A DATA ANALATICS APPROACH TO THE CYBER CRIME UNDERGROUND ECONOMY

M.Sravan Kumar Babu[1], A.Chandana[2], A.Anusha[3], K.Harika[4], P.Jhansi[5]

[1]Assistant Professor, Department of CSE(CS), MallaReddy Engineering College for Women, Hyderabad, TS, India.

Email: sravan.mokkapati@gmail.com

[2,3,4,5]UG Students, Department of CSE(CS), MallaReddy Engineering College for Women, Hyderabad, TS, India.

ABSTRACT:
Despite the rapid escalation of cyber threats, there has still been little research into the foundations of the subject or methodologies that could serve to guide Information Systems researchers and practitioners who deal with cyber security. In addition, little is known about Crime-as-a-Service (CaaS), a criminal business model that underpins the cybercrime underground. This research gap and the practical cybercrime problems we face have motivated us to investigate the cybercrime underground economy by taking a data analytics approach from a design science perspective. To achieve this goal, we proposed data analysis framework for analyzing the cybercrime underground, CaaS and crime ware definitions, and an associated classification model. In addition, we develop an example application to demonstrate how the proposed framework and classification model could be implemented in practice. We then use this application to investigate the cybercrime underground economy by analyzing a large dataset obtained from the online hacking community. By taking a design science research approach, this study contributes to the design artifacts, foundations, and methodologies in this area. Moreover, it provides useful practical insights to practitioners by suggesting guidelines as to how governments and organizations in all industries can prepare for attacks by the cybercrime underground.

Keywords: IDS, CNN, ANN, SVM, Digital terror.

1. INTRODUCTION

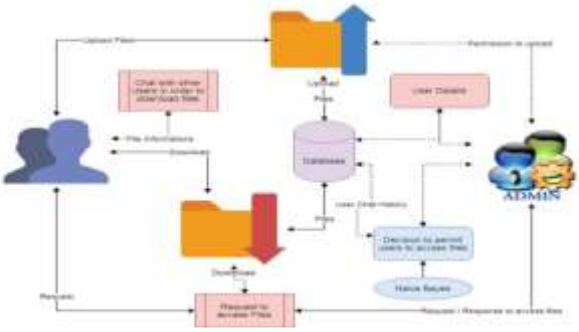
As the threat posed by major cyber attacks (e.g., ransom ware and distributed denial of service (DDoS)) and cybercrime has risen, people, governing organizations, and governments have rushed to devise countermeasures. Want to Cry ransomware was responsible for about 45,000 assaults in nearly 100 countries in 2017 [1]. The growing impact of cybercrime has prompted leadership to boost its top-secret expenditures. Global cyber attacks (such as Want to Cry and Peaty) are carried out by highly organized criminal gangs, and many recent efforts have been carried out by organized or national level crime groups. In general, criminal groups use the cybercrime black market to acquire and sell hacking tools and services, and attackers share a variety of hacking-related data. As a result, the cybercrime underground has emerged as an unique form of organization that both administers black marketplaces and facilitates cybercrime plots. Because well planned cybercrime necessitates the existence and operation of an internet network, it is heavily reliant on closed antiestablishment communities (e.g., Hack forums and Crackingzilla). Because of the secrecy provided by these closed groups, cybercrime networks are structured differently from conventional Mafia-style hierarchies [4], which are vertical, resolute, inflexible, and fixed. Cybercrime networks, in contrast, are lateral, diffuse, fluid, and dynamic. Because the internet is a web of networks [5,] the threat posed by the wage growth of highly professional network-based cybercrime business models such as Crime ware-as-a-Service (CaaS) is mostly unseen to governments, governing bodies, and the general public.

2. LITERATURE SURVEY

To detect drive-by-download assaults, techniques that evaluate web pages for harmful information in a virtual or emulated environment have been developed. Crawlerbased techniques that evaluated billions of online pages were used to investigate the prevalence of rogue web sites. Another research looked at drive-by attacks using infiltration and revealed information on the compromised web servers utilised in the assaults as well as the security posture of possible victims. Lack of understanding and attention to browser and other security indicators are examples of phishing schemes. Several techniques for detecting phishing sites have been proposed, including evaluating page content, layout, and other abnormalities. Furthermore, research have been conducted to examine the methods of operation of the criminal activities behind phishing as well as the efficiency of phishing countermeasures. The underground economy of the Internet is examined through the listed pricing of web forums and IRC chat rooms. Holz et al. investigated botnet drop zones, which are used to store stolen information from victims. Stone-Gross et al. took over the Torpig botnet, analysed the data exfiltrated from infected machines, and calculated the worth of the compromised financial information. The underworld of large-scale spam campaigns was investigated. The article investigated an underground forum used by spammers to exchange products and services, as well as the intricacy of coordinating spam operations. Christin et al. investigated another form of fraud called as One Click Fraud. The scam operates through intimidation, threatening unwary website users with shame unless they pay for a nonexistent service. The authors provided an economic model to assess the number of people who must fall victim to the scam in order for it to be commercially feasible, and they projected losses ranging from tens to hundreds of thousands of dollars.

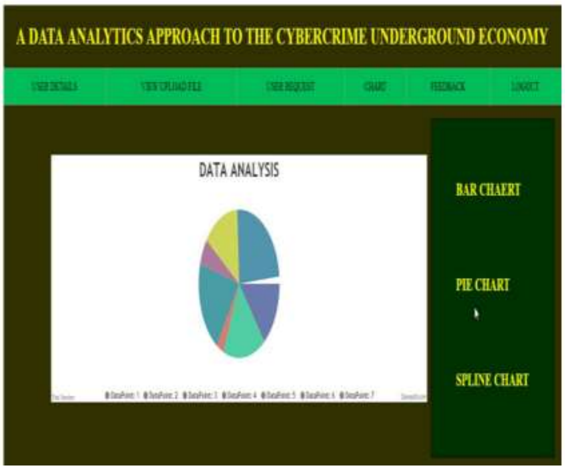
PROPOSED SYSTEM

Our data analysis framework's objective is to perform a big-picture examination of the cybercrime underground by encompassing all aspects of data analysis from start to finish. This structure is made up of four steps: (1) setting goals; (2) identifying sources; (3) deciding on analytical techniques; and (4) putting the application into action. Because this study highlights the relevance of RAT in understanding the cybercrime underground, the following RAT-based definitions are important to this framework: The RAT components may be found in all of the steps 1–4. A. Step 1: Establishing Objectives The first step is to define the analytical conceptual scope. This phase specifically specifies the analysis context, particularly the objectives and aims. We examined the cybercrime underground, which functions as a secretive group, to get a thorough grasp of current CaaS research. As a result, the suggested framework's objective is to "examine the cybercrime underground economy." B. Step 2: Locating Sources Based on the goals established in Step 1, the second step is to select data sources. This phase should take into account what data is required and where it may be acquired. We consider data on the cybercrime underground community since the objective of this study is to investigate the cybercrime underground. As a result, we gathered such information from the community and got a malware database from a prominent worldwide cyber security research organisation. We utilised a selfdeveloped crawler that can overcome captchas and anti-crawling scripts to obtain the essential data because fraudsters frequently change their IP addresses and use anti-crawling scripts to disguise their communications. We gathered a total of 2,672,091 postings offering CaaS or crimeware from a big hacking community site (www.hackforums.net) with over 578,000 users and over 40 million posts between August 2008 and October 2017. We also gathered 16,172 user profiles of vendors and potential purchasers based on their contact histories, as well as pricing and transaction-related questions and responses. Instead of standard e-commerce platforms, the black market employs classic forum threads (e.g., bulletin boards) (e.g., eBay, and Amazon). Sellers, for example, post threads in marketplace forums to offer things, and potential purchasers remark on these topics. Converting this unstructured data into structured data was perhaps one of the most difficult difficulties. We utilised a number of text mining algorithms to extract the key elements, such as named entity recognition to extract business names (see Section IVC(2)), because the product features, pricing, and descriptions were described inside lengthy texts. We had to build a lexicon for usage during a preprocessing phase because these documents had many typographical mistakes and jargon phrases. In addition, we received from a cybersecurity business a malware database including approximately 53,815 records spanning cybercrime incidents between May 11, 2010 and January 13, 2014. This one-of-a-kind dataset bolstered our research by giving real-world evidence from a fresh perspective.



RESULTS AND DISCUSSION

This study adds to the body of knowledge by demonstrating new approaches to the problems cybercrime and social media researchers face Despite the increasing importance of data analysis, researchers have been slow to recognize the advantages of new and more powerful data driven analysis methods. We have applied several modern techniques, such as machine learning, key phrase extraction, and natural language processing, in this area, thereby encouraging future research to be more systematic and empirical. In addition, our results suggest that combining natural language processing and machine learning approaches is a suitable way to study closed communities whose members frequently use jargon or obscure expert language. Although our study has made several significant findings, it nevertheless has several limitations that will need to be addressed in future studies. These will be able to add more analysis and significant further insights. First, we only collected data from the largest hacking community and did not consider other hacking communities. Future studies will therefore need to generalize our findings by investigating a wider range of hacking communities. Second, this study has focused on the CaaS and crime ware available in the cybercrime underground, but much in depth analysis remains to be done on the configurations of cybercrime networks. Future research could cluster keywords and threats by industry to provide a deeper understanding of the potential vulnerabilities, and it could attempt to discover the network effects involved or the leaders of the cybercrime underground.



3. CONCLUSION

We have focused mainly on building and evaluating artifacts rather than on developing and justifying theory: actions are usually considered to be the main focus of behavioral science. We have therefore proposed two artifacts: a data analysis framework and a classification model. We have also conducted an ex-ante evaluation of our classification model's accuracy and an ex-post evaluation of its implementation using example applications. In line with the initiation perspective of DSR, these four example applications demonstrate the range of potential practical applications available to future researchers and practitioners. Unlike previous studies that have presented general discussions of a broad range of cybercrime; our study has focused primarily on CaaS and crime ware from an RAT perspective. We have also proposed sets of definitions for different types of CaaS (phishing, brute force attack, DDoS attack, and spamming, creeping, and VPN services) and crime ware (drive-by download, botnets, exploits, ransom ware, root kits, Trojans, creepers, and proxies) based on definitions taken from both the academic and business practice literature. Based on these, we have built an RATbased classification model. This study emphasizes the importance of RAT for investigating the cybercrime underground, so these RAT-based definitions are critically important parts of our framework. In addition, unlike prior. research that discussed the cybercrime underground economy without attempting to analyze the data, we have analyzed large-scale datasets obtained from the underground community. Looking at the CaaS and crime ware trends, our results show that the prevalence of botnets (attack-related crime ware) and VPNs (preventive measures, related to CaaS) has increased in 2017. This indicates that attackers consider both the preventive measures taken by organizations and their vulnerabilities. The most common potential target organizations are technology companies (28%), followed by content (22%), finance (20%), e-commerce (12%), and telecommunication (10%) companies. This indicates that a wide variety of companies in a range of industries are becoming potential targets for attackers, having become more vulnerable due to their greater reliance on technology.

REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.

[2] R. Christopher, “Port scanning techniques and the defense against them,” SANS Institute, 2001.

[3] M. Baykara, R. Das¸, and I. Karado ˘gan, “Bilgi g ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi,” in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, “Practical automated detection of stealthy portscans,” Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, “Surveillance detection in high bandwidth environments,” in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.

[6] K. Ibrahimi and M. Ouaddane, “Management of intrusion detection systems based-kdd99: Analysis with lda and pca,” in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[7] N. Moustafa and J. Slay, “The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems,” in Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, “Detection and classification of malicious patterns in network traffic using benford’s law,” in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.

[9] S. M. Almansob and S. S. Lomte, “Addressing challenges for intrusion detection system using naive bayes and pca algorithm,” in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, “Combined analysis of support vector machine and principle component analysis for ids,” in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.