NetSet: A Set Visualization Tool for Network Metadata Exploration and Threat Hunting

Brett Fouss
MIT Lincoln Laboratory
Lexington, MA

Dennis Ross
MIT Lincoln Laboratory
Lexington, MA

Shannon Robinson
Tufts University
Medford, MA

Kenneth Alperin
MIT Lincoln Laboratory
Lexington, MA

***Abstract*—The research and development of effective visualizations and visualization tools are necessary to achieve comprehensive cyber situational awareness, and are a growing need in cyber security [1], [2]. One common desire is the ability to quickly view and synthesize large complex datasets and understand the underlying set membership of the elements. Advances in set-typed data visualization have not yet been broadly applied to cyber data for operational use. To this end, we created NetSet, a network metadata analysis tool that leverages Lex et al.'s UpSet technique for visualizing intersecting sets [3] along with new capabilities for temporal awareness. We argue that NetSet is an effective and perceptually sound tool for set membership analysis tasks via a common data analysis use case, highlighting the need for novel set-typed data visualization techniques in cyber security.**

I. Introduction

Cyber defense is an increasingly urgent requirement for commercial and government entities requiring fast and accurate analytic strategies to support cyber situational awareness (cyber SA). Many cyber SA tasks can be described as set membership tasks, where the unions and intersections have implications for potential adversarial events [4]. Network traffic analysis presents such a use case as it generates large amounts of data with many possibly overlapping features. Common tools, such as Vern Paxson's *Bro*, provide mechanisms for cyber analysts to organize and examine this network data and metadata, but not to explicitly explore set memberships.

For set relationships, data should be explored not only as it is distributed across each property of network metadata (e.g. IP addresses, alert signatures, etc.), but also the combinations and aggregations of those properties. These combinations can highlight trends that may otherwise not be observed in the large volume of data. This frames metadata analysis as a set membership task.

Currently, many technologies for analyzing set membership are labor intensive or require a significant degree of technical expertise and pivoting capability [2], [5]. Data visualization is an effective way for cyber analysts to avoid these pitfalls, but there has not been a concerted effort to apply visualization to set membership tasks because they become perceptually unwieldy as the number of sets increase while not providing a simple way to describe element attributes in detail.

We present a network metadata analysis tool called NetSet as a solution. NetSet uses Lex et al.'s UpSet technique to provide a clear, effective and scalable visualization of combinations of categorical or binned properties inside Bro cyber data sets. Further context is built using widgets to help cyber analysts launch investigations within NetSet.

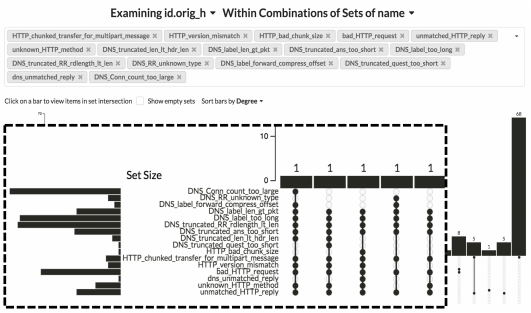

Fig. 1. NetSet's UpSet-style set visualization on `bro-weird` data filtered with DNS and HTTP alerts highlighting hosts with many different alerts

II. NetSet Dashboard Prototype

With the current need for set visualization of cyber data, we created a dashboard prototype version of NetSet. The main aspect of the original UpSet visualization that we incorporated in our tool was the structure of the set view. The set view addresses the difficulty of displaying the "combinatorial explosion" of set intersections and unions. In this view, a combination matrix lists all of the intersections as full circles connected by a line- these may be aggregated to show unions as well. Following UpSet design, the cardinality of each set intersection is plotted using a bar chart. A separate bar chart shows the size of each set.

In addition to the UpSet-style set component, NetSet adds components intended for cyber use cases. First, an element view provides further insight into the members of the set intersections. By clicking on the bars in the table view, users can view a paginated list of all records. Specifically, the source and destination IP address and port information, a time stamp, and other fields depending on the log type are shown. These drill-down data points are used to initiate and support investigations.

NetSet also includes a component to visualize data over time. By clicking on a set member in the element view, the user can see what the set membership of the element was at each time interval. This is represented by displaying the same vertical column of circles used in the set view along a zoomable axis. This extends NetSet's capabilities by displaying an event over a time window that allows users to investigate the temporal evolution of the event's set membership.

III. Triaging Potential Indicators of Compromise

To show how NetSet can be used to examine potential indicators of compromise (IOCs) and IPs to explore further,

we deployed it on Bro log data from the 2012 Mid-Atlantic Collegiate Cyber Defense Competition (MACCDC) [6].

All of the Bro data are loaded, but we focus on `bro-weird` that provides metadata associated with unexpected network activity for this exploration. Analysts generally start by discovering misconfiguration or malicious activity within a network according to rules and alerts, making `bro-weird` an effective stand-in for an initial hunt dataset. We configure NetSet to load the entire MACCDC network Bro data which spans 8:30 AM on March 16, 2012 to 4:55 PM on March 17, 2012 in about 4.4 GBs of logs. We then drilled down into the `bro-weird` dataset for this investigation.

Analysts might first compare IP addresses to rule-based alerts. We configure NetSet to encode source IPs as set members and `bro-weird` alert types as sets. Data is then aggregated by IP, grouping by shared combinations of observed suspicious activity. Due to the large amount of `bro-weird` records, effort would be made to reduce the search space allowing the investigation to focus on particular set-encoded alerts. To simulate this reduction, `bro-weird` entries not pertaining to DNS and HTTP protocols are filtered out by NetSet, as shown in the highlighted area of Fig. 1.

Sorting by set intersection degree allows analysts to notice that most IPs with some combination of these alerts were only associated with four or fewer types of `bro-weird` logs. However, a few IPs had unique combinations of eight or more types of `bro-weird` alerts. NetSet allows a drill down to view all records associated with these hosts, as well as view the evolution of set membership of the hosts over time. Fig. 2 shows the temporal set visualization component for an IP chosen because of its large number of suspicious behaviors. Periodic intervals of activity on March 16 are immediately obvious. Such behavior is suspicious and warrants additional investigation by an analyst.

This workflow would help an analyst quickly triage a large dataset, finding hosts with high degrees of `bro-weird` DNS and HTTP related alerts with which to launch further investigation. This workflow could be easily replicated for other alerting schema to discover trends of interest.

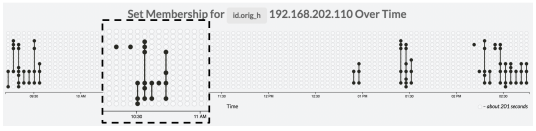

Fig. 2. Periodicity of `bro-weird` alerts on a single host

IV. Discussion

Performing a similar network analysis using current techniques would prove difficult [1], [5]. Given a specified list of set encodings, NetSet automatically enumerates all possible combinations of those properties and assigns membership to each member, omitting combinations not present in the data. To achieve similar capability with present tools an analyst would have to write a join query based on *each* combination of selected attribute values. Writing such join queries is time-intensive and prone to error, and the number of queries increase exponentially as more selected attributes are chosen.

Scaling is also a concern as common set visualizations (e.g. Venn diagrams) can only show set intersection of up to three or four sets easily [3]. As shown in Fig. 3, more advanced techniques such as parallel sets [7] effectively break down a few properties of set members, but often don't show how set members map to combinations of properties.

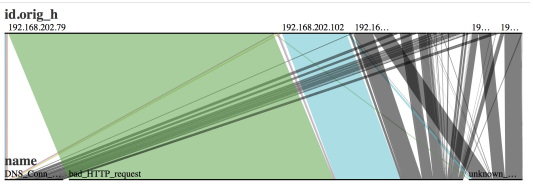

Fig. 3. Parallel set displaying mapping of `id.orig_h` to `name` fields within MADCCDC `bro-weird` logs. Attention is drawn to most common mappings, but most set membership data is obscured

Although common set visualizations often fail meet the needs of cyber analysts, NetSet demonstrates how novel set visualization research might be useful in cyber use cases. Building on UpSet's effective set intersection visualization, NetSet adds a tabular network metadata view, as well as a temporal breakdown of set membership.

Here, we use `bro-weird` logs as a proxy for rule-based alert data to simulate a cyber hunt work flow. However, NetSet easily supports any investigation into set membership within any network metadata (e.g. connection information, status codes in web traffic).

Future work might include a human factors evaluation to assess the usability of NetSet. Further research should also be done into how other novel set visualization techniques such as PowerSet [8] can be applied to cyber use cases.

References

[1] D. Staheli, T. Yu, R. J. Crouser, S. Damodaran, K. Nam, D. O'Gwynn, S. McKenna, and L. Harrison, "Visualization evaluation for cyber security: Trends and future directions," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, VizSec '14, (New York, NY, USA), pp. 49–56, ACM, 2014.

[2] L. Franklin, M. Pirrung, L. Blaha, M. Dowling, and M. Feng, "Toward a visualization-supported workflow for cyber alert management using threat models and human-centered design," in *Visualization for Cyber Security (VizSec), 2017 IEEE Symposium on*, pp. 1–8, IEEE, 2017.

[3] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister, "Upset: visualization of intersecting sets," *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, pp. 1983–1992, 2014.

[4] S. Jajodia, S. Noel, P. Kalapa, M. Albanese, and J. Williams, "Cauldron mission-centric cyber situational awareness with defense in depth.," in *MILCOM*, pp. 1339–1344, 2011.

[5] M. Albanese, N. Cooke, G. Coty, D. Hall, C. Healey, S. Jajodia, P. Liu, M. D. McNeese, P. Ning, D. Reeves, *et al.*, "Computer-aided human centric cyber situation awareness," in *Theory and Models for Cyber Situation Awareness*, pp. 3–25, Springer, 2017.

[6] M. Sconzo, "Samples of security related data."

[7] F. Bendix, R. Kosara, and H. Hauser, "Parallel sets: visual analysis of categorical data," in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pp. 133–140, IEEE, 2005.

[8] B. Alsallakh and L. Ren, "Powerset: A comprehensive visualization of set intersections," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 361–370, 2017.