JOURNAL OF DATA SCIENCE | Vol.2024:12
eISSN:2805-5160

A Data Analytics Strategy to the Underground Economy of Cybercrime

Varun PR[1], Usha Sree[1], Malathy Batumalay[2]

[1]Dayananda Sagar Academy of Technology and Management Bangalore, Karnataka 560082
[2]Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia

Email: varunprgowda07@gmail.com, ushasree19832000@gmail.com

Abstract

Massive cyberattacks and cybercrimes such as ransomware and distributed denial of service (DDoS) attacks have become more and more of a danger, and individuals, organizations, and governments have struggled to discover effective means to defend against them. In 2017, the WannaCry ransomware was to blame for roughly 45,000 strikes across almost 100 nations. Governments have come under pressure to enhance their cybersecurity spending as a result of the increasing impact of cybercrime. In his fiscal year 2017 budget, United States President Barack Obama suggested investing more than $19 billion in cybersecurity, a rise of more than 35% from 2016. Thus, a new sort of organization known as the "cybercrime underground" has developed, one that both runs underground markets and fosters the growth of cybercrime conspiracies. The threat posed by the emergence of highly skilled network-based cybercrime business models, such as Crimeware-as-a-Service (CaaS), is mostly invisible to governments, organizations, and individuals because cybercrime networks are lateral, diffuse, fluid, and dynamic. Although cyber dangers are rapidly increasing, there hasn't been much research done on the methodology or theoretical underpinnings of the field that could help inform information systems researchers and practitioners who work in the field of cyber security. Additionally, little is understood about Crime-as-a-Service (CaaS), the illegal business model that supports the underground world of cybercrime.

Keywords

CAAS, Cyberattacks, classification, ransomware

Introduction

As the threat posed by serious cyberattacks (such as ransomware and distributed denial of service (DDoS)) and cybercrime has escalated, people, governing bodies, and governments have hurried to devise solutions. Want to Cry ransomware was responsible for around 45,000 attacks in nearly 100 countries in 2017. As a result of the growing impact of cybercrime, leadership has raised its top-secret spending. International cyberattacks are carried out by highly structured criminal gangs (such as Want to Cry and Peaty), and many recent attempts have been

Submission: 19 May 2024; **Acceptance:** 23 June 2024

undertaken by structured or national-level crime groups. Attackers share a variety of hacking-related data, and criminal organizations routinely buy and sell hacking tools and services on the criminal black market (J. C. Wong and O. Solon., 2017). As a result, the underground cybercrime market has evolved into a new form of organization that handles underground markets and promotes cybercrime schemes. Because it necessitates the building and operation of an internet network, well-planned cybercrime primarily relies on closed antiestablishment networks (such as Hack forums and Crackingzilla). Because of the confidentiality provided by these closed organizations, cybercrime networks are organized differently than typical Mafia-style hierarchies, which are vertical, determined, inflexible, and fixed (The White House, 2016). Cybercrime networks, on the other hand, are lateral, diffuse, fluid, and dynamic. Because the internet is a web of networks, the threat posed by highly skilled network-based cybercrime business models such as Crimeware-as-a-Service (CaaS) is generally unknown to governments, governing organisations, and the general public.

This study of ransomware encrypts a user's data and then demands money to decrypt it. This attack made use of malicious software known as "WanaCrypt0r 2.0" or WannaCry, which makes use of a Windows flaw (A. K. Sood and R. J. Enbody, 2013). The has developed the Cybersecurity National Action Plan (CNAP), a comprehensive strategy, to handle the nation's growing cybersecurity challenges. R J Enbody shows that Crimeware-as-a-service (CaaS) is now a significant part of the shadow market (S. W. Brenner, 2002). By making cybercrime more organized, automated, and accessible to criminals with minimal technical capabilities, CaaS gives it a new dimension. This literature review intends to study and analyze the body of knowledge already available on organized cybercrime, with an emphasis on its causes, methods, effects on society, and defenses proposes the fundamental aspects and components that make up a systematic and cohesive framework for creating creative artifacts to solve particular issues or overcome particular obstacles are referred to as the anatomy of a design theory (K. Hughes.,1994)

Because it functions in the virtual realm, with various spatial and temporal restrictions, separating it from other crimes occurring in the actual world, cybercrime has undergone a revolutionary transition, going from product-oriented to service-oriented (S. Gregor and A. R. Hevner, 2013). Due to new technology developments giving organized cybercriminal gangs previously unheard-of opportunities for exploitation, the cybercrime underground has grown into a covert black market. A very sophisticated business model underpins the subterranean economy of cybercrime (R. Hevner, S. T., 2004). The "business model used in the underground market where illegal services are provided to underground buyers to help them conduct cybercrimes, such as attacks, infections, and money laundering in an automated manner" is known as "CaaS." In contrast to crimeware, which is a do-it-yourself product, CaaS is therefore characterized to as a do-it-for-me service. Customers of CaaS do not need to run a hacking server or possess advanced hacking abilities because the service is created for beginners (K. Peffers, 2007). As a result, the CaaS business model can include the following roles: creating a hacking program, carrying out an attack, ordering an attack, supplying an attack server (infrastructure), and laundering the proceeds (S. Gregor, 2002).

Methodology

Our data analysis methodology aims to encompass all stages of data analysis from the beginning to the finish to perform a broad examination of the criminal underworld. This

framework consists of four steps: (1) goal definition; (2) source identification; (3) method selection; and (4) application implementation. The proposed RAT-based definitions are essential to this framework since this study emphasizes the significance of RAT for studying cybercrime underground: The RAT elements are present in each of steps 1-2.

Setting goals is the first step (S. Gregor and D. Jones, 2007). Identifying the conceptual range of the investigation is the first stage. This stage specifically describes the context of the analysis, including the objectives and aims. We looked into the closed community of cybercrime underground in order to acquire a thorough understanding of the present CaaS research. The proposed approach therefore aims to "investigate the cybercrime underground economy."

Step Two: Finding Sources Based on the objectives set forth in Step 1, the second step is to determine the data sources. This stage should take where and how to get the data into account. We take into account information on the cybercrime underground community as the aim of this study is to explore it. Therefore, we secured a malware database from a top worldwide cyber security research business and collected such data from the community itself. We employed a self-developed crawler that can overcome captcha's and anticrawling scripts to acquire the essential data because fraudsters frequently change their IP addresses and use those tools to hide their interactions. 2,672,091 posts advertising CaaS or crimeware were gathered by us between August 2008 and October 2017, from www.hackforums.net, a popular hacking forum with over 578,000 users and 40 million postings.

Dataset Upload & Analysis: Using this module, we will upload a dataset and then use analysis methods to detect various cybercrimes and their counts, as well as clean the dataset by deleting missing values. Dataset Processing & Analytical Methods: Using this module, we'll encrypt attack labels with integer IDs before dividing the dataset into train and test portions. The program used 80% of the training dataset to train the Nave Bayes classification algorithm and 20% of the test dataset to assess its performance.

Run the Naive Bayes Classification Model: Using this module, we will train the classification algorithm with a dataset that contains more than 80% of the data, and then we will create a prediction model. Classification Performance Graph: To determine the performance accuracy, we will use this module to plot the Naive Bayes accuracy and precision graph. Cybercrime Prediction: Using this module, we will input a sample dataset from a network of cybercriminals, and a classification model will then determine whether the sample data contains any criminal activity signatures.

Based on their communication histories, prices, and questions and answers regarding the transactions, we also gathered 16,172 user profiles of sellers and potential purchasers. Instead of the usual e-commerce platforms (like eBay and Amazon), the black market uses conventional forum discussions (like those seen on bulletin boards). For instance, in forums for marketplaces, vendors post topics to advertise their wares, and potential customers respond to these discussions. Therefore, transforming this unstructured data into structured data posed one of the biggest hurdles.

System Design

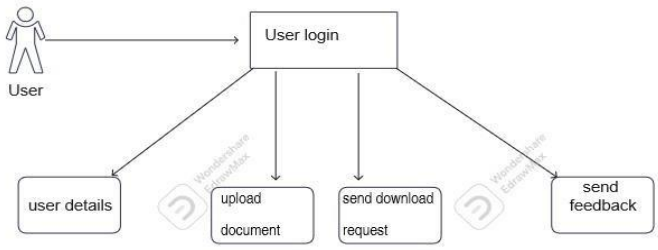

Figure 1. User Architecture

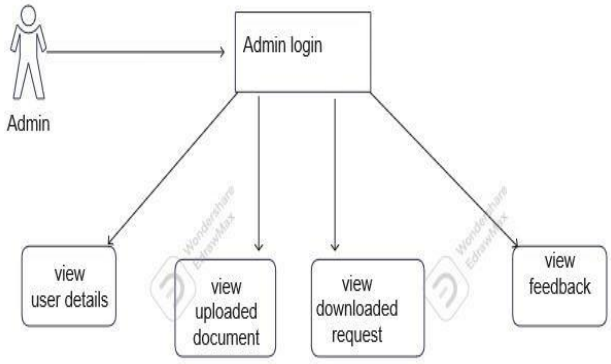

Figure 2. Admin Architecture

Login Module

The user can create a username and password for authentication so they can upload files, issue requests, and provide comments. Administrators can view user data, keep tabs on user activities, and get requests and feedback.

Upload files

Users are allowed to upload files with the specified tags. An admin must authorize a file before it can be published or accessed by other users. Although executable files are prohibited, data sets may be submitted and used to generate analytics.

File download

When the administrator has given permission, the files can be downloaded after being requested and authorized. From the user conversation, the decision to authorize files can be drawn. A decision is made by the administrator on user approval and file downloads. Additional actions are allowed depending on the users.

Graphical representation

The analytical process based on the user-uploaded document will be displayed in this module. It will also give a graph and a list of the different types of attacks already recorded in the dataset.

Results

This study contributes to the body of knowledge by illuminating fresh solutions to the challenges faced by social media and cybercrime researchers. Researchers have been hesitant to recognize the benefits of new and more potent data-driven analysis methodologies despite the growing importance of data analysis. We have used several contemporary methods in this field, including machine learning, keyphrase extraction, and natural language processing. This has inspired more organized and empirical future studies.

Although our study produced several important findings, it also had several shortcomings that will need to be addressed in subsequent research. These will be able to provide additional analyses and important new insights. First, we did not take into account other hacker communities and solely gathered data from the biggest hacking community. Therefore, future research will need to generalize our findings by looking into a wider variety of hacking communities. Second, while this study has concentrated on the CaaS and criminal software available on the dark web, there is still much in-depth research to be done on the setups of criminal networks.

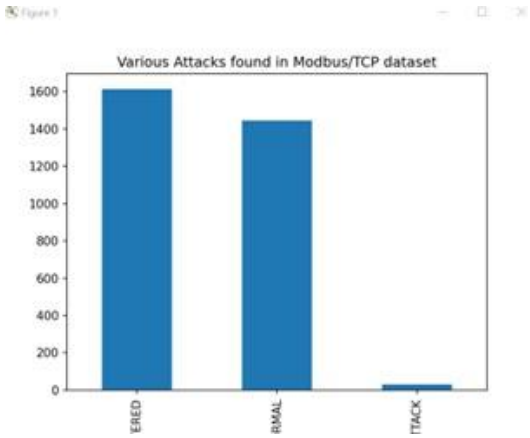

Figure 3. Preprocess Data

After logging in, the user should upload the data document with all the attack specifics so that the system can predict the different types of attacks.

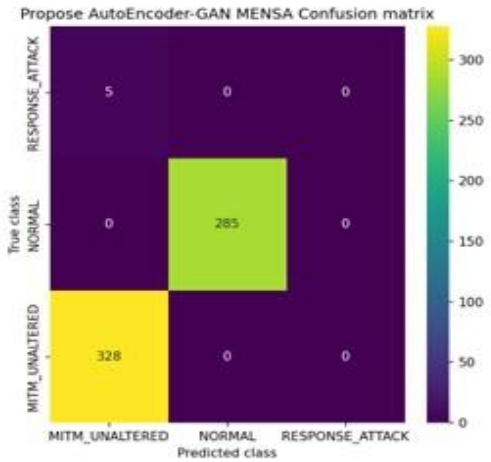

Figure 4. Encoder Gan Heat Map

The proposed autoencoder GAN MENSA confusion matrix will assess the datasets and predict the heat map.

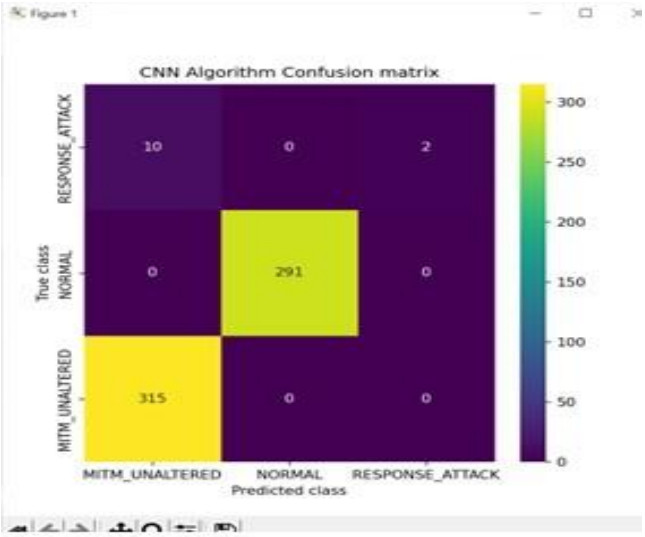

Figure 5.CNN Heat Map

The proposed CNN algorithm confusion matrix will assess the datasets and predict the heat map.

Conclusion

Finally, there are significant societal ramifications for this study. Nation-sponsored terrorists have been threatening the world with cyberterrorism and cyberwar over the past few years. Cyberterrorism, according to Pollitt, is "the premeditated, politically motivated attack against information, computer systems, computer programs, and data which results in violence against non-combatant targets by subnational groups or clandestine agents." Cyberterrorists have political motivations, in contrast to most cybercriminals, who are primarily motivated by financial gain. Governments should, for instance, improve their quick responses to dangers like cyberespionage and cyberterrorism to better protect their citizens in online virtual environments. Therefore, this issue has significant ramifications for the requirement of a worldwide cyber defense to uphold a cyber-safe environment.

References

J. C. Wong and O. Solon. (2017, May 12). Massive ransomware cyber-attack hits nearly 100 countries around the world. [Online]. Available: https://www.theguardian.com/technology/2017/may/12/global cyberattack-ransomware-nsa-uk-nhs

“FACT SHEET: Cybersecurity National Action Plan,” ed: The White House, 2016.

A. K. Sood and R. J. Enbody, “Crimeware-as-a-service—A survey of commoditized crimeware in the underground market,” Int. J. Crit. Infr. Prot., vol. 6, no. 1, pp. 28–38, 2013.

S. W. Brenner, “Organized Cybercrime? How Cyberspace May Affect the Structure of Criminal Relationships,” N. C. J. Law & Technol., vol. 4, no. 1, pp. 1-50, 2002.

K. Hughes, “Entering the world-wide web,” ACM SIGWEB Newsl., vol. 3, no. 1, pp. 4–8, 1994.

S. Gregor and A. R. Hevner, “Positioning and Presenting Design Science Research for Maximum Impact,” MIS Quart., vol. 37, no. 2, pp. 337-356, 2013.

A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design Science in Information Systems Research,” MIS Quart., vol. 28, no. 4, pp. 75- 105, 2004.

K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A Design Science Research Methodology for Information Systems Research,” J. Manag. Inf. Syst., vol. 24, no. 3, pp. 45–77, 2007.

S. Gregor, “Design theory in information systems,” Aust. J. Inf. Syst., vol. 10, no. 1, pp. 14–22, 2002.

S. Gregor and D. Jones, The Anatomy of a Design Theory, J. the Assoc. Inf. Syst., vol. 8, no. 5, pp. 313 335, 2007.