IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006069 |

A Data Analytics Approach to the Cybercrime Underground Economy

Sk.Shahina, K.Varsha Ojaswini, N.Bhanu Teja, K.Sadhana, P.Bhargav Nadh

Assistant Professor, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE, Tirumala Engineering College, NRT, Andhra Pradesh, India

ABSTRACT:-Despite the fast rise of cyber dangers, little study has been conducted into the subject's foundations or techniques that might help to assist Information Systems researchers and practitioners who deal with cyber security. Furthermore, nothing is known about Crime- as-a-Service (CaaS), a criminal business model that serves as the foundation of the cybercrime underground. This research vacuum, as well as the actual cybercrime issues we confront, compelled us to explore the cybercrime underground economy using a data analytics method from a design science viewpoint. To that end, we propose (1) a data analytic methodology for investigating the cybercrime underground, (2) definitions of CaaS and criminal ware, and (3) an associated classification model. Furthermore, we (4) provide an example application to show how the suggested framework and classification model may be used in practise. The programme is then used to study the cybercrime underground economy by analysing a big dataset acquired from the internet hacking community. This work contributes to the design artefacts, foundations, and techniques in this field by employing a design science research strategy. Furthermore, it provides practitioners with important practical insights by proposing suggestions for how governments and companies in various industries may prepare for cybercrime underground assaults.

KEYWORDS *:* - Crimeware-as-a-Service, crimeware, underground economy, hacking community, machine learning.

I INTRODUCTION

As the threat posed by major cyberattacks (e.g., ransomware and distributed denial of service (DDoS)) and cybercrime has risen, people, governing organisations, and governments have rushed to devise countermeasures. Wanna 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 cyberattacks (such as Wanna Cry and Petya) are carried out by highly organised criminal gangs, and many recent efforts have been carried out by organised 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 organisation 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 anti- establishment communities (e.g., Hackforums and Crackingzilla). Because of the secrecy provided by these closed groups, cybercrime networks are structured differently from conventional Mafia-style heirarchies [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 Crimeware-as-a-Service (CaaS) is mostly unseen to governments, governing bodies, and the general public.

II. 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. Crawler- based 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

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006069 |

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.

III 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 self- developed 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 IV- C(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.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006069 |

IV. IMPLEMENTATION

SYSTEM ARCHITECTURE

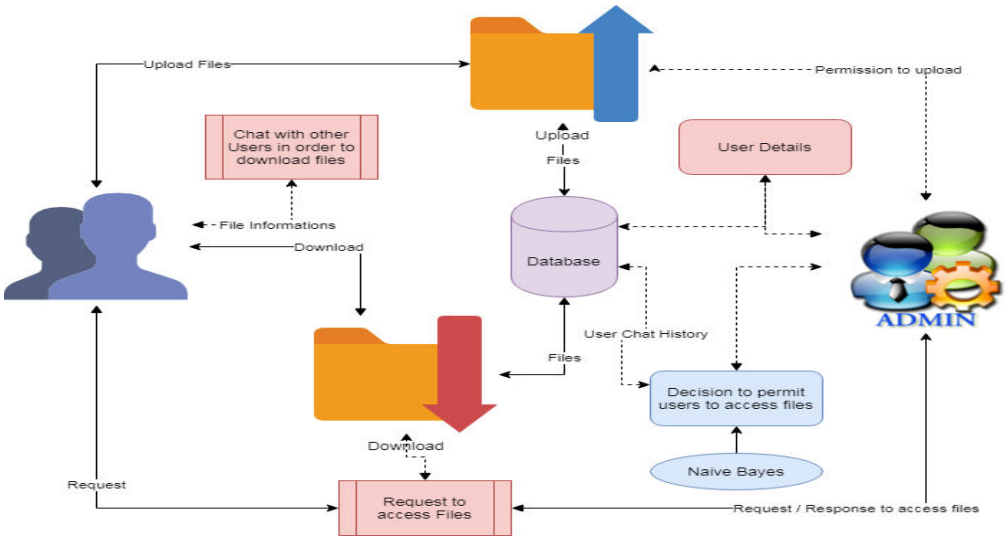

Fig 1:System Architecture

MODULES IN THE SYSTEM ARCHITECTURE:

1. Files to Upload

Users are permitted to upload files with the specified tags. When a file is uploaded, it is forwarded to admin for approval before it can be published or seen by other users. These submitted materials can be in any format, including documents, music, and video, but executable (.exe) files are not permitted.

2.Observation of Conversations

Users are permitted to communicate with one another. The administrator could keep an eye on this. The malevolent conversion enjoys threatening the data. In order to defend cybercrime and prevent the formation of a cybercrime community. This is possible with the aid of a classification method known as naive Bayes classification.

3.File Downloads

The files may be downloaded by requesting them, and once authorised by the administrator, they can be downloaded. The choice to authorise files can be derived from the user discussion. The administrator takes action on download files and user approval status. Based on the users, further activities are permitted.

4. Graphic Representations

The approvals and disapprovals are used to compute the analyses of proposed systems. This can be quantified using graphical notations such as a pie chart, a bar chart, or a line chart. The data can be presented in a dynamical format.

ALGORITHM:

The Naive Bayes Classifier is an algorithm.

Naive Bayes is a classification algorithm that may be used to solve binary (two-class) and multi-class classification issues. When presented using binary or categorical input values, the approach is simplest to grasp.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006069 |

The method is known as naive Bayes or stupid Bayes because the computation of the probability for each hypothesis is reduced to make it tractable. Instead of attempting to determine the values of each attribute value P(d1, d2, d3|h), they are considered to be conditionally independent given the goal value and computed as P(d1|h) * P(d2|H), and so on.

This is a very strong assumption that is highly unlikely to be true in real-world data, namely that the characteristics do not interact. Despite this, the method works very well on data where this assumption is not true.

Use a Naive Bayes Model to Make Predictions

The Bayes theorem may be used to create predictions for fresh data given a naïve Bayes model.

P(d|h) * P(h) MAP(h) = max(P(d|h) * P(h))
Using our previous example, if we created a new instance with the weather set to sunny, we can calculate:
P(weather=sunny|class=go-out) * P(weather=sunny|class=go-out)
P(weather=sunny|class=stay-home) * P(class=stay-home)
We can select the class with the highest computed value. By normalising these values, we may convert them to probabilities:
P(go-out|weather=sunny) = go-out / (go-out
+ stay-home) P(stay-home|weather=sunny) = stay-home / (go-out + stay-home) P(stay- home|weather=sunny) = stay-home / (go-out + stay-home)

We might extend the above example if we had more input variables. Assume we have a "vehicle" attribute with the values "functioning" and "broken." This probability can be multiplied into the equation.
The following is the computation for the “go- out” class label, with the automobile input variable set to “working”:

go-out = P(weather=sunny|class=go-out) x P(car=working|class=go-out) x P(class=go-out) x P(class=go-out) x P(class=go-out) x P(class=go-out) x P(class=go-out) x P(class=

V. CONCLUSION

Because this study uses a DSR method, we have concentrated mostly on creating and assessing artefacts rather than formulating and supporting theory: actions are typically seen as the primary emphasis of behavioural science. As a result, two artefacts have been proposed: a data analysis framework and a categorization model. We also performed an ex-ante evaluation of the accuracy of our classification model and an ex-post evaluation of its implementation using sample applications. In accordance with the DSR's initiating viewpoint, these four sample applications show the breadth of potential practical applications open to future researchers and practitioners. Unlike earlier research, which offered generic talks of a wide variety of cybercrime, our study concentrated mainly on CaaS and criminal ware from a RAT standpoint. We also proposed definition sets for various types of CaaS (phishing, brute force attack, DDoS attack, and spamming, crypting, and VPN services) and crime ware (drive-by download, botnets, exploits, ransomware, rootkits, Trojans, crypters, and proxies) based on definitions from academic and business practise literature. On the basis of this, we created a RAT-based categorization model. Because this study highlights the relevance of RAT in analysing the cybercrime underground, these RAT-based criteria are key components of our approach. Furthermore, unlike previous studies that addressed the cybercrime underground economy without attempting to evaluate the data, we examined large-scale datasets acquired from the underground community. In terms of CaaS and crimeware trends, our findings reveal that the prevalence of botnets (attack-related crimeware) and VPNs (preventive measures linked to CaaS) rose in 2017. This implies that attackers take into account both companies' preventative measures and their weaknesses. Technology businesses (28%) are the most prevalent prospective target organisations, followed by content (22%), finance (20%), e-commerce (12%), and telecommunications (10%) enterprises. This suggests that a wide range of organisations in a variety of industries are becoming possible targets

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006069 |

for attackers, since they have grown more susceptible as a result of their increased dependence on technology.

REFERENCES

[1] O. Solon and J. C. Wong. (May 12, 2017) A massive ransomware cyber-attack has impacted almost 100 nations worldwide. [Online]. https://www.theguardian.com/technology/20 17/05/12/global-cyber-attack-ransomware- nsa-uk-nhs

[2] The White House, "FACT SHEET: Cybersecurity National Action Plan," 2016.

[3] A. K. Sood and R. J. Enbody, "Crimeware-as-a-Service—A Survey of Commoditized Crimeware on the Underground Market," International Journal of Critical Infrastructural Protection, vol. 6, no. 1, pp. 28–38, 2013.

[4]S. W. Brenner, "Organized Cybercrime? How Cyberspace May Affect the Structure of Criminal Relationships," New Caledonia Journal of Law and Technology, vol. 4, no. 1, pp. 1-50, 2002.

[5] K. Hughes, "Introduction to the World Wide Web," ACM SIGWEB Newsl., vol. 3, no. 1, pp. 4–8, 1994.

[6] S. Gregor and A. R. Hevner, "Positioning and Presenting Design Science Research for Maximum Impact," MIS Quarterly, vol. 37, no. 2, 2013, pp. 337-356.

[7] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research," MIS Quarterly, vol. 28, no. 4, 2004, pp. 75–105.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details