Towards Network Science Enhanced Cyber Situational Awareness

Geoffrey B. Dobson, Timothy J. Shimeall, Kathleen M. Carley

Carnegie Mellon University, USA

ABSTRACT

A dynamic network analysis is conducted on network flow data to demonstrate an improvement in cyber situational awareness. The analysis begins by collecting network-level data (density, network centralization total degree, and fragmentation) on samples of network flow data using the SiLK collection and analysis suite. The next phase categorized the data into four types: autonomic inflow, autonomic outflow, human inflow, and human outflow. Using the CASOS tool ORA, a series of dynamic network analyses were performed on each hour of the data. The results showed variations between the autonomic and human traffic that can be used by firms to gain more detailed understanding on how traffic behaves on their computer networks. The more granular profiles of operations permit separate understanding of automated and manual processes. The network science techniques provide a basis for providing these improvements in a systematic and repeatable manner.

Keyword: dynamic network analysis, cyber security, network traffic analysis, cyber situational awareness

1 INTRODUCTION

In December 2016, the National Cybersecurity and Communications Integration Center released "Grizzly Steppe – Russian Malicious Cyber Activity", which provided a thorough analysis of the Russian activities aimed at disrupting the United States presidential election. The analysts concluded that two separate advanced persistent threats conducted a spear phishing campaign, installed malware, and exfiltrated data (NCCIC and FBI, 2016). The previous year, in February 2015, FireEye released a report of the cyber attack methods employed in the ongoing Syrian civil war. The report is striking in its similarity to the NCCIC report. The authors concluded that cyber operatives conducted a spear phishing campaign, installed malware, and exfiltrated data (Regalado et al., 2015). The methods employed are well known, and easily defensible. In many cases, it's quite simple to look back, and see advance indicators of the cyber attacks. So why do these attacks continue to occur, at an alarming frequency? The answer lies in the difficulty of creating normal network operational data profiles that can be assessed in real-time, in order to detect anomalies.

We ask – *does binning data, and then calculating graph level measures, provide a more granular picture of the network so that anomaly detection is easier to accomplish? If so, can we create "normally operating" network science-based signatures? How can organizations use these insights, incorporating their known patterns of life, to achieve enhanced cyber situational awareness?* Our primary contribution with this research is to address these questions using a large empirical data set. This work takes a network science approach to the goal of gaining cyber situational awareness. This study conducted a dynamic network analysis on a full month of netflow records captured on a live network used for corporate operations. That data is binned into different categories in order to analyze human-generated traffic separate from autonomic traffic. Three network measures are taken on the data at every hour: network density, total degree network centralization, and fragmentation. These measures are chosen based on their potential for detecting anomalous traffic.

Density is defined as the ratio of the number of links versus the maximum possible links for a network. In a computer network context, this could be interpreted as an activity level amongst internal and external IP addresses. That is, are the internal network computers mostly communicating with different external IP addresses (lower density) or the same external IP addresses (higher density)? Total degree network centralization is defined as the average total-degree centrality of each node in a square network. In a computer network context, this could be interpreted as a concentration level

amongst IP addresses. That is, are the internal IP addresses receiving information from different IP addresses (lower total degree centralization) or similar external IP addresses (higher total degree centralization)? Fragmentation is defined as the proportion of nodes in a network that are disconnected. In a computer network context, this could be interpreted as grouping of IP addresses. That is, are the IP addresses forming many different components (higher fragmentation) or just a few components (lower fragmentation)? In a completely connected network, fragmentation value will be zero. For any of these three network level measures, a large deviation from the norm, could be a simple way to warn of anomalous activity, not recognized by other detection systems. The output of the dynamic network analysis can inform organizational "normal operating behavior" profiles. Then, known patterns of life such as hours of operation, specialized operations, and network upgrades allow leadership to anticipate potential deviations from the norm.

This paper will be organized as follows. Section 2 provides background and motivation for this research. Section 3 describes the method of binning netflow data and then conducting hourly dynamic network analysis. Section 4 presents the results of analysis. Section 5 discusses takeaways, limitations, and future work, followed by the conclusion in Section 6.

2 BACKGROUND

2.1 Defining Cyber Situational Awareness

What is cyber situational awareness? Many attempts at definitions have been made. Tadda and Salerno applied general situation awareness reference and process models to the cyber domain (Tadda et al., 2010). Barford et al described cyber situational awareness for cyber defense as a three phase process (recognition, comprehension, and projection) and went on to give 11 varying viewpoints about what makes up cyber SA (Barford et al., 2010). More recently, Onwubiko says cyber SA "encompasses people (operator/team), processes, and technology required to gain awareness of historic, current and impending (future) situations in cyber, the comprehension of such situations, and using those understandings to estimate how current situations may change, and through those predict future situations and the resolution of the current situation, and the enablement of control to protect the systems from future projected incidents" (Onwubiko, 2016). This definition aligns closest with what this research provides, specifically predicting future situations based on a changing set of current situations.

2.2 Manual and Automated Analysis for Situational Awareness

Many researchers are working on various aspects of cyber situational awareness. Franke and Brynielsson conducted s systematic review of 102 research papers in cyber situational awareness, and separated the literature into four broad categories, one of which is "Experiments, methodology, technology". A key finding was that more empirical research should be conducted (Franke et al., 2014). Klein, Tolle, and Martini proposed a "holistic approach to Cyber Defense" based on the well-known OODA loop (Observe, Orient, Decide, Act). Cyber defense should align all activities of the organization, and by applying the OODA loop, a continuous process ensues, constantly updating the data and its processing (Klein et al., 2011). The research in this paper supports this goal by providing methods to identify relevant data to be observed and acted on. Morris et al created an ontology driven framework for enterprise mission assurance. They argue that typically, network sensor-based systems "rarely take into account the context, patterns, and relationships between the threats and other cyber and real-world entities" (Morris et al, 2011). The work in this paper builds on this line of reasoning by defining the known organizational patterns of life, to more accurately predict network anomalies.

Cyber situational awareness is still in its infancy. When asked about the status of cyber situational awareness at the U.S. Department of Health and Human Services, in June of 2014, HHS Chief Information Security Officer Kevin Chearest made the following observation: "There was no situational awareness at the department" (Jackson, 2014). The following year, The Health Information Trust Alliance (HITRUST) conducted a study of the state of cyber situational awareness for healthcare organizations across the United States and concluded that there is an industry wide lack of tools for true cyber situational awareness. The report stated: "This lack of awareness leads many organizations to expend resources and rely heavily on indicators of compromise (IOC) to determine if a breach … has already occurred … this approach is retrospective in nature and introduces inefficiencies" (Darkreading, 2015). The work in this paper is aimed at creating techniques to be integrated into toolsets that will first model the normal operating behavior specific to an organization's network, and then detect changes to the network, as close to real-time as possible. Dynamic network analysis technique would best serve as part of a suite of complex cyber situational awareness tools, most likely with a dashboard for human machine interaction.

Bradshaw, et al introduced "Sol" in 2012, an agent-based framework that combines advanced modeling techniques and the latest visualization

techniques to inform human analysts team interactions. The visualization is based on cockpit flight control dashboards, where the pilot is presented with all of the information necessary to successfully operate the aircraft. The authors specifically call out a key challenge: "the difficulty in finding the flight performances model for network analysis. Whereas the primary task of the pilot is to fly effectively within the known parameters of a fixed aerodynamic model, the job of the NOC analyst is to understand emerging threats accurately against the moving target of a network that is constantly changing" (Bradshaw et al., 2012). The approach taken in this paper could be characterized in exactly that manner: how does an organization find its appropriate cyber performance model? The network science measures should be evaluated and potentially included in a comprehensive cyber situational awareness dashboard.

2.3 What are the Meaningful Signals?

When a pilot climbs into a cockpit, he/she will recognize the displays, because they are nearly all the same no matter the size or model of the airplane. Organizations of varying sizes and functions will likely need different dashboards for cyber situational awareness. These dashboards will likely combine volumetric data, network data, behavioral characteristics, and emerging trend predictions, with both statistical and heuristic summary processing capabilities.

Cheng et al proposed a six layer cyber common operational picture (Cyber-COP) approach to detection of cyberattacks. Their design aims to methodically define a Cyber-COP architecture, so that a software toolkit can be deployed to provide comprehensive cyber situational awareness. The toolkit relies on defining attack graphs and then overlaying the attack graphs on the Cyber-COP based on deterministic sequences of attacks onto network nodes (Cheng et al., 2010). Dynamic network measures take a different approach, which can complement a comprehensive Cyber-COP, by attempting to detect stochastic attack traffic.

Some research aims at defining early or real-time warning systems for cyber situational awareness by defining normal signatures of computer behavior and thereby anomaly detection schemes where the behavior goes outside of the norm. This is very difficult based on the amount of data which must be processed, and the shifting nature of normal behavior in computer networks. Lee, Lee, and Kim created a Knowledge-Based Real-Time Cyber-Threat Early Warning System that received data from three systems and alerted when hourly values were above a critical value (Lee et al., 2006). The

warning system relies on volumetric data whereas the dynamic network approach relies on graph-level measures of the traffic.

Diettrich et al developed the TaskTracer system, a Microsoft Windows add-in that captures user events and creates workflow discovery system. They apply machine learning techniques to search for information flow subgraphs, effectively finding known organizational workflows. A discovered shortcoming of the approach occurs when multiple ways (information flows) of completing work tasks are present, limiting the cyber defense capability of this approach (Diettrich et al., 2010).

Friedberg, Skopik, and Fiedler developed a system agnostic dynamic tool called automatic event correlation for incident detection (AECID). It creates fingerprints of log files based on pattern detection, and determines if the new pattern is valid, or anomalous (Friedberg et al., 2015). This novel approach is limited by the system's ability to correctly determine valid or anomalous patterns.

2.4 The Dynamic Network Approach to Enhancing Cyber Situational Awareness

In this paper, we present the following three assumptions: Assumption one – Different types of netflow data will present different network measures. Assumption two – All flows, when analyzed in a temporal manner, will show strong characteristics of periodicity, providing normally operating network behavior profiles. Assumption three – By incorporating patterns of life into the analysis, cyber situational awareness is enhanced. In the results section we'll discuss whether or not the methods were able to meet our assumptions, and then in the discussion section, we'll note limitations and opportunities for future work.

This paper describes an exploratory analysis performed using the dynamic network approach. This approach exploits three key assumptions: that netflow records for different categories of traffic will present different network measures; that normal profiles of network operation will show strong characteristics of periodicity when analyzed in a temporal manner; and that incorporating patterns built on the measures and periodic characteristics will enhance cyber situational awareness for an organization. These assumptions lead to an analysis method that is described in the next section, and then the following section presents and discusses results of applying this method to the networks for a moderate-sized organization. The conclusion section revisits these assumptions in light of the experience gained by applying the method.

3 METHOD

This research addresses a gap in research by taking a network science approach to the field of cyber situational awareness. The approach observes three network level measures temporally. The three measures are: density, total degree network centralization, and fragmentation. These are measured via a three step method. The first step retrieves flow records into the four bins using simple criteria to separate human and autonomic traffic. These records are exported in comma-separated-value format. Step two imports the saved files into the CASOS tool ORA (Altman et al., 2017) and converts them to DyNetML files, allowing for network data inspection. Step three conducts a dynamic network analysis on the four datasets measuring density, total degree centralization and fragmentation on an hourly basis.

3.1 Step 1 – Filtering Flow Records on SiLK Collector

The System for internet-Level Knowledge (SiLK)[1] is a suite of tools for network flow data collection and analysis, developed by the CERT program of Carnegie Mellon University's Software Engineering Institute. The tools aggregate the headers of related packets of internet protocol traffic traversing the network, and give the analyst capability to both rapidly query for specific traffic of interest, and to flexibly summarize and trend characteristics of that traffic. A virtual machine is deployed to the network, pulling traffic to summarize from a span port of a router or switch, often a network boundary router. The YAF (Yet Another Flowmeter) tool turns packets into flow records. Each flow record contains a variety of information, as shown in Table 1, with the most useful usually being: source/destination address, source/destination port, transport protocol, bytes, and duration. These flow records are then passed to collection tools that store these records in a repository of binary flat files for later querying. An example setup is shown in Figure 1.

[1] See http://tools.netsa.cert.org/silk/docs.html for a description of this tool suite and documentation of each tool.

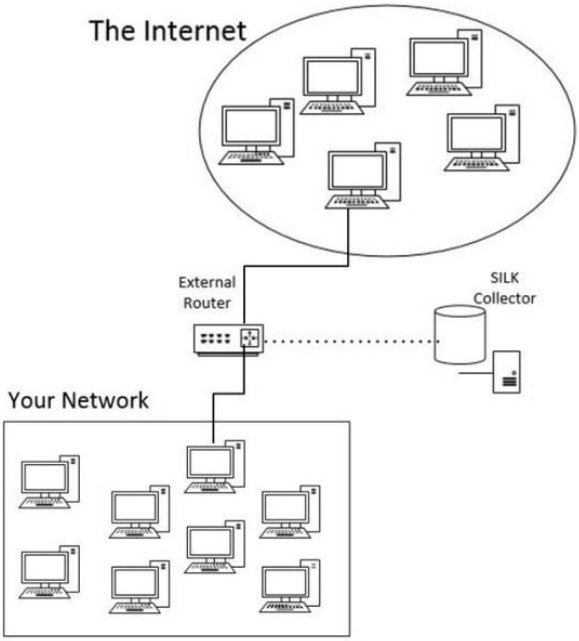

Figure 1. SiLK Collector setup

Table 1. List of Flow Fields

Name	Description
sIP	source IP address
dIP	destination IP address
sPort	source port for TCP and UDP, or equivalent
dPort	destination port for TCP and UDP, or equivalent
protocol	IP transport protocol
packets	packet count
bytes	byte count
flags	bit-wise OR of TCP flags over all packets
sTime	starting time of flow in millisecond resolution
duration	duration of flow in millisecond resolution
eTime	end time of flow in millisecond resolution
sensor	name or ID of sensor at the collection point
class	class of sensor at the collection point
type	type of sensor at the collection point
iType	the ICMP type value for ICMP or ICMPv6 flows and empty for non-ICMP flows
iCode	the ICMP code value for ICMP or ICMPv6 flows and empty for non-ICMP flows.
in	router SNMP input interface or vlanId
out	router SNMP output interface or postVlanId
nhIP	router next hop IP
initialFlags	TCP flags on first packet in the flow
sessionFlags	bit-wise OR of TCP flags over all packets except the first in the flow
attributes	collection attributes (such as active timeout)
application	guess as to the content of the flow

This paper presents results with data gathered from an operational corporate network that has between 1,000 and 5,000 employees and approximately 3,000 network connected devices. The network is setup such that each computer is able to freely access any IP address on the internet. Each flow record will have a source IP address and destination IP address, one being on the corporate network, and one outside of the corporate network (purely internal traffic and transit traffic between external addresses are excluded from this data). Therefore, at step three, we are analyzing a bipartite, unimodal network.

A key concept of this method, and the primary concern of step one, is the deliberate binning of flow records. The way that flow records should be binned is an internal organizational decision based on a mixture of experience and intuition. This analysis bins the data into four categories: autonomic inflow, autonomic outflow, human inflow, and human outflow. This serves several purposes. First, it decreases the size of data to be

analyzed at any given time. Second, it allows reasoning about the difference between bins. Should, for example, a systems upgrade will be occurring overnight, the analysis would predict anomalous looking autonomic traffic. Third, and most importantly, binning the data provides a more granular dynamic network analysis in step three. Table 2 summarizes the four bins of flow data and the associated query conditions (referencing SiLK's rwfilter tool, but source and destination options would need to be added in operation) used to retrieve the data.

Table 2. Flow Bin Descriptions and SiLK Query

Flow Bin	Description	Partial SiLK Query
Autonomic Inflow	Traffic flowing into the network automatically initiated by software	rwfilter --type=in, inweb --bytes=1-96 --packets=1-2
Autonomic Outflow	Traffic flowing out of the network automatically initiated by software	rwfilter --type=out, outweb --bytes=1-96 --packets=1
Human Inflow	Traffic flowing into the network as a result of human behavior	rwfilter --type=in, inweb --bytes=97- --flags-all=AS/SA --packets=3-
Human Outflow	Traffic flowing out of the network as a result of human behavior	rwfilter--type=out, outweb --bytes=97- --flags-all=AS/SA --packets=2-

Each query is run to retrieve a separate day's data, with the day embedded in the name of the output file. One day of flow records provides a logical start and stop point for the dynamic network analysis in step three. After running the queries, and downloading each file to comma-separated-value format, this process produces 120 files (30 days with one file for each of the four bins) totaling approximately 119,000,000 nodes and 36,000,000 links.

3.2 Step Two – Import to ORA and Inspect Data

This step imports each comma-separated-value file into ORA as an agent-by-agent dynamic meta-network. Each IP address is considered an agent. Since each NetFlow record includes a timestamp, the import process recognizes this meta-network as dynamic. The import process uses the start time field in the netflow records to automatically aggregate by hour. This way, the 30 files from each day of June each contain 24 hourly networks. There are 2,328,504 total nodes and 35,516,517 total links in the meta-networks. These are very sparse networks in that there are a relatively small number of links represented per hour. Most nodes only connect with one

other node. Few nodes connect with many nodes. Table 3 describes a sample hourly meta-network.

Table 3. Sample Meta-Network

Item	Details
Flow Bin	Autonomic Inflow
Date	June 15, 2016
Hour	11:00 AM
Nodes	31,969
Links	140,897
Density	.000138
Total-Degree Centralization	.000577
Fragmentation	.000118

3.3 Step Three – Conduct Dynamic Network Analysis

The dynamic network analysis uses three measures: density, total degree network centralization, and fragmentation. Density is defined as the ratio of the number of links versus the maximum possible links for a network. In a computer network context, this could be interpreted as an activity level between internal and external IP addresses. That is, whether the internal network computers mostly communicate with different external IP addresses (lower density) or the same external IP addresses (higher density). Total degree network centralization is defined as the total-degree centrality of each node in a unimodal network. In a computer network context, this could be interpreted as a concentration level among IP addresses. That is, whether the internal IP addresses receive information from different IP addresses (lower total degree centralization) or similar external IP addresses (higher total degree centralization). Fragmentation is defined as the proportion of nodes in a network that are disconnected. In a computer network context, this could be interpreted as grouping of IP addresses. That is, whether the IP addresses form many different graph components (higher fragmentation) or just a few components (lower fragmentation). In a completely connected network, fragmentation value will be zero.

After importing all 30 days of a flow type, the analysis runs each bin of flow records through the built-in ORA algorithms for density, total degree centralization, and fragmentation, with the hourly network results saved for calculating averages and standard deviation values. This provides the normally operating network behavior, suitable for inclusion in a dashboard for situation awareness based on network science. The formula for each measure is depicted in Table 4.

Table 4. Dynamic Network Analysis Formulas

Network Measure	Formula
Density	Let A be the binary input network with n columns Density = sum(A) /n*(n-1)
Total-Degree Centralization	let A be the input unimodal network with N nodes let v = vector of Total-Degree Centrality values for network A let c be the Centralization value for vector v Total-Degree Centralization = c / (N-2)
Fragmentation	let A be the unimodal network with N nodes let S_k be the number of nodes in the kth weak component of network A let $S = (\sum S_k(S_k - 1))/N(N - 1)$ Fragmentation = 1 - S

4 RESULTS

Assumption 1 - Different categories of netflow data will present different network measures

This section examines the results of binning the large amounts of flow data into different categories and then performing a dynamic network analysis on the binned data. Table 5 lays out the statistics for each measure.

Table 5. Network Measures of Binned Data

NetFlow Type	Avg. Density	Std. Dev	Avg. Centralization, Total Degree	Std. Dev	Avg. Fragmentation	Std. Dev
Autonomic In	0.000243	0.000081	0.000405	0.000621	0.000117	0.000142
Autonomic Out	0.000189	0.000080	0.000981	0.000265	0.004120	0.003238
Human In	0.000096	0.000034	0.000872	0.000321	0.054300	0.048000
Human Out	0.000130	0.000049	0.001075	0.000299	0.002358	0.001891

The Table 5 statistics support the first assumption. The average density of autonomic inflow is much greater than human inflow. That is, when measuring hourly samples of autonomic inflow, the density of the network appears to be approximately 2.5 times greater than the density of human inflow. When considering differences in human versus autonomic flow, the greatest difference appears between fragmentation values of autonomic inflow versus human inflow. This data shows that human inflow

fragmentation presents values approximately 464 times greater than autonomic inflow fragmentation values. The autonomic versus human flow analysis that presents the least difference is found when analyzing the average total degree centralization of the hourly networks between autonomic outflow and human outflow. Human outflow average total degree centralization is approximately 1.1 times greater than autonomic outflow average total degree centralization. The average values of both data bins appear well within one standard deviation of each other.

Assumption 2 - Normal profiles of network operations will show strong characteristics of periodicity

This section examines the assumption of binned data presenting strong indications of periodicity. After loading the datasets into ORA, periodicity was tested via Fourier transform applied to the sequence of measures for each meta-network. Unsurprisingly, all meta-networks showed strong evidence of both twenty four hour and seven day periods. This is clearly shown in the three Figures below: 2, 3, and 4. Throughout all meta-networks, other weaker periodicities were detected, but none were consistent across multiple measures.

The next three figures show the entire month (every hour) human outflow bin over all three measures.

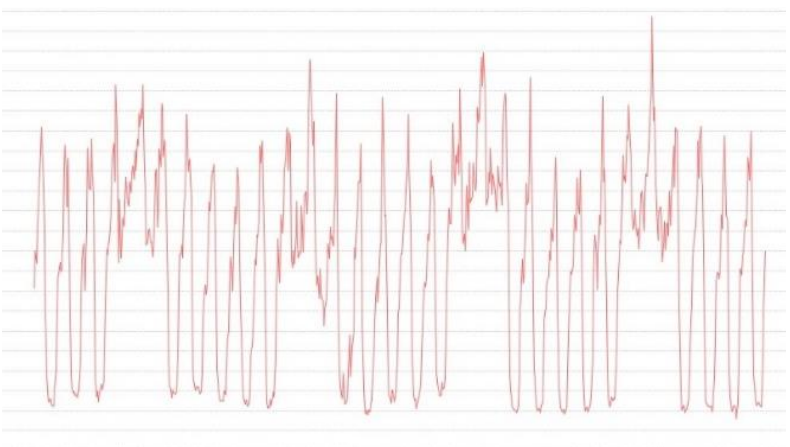

Figure 2. Human Outflow Hourly Density Values

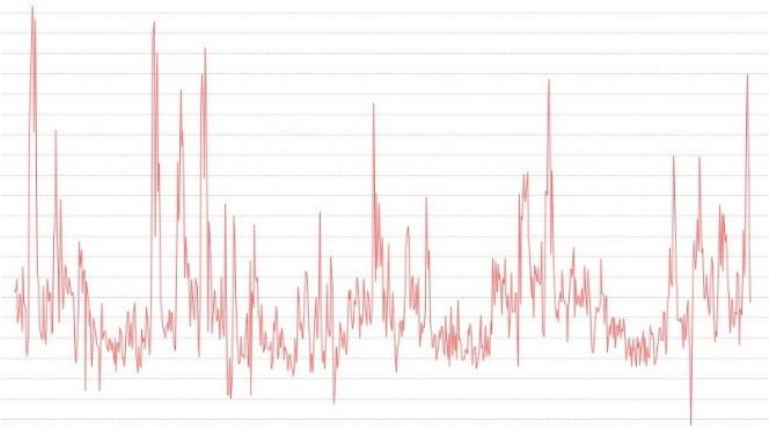

Figure 3. Human Outflow Hourly Network Centralization Values

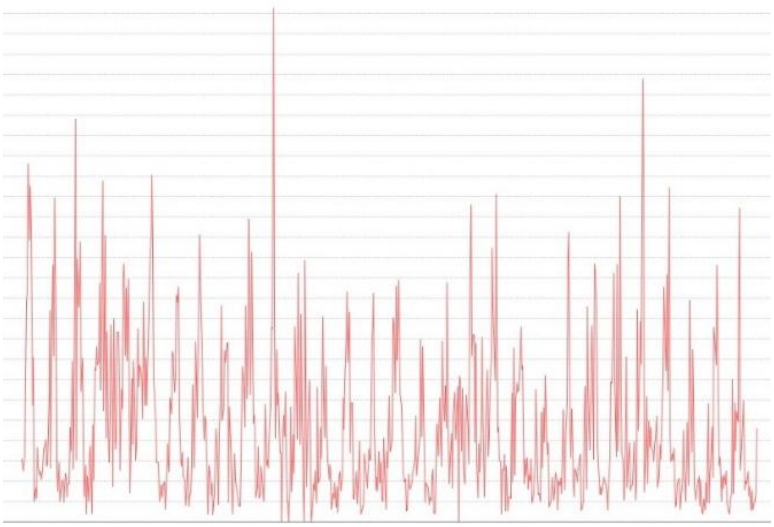

Figure 4. Human Outflow Hourly Fragmentation Values

In all three figures, periodic patterns clearly emerge, supporting assumption two. The density values visually change the most from weekday to weekend. Deviations from normal behavior are apparent in the spikes that appear. This data can be used by the organization to create the operating signatures with which a cyber situational awareness dashboard would be normalized.

Assumption 3 - By incorporating patterns built on dynamic graph measures and periodic characteristics, cyber situational awareness is enhanced

The final takeaway from this study is the targeted analysis provided through incorporating patterns of autonomic vs. manual traffic. A question a network administrator might ask is: "Are we concerned that the network density is .000206?" A better question to ask is: "Are we concerned that autonomic outflow network density is .000206, at 3:00 PM on a Tuesday?" The only way to gain true cyber situational awareness is to understand the data in a given context, of a given type, and at a given time. This is supported by the processed data in two ways. First, Figure 5 displays a clear difference in autonomic inflow network density over the weekends, supporting assumption three.

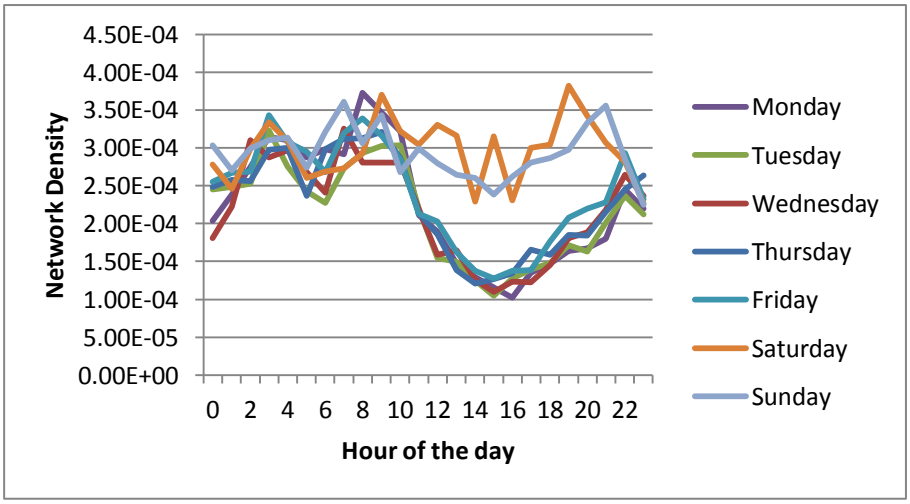

Figure 5. Autonomic Inflow Network Density Days of the Week Analysis

Next, in Figure 6, a clear difference emerges in different bins of netflow data, over the same day of the week, and same network measure. This picture shows that the network density of human outflow is far different than the other three bins. Also, a pattern emerges that will be more useful in cyber situational awareness. The network density of human outflow varies greatly during the daylight hours versus the nighttime hours, as may be predicted. Unexpectedly, the human inflow does not vary much between daylight and nighttime hours. Therefore a dashboard for cyber situational awareness should take this key difference into account. An anomaly would be much easier to detect on the human outflow data.

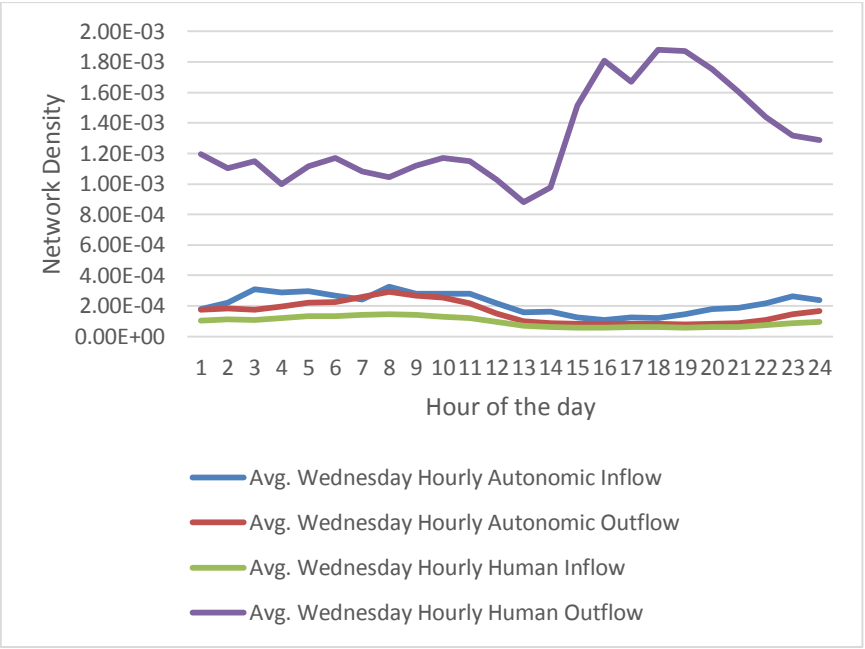

Figure 6. Network Density Binned Data Analysis

5 DISCUSSION

This section discusses limitations of the current work and opportunities for future research.

5.1 Uncertainty in Differences of "Autonomic" vs "Human"

The SiLK queries displayed in Table 2 are a first attempt at categorizing the netflow into "autonomic" and "human" based on several experienced analyst opinions. Most autonomic and human inflow will be captured in these queries, however there will be some that overlap or are left out. This research did not focus on testing whether or not those queries were correct, but rather served as a proof of concept in the potential of that technique. More importantly, correctness of the queries supports use those categories for Assumption 3, permitting security evaluation regarding the mix of behavior. Questions addressing expected behavior of networked systems could be analyzed with binned data in a more thorough manner than non-binned data. For example, a large maintenance operation on routing systems might produce anomalous traffic on the autonomic outflow data. Future

work is needed in defining query criteria for netflow traffic that is truly autonomic versus human generated.

5.2 Anomalous Traffic in the Dataset

A challenge for any attempt at creating "normal operating behavior" is the potential for anomalous data to be captured in the baseline data. This is certainly a concern for this empirical study. The data examined were unfiltered netflow records from every hour of every day of June. There is almost certainly some anomaly data included in the baseline. For this work, we think that enough normal data is included, and the organization is unaware of any major issues during the collection period. Future work could address this concern, attempting to discard data outside of a given threshold.

6 CONCLUSION

This paper describes a study of a dynamic network analysis on a full month of netflow data, binned by four data types: autonomic inflow, autonomic outflow, human inflow, and human outflow. This empirical study aimed at evaluating the efficacy of utilizing network measures to enhance cyber situational awareness. A full month of operational corporate netflow was analyzed, 42.13 GB in size. All of the analytic assumptions were validated by the study results. First, binning data provided observable differences in the network measures. Second, a dynamic network analysis showed clear patterns of periodicity, providing the potential for normal network operating profiles to be created. Third, the data was analyzed in a way that known patterns of life could be incorporated. This research could be expanded in several ways. At present the data must be queried and saved, and then integrated into a separate tool. A software toolset that automatically queried, processed, and analyzed the data would be useful, moving towards real-time network measure situational awareness. Also, evaluation of other network measures would prove quite useful.

7 REFERENCES

Altman, N., Carley, K.M., & Reminga, J. (2017). ORA User's Guide 2017. *Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report CMU-ISR-17-100*

Barford, P., Dacier, M., Dietterich, T. G., Fredrikson, M., Giffin, J., Jajodia, S., Jha, S., Li, J., Liu, P., Ning, P., Ou, X., Song, D., Strater, L., Swarup, V., Tadda, G., Wang, C., & Yen, J. (2010). Cyber SA: Situational Awareness for Cyber Defense. *Cyber Situational Awareness*, *46*(1), 3-13.

Bradshaw, J. M., Carvalho, M., Bunch, L., Eskridge, T., Feltovich, P. J., Johnson, M., & Kidwell, D. (2012). Sol: An agent-based framework for cyber situation awareness. *KI-Künstliche Intelligenz*, *26*(2), 127-140.

Cheng, Y., Sagduyu, Y., Deng, J., Li, J., & Liu, P. (2012, May). Integrated situational awareness for cyber attack detection, analysis, and mitigation. In *SPIE Defense, Security, and Sensing*(pp. 83850N-83850N). International Society for Optics and Photonics.

Dark Reading. (2015, March 4). Healthcare Organizations Lack Tools for Cyber Situational Awareness and Threat Assessment. Retrieved February 5, 2017, from http://www.darkreading.com/analytics/healthcare-organizations-lack-tools-for-cyber-situational-awareness-and-threat-assessment/d/d-id/1319344

Dietterich, T. G., Bao, X., Keiser, V., & Shen, J. (2010). Machine learning methods for high level cyber situation awareness. *Cyber Situational Awareness*, 227-247.

Franke, U., & Brynielsson, J. (2014). Cyber situational awareness–a systematic review of the literature. *Computers & Security*, *46*, 18-31.

Friedberg, I., Skopik, F., & Fiedler, R. (2015). Cyber situational awareness through network anomaly detection: state of the art and new approaches. *e & i Elektrotechnik und Informationstechnik*, *132*(2), 101-105.

Jackson, W. (2014, July 25). HHS and health care sector expand cybersecurity info sharing. Retrieved February 5, 2017, from https://gcn.com/articles/2014/06/25/hhs-cybersecurity-hitrust.aspx

Klein, G., Tolle, J., & Martini, P. (2011, August). From detection to reaction-a holistic approach to cyber defense. In *Defense Science Research Conference and Expo (DSR), 2011* (pp. 1-4). IEEE.

Lee, S., Lee, D. H., & Kim, K. J. (2006, December). A conceptual design of knowledge-based real-time cyber-threat early warning system. In *International Symposium on Parallel and Distributed Processing and Applications* (pp. 1006-1017). Springer, Berlin, Heidelberg.

Morris, T. I., Mayron, L. M., Smith, W. B., Knepper, M. M., Ita, R., & Fox, K. L. (2011, February). A perceptually-relevant model-based cyber threat prediction method for enterprise mission assurance. In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2011 IEEE First International Multi-Disciplinary Conference on* (pp. 60-65). IEEE.

NCCIC, & FBI. (2016). GRIZZLY STEPPE – Russian Malicious Cyber Activity. Joint Analysis Report. Retrieved February 5, 2017, from https://www.us-cert.gov/sites/default/files/publications/JAR_16-20296A_GRIZZLY%20STEPPE-2016-1229.pdf.

Reference Number: JAR-16-20296A

Onwubiko, C. (2016). Understanding Cyber Situation Awareness. International Journal on Cyber Situational Awareness, Vol. 1, No. 1, pp11-30.

Regalado, D., Villeneuve, N., & Railton, J. S. (2015). BEHIND THE SYRIAN CONFLICT'S DIGITAL FRONT LINES. FireEye Special Report. Retrieved February 5, 2017, from https://www.fireeye.com/content/dam/fireeye-www/global/en/current-threats/pdfs/rpt-behind-the-syria-conflict.pdf.

Tadda, G. P., & Salerno, J. S. (2010). Overview of cyber situation awareness. *Cyber situational awareness*, 15-35.

BIOGRAPHICAL NOTES

Geoffrey B. Dobson is a member of the Technical Staff with the CERT Cyber Workforce Development Group at the Software Engineering Institute (SEI) and graduate student with the Center for Computational Analysis of Social and Organizational Systems in Carnegie Mellon University's School of Computer Science. Mr. Dobson works with Department of Defense cyber teams in order to model and simulate realistic threats to operations. His work focuses on the team behaviors that lead to effective mission outcomes improving cyber situational awareness for senior military leaders. Mr. Dobson is an Air Force Reserve Cyber Operations officer and holds the rank of Major. He holds a B.A. in Mathematics from Saint Vincent College, a B.S. in Computer Engineering from the University of Pittsburgh, and a Master of Information Systems Management from Carnegie Mellon University.

Dr. Timothy J. Shimeall is a Senior Member of the Technical Staff with the CERT Network Situational Awareness Group at the Software Engineering Institute (SEI). Dr. Shimeall's work draws heavily on data from a variety of large (one million hosts or more) networks. Dr. Shimeall is responsible for overseeing and participating in the development of analysis methods in the area of network systems security and survivability. This work includes development of methods to identify trends in security-related network behavior and in development of methods to detect behavior that provides insight into network intruders and their behavior. Before joining the SEI, Dr. Shimeall was an Associate Professor at the Naval Postgraduate School in Monterey, CA. He received his Ph.D. degree in Information and Computer Science from the University of California, Irvine.

Dr. Kathleen M. Carley is a Professor of Computer Science at the Institute for Software Research, IEEE Fellow, and Director of the Center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University. Dr. Carley's research combines cognitive science, sociology, and computer science to address complex social and organizational issues. Her group has developed tools for extracting sentiment, social and semantic networks from social media and other textual data (AutoMap), simulating epidemiological models (BioWar), simulating convert networks (DyNet), and simulating changes in beliefs and practice given information campaigns (Construct). Dr. Carley received her SB degree in Economics and Political Science from M.I.T., and a Ph.D. degree in Sociology from Harvard University.

REFERENCE

Reference to this paper should be made as follows: Dobson, G. B., Shimeall, T. J., and Carley, K. M. (2017). Towards Network Science Enhanced Cyber Situational Awareness. *International Journal on Cyber Situational Awareness*, Vol. 1, No. 1, pp11-30.