Incremental Hacker Forum Exploit Collection and Classification for Proactive Cyber Threat Intelligence: An Exploratory Study

Ryan Williams
Management Information Systems
University of Arizona
Tucson, AZ, United States
ryanwilliams12@email.arizona.edu

Sagar Samtani
Information Systems and Decision Sciences
University of South Florida
Tampa, FL, United States
sagars@email.arizona.edu

Mark Patton
Management Information Systems
University of Arizona
Tucson, AZ, United States
mpatton@email.arizona.edu

Hsinchun Chen
Management Information Systems
University of Arizona
Tucson, AZ, United States
hchen@eller.arizona.edu

***Abstract*— Cyber threats have emerged as a key societal concern. To counter the growing threat of cyber-attacks, organizations, in recent years, have begun investing heavily in developing Cyber Threat Intelligence (CTI). Fundamentally a data driven process, many organizations have traditionally collected and analyzed data from internal log files, resulting in reactive CTI. The online hacker community can offer significant proactive CTI value by alerting organizations to threats they were not previously aware of. Amongst various platforms, forums provide the richest metadata, data permanence, and tens of thousands of freely available Tools, Techniques, and Procedures (TTP). However, forums often employ anti-crawling measures such as authentication, throttling, and obfuscation. Such limitations have restricted many researchers to batch collections. This exploratory study aims to (1) design a novel web crawler augmented with numerous anti-crawling countermeasures to collect hacker exploits on an ongoing basis, (2) employ a state-of-the-art deep learning approach, Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN), to automatically classify exploits into pre-defined categories on-the-fly, and (3) develop interactive visualizations enabling CTI practitioners and researchers to explore collected exploits for proactive, timely CTI. The results of this study indicate, among other findings, that system and network exploits are shared significantly more than other exploit types.**

Keywords—cyber threat intelligence, hacker forum, CTI, web crawling, recurrent neural network, hacker exploits

I. INTRODUCTION

Cyber threats continue to evolve in hopes to subvert security measures employed by security professionals. Amongst possible victims, organizations are particularly high value targets for cyber-attacks due to the valuable data they manage. The average organizational cost of a data breach for U.S. companies in fiscal year 2017 was $7.35 million [1]. Breaches can be caused by any number of different vulnerabilities. While system glitches and human negligence can be contributing factors, attacks from malicious criminals are more prominent, contributing up to 47% of breaches [1].

To counter the growing threat of cyber-attacks, organizations in recent years have begun investing heavily in developing Cyber Threat Intelligence (CTI). Fundamentally a data driven process, many organizations have traditionally collected and analyzed data from internal systems such as security information and event management systems (SIEMs), log files, Network Intrusion Detection and Prevention Systems (NIDS/NIPS), and others to provide insights into emerging threats and key threat actors. Despite their value and prevalence, data collected from internal systems are considered reactive CTI.

Open Source Intelligence (OSINT), or intelligence collected from publicly available sources, can offer significant value to proactive CTI by alerting organizations to threats they were not previously aware of [2]. One emerging OSINT data source is the online hacker community. Comprised of hacker forums, DarkNet Markets, carding shops, and Internet-Relay-Chat (IRC), the online hacker community enables millions of hackers from multiple geo-political regions such as China, US, and Russia to share malicious tools and knowledge. Amongst the four, forums provide the richest metadata, data permanence, and freely available Tools, Techniques, and

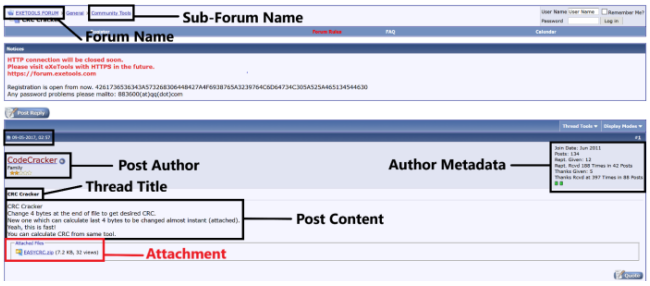

Fig. 1. Example Hacker Forum Posting

Procedures (TTP) [3].

Users on hacker forums regularly share cyber-attack assets such as attachments, source code, and tutorials [3]. One well-known example is the BlackPOS malware, used in Target and Home Depot breaches. Shared in hacker forums months before the attacks, BlackPOS remains available in forums for hackers to freely download [4]. As such, it continues to pose a threat to other organizations with similar vulnerabilities. Fig. 1 provides another example of a hacker attaching a CRC cracker.

Identifying and analyzing attachments found in hacker forums can directly inform proactive CTI and mitigation. Altogether, hundreds of hacker forums exist that accommodate hundreds of thousands of members who generate millions of posts containing tens of thousands of malicious assets. Despite their CTI value, forums are notoriously difficult to collect compared to traditional OSINT social media data (e.g., Twitter). Forums often employ numerous anti-crawling measures such as authentication, throttling, and obfuscation. Such limitations have restricted many researchers to batch forum collections, either manually or automatically. While this approach may be suitable for other OSINT applications, CTI is a domain where timeliness is of utmost importance. Hackers develop new exploits at staggering rates. As a result, the value of knowledge for existing exploits rapidly decays.

Given the volume, variety, and velocity of hacker forum data, this exploratory study aims to (1) design a novel web crawler augmented with numerous anti-crawling countermeasures to collect hacker exploits on an ongoing basis, (2) employ a state-of-the-art deep learning approach, Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN), to automatically classify exploits into pre-defined categories on-the-fly, and (3) develop interactive visualizations enabling CTI practitioners and researchers to explore collected exploits for proactive, timely CTI.

The remainder of this paper is organized as follows. First, we review literature identifying prevailing web forum crawling methods and existing approaches to collecting hacker forums. Following our review, we identify key research gaps and pose research questions for study. Section IV details our research design, section V summarizes our experiment results, collected data, and highlights our visualizations, and section VI points to some promising directions for future research and concludes this work.

II. Literature Review

Two streams of literature are reviewed to form the basis of this research. First, web forum crawling is reviewed to identify techniques in navigating and indexing web-based forums. Last, previous hacker forum collection efforts are studied to examine prevailing efforts to collect hacker forum data.

A. Web Forums: Characteristics and Crawling Strategies

Internet forums are online message boards designed for users to interact with each other through discussion of various topics. While many forum frameworks exist (e.g., vBulletin, phpBB), all forums follow a similar tree structure. A forum contains multiple sub-forums, each dedicated to a specific topic. Within sub-forums, users can start threads, or discussions related to a specific sub-topic. Users converse in threads by creating posts. Each post contains the username of the poster, the date the post was created, and the post content. Depending on the underlying forum technology, users can embed text, images, videos, links, source code, and attachments into their posts. Forums are naturally archiving; each post remains permanently accessible unless forum administrators or original posters remove it.

Like other web platforms, forums are built using HTML. Consequently, it is possible to navigate and collect their contents via web crawlers. Web crawlers are software programs that traverse the internet by following hyperlinks and collect web pages using the HTTP protocol [5]. Their usual intention is to create a local index of web pages [5]. Traditional web crawlers adopt a Breadth-First Search (BFS) strategy to navigate hyperlinks [6]. However, the unique characteristics and structure of forums make generic web crawlers employing this strategy inefficient for data collection [6]. The tree structure of forums often results in duplicate pages/links, uninformative pages, and page-flipping links. To counter these issues, past studies have primarily employed URL parsing and analytics combined with regular expressions to target specific areas of web forums for collection [6][7].

B. Hacker Forum Collection Efforts

While hacker forums are built on similar frameworks as traditional forums, they employ an array of anti-crawling measures designed to hinder large-scale, automated collection. Common mechanisms include authentication, turing tests, throttling, obfuscation, and network traffic analysis [8]. Some forums also implement drive-by malware to infect users as they browse the site. Efforts have been made to subvert anti-crawling measures. For example, human intervention can help crawlers pass authentication tests such as CAPTCHA, artificially limiting the speed at which the crawler requests new web pages can prevent the targeted site from becoming suspicious, and manually creating a new session by restarting the crawling process frequently can allow continuous collection [8]. Table 1 summarizes selected studies leveraging these techniques to collect hacker forums for CTI applications.

TABLE I. Selected Previous Hacker Forum Collection Efforts

Authors	Forums Crawled	Data Collected	Collection Procedure
Macdonald et al., 2015	**1**	**Unspecified number** forum posts	Batch
Samtani et al., 2015	**5**	**3,251** attachments; **14,944** source code files; **671,633** forum posts	Batch
Benjamin et al., 2015	**10**	**99,353** forum posts	Batch
Nunes et al., 2016	**21**	**162,872** forum posts	Batch
Grisham et al., 2017	**4**	**43,462** attachments; **481,922** forum posts	Batch

Previous efforts to collect hacker forums have utilized batch collection methods to crawl and gather data. This means the forum is collected all at once, often without any intention of re-crawling the forum later on to collect newly generated posts.

Following collection, these studies have analyzed assets [3][5][9] and user posts [10][11][12] to develop CTI. Should the authors wish to gather newer data to develop more up-to-date CTI insights, their strategy would require them to download the entire forum again and analyze only the postings which were made between the previous and current collections. This strategy is inefficient, time-consuming, and hinders the development of valuable CTI (e.g., emerging threats, hackers). Moreover, it increases the researcher's likelihood of being discovered, blocked, and targeted by the hackers. These drawbacks motivate the development of novel incremental crawling and on-the-fly analysis of hacker forum data.

III. Research Gaps and Questions

Several research gaps were identified from our literature review. First, current CTI efforts focus heavily on the use of internal information to generate threat intelligence. This means current security measures are often handled reactively instead of proactively. Second, web forum characteristics make traditional web crawling techniques ineffective for efficient navigation and indexing. Finally, previous hacker forum collection efforts have been primarily focused on batch collection and processing of hacker forum data. As threats evolve over time, these static collections become less insightful. With these research gaps in mind, the following research questions have been proposed to guide the study:

- How can an incremental crawling strategy bypassing hacker forum anti-crawling measures be developed to collect hacker forum attachments on an ongoing basis?
- How can hacker attachments be collected and analyzed simultaneously?
- How can visualizations be developed that display up-to-date hacker forum data which provide proactive CTI?

IV. Research Design

We developed a four-phased research design (Fig. 2) to answer the proposed research questions. The four phases include forum identification and crawler initialization, incremental hacker forum crawling, attachment analysis, and data storage and visualization. Each phase is detailed in the following sub-sections.

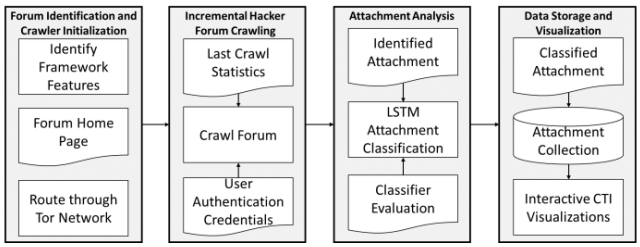

Fig. 2. Research Design

A. Forum Identification and Crawler Initialization

This phase focuses on the initial crawler configuration. Our literature review revealed that a traditional web crawling approach cannot directly apply to crawling hacker forums. Moreover, an incremental approach to asset collection provided a number of challenges. It is not enough that the crawler be able to collect every attachment on a hacker forum, it must also revisit that hacker forum and only collect new attachments that it has not seen before. To accomplish this, the crawler utilized the underlying forum framework (HTML structure) and metadata (e.g. post date) to target specific areas of the forum that needed to be crawled. Each forum framework has a unique HTML structure and naming schemes. Manual exploration of 74 hacker forums revealed that 21/74 (28.38%) were vBulletin, 15/74 (20.27%) XenForo, 9/74 (12.16%) Invision, 6/74 (8.10%) MyBB, and 5/74 (6.76%) phpBB. Based on this finding, the crawler was tailored for the vBulletin framework. Since forums generally follow a similar structure, this crawler could be modified to crawl other frameworks by changing the specific HTML tags it searches for on each page.

A total of 10 hacker forums were identified for collection: OpenSC, Garage4hackers, Hacksden, AntiOnline, Crackingzilla, WebCracking, SafeSkyHacks, Ashiyane, Hack, and Haker. Beyond their usage of vBulletin these forums were chosen because they allow users to embed attachments directly into their posts and can be accessed without monetary payment. If the forum requires an account to access its content, the user can create an account and provide the credentials to the crawler. All traffic is routed through Tor, a network of servers running specialized software that provide anonymity for the user. This ensures that the crawler remains anonymous while also allowing it to access content hidden from the surface web. After the connection to Tor is successfully established, the crawler begins searching for attachments.

The crawler itself is written in the Python programming language. Python provides a great number of libraries dedicated to web crawling, HTML parsing, data science, and other important functionalities needed for this study. The bulk of the crawler's functionality is provided by the "requests_html" library for HTTP requests, the "BeautifulSoup" library for HTML parsing, and the "keras" [13] library for classification.

B. Incremental Hacker Forum Crawling

This phase of the research design focuses on the incremental approach to forum crawling. The crawler uses a Depth-First Search (DFS) approach to collecting hacker forums. As opposed to BFS, DFS exhaustively explores each branch before moving on. In the context of hacker forums, the crawler starts with one sub-forum and crawls each topic and post within that sub-forum before moving on to the next sub-forum. Based on the structure and naming schemes of the vBulletin framework's HTML code, the crawler targets specific links on the page and follow a systematic approach to attachment collection.

Starting at the home page, the crawler will identify all of the forum's sub-forums and begin crawling each sub-forum one by one. If a sub-forum is contained within another, which is common with vBulletin forums, the crawler will also collect those. As forums are characterized by their data richness, a great deal of information that can be parsed from the HTML code and utilized to make the crawler more efficient than traditional BFS crawlers. For instance, within each sub-forum,

user-made topics are already listed in chronological order by most recent activity. With this information, the crawler can quickly determine when the sub-forum was last active and compare that date to the date that sub-forum was last crawled. If there is any new activity, the crawler only parses the new user topics and moves to the next sub-forum. Within each topic, posts are also ordered chronologically, with the oldest posts listed first. To maximize efficiency, the crawler starts at the very last post within a thread. With this strategy, the crawler can quickly identify if any new posts have been made in topics it has already crawled.

Another piece of information that forums usually provide is whether an attachment exists within a user topic. As the focus of this research is collecting attachments in hacker forums, knowing which user topics contain attachments significantly increases the efficiency of the crawler. Once a new topic is found and it is discovered to contain an attachment, the crawler will enter that topic and begin parsing through user posts to find and collect the attachment's information.

C. Attachment Analysis

This phase is composed of two major sections. The first section focuses on training a Recurrent Neural Network (RNN) for classifying hacker forum attachments. An RNN was chosen for classification because of its strong performance in classifying hacker forum text [9]. To determine the best performing RNN for hacker attachment classification, three RNN variants were benchmarked on a gold-standard set. These include a standard RNN, a Gated Recurrent Unit (GRU) RNN, and an LSTM RNN. The data used to train and test these RNNs was provided by previous research focused on collecting and classifying attachments on hacker forums [3]. The gold-standard set is a collection of 15,123 hacker forum attachments and their associated metadata (e.g. sub-forum name, author name, post content) retrieved from the AZSecure Hacker Assets Portal [4][14]. The developers of the portal authors used Latent Dirichlet Allocation (LDA) to identify themes for each attachment post. Based on the LDA results, attachments were labeled by exploit type [3][9]. The final exploit categories were system, web, network, database, and mobile.

The second section of this phase occurs simultaneously with forum crawling. As the crawler collects new attachments, its metadata is immediately passed to the trained LSTM RNN for classification. Once the LSTM RNN has classified the attachment, the information is parsed into a database for storage and further analysis.

D. Data Storage and Visualization

The last phase of the research design focuses on extracting valuable information that an up-to-date hacker forum collection can provide CTI. Once attachments have been collected, classified, and stored, this information can be analyzed to gain further insights. This study focuses on two main areas for analysis, author activity and exploit postings. To accomplish this, Tableau is used to visualize characteristics within the data and help identify emerging threats.

V. Results And Discussion

A. Attachment Classification Results

All RNN classifier variants were trained using hold-out validation, where the model was trained on 80% of the data and tested on the remaining 20%. Model performance was evaluated using the precision, recall, F-Measure metrics. Benchmarking in this fashion is a commonly accepted practice in computer science literature. Results of our experiment are summarized in Table 2.

TABLE II. RNN Models Benchmark

Model	Precision	Recall	F-Measure
LSTM	97%	98%	98%
GRU	96%	96%	96%
RNN	90%	90%	90%

Overall, the basic RNN, GRU RNN, and LSTM RNN achieved F-Measures of 90%, 96%, and 98% respectively. Given LSTM's superior performance, we leveraged the trained LSTM RNN for classifying new attachments collected by the crawler.

B. Data Collection Summary

For this study, attachments were only collected if they were directly embedded in the user's post. External attachments that are hosted on third-party sites were ignored because contextual information (e.g. file name) that the RNN uses during classification was not readily available. It was also not guaranteed that the hyperlinks provided lead to the supposed exploit. Table 3 summarizes the 10 forums that were crawled and collected.

TABLE III. Forum Collection Statistics

Forum	Language	# of Sub Forums	# of Threads	# of Posts	# of Assets
Hacksden	English	70	10,359	61,534	77
Crackingzilla	English	61	11,451	167,206	1
Garage4Hackers	English	47	1,544	8,620	51
OpenSC	English	56	22,897	184,211	1,179
AntiOnline	English	39	14,771	160,897	77
WebCracking	English	109	7,832	92,025	7
SafeSkyHacks	English	100	12,780	31,733	89
Ashiyane	Arabic	49	65,251	538,708	1,388
Hack	Polish	52	10,452	63,515	52
Haker	Polish	34	924	9,374	9
Total:	**-**	**617**	**158,261**	**1,317,823**	**2,930**

Ashiyane contained significantly more threads, posts, and attachments. OpenSC also contained significantly more attachments than most of the other forums despite its average number of threads and posts. A few forums contained a surprising lack of post attachments. Crackingzilla, WebCracking, and Haker hosted less than 10 attachments each while containing a large amount of threads and posts. This could point to different policies on sharing attachments that prevent attachments being directly embedded within posts.

In total, 2,930 attachments were collected and classified from the 10 hacker forums. System exploits make up a

majority of the attachments shared on hacker forums at 59.32%. These attachments cover a wide range of different exploits including crypters, keyloggers, and Remote Access Trojans (RATs). Network exploits also make up a significant chunk of the shared attachments at 31.06%. This category includes exploits such as botnets and Distributed Denial of Service (DDoS) attacks. Web (e.g. XSS, SQL injection), database (e.g. MySQL attack), and mobile (e.g. Android attack) exploits comprise a relatively small number of attachments being shared at 9.6% combined.

C. *Threat Identification*

With the attachments collected and classified, a wide variety of insights can be extracted from the data. For this study, visualizations created in Tableau were utilized to identify trends and emerging threats in hacker forums by perform an analysis on two primary areas:

- Exploit postings – What exploits types are being shared, when are exploits being shared, and what individual exploits have been shared recently
- Author activity – Which users are the most active, which forums are the most active, and what forums share the most of each exploit type

1) Individual Exploit Postings

Proactive CTI depends on timely information. Fig. 3 provides an example of a dashboard that lists individual exploits that have been collected by the forum crawler. The exploits are displayed on a timeline that can be adjusted to display information from the last month or show trends over multiple years. Fig. 3 shows all of the exploits that were shared from the beginning of 2015 to the middle of 2017. The figure shows that exploit postings are not as frequent now as they were in 2015. When using the dashboard, the user can also hover over each exploit to reveal more information including exploit name, author, forum/sub-forum/thread, and the URL of the attachment download. This information becomes significantly more important for recently shared exploits, where users can see exactly what new threats are being shared in hacker forums. As the incremental crawler continues to collect new exploits, analysis would be immediately available for proactive threat intelligence.

2) Author Activity All-Time

While it is important to consider the specific exploits being shared in hacker forums, it can also be valuable to look at which communities are the most active and how many authors are contributing to exploit dissemination. This information can point to what types of exploits each forum shares the most. Forums like Ashiyane contain postings of many different exploit types, while OpenSC and Garage4Hackers focus primarily on system and network exploits. While system and network exploits are the most common postings, this type of analysis becomes more valuable when forums are identified that focus primarily on the lesser shared exploit types such as web. These unique forums can be closely monitored for lesser utilized but potentially overlooked threats. Fig. 4 displays the all-time most active authors, based on number of attachments posted.

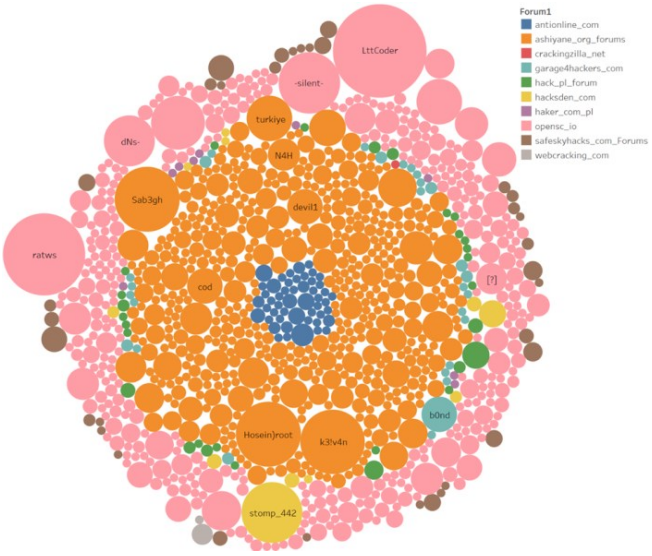

Fig. 4. Author Activity (All-Time)

The size of the author's bubble indicates how many exploits they have shared. The color of the bubble shows what forum that author was posting in. As seen previously, OpenSC and Ashiyane are by far the most active hacker forums in both number of active authors and number of exploits shared. Both of these forums have many significant contributors, but this is not the case for all forums. Hacksden appears to have one primary contributor for exploit postings, while AntiOnline has more active authors but none of them offer significant contributions.

3) Author Activity by Year and Exploit

By combining both exploit data and author activity we can gain further insights into which exploits, authors, and forums should warrant more attention. Fig. 5 provides a breakdown of author activity by year and exploit type. This visualization reinforces previous findings while also providing new insights not apparent from the other visualizations. System and network exploits continue to remain the most posted attachments within hacker forums. On top of this, there were many more exploit postings in 2015 than in the years after. Surprisingly, while Ashiyane has been expectedly active, OpenSC has been relatively dormant for the past few years. This visualization, in particular, provides valuable insights into what exploit types

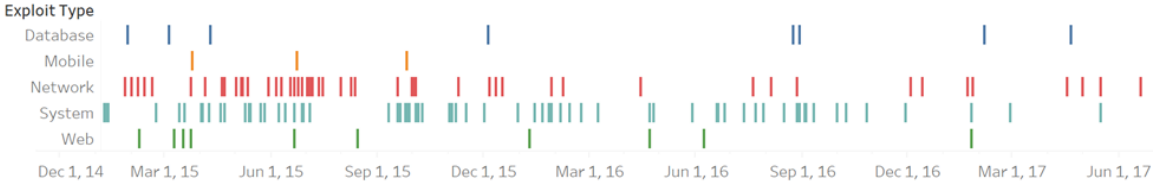

Fig. 3. Individual Exploit Postings (2015 – 2017)

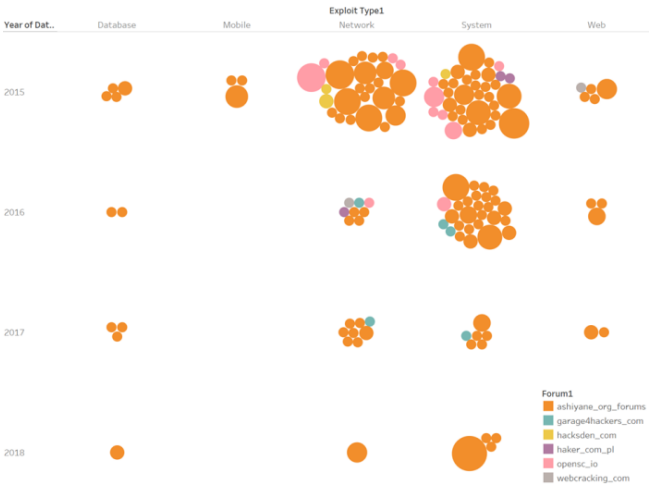

Fig. 5. Author Activity by Year and Exploit (2015 – 2018)

are being shared the most and which forums are currently the most active in providing those exploits.

VI. Conclusion and Future Directions

Cyber threats are an ever-growing problem. Organizations continue to amass large amounts of personal information on users and customers. This information can be extremely valuable to hackers looking to sell it on the dark web. In order to protect user information, organizations must rely on more than internal intelligence to mitigate threats.

In this study, we developed a novel incremental crawling approach designed to gather hacker forum attachments on an ongoing basis. This incremental crawler bypasses several of the common anti-crawling measures (e.g., authentication, throttling rates) while simultaneously navigating complex forum structures via a DFS strategy. The incremental crawler, combined with an LSTM RNN, is designed to provide proactive CTI through the collection and classification of hacker forum attachments. The results of this study indicate the majority of exploits shared on hacker forums target system and network vulnerabilities, making up 90.38% of the 2,930 collected attachments. This study also develops a visualization system which enables users to extract value from the exploit collection. Specifically, exploit postings and author activity are analyzed to gain insights into which exploit types, authors, and forums should be more heavily observed for proactive CTI.

This research can be expanded in several directions. First, future work can repurpose the crawler for other frameworks. Second, the crawler can collect new assets beyond attachments such as source code and tutorials on how hackers execute certain attacks. These different assets could be collected and classified with a similar approach this study took to attachments. Finally, more in-depth analysis and visualization of the data can be performed to provide greater insights. Each extension would help organizations receive up-to-date and deeper insights to improve their cybersecurity posture and contribute to a safer and more secure society.

Acknowledgment

This work was supported in part by the National Science Foundation (NSF) DUE-1303362 (SFS), SES-1314631 (SaTC), ACI-1443019 (DIBBs), and 1719477 (EAGER).

References

[1] Ponemon Institute LLC. (2017). 2017 Cost of Data Breach Study, (June), 1–34. Retrieved from https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=SEL03130WWEN&

[2] Bromiley, M. (2016). Threat Intelligence: What It Is, and How to Use It Effectively. SANS Institute InfoSec Reading Room, 15.

[3] Samtani, S., Chinn, R., & Chen, H. (2015). Exploring Hacker Assets in Underground Forums. IEEE.

[4] Samtani, S., Chinn, K., Larson, C., & Chen, H. (2016). AZSecure Hacker Assets Portal: Cyber Threat Intelligence and Malware Analysis. IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data, ISI. https://doi.org/10.1109/ISI.2016.7745437

[5] Fu, T., Abbasi, A., & Chen, H. (2010). A Focused Crawler for Dark Web Forums. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. https://doi.org/10.1002/asi

[6] Jiang, J., Song, X., Yu, N., & Lin, C.-Y. (2014). FoCUS?: Learning to Crawl Web Forums. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 25(6), 1293–1306.

[7] Pavkovic, M., & Protic, J. (2013). Intelligent crawler for web forums based on improved regular expressions. 2013 21st Telecommunications Forum Telfor, TELFOR 2013 - Proceedings of Papers, 817–820. https://doi.org/10.1109/TELFOR.2013.6716355

[8] Baravalle, A., Lopez, M. S., Lee, S. W., & Kingdom, U. (2016). Mining the Dark Web. IEEE International Conference on Data Mining Workshops, 350–356. https://doi.org/10.1109/ICDMW.2016.149

[9] Grisham, J., Samtani, S., Patton, M., & Chen, H. (2017). Identifying Mobile Malware and Key Threat Actors in Online Hacker Forums for Proactive Cyber Threat Intelligence. IEEE International Conference on Intelligence and Security Informatics: Security and Big Data, ISI. https://doi.org/10.1109/ISI.2017.8004867

[10] Benjamin, V., Li, W., Holt, T., & Chen, H. (2015). Exploring Threats and Vulnerabilities in Hacker Web: Forums, IRC and Carding Shops. IEEE International Conference on Intelligence and Security Informatics: Securing the World through an Alignment of Technology, Intelligence, Humans and Organizations, ISI. https://doi.org/10.1109/ISI.2015.7165944

[11] Macdonald, M., Frank, R., Mei, J., & Monk, B. (2015). Identifying Digital Threats in a Hacker Web Forum. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 926–933. https://doi.org/10.1145/2808797.2808878

[12] Nunes, E., Diab, A., Gunn, A., Marin, E., Mishra, V., Paliath, V., … Shakarian, P. (2016). Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence. IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data, ISI. https://doi.org/10.1109/ISI.2016.77454

[13] F. Chollet, Keras, (2015), GitHub repository, https://github.com/keras-team/keras

[14] Sagar Samtani, Ryan Chinn, Hsinchun Chen & Jay F. Nunamaker Jr. (2017) Exploring Emerging Hacker Assets and Key Hackers for Proactive Cyber Threat Intelligence, Journal of Management Information Systems, 34:4, 1023-1053, DOI: 10.1080/07421222.2017.1394049