International Conference on Recent Advancement in Science & Technology- 2020
(ICRAST-2020)

A Review on Data Analytics Approach to the Cybercrime

Tripti Sahu
Computer Science & Engineering
Bansal Institute of Engineering & Technology, Lucknow – 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 propose (1) a data analysis framework for analyzing the cybercrime underground, (2) CaaS and crime ware definitions, and (3) an associated classification model. In addition, we (4) 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: Naive Bayes Classifier & Cyber Crime.

1. Introduction:

As the threat posed by massive cyberattacks (e.g., ransomware and distributed denial of service attacks (DDoS)) and cybercrimes has grown, individuals, organizations, and governments have struggled to find ways to defend against them. In 2017, ransomware known as WannaCry was responsible for nearly 45,000 attacks in almost 100 countries [1]. The explosive impact of cybercrime has put governments under pressure to increase their cybersecurity budgets. United States President Barack Obama proposed spending over $19 billion on cybersecurity as part of his fiscal year 2017 budget, an increase of more than 35% since 2016[2].

Global cyberattacks (such as WannaCry and Petya) are executed by highly organized criminal groups, and organized or national-level crime groups have been behind many recent attacks. Typically, criminal groups buy and sell hacking tools and services on the cybercrime black market, wherein attackers share a range of hacking-related information. This online underground market is operated by groups of attackers, and it in turn supports the underground cybercrime economy [3]. The cybercrime underground has thus emerged as a new type of organization that both operates black markets and enables cybercrime conspiracies to flourish.

Because organized cybercrime requires an online network to exist and to conduct its attacks, it is highly dependent on closed underground communities (e.g., Hackforums and Crackingzilla). The anonymity these closed groups offer means that cybercrime networks are structured differently than traditional Mafia-style heirarchies [4], which are vertical, concentrated, rigid, and fixed. In contrast, cybercrime networks are lateral, diffuse fluid, and evolving. Since cyberspace is a network of networks [5], the threat posed by the rise of highly professional network-based cybercrime business models, such as Crimeware-as-a-Service (CaaS), remains mostly invisible to governments, organizations and individuals.

Even though Information Systems (IS) researchers and practitioners are taking an increasing interest in cybercrime, due to the critical issues arising from the rapid increase in cyber threats, few have attempted to put this new interest on a solid foundation or develop suitable methodologies. Previous studies have not analyzed the underground economy behind cybercrime in depth. Furthermore, little is known about CaaS, one of the primary business models behind the cybercrime underground. There is an overall lack of understanding, both in research and practice, of the nature of this underground and the mechanisms underlying it.

This research gap, and the practical problems faced by cybercriminals, motivates our study. We take a data analytics approach and investigate the cybercrime economy from a design science perspective. To achieve this goal, we (1) propose a data analysis framework for analyzing the cybercrime underground to guide researchers and practitioners; (2) define CaaS and crimeware to better reflect their features from both academic research and business practice perspectives; (3) use this to build a classification model for CaaS and crimeware; and (4) build an application to demonstrate how the proposed framework and classification model could be implemented in practice. We then evaluate this application by applying it in a case study, namely investigating the cybercrime economy by analyzing a large dataset from the online hacking community.

2. Review Literature:

Although both academics and practitioners have recently started to devote more attention to CaaS, its fast growing nature has prevented them from reaching consensus on how to define different types of CaaS and crimeware. As a result most of the academic research has borrowed the definitions used by the business practice literature leading to widely varying interpretations in different disciplines. Given this ambiguity, we approach categorizing CaaS and crimeware from an RAT perspective (considering vulnerabilities as suitable targets and preventive measures as capable guardians against crime) in a cybercrime underground context. In addition, we redefine CaaS and crimeware based on the definitions used in existing research and practice.

3. Classification of Crimeware Services and Products

The definitions of CaaS and crimeware used in the academic and business practices literature, which form a basis for our classification model, suitable for the IS field. We reclassify CaaS and crimeware in terms of the suitable targets (attack strategy/mode) and absence of capable guardians (preventive measures) in a cybercrime underground context.

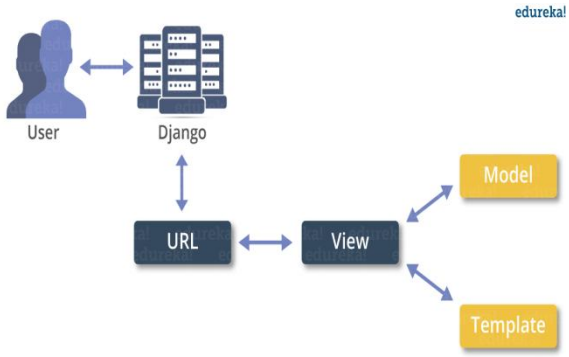
The different attack strategies/modes are associated with RAT's suitable targets because vulnerable organizations products, and services may suffer from attacks using a variety of strategies. In contrast, preventive measures are associated with RAT's absence of capable guardians because encryption and VPN services, crypters, and proxies are intended to neutralize preventive measures by bypassing anti-virus and log monitoring software.

4. Brute Force Attack Services

A brute force attack is an attempt to log in to an account and steal it by repeatedly trying random passwords. Such attacks often target less specific targets than phishing or social engineering. For example, an attacker may try to log in using one of the system's default usernames (e.g., "root" or "admin") by systematically trying all possible passwords. We thus define a brute force attack service as a service that hacks accounts by trying all possible passwords.

5. Python

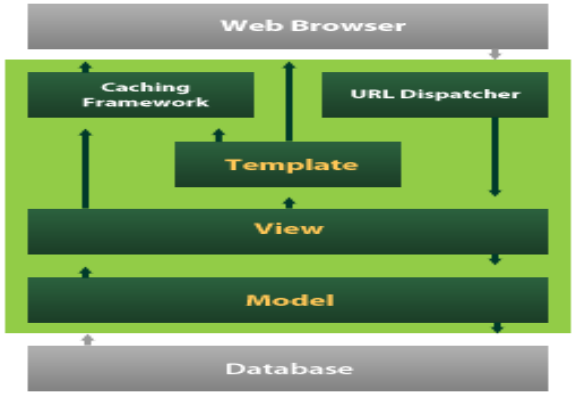
Python is a general-purpose interpreted, interactive, object oriented, and high-level programming language. An interpreted language Python has a design philosophy that emphasizes code readability (notably using whitespace indentation to delimit code blocks rather than curly brackets or keywords), and a syntax that allows programmers to express concepts in fewer lines of code than might be used in languages such as C++or Java. It provides constructs that enable clear programming on both small and large scales. Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of its variant implementations. C Python is managed by the non-profit Python Software Foundation. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative functional and procedural, and has a large and comprehensive standard library.



6. Django

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

Django's primary goal is to ease the creation of complex, database driven websites. Django emphasizes reusability and "pluggability" of components rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models.



Django also provides an optional administrative create, read update and delete interface that is generated dynamically through introspection and configured via admin models.

7. Objectives

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the

management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

8. Proposed System

The goal of our data analysis framework is to conduct a big picture investigation of the cybercrime underground by covering all phases of data analysis from the beginning to the end. This framework comprises four steps: (1) defining goals; (2) identifying sources; (3) selecting analytical methods; and (4) implementing an application. Because this study emphasizes the importance of RAT for analyzing the cybercrime underground the proposed RAT based definitions are critical to this framework Steps 1–4 all contain the RAT elements.

9. Naive Bayes Classifier

Naive Bayes is the Algorithm in Information Technology. Naïve Bayes a classification algorithm for binary (two class) and multi-class classification problems. The technique is easiest to understand when described using binary or categorical input values.

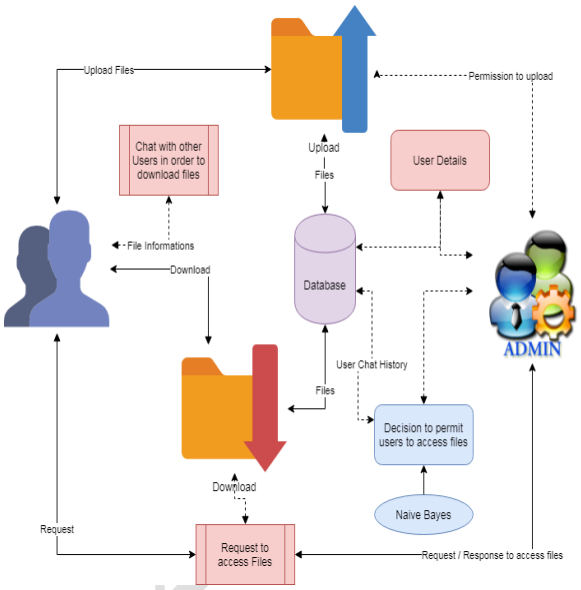
It is called *naive Bayes* or *idiot Bayes* because the calculation of the probabilities for each hypothesis is simplified to make their calculation tractable. Rather than attempting to calculate the values of each attribute value P(d1, d2, d3|h), they are assumed to be conditionally independent given the target value and calculated as P(d1|h) * P(d2|H) and so on.

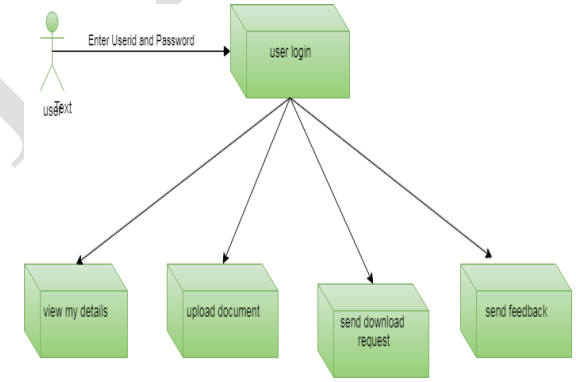
This is a very strong assumption that is most unlikel in real data, i.e. that the attributes do not interact. Nevertheless the approach performs surprisingly well on data where this assumption does not hold.

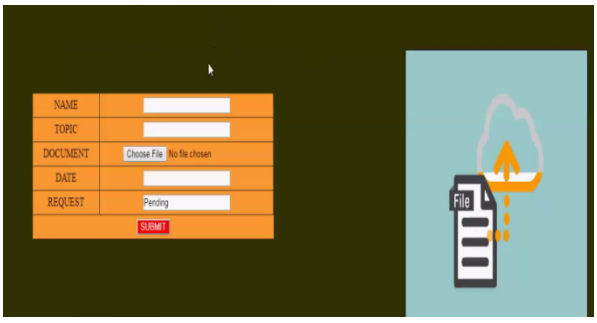
10. Architecture Diagram

These results can be intuitively understood, enhancing our understanding of how CaaS and crimeware change over time. First bar graphs show which of the selected keywords were most used within the given period. Second, daily trend graphs show the frequencies with which particular CaaS and crimeware items are mentioned. These both serve to highlight the changes in cybercrime market trends over time. Although this application is based on the proposed classifications it also allows new CaaS and crimeware items to be added that have not yet been classified.

The goal of our data analysis framework is to conduct a big-picture investigation of the cybercrime underground by covering all phases of data analysis from the beginning to the end. This framework comprises four steps: (1) defining goals; (2) identifying sources; (3) selecting analytical methods; and (4) implementing an application. Because this study emphasizes the importance of RAT for analyzing the cybercrime underground the proposed RAT based definitions are critical to this framework: Steps 1–4 all contain the RAT elements.







11. 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, crypting and VPN services) and crime ware (drive by download, botnets, exploits ransomware, rootkits Trojans, crypters, and proxies) based on definitions taken from both the academic and business practice literature. Based on these, we have built an RAT-based 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 crimeware trends our results show that the prevalence of botnets (attack-related crimeware) 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] J. C. Wong and O. Solon. (2017, May 12). *Massive ransom ware cyber-attack hits nearly 100 countries around the world*. [Online]. Available: https://www.theguardian.com/technology/2017/may/12/global-cyber-attack-ransomware-nsa-uk-nhs

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

[3] 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.

[4] 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.

[5] K. Hughes, "Entering 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 Quart.*, vol. 37, no. 2, pp. 337-356, 2013.

[7] 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.

[8] 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.

[9] S. Gregor, "Design theory in information systems," *Aust. J. Inf. Syst.*, vol. 10, no. 1, pp. 14–22, 2002.

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

[11]M. Yar, "The Novelty of 'Cybercrime': An Assessment in Light of Routine Activity Theory," *Eur. J. Criminol.*, vol. 2, no. 4, pp. 407–427, 2005.

[12] K.-K. R. Choo, "Organised Crime Groups in Cyberspace: a Typology," *Trends in Organized Crime*, vol. 11, no. 3, pp. 270–295, 2008.

[13]L. E. Cohen and M. Felson, "Social Change and Crime Rate Trends: A Routine Activity Approach," *Am. Sociol. Rev.*, vol. 44, pp. 588–608, 1979.