Poster: Is a Picture Worth a Thousand Alerts?

Nicklaus A. Giacobe
The Pennsylvania State University
College of Information Sciences and Technology
101 Information Sciences and Technology Building
University Park, PA 16802
nxg13@ist.psu.edu

1. INTRODUCTION

This poster presents preliminary work in the adoption of a geovisual analytic toolkit for the purpose of representing organization-level cyber security data. This data is generally represented in text form and static visualizations such as histograms and simple pie, line or bar charts. We use the GeoViz Toolkit to represent vulnerability, network flow, intrusion detection and OS security log data fabricated for the 2011 Visual Analytics Science and Technology (VAST) Conference. The raw data was processed using custom scripts, written in Perl, to identify first order entities such as host computers and users. Next, we tabulate the number and type of events that happen to these entities with an hourly analysis period. We create a map of these first-order entities and represent the events to the end user with coordinated geographical and non-geographical visual analytic interfaces from the GeoViz Toolkit [1]. In this poster, we extend the work that we presented for the 2011 VAST Challenge [2] by adding network flow and server log data to our original assessment of IDS alerts and vulnerability scans. We conclude this as a work-in progress and discuss options for human subject evaluation of cyber security interfaces to measure their impact on situation awareness.

2. 2011 VAST MINI CHALLENGE 2

The VAST Conference issues visual analytics challenges as part of its annual conference. In 2011, three different mini-challenges were issued including microblog, cyber security and text analysis tasks. Mini-Challenge 2 sought visual analytic tools and techniques to evaluate three days' worth of cyber security data for a fictitious mid-sized company that has several servers and about 150 workstations [3]. The data included vulnerability scan output from a single pass with NESSUS as well as three days of output from Snort IDS, Cisco ASA Firewall and a Windows Server Security Log. Competitors were asked to analyze the data with off-the-shelf or customized visual analytic tools. Several cyber security events were imbedded in this fabricated data including: obviously vulnerable computers; a denial of service attack against the web server; infected hosts scanning the network for other machines to compromise; and unauthorized access in the form of new hosts being added to the network.

3. PARSING THE RAW DATA

We follow a data fusion model of extracting the first order entities from the data [4]. These entities are generally individual host computers represented by their hostname or IP address or computer users represented by their usernames. We then tabulate the number of times an event happens (IDS alert, network connection to another host, or a Windows security event) and tabulate these for each hour.

We wrote several custom scripts in Perl to parse the provided TXT, CSV and XML files. Each script parses the data from one source of cyber data, event-by event, and tabulates the metadata representations based on interesting metrics. For the NESSUS scan, we total the CVSS Base score (a representation of the severity of the vulnerability) for all of the vulnerabilities found on each host. For the IDS alerts, we total the number of alerts of each type and add these totals to both the source and destination hosts indicated in the alert. For the firewall, we tabulate only the number of new connections created and total for both the source and destination hosts indicated by the record. Finally, for the server logs, we tabulate security log events by event ID and by IP address, matching hostnames to IP addresses as we go. We also identify user accounts and tabulate pertinent events (log on/log off) for individual users. Output from each script is a flat file CSV or dBase formatted file.

4. FABRICATING A NETWORK MAP

Geovisual analytic tools like the GeoViz Toolkit require a geographic base layer for the presentation of the data. The .shp file format from ArcGIS 10 (by ESRI - Environmental System Research Institute) is commonly used. The .shp file records the shapes of geographic features and the corresponding .dbf file provides the numerical values of each data type that applies to that geographic space. However, there is no true geography provided with this data. So, we substitute a spatialization for the geographic shapes instead. Each subnet is represented as a block of connected sub-blocks. For the VAST Challenge, we created these shapes by hand in ArcMap, but we now create shape files programmatically in Matlab. The order in which each polygon (in our case only a square) is added to the shape file must match the order that the data records appear in the rows of the corresponding .dbf. This order is the connection between the .shp and the fields of the .dbf.

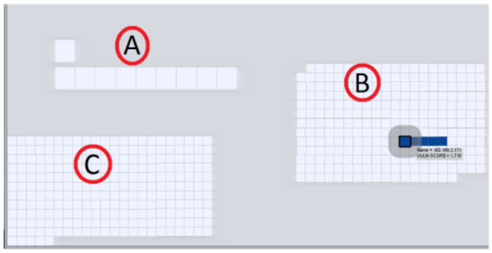

Figure 1. Shape File displaying vulnerability scores with A: servers, B: workstations and C: unused server IP space

The size of the shape is selected to indicate the relative importance of the entity. For example, a server might have a shape four times the size of a workstation (see Figure 1 (A) Servers vs. (B) Workstations). This allows the analyst the ability to quickly identify important nodes. While there may only be six servers on the server subnet, it is important to represent all of the possible addresses in the subnet. This allows for representation of attempted connections to non-existent hosts (Figure 1 C: Unused server IP space) and the discovery of newly added nodes (possibly unauthorized) to the network.

5. CYBER DATA IN THE GVT

Three different views in the GeoViz Toolkit provide the primary visualizations that support the analysis. They are the Parallel Coordinate Plot, Geomap and Histogram (see Figure 2). We use the parallel coordinate plot to represent time-series data (same data type over several time intervals) and re-configure it to compare values of different data types in the same time interval. The geomap represents either a single data value or two values simultaneously in different colors. In the histogram, selection of groups of values highlights and emphasizes those data in the other coordinated views.

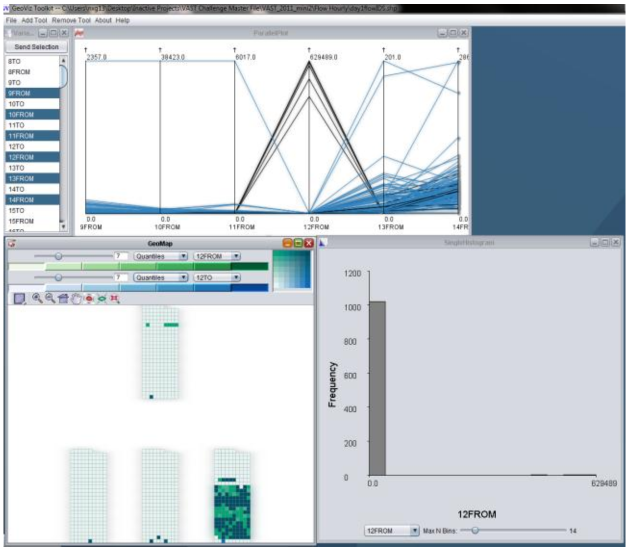

Figure 2. The GeoViz Toolkit with parallel coordinate plot (top), geomap (bottom-left) and histogram (bottom right).

6. PROPOSED EVALUATION

Cyber security data fusion and visualization techniques often promise to increase situation awareness. While this goal is important, evaluation is often left incomplete. This work-in-progress has identified that measurement of situation awareness in cyber security is a non-trivial task. Through interviews with practicing cyber security professionals [5], we hope to identify the important cues that are required for analysts to recognize. These findings are the foundation for the development of an implementation of a freeze-probe technique specific to the cyber security domain similar to SAGAT for military aviation [6].

The testing of the measurement technique will require collection of in-trial freeze probes and a variety of other pre- and post-trial questionnaires and assessments such as SART [7], NASA-TLX [8], Human Performance Scoring [9] and subject demographics. A second text-based interface will be used as a baseline for comparison [10]. Two separate pools of research subjects ("novices" and "experts") will be recruited. We hope to find positive correlation between our cyber-SAGAT, the Human Performance Score and the visual analytic interface, regardless of the subject's cyber experience.

7. ACKNOWLEDGMENTS

This work was partially supported by U.S. Army Research Office (ARO) MURI Grant "Computer Aided Human Centric Cyber Situation Awareness" W911-NF-09-1-0525.

8. REFERENCES

[1] F. Hardisty and A. Robinson, "The geoviz toolkit: using component-oriented coordination methods for geographic visualization and analysis," International Journal of Geographical Information Science, vol. 25, pp. 191-210, 2011.

[2] N. Giacobe and S. Xu, "Geovisual Analytics for Cyber Security: Adopting the GeoViz Toolkit - VAST 2011 Mini Challenge 2 Award: "Innovative Tool Adaptation"," presented at the IEEE Symposium on Visual Analytics Science and Technology, 2011.

[3] M. W. Georges Grinstein, Kristin Liggett and Danko Nebesh. (2011, 3/26/2012). 2011 IEEE VAST Challenge, Mini-Challenge 2. Available: http://hcil.cs.umd.edu/localphp/hcil/vast11/

[4] N. A. Giacobe, "Data fusion in cyber security: first order entity extraction from common cyber data," presented at the Proc. SPIE 8408, 84080E (2012), Baltimore, MD, 2012.

[5] M. Tyworth, Giacobe, Nicklaus A., Mancuso, Vincent, "Cyber situation awareness as distributed socio-cognitive work," presented at the Proc. SPIE 8408, 84080F, Baltimore, MD, 2012.

[6] M. R. Endsley, "Situation awareness global assessment technique (SAGAT)," presented at the Aerospace and Electronics Conference, 1988. NAECON 1988., Proceedings of the IEEE 1988 National, 1988.

[7] R. Taylor, "Situational awareness rating technique (SART): The development of a tool for aircrew systems design," the Situational Awareness in Aerospace Operations AGARDCP478, 1990.

[8] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," Human mental workload, vol. 1, pp. 139-183, 1988.

[9] A. R. Wellens and D. Ergener, "The CITIES Game," Simulation & Gaming, vol. 19, p. 304, 1988.

[10] D. M. Vincent Mancuso, Nicklaus Giacobe, Michael McNeese and Michael Tyworth, "idsNETS: An Experimental Platform to Study Situation Awareness for Intrusion Detection Analysts," in 2nd IEEE Conference on Cognitive Methods in Situation Awareness and Decision Support, New Orleans, LA, 2012.