Dark-Net Ecosystem Cyber-Threat Intelligence (CTI) Tool

Nolan Arnold, Mohammadreza Ebrahimi, Ning Zhang, Ben Lazarine, Mark Patton, Hsinchun Chen
Department of Management Information Systems
University of Arizona
{nolanarnold, ebrahimi, zhangning, benlazarine, mpatton}
@email.arizona.edu; hchen@eller.arizona.edu

Sagar Samtani
Department of Information Systems and Decision Sciences
University of South Florida
Tampa, Florida, USA
ssamtani@usf.edu

***Abstract*— The frequency and costs of cyber-attacks are increasing each year. By the end of 2019, the total cost of data breaches is expected to reach $2.1 trillion through the ever-growing online presence of enterprises and their consumers. The tools to perform these attacks and the breached data can often be purchased within the Dark-net. Many of the threat actors within this realm use its various platforms to broker, discuss, and strategize these cyber-threat assets. To combat these attacks, researchers are developing Cyber-Threat Intelligence (CTI) tools to proactively monitor the ever-growing online hacker community. This paper will detail the creation and use of a CTI tool that leverages a social network to identify cyber-threats across major Dark-net data sources. Through this network, emerging threats can be quickly identified so proactive or reactive security measures can be implemented.**

Keywords*— *Dark-Net Ecosystem, Dark-Net Markets, Dark-Net Forums, Cyber-Threat Intelligence Tools, Dark-Net Network Visualization

I. Introduction

DATA breaches are becoming more frequent with the "*rapid digitization of consumers' lives and enterprise records that will increase the cost of data breaches to $2.1 trillion globally by 2019*" [1]. The risk of being exploited through hacker assets increases with the size of the internet of things through the advent of mobile computing and an expected 46 trillion connected devices in use by 2021 [1].

These hacker assets can be purchased and discussed within communities of threat actors across the Dark-net. Despite the growing interest in the Dark-net, the scope and scale of discoverable malicious hacker assets is unclear.

This study aims to create a versatile cyber-threat intelligence tool through a comprehensive multi-node network that identifies threats across major Dark-net data sources. Cyber-threat assets will be linked to threat-actors using text features found across multiple Dark-net data collections performed by The University of Arizona's Artificial Intelligence Lab.

By examining these features and connections, this ecosystem can be scaled while emerging threats can be quickly identified so proactive or reactive security measures can be implemented based on the threat landscape.

This report will cover the literature review, research design, analytical approaches, results and future directions for this project.

II. Literature Review

This project's literature review will cover research efforts made within the areas of cyber-threat intelligence (CTI) and social network analyses in regard to the Dark-net, and the identified research gaps and questions.

A. Social Network Analysis

Examining the structure of the Dark-net ecosystem is an important exercise for gaining domain knowledge of these threat communities. The underlying processes across these sites such as trading, and vetting remain poorly understood. Using an Event Analysis of Systemic Teamwork (EAST) approach can be beneficial in exploring the implications in trying to understand the complex ecosystem while trying to identify vulnerabilities for potential disruption [2].

This approach for understanding the Dark-net ecosystem from a domain perspective can be useful when applying a social network-based analysis across multiple datasets. However, EAST's activity centric approach does not prioritize the content of these interactions, thus performing a social network analysis for this domain may expand upon these findings for the creation of actionable CTI.

The area of deep web social network visualization is a relativity untapped field. Researchers have mainly focused on plotting data more towards a surface level, such as mapping .onion sites via how they are connected to one another through URLs. An example of this is Hyperion Grey's Dark Web Map [3]. This overly expansive approach is a useful exercise for getting an idea of how this decentralized network operates in its entirety. However, such approaches rarely include the actual content of the sites. This lack of granularity prevents researchers from producing even moderately actionable CTI findings.

B. Cyber-Threat Intelligence

The secondary focus of this project is the identification of cyber-threats through breach forensics across the Dark-net. However, the area of hacker related assets on Dark-net Markets (DNMs) has received less attention than other parts of these digital black markets. This is due to the online economy for narcotics being often prioritized by law enforcement as it is believed to be a more widespread issue. Nevertheless, the relationship between vendors, buyers, and post authors remains consistent across these areas [4].

Details within product listings, feedback and threads can paint an accurate picture of the emerging threats and actors

978-1-7281-2504-6/19/$31.00 ©2019 IEEE

within this realm. Through a text centric focus, several different approaches can be used to find relevant information on the prevalence of malicious tools, services, actors, and breach data.

The resource that was the main influence for this area of the project was an analysis conducted by Ryan Compton in 2015 that aimed to examine co-occurrence relationships between threat actors and the products they offered [5]. This method used a stochastic block model-based hierarchal edge bundling to generate a visualization of the evolution product network.

C. Research Gaps & Questions

Two gaps were identified following this literature review. Within the scope of Dark-net analyses, the examination of cyber-threat assets and the communication surrounding them across multiple data sources has remained relatively unexplored through analytical approaches. The second gap that was found regarded the extent of pre and post obtainable data following breaches and their victims remains unclear. Based on these gaps, the following questions are posed;

1) What is the extent of communication and cyber-threats across hacker communities?
2) Which types of cyber-threat assets is the Dark-net ecosystem trending towards?

III. Research Design

The research design for this project followed four major phases; data collection, threat-identification, threat profiling and visualization. Figure 1 depicts this research design.

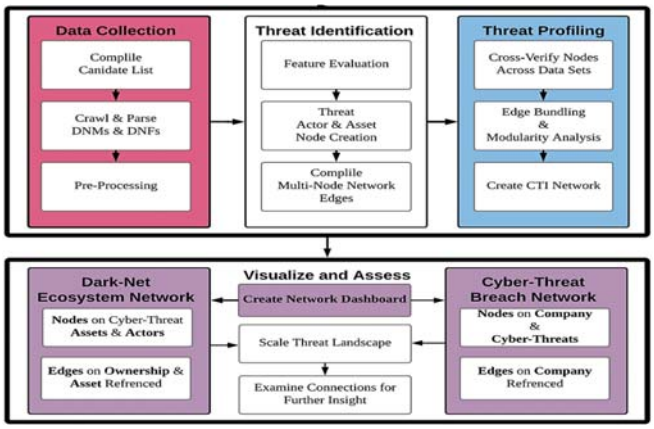

Fig. 1. Research Design*: Dark-Net Ecosystem Network Analysis*

A. Data Collection

The data collection phase began with an initial draft of potential Dark-net markets (DNMs), Dark-net Forums (DNFs) and exploit databases to be collected that was based on each site's number of listings, threads, threat actors as well as the number of cyber-threat related listings. From this list, python-based web crawlers were developed to collect each site.

Our team's crawlers initially faced issues from several sites' crawling-prevention measures. This issue was solved by using ten crawlers all linked to different accounts. With these ten crawlers running in unison, each sites' logs were congested enough to prevent the occurrence of captcha codes.

This collection spanned from October 2017 – January 2019 that includes the five largest DNMs that provided 224,270 product listings and 7,911 vendors, the two largest exploit databases (*Exploit DB/0day.Today*) providing 43,678 exploit listings and three major DNFs providing 204,001 threads and 14,196 authors. A total of 112 categories are found across these ten datasets.

1) Data Sources

The two major data sources for this Social Network analysis of Dark-net communities is Dark-net forums as the primary data source for observing threat actor interactions and Dark-net markets as the cyber-threat asset product network. The second network produced is focused on the relationship between cyber-threat assets and exploit database listings linked by companies to make a breach forensic network. Below are brief explanations of the data sources that were used to create these multi-node networks.

a) Dark-Net Forums

These community hubs are hot beds for threat actor communication and activity across various CTI data sources ranging from DNMs, carding shops, independent vendors, contractual hackers and much more. Dark-net forums are text rich and contain contextual and time series data that can be linked to other data sources for additional knowledge gain.

b) Dark-Net Markets

These deep-web-based e-commerce sites function primarily as black markets where users create profiles to buy or sell drugs, breach data, forged documents, currency, hacking tools, guides, as well as many other illegal and legal goods. In order to keep transactions anonymized and integrable, Dark-net markets typically adopt an escrow system where cryptocurrency is laundered through the site and is released following buyer feedback via product reviews.

c) Exploit Databases

Within both the surface web and deep web are databases dedicated to revealing exploits discovered across various operating systems, applications, sites, etc. These databases include exploits, shellcode, 0days and much more. Most of these listings in their entirety are free for users while the lesser known and more visceral listings are available for purchase via cryptocurrency. Table I details the DNF & DNM datasets used to create the multi-node social network.

Table I. Network Percentages by Site

Site	Type	Language	Total %	Actors
Rutor	DNF	RU	28.8%	8,318
Dream	DNM	EN	27.77%	2,958
Wallstreet	DNF	EN	20.66%	5,382
TradeRoute	DNM	EN	18.71%	4,201
FrenchDeepWeb	DNM	FR	1.33%	261
Silk3	DNF	EN	1.96%	429
Tochka	DNM	EN	0.83%	213
Valhalla	DNM	EN/FIN	0.27%	79

A. Threat Identification

The decision to use a multi-node approach was made in order to maintain a robust network that is granular and modular for any additional datasets. This made the edge creation stage exhaustive in order to identify any connections that were not immediately apparent within the initial testbed. DNF data was used as the base of this network due to thread authors being able to provide versatile content that can add value were other data-sets may lack.

An example that inspired this project was the universal absence of time series data across DNM listings that stunted the CTI credibility of the data. However, many DNF threads are used as a promotion or review platform for specific products for sale on their respective DNMs. By linking these datasets, not only can time data be added to product data, but data can also be derived via sentiment, credibility, buyer names as well as an estimated quantity sold. Table 2 details the data dictionary used to create the multi-node network between DNMs and DNFs.

The breach forensics network aims to identify if the start and end points of recent data breaches can be linked to Dark-net data sources. Start points include potential exploits or user authentication means that could be used to execute a breach

TABLE II. NETWORK MAKEUP BY SITE

Type	Attribute	Value	Connection
CTI Asset	Category	Feature	Category
	Subforum	Feature	Category, company
	threadTitle	Node, Edge	sellerName, productName
	Postdate	Feature	Product
	productName	Node, Edge	authorName, company productName, buyerName,
	flatContent	Feature	sellerName, buyerName, authorName, company
Threat Actor	authorName	Node, Edge	sellerName, buyerName
	dateJoined	Feature	buyerName
	sellerName	Node, Edge	authorName, productName
	lastActive	Feature	buyerName
	buyerName	Node, Edge	authorName, productName, buyerName

= Feature = Node & Edge

Due to the obscurity and lack of consistency within listing names, classification rendered few results. Using SQL queries, text from each DNM product and DNF thread could be accurately pulled to discover company names. 132 major companies were identified across the three data sets, with notable industries such as banking, e-commerce, airlines and much more. Examples of some of the more prevalent companies are Amazon, PayPal, and Microsoft.

B. Threat Profiling

1) Dark-Net Ecosystem Social Network

From working extensively with the data to create every possible relationship, an initial outline of the graph was made that would better align with the CTI relevant features that needed to be included for this project's analysis. Because of this, the threat profiling stage acted as a secondary data preprocessing phase to eliminate any unneeded noise that arose following the node and edge creation tasks.

This preprocessing was done by plotting the node and edge data into Gephi and running initial network statistics to generate the degree, rank, closeness and betweenness for each node. Through focusing on relevant threat actors while using each node's statistics, the initial graph was filtered down from 450,378 to 167,763 nodes. Most of the eliminated nodes stemmed from products, or threads with minimal community involvement (few authors connected) or having no apparent CTI value (category associated with narcotics or legal goods).

This final dataset created three networks, each with a varying focus. 'All' represents the network as a whole, 'Threat-Asset' represents products and threads that fall into one of the 12 meta CTI categories, and 'Breach' that represents products and threads with a company name found within the data. Table III summarizes these networks.

TABLE III. NETWORK STATISTICS: DARKNET ECOSYSTEM SOCIAL NETWORK

Value	All	Threat-Asset	Breach
Nodes	167,763	38,894	12,001
Edges	219,412	52,859	14,755
Products	129,289	26,680	6,834
Threads	16,611	4,609	1,039
Vendors	7,851	7,758	2,223
Authors	13,990	13,985	1,919
Top Features	Cross Market Vendors = **523** Author/Vendor Matches = **3,502**	Fraud P.I.I Hacking Accounts	PayPal Microsoft Apple Amazon

2) Cyber-Threat Breach Forensic Network

To create a hierarchical network, nodes and edges were created based on categorical queries that pulled only relevant CTI data that provided 97 nodes across the 5 DNMs, 3 DNFs and 2 Exploit Databases. The data used for this project consisted of:

- 38,894 Asset Listings
- 36,606 Exploit Listings
- 236 Threat Actors
- 12 DNM Categories
- 18 Exploit Categories
- 67 At Risk Companies

Edges were based on the company name pulled for each listing while weight was based on frequency of each relationship (i.e. company -> category). Nodes were ordered and heat-mapped on calculated degree.

C. Visualization

In order to maintain the desired finesse of a practical and actionable CTI tool, the ability to add any missed or new data would need to be a vital feature. Through this consideration, an interactive network dashboard was built using Gephi. This powerful open sourced tool is the best user-friendly network analysis and visualization tool for big data. Through creating a thorough dashboard, additional data can be streamed directly into a network and exported for fast targeted CTI analyses.

IV. Analytical Approach

Through the research design for this project, two network visualizations were created to examine the relationships between cyber-threat assets, companies, exploits and the threat actors profiting from them.

A. Dark-Net Ecosystem Social Network

The Dark-net ecosystem social network is the primary analysis for this project. An expansive approach was used initially in forming this network that used the most data possible to create an accurate and granular depiction of this major facet of the Dark-net. With this goal, the network was filtered down in three steps, each one furthering the central CTI focus of this project. Below are the three views of this network.

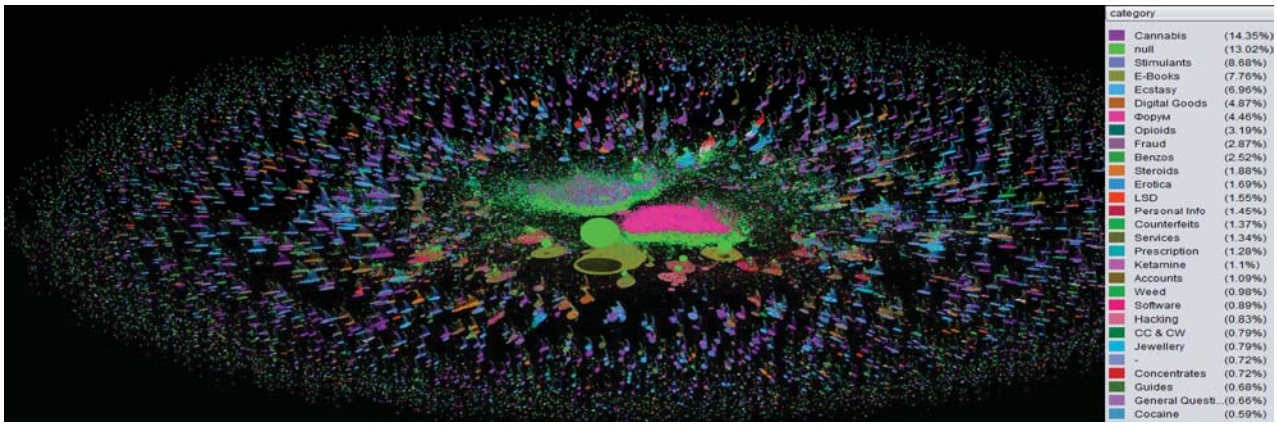

Fig. 2. Analytical Approach*: Darknet Ecosystem Social Network. - (Green nodes are threat-actors; other colors are categories)*

1) Dark-Net Ecosystem Network

This first network view includes all categories of products and threads while only filtering out threat actors that were weakly connected, low contributing or attached to non-relevant nodes in the product network. This network view works less as a targeted cyber-threat analysis tool, but rather a tool to study the interaction and overlap between threat assets and actors within the Dark-net ecosystem. The visualization for this network can be seen in Figure 2.

2) Dark-Net Cyber-Threat Assets Network

The second network extracted shows the cyber-threat asset focus that was the original foundation of this project's CTI goals. The 112 categories have been filtered down to the 12 most relevant CTI fields. These 12 have been condensed to high-level meta categories such as fraud, hacking, and accounts. The main value of this graph is as a tool for target analysis for specific forms of cyber-threat assets. This network can be seen in Figure 3.

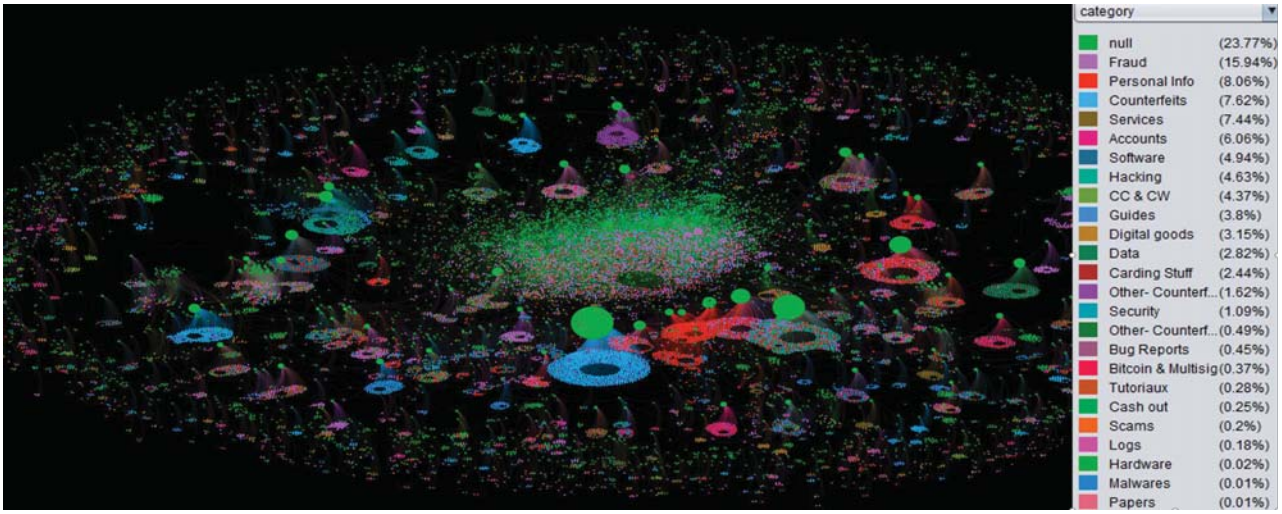

Fig. 3. Analytical Approach*: Darknet Ecosystem Cyber-Threat Network (Green nodes are threat-actors; other colors are security categories)*

2) Breach Network

The final social network filtered is concerned with the prevalence of company related names found within DNM products, DNF threads and their attached threat actors. This social network's main strength is that it leverages various companies' presence across the Dark-net ecosystem.

This shows the spread of information as well as number of involved threat-actors. This network will feed directly into the breach forensics analysis to show a more quantitative approach for cyber-threats on a company by company basis. The visualization for this network can be seen in Figure 4.

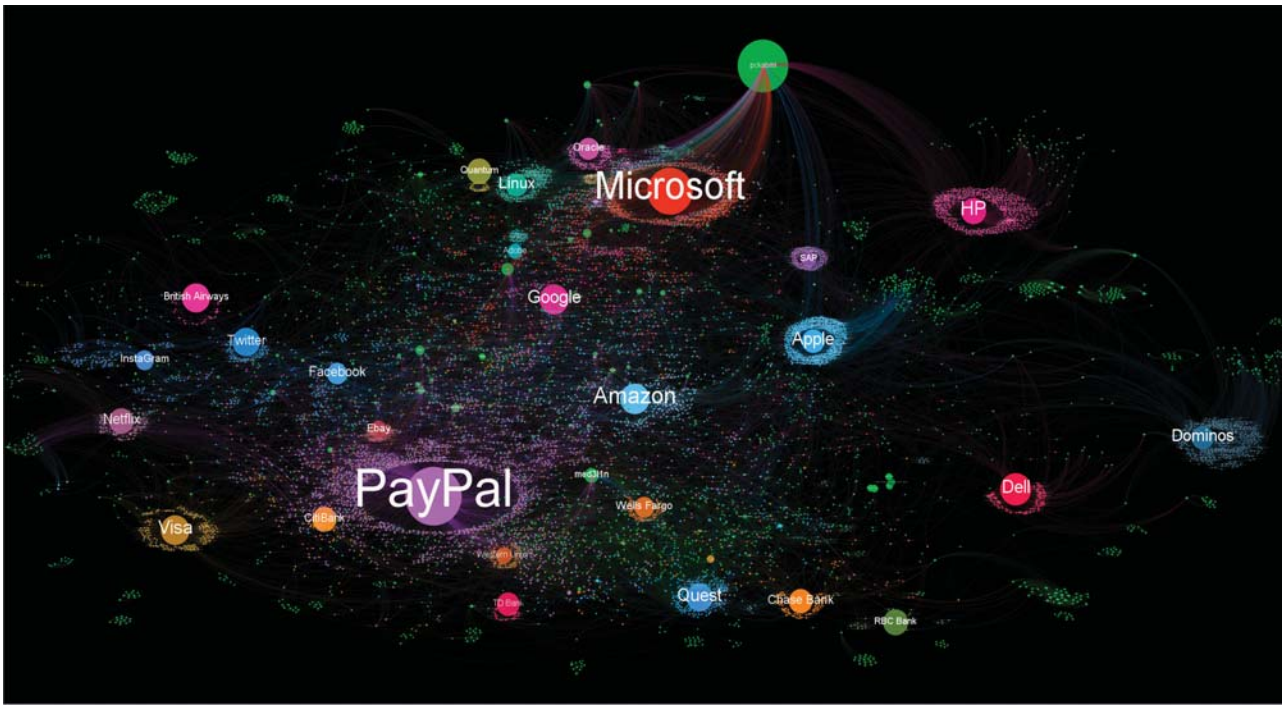

Fig. 4. Analytical Approach*: Darknet Ecosystem Breach Network (Green nodes are threat-actors; other colors are company related products or threads)*

B. Breach Forensics Network

The second analysis produced focused on the relationship between cyber-threat assets and exploits linked by companies. This network used companies as the outer nodes while the inner nodes are attributed to the category of cyber-threat asset and exploit.

Nodes are sized and heat-mapped based on quantity. Edges are formed on company names pulled within each listing while weight is scaled by the frequency of each relationship. Figure 5 depicts this network.

V. Results & Future Direction

A. Results

Fraud is the most populated category through its connection to 47 of the 67 companies used within this edge bundled network. Breached accounts and hacking tools follow closely behind as cyber-threats that are currently available to be used to attack companies' and their customers. Through looking at a notable breached company, PayPal has a weight 238% larger than the next highest company node with listings spanning across all 12 DNM categories and five exploit categories.

Multiple inferences can be made from examining the results within these networks. Some of the more matured breaches have a larger spread within DNMs that span across several different categories between cyber-threat assets and exploits. However, by examining the newer breached companies with few connections, inferences regarding initial threat actors behind the breach or early buyers of their data can be made.

B. Future Directions

With the current dashboard, backend and queries built, additional data from an updated collection or new source can be added quickly to have the results graphed within minutes. The best feature to future proof this tool would be to couple it with an automated crawler and parser that would incrementally stream new collections directly into these networks.

Currently, a stand-alone scalable dashboard has been developed that allows users to quickly visualize and query via search or click. A 3D navigation dashboard is now being developed in Cytoscape for a more exploratory and interactive user experience. With 3D navigation as an option, the digital targeting process can be automated by following the path of a set connections with each node's data being placed in a coherent structure to build a narrative for any targeted analysis.

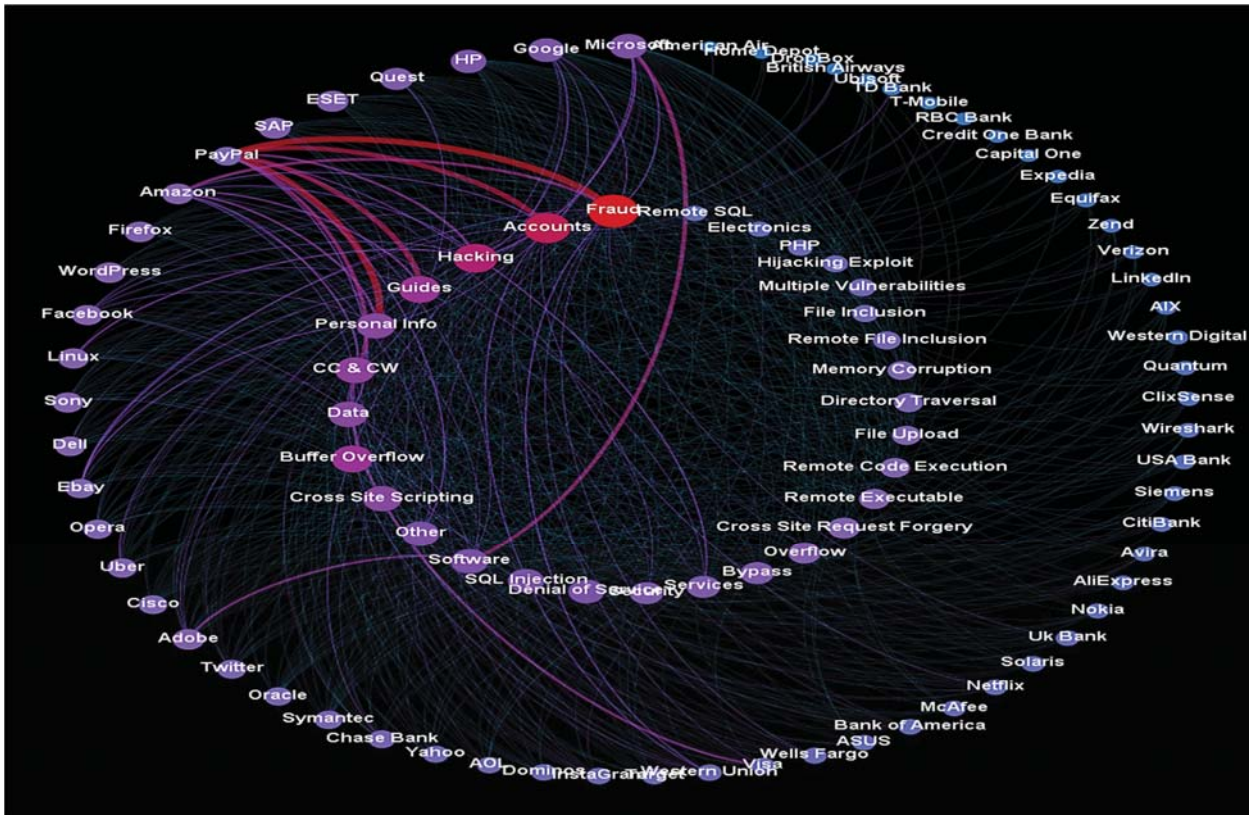

Fig. 5. Analytical Approach*: Cyber-Threat Breach Network. (Sized and heat-mapped based on frequency)*

VI. Conclusion

The information conveyed across these networks creates a coherent depiction of this major facet within the Dark-net ecosystem. By examining the connections within each network, researchers will be able to find CTI related findings that may have been overlooked through initial analyses that take an overly expansive or focused approach.

Through this dashboard, CTI researchers can perform:

- Fast integrations of additional data.
- Targeted analyses across multiple datasets.
- Automated approaches to digital targeting process.

The effectiveness of this CTI tool needs to be evaluated outside of this project. However, by examining the results within this study, the economy of the Dark-net and breach victims can be better scaled and examined for granular analyses. The versatility of this tool will guarantee its relevancy and efficiency for any future research within the often-hectic area of cyber-threat intelligence.

VII. ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. NSF DUE-1303362 (SFS), NSF SES-1314631 (SaTC), and NSF CNS-1850362.

References

[1] J. Moar, "Cyber Crime and the Internet of Threats*"* Juniper, *2017*.
[2] B. Lane, D et al. "The Dark Side of the Net: Event Analysis of Systemic Teamwork (EAST) Applied to Illicit Trading on a Darknet Market." Proceedings of the Human Factors and Ergonomics Society Annual Meeting Vol.62 Iss.1. Sage Journals, (2018).
[3] M. Hasse, "Dark Web Map" Hyperion Grey, (2019).
[4] T. Kaskela, A. Oksanen, "Seller's reputation and capacity on the illicit drug markets: 11-month study on the Finnish version of the Silk Road" Research Gate, (2017).
[5] R. Compton, "Darknet Market Basket Analysis" (2015).
[6] H. Lawrence, A. Hughes, R. Tonic, C. Zou, "D-miner: A Framework for Mining, Searching, Visualizing, and Alerting on Darknet Events" IEEE. University of Central Florida, (2017).
[7] T. Reksna, D. Garlaschelli, F. Takes, "Complex Analysis of Darnket Black Market Forum Structure." Huygens-Kamerlingh Onnes Lab. Leiden University, (2017).
[8] J. Robertson, A. Diab, E. Marin, E. Nunes, V. Paliath, , J. Shakarian, "Darkweb Cyber Threat Intelligence Mining." Cambrige University Press, (2017).
[9] A. Al-Iman, B. Abdulmajeed, "The NPS Phenomenon and the Deep Web: Internet Snapshots of the Darknet and Potentials of Data Mining." Global Journal of Health Science, (2017).
[10] T. Holt, O. Smirnova, A. Hutchings, "Examining signals of trust in criminal markets online." Oxford Journal of Cyber Security, (2016).
[11] M. Eric, J. Michael, M. Rohan, "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains.*"* Lockheed. (2016).
[12] V. Ray, "Exploring the Cybercrime Underground: Part 4 – Darknet Markets." Poloalto Networks, (2017).
[13] S. Samtani, H. Chen, "Identifying Hacker's Targets for Proactive Cyber Threat Intelligence: A Deep Structured Semantic Model Approach", University of Arizona, (2017)