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

A Data Analytics Approach to the Cybercrime Underground Economy

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.

Keywords: Data Analysis, Cyber Crime, Python, VPN Services.

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

2. Background

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,

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.

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.

Crypting Services

Crypter encrypts programs or source code to avoid detection and tracking and thus bypass anti-virus software [30]. Like other hacking services, encryption is sold as a service because crypters require a certain level of skill to use. The goal of such a service is to neutralize the preventive measures put in place by organizations and anti-virus software preventing hacking programs from being caught or allowing them to be left behind to collect information. We define an crypting service as a service that encrypts malicious code by using a crypter to bypass anti-virus software.

VPN Services

All Rights Reserved © 2020 IJRDASE

Networks connect different entities and private networks only allow access by closed communities of authorized users [31]. The most secure way to access the Internet is using a VPN, because it hides all user information (e.g., identity and IP address). Because attackers use VPN services to avoid tracking or IP blocks, they are categorized as CaaS-related preventive measures. We thus define a VPN service as a service that provides a secure connection to the Internet via a virtual private network.

A. K. Sood and R. J Enbody, 2013, Crimeware as-a-service (CaaS) has become a prominent component of the underground economy. CaaS provides a new dimension to cyber crime by making it more organized, automated, and accessible to criminals with limited technical skills. This paper dissects CaaS and explains the essence of the underground economy that has grown around it. The paper also describes the various crimeware services that are provided in the underground market.

3. Methodology

This work depicts about the prerequisites. It determines the equipment and programming prerequisite that are needed for software to keeping in mind the end goal, to run the application appropriately. The Software Requirement Specification (SRS) is clarified in point of interest, which incorporates outline of this exposition and additionally the functional and non-practical necessity of this thesis.

General Description

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.

Users Perspective

The Characteristic of this task work is to give information adaptability security while sharing information through cloud. It gives a proficient approach to share information through cloud.

Feasibility Study

Believability is the determination of paying little respect to whether an undertaking justifies action. The framework followed in building their strength is called acceptability Study, these kind of study if a task could and ought to be taken.

Three key thoughts included in the likelihood examination are:

- Technical Feasibility
- Economic Feasibility
- Operational Feasibility

Technical Feasibility

Here it is considered with determining hardware and programming, this will effective fulfill the client necessity the specialized requires of the framework should shift significantly yet may incorporate

- The office to create yields in asecified time.
- Reaction time under particular states.
- Capacity to deal with a particular segment of exchange at a specific pace.

Economic Feasibility

Budgetary examination is the often used system for assessing the feasibility of a projected structure. This is more usually acknowledged as cost/favorable position examination. The method is to center the focal points and trusts are typical casing a projected structure and a difference them and charges. These points of interest surpass costs; a choice is engaged to diagram and realize the system will must be prepared if there is to have a probability of being embraced. There is a consistent attempt that upgrades in exactness at all time of the system life cycle.

Operational Feasibility

It is for the most part identified with human association and supporting angles. The focuses are considered:

What alterations will be carried through the framework?

- What authoritative shapes are dispersed?
- What new aptitudes will be needed?
- Do the current framework employee's individuals have these aptitudes?
- If not, would they be able to be prepared over the span of time?

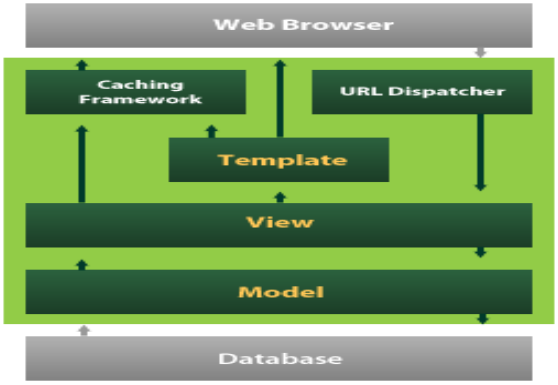
Python

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 white space 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

All Rights Reserved © 2020 IJRDASE

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.



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.
3.When the data is entered it will check for its validity. Data can be entered with the help of screens.

Modules:

- **Upload Files**
  Users are allowed to upload the files with the tags given. Once the file is uploaded, then it is sent to approval from admin to publish or make view to other users. These uploaded files can be in any form document, audio or video but not allowed to upload the executable (.exe) files.
- **Conversation Monitoring**
  Users are allowed to communicate among the other users. This could be monitor by the admin. The malicious conversion likes to threaten the data. In order to protect the cybercrime and prevents from forming cybercrime community. This can be achieved by the help of classification algorithm named naïve Bayes classification.
- **Download Files**
  The files can be downloading by requesting for the file and once admin approved the files then can be downloadable. The decision to approve files can be taken from the conversation between users. Admin takes the action on download files and approvable status of users. The users are allowed further actions based on the users.
- **Graphical Representations**
  The analyses of proposed systems are calculated based on the approvals and disapprovals. This can be measured with the help of graphical notations such as pie chart, bar chart and line chart. The data can be given in a dynamical data.

Algorithm

- **Naive Bayes Classifier**

Naive Bayes is 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.

- **Make Predictions with a Naive Bayes Model**

Given a naive Bayes model, you can make predictions for new data using Bayes theorem.

MAP (h) = max (P(d|h) * P(h))

Using our example above, if we had a new instance with the *weather* of *sunny*, we can calculate:

go-out = P (weather=sunny|class=go-out) *P(class=go-out) stay-home = P (weather=sunny|class=stay home) * P(class=stay home)

We can choose the class that has the largest calculated value. We can turn these values into probabilities by normalizing those as follows:

P (go-out |weather=sunny) = go-out / (go-out + stay-home)
P (stay-home|weather=sunny) = stay-home / (go-out + stay-home)

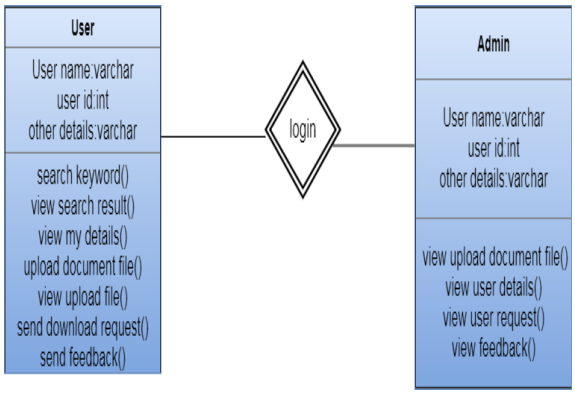
If we had more input variables we could extend the above example. For example, pretend we have a "*car*" attribute with the values "*working*" and "*broken*". We can multiply this probability into the equation.

For example below is the calculation for the "go-out" class label with the addition of the car input variable set to "working":

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

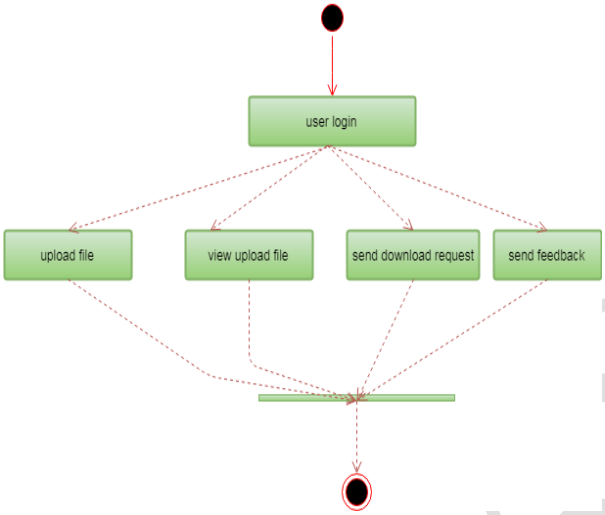
All Rights Reserved © 2020 IJRDASE

Class Diagram



Activity Diagram

User



System Test

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

Types Of tests

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

4. Result

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.

- Defining Goals The first step is to identify the conceptual scope of the analysis. Specifically this step identifies the analysis context namely the objectives and goals. To gain an in-depth understanding of the current CaaS research, we investigated the cybercrime underground, which operates as a closed community. Thus, the goal of the proposed framework is to "investigate the cybercrime underground economy."

- Identifying Sources the second step is to identify the data sources, based on the goals defined by Step 1. This step should consider what data is needed and where it can be obtained. Since the goal of this study is to investigate the cybercrime underground we consider data on the cybercrime underground community. We therefore collected

All Rights Reserved © 2020 IJRDASE

such data from the community itself and obtained a malware database from a leading global cyber security research firm. Because cybercriminals often change their IP addresses and use anti-crawling scripts to conceal their communications, we used a self-developed crawler that can resolve captchas and anti-crawling scripts to gather the necessary data. We collected a total of 2,672,091 posts selling CaaS or crimeware, made between August 2008 and October 2017, from a large hacking community site (www.hackforums.net) with over 578,000 members and more than 40 million posts.

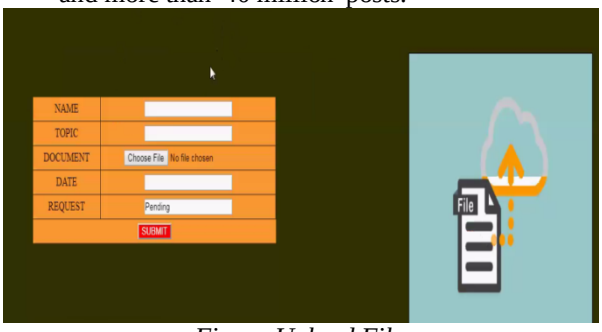


Figure Upload File

Figure User uploaded documents

5. 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.

6. Future Work:

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 crimeware 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.

References

[14]M. Felson, "Routine Activities and Crime Prevention in the Developing Metropolis," *Criminol.*,vol. 25, no. 4, pp. 911–932, 1987.

[15]F. Mouton, M. M. Malan, K. K. Kimppa, and H. S. Venter. "Necessity for ethics in social engineering research," *Comput.Security*, vol. 55, 114–127, 2015.

[16]A. S. Rakitianskaia, M. S. Olivier, and A. K. Cooper, "Nature and Forensic Investigation of Crime in Second Life," in *10th Annual Inf. Security South Afr. Conf.*, 2011.

[17]A. van der Merwe, M. Loock, and M. Dabrowski, "Characteristics and Responsibilities Involved in a Phishing Attack," in *Proc., 4th Int. Symp.on information and*

All Rights Reserved © 2020 IJRDASE

communication technologies, 2005, pp. 249–254: Trinity College Dublin.
[18]L. Volonino, R. Anzaldua, and J. Godwin, *Computer Forensics: Principles and Practices*. Prentice-Hall, Inc., 2006.
[19]G. Álvarez, F. Montoya, M. Romera, and G. Pastor, "Cryptanalyzing a Discrete-Time Chaos Synchronization Secure Communication System," *Chaos, Solitons & Fractals*, vol. 21, no. 3, pp. 689–694, 2004.
[20]M. Goncharov. (2014). *Russian Underground Revisited*.[Online]. Available: https://www.trendmicro.de/cloud-content/us/pdfs/security-intelligence/white-papers/wp-russian-underground-revisited.pdf
[21]V. Bezmalyi. (2014, Oct. 1). *Why Phishing Works and How to Avoid It*. [Online]. Available: https://blog.kaspersky.com/how-to-avoid-phishing/6145/
[22]C. Ng. (2014, May 21). *What's the Difference between Hacking and Phishing?* [Online]. Available: https://blog.varonis.com/whats-difference-hacking-phishing/
[23]P. Shankdhar. (2017, May 29). *Popular Tools for Brute-force Attacks*.[Online]. Available: http://resources.infosecinstitute.com/popular-tools-for-brute-force-attacks
[24]J. Mirkovic, G. Prier, and P. Reiher, "Source-end DDoS Defense," in *Second IEEE Int. Symp. on Network Computing and Applications, 2003. NCA 2003.*, 2003, pp. 171–178: IEEE Comput. Soc.
[25]A. Singh and D. Juneja, "Agent Based Preventive Measure for UDP Flood Attack in DDoS Attacks," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 8, pp. 3405–3411, 2010.
[26]D. McMillen. (2016, Mar. 24). *Why Botnets Remain the Go-To Weapon for Cybercriminals*. [Online]. Available: https://securityintelligence.com/why-botnets-remain-the-go-to-weapon-for-cybercriminals/

ICRAST 2020

All Rights Reserved © 2020 IJRDASE