pipeline. The visualizations are related to classical views like timelines and graphs, but optimized for the represented data. The focus group are domain experts who want to collaborate on a security project. With AnNetTe we were able to find more than 12 events and verify and discuss them with proof from the dataset.

*e-mail: csm, cong.guo, hongwei.ai, xiaoru.yuan@pku.edu.cn
†e-mail: merklefn, schaefha@studi.informatik.uni-stuttgart.de
‡e-mail: Thomas.Ertl@vis.uni-stuttgart.de

2 ANNETTE

Our tool AnNetTe was developed using the d3 libary (http://d3js.org/) for javascript. Seperate databases for each visualization provide faster and reduce data traffic. A mayor goal besides high functionality was to enable easy collaboration. AnNetTe is divided into three visualizations on different levels of detail. In the overview users select an time, in the ring graph they can choose a group of IPs and finaly in the connection river they can verify their event and submit it.

2.1 Timeline Overview

The overview visualization of AnNetTe shows 3 different types of timelines (Figure 1 a). There are 3 different types of timelines. The first part of the Overview consists of 4 static timelines of different accumulated variables. We use the summed status value and CPU load of the health dataset, the count of denial logs in the IPS system as well as the summed total bytes. The second part of the Overview shows 4 calculated anomaly detection timelines. The used anomaly detection is a sample entropy, which gives a measure about the increasing diversity or disorder in the development over time [2]. The sample entropy is used on categorical datasets like IPs and ports of source and destination. The third part of the overview provides a zoom function. For the time selected it shows the details of all 4 entropy timelines. The timelines in the detailed view are overlaying to provide easier comparison.

2.2 Ring graph

The ring graph shows the connections of the selected time structured and grouped by subnet (Figure 1 b). The color of each connection and ring part is used to encode the subnet as well as the source of each connection. Since internal IPs are in blue color and external ones differ from green to red, the connection type is obvious. For clutter reduction IPs can be excluded from the graph by interaction. A faster overview tis provided by an animation of the ring graph over the time selected. In figure 2 you can see ring graphs of everyday summary.

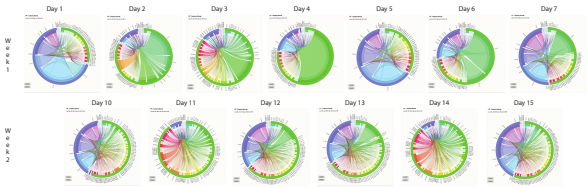

Figure 2: Ring graphs for everyday event summary.

2.3 Connection River

The connection river shows details as one fingerprint adapted from a paralell coordinates view. The view has three axis for source and destination each. The first axis shows the IPs the second the connection time and the third the ports used. In (Figure 1 c) you see each line represents a connection flowing from the three source axes to the symmetrical destination axes. Between the two port axes one of the connection variables is displayed as the height of the flow lines. This varibale can be chosen from the preview of each variables connection river below the visualization (Figure 1 d). The bar of colored squares besides the river shows the relative stage of each accumulated health attribute for internal IPs. For each IP and time connection that has been protocoled in the IPS log a red line is drawn in the connection river. This line ends at the port axes to signify the denial of this connection. In the connection river the user can also exclude IPs. If he selects instead all other IPs as well as their destinations will be excluded. The user can focus the visualization to on his event and get a singular fingerprint of his data.

3 Results

With AnNetTe we identified more than 12 events. Common event types are DOS attacks, network scans and port scans. We will describe finding one event with he help AnNetTe.

3.1 Example event - DOS and Scan

(Figure 1) presents all three levels of the event. To find it we were searching for peaks in the total bytes timeline. In the detailed view we also see a source port outlier. We select this anomaly. The ring graph shows 5 dominant external IPs. In the river many source ports and the high bytes numbers as the heigt are displayed. Because of the red lines we separate the river. First we select the 5 dominant IPs and see the DDOS attack with high unhealthy status of the victim (Figure 1 e). Then we select the IP with red lines and see a port scan attack on one server (Figure 1 f). It shows a high number of destination ports and IPS warnings. The user could now store both events in the system and wait for the opinion of his partners.

3.2 Overall observations

By bringing together all events visually and discussing them in AnNet we were able to find overall relations. The pattern of health got worse over most nights and relaxed again over daytime. Some attacking IPs belonged to an organized group. Other events could be seen in AnNetTe but not be totaly identified. In figure 3 you can see on monday after the regular work time external connections diminish and disappear later. The remaining main activity is the health monitor 172.10.0.6 which protocols strong unhealthyness. Considering the pattern of the following nights, this is could be a network breakdown.

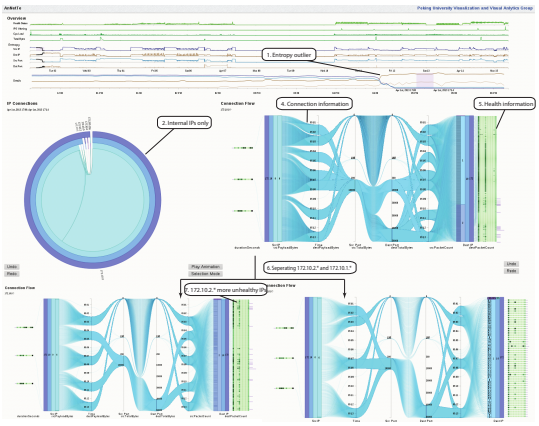

Figure 3: Example of suspicious beahviour found with AnNetTe

4 Collaboration Platform

Crowd sourcing has been approved by previous papers on its general benefits [3]. AnNetTes aim is to do security analysis as a collaborative work of domain experts. To collaborate users can visit the website of our tool. To submit an event they give information such as the event type and description in addtition to the selected data like time, IPs and a snapshot which is added automatically. After sumbission the event is shown as dot in the timeline. Others can select the events and it will be loaded in all three visualizations. After reviewing they can comment and agree or disagree.

5 Conclusion

Although network security datasets are abstract and require expert domain knowledge, we find, that users can work together to make it more comprehensive and create deeper insights. With AnNetTe we created a simplified but intuitive way of visualizing and interacting wth the data. By usign a fingerprint visualization in the detailed level, we provided easy recognition for the collaboration on previous events.

Acknowledgements

The authors thank Zhuo Zhang from QIHU360 for providing the valuable suggestions and feedback from security domain. We also thank the Baden-Wuerttemberg Stiftung for supporting this collaboration with a scholarship. This work is supported by National NSFC Project (No. 60903062 and 61170204) and National NSFC Key Project (No. 61232012).

References

[1] IEEE. http://www.vacommunity.org/vast

[2] A. Lakhina, M. Crovella, and C. Diot. Mining anomalies using traffic feature distribution. *SIGCOMM '05*, Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications:217–228, 2005.

[3] W. Willett, J. Heer, and M. Agrawala. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. *ACM*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems:203–2012, 2010.