DATA SECURITY APPROACH ON CYBERCRIME USING WEB VULNERABILITY

A FINAL YEAR CAPSTONE DESIGN PROJECT

P.VELMURUGADASS
Department of computer Science and engineering
Kalasalingam academy of Research and education,
Krishnankoil, India
p.velmurugadass@klu.ac.in

K.VISWAS AYYAPPA
Department of computer Science and engineering
Kalasalingam academy of Research and education,
Krishnankoil, India
9918004167@klu.ac.in

G.VENKATA SRINIVAS
Department of computer Science and engineering
Kalasalingam academy of Research and education,
Krishnankoil, India
9918004156@klu.ac.in

C.HRUTHINATH KUMAR
Department of computer Science and engineering
Kalasalingam academy of Research and education,
Krishnankoil, India
9918004150@klu.ac.in

Abstract: Internet could be a major supply of spreading terrorist act through speeches and videos. Terrorist organizations use net particularly social networks to brain wash people and promote terrorist activities through provocative websites that inspire helpless individuals to affix terrorist organizations. Therefore, here we have a tendency to propose Associate in Nursing economical net data processing system to observe such net properties and flag them mechanically for human review. websites are created from hypertext mark-up language (Hypertext markup language). In numerous arrangements and have pictures, texts etc., intermixed on a single website. Here, we have a tendency to used data processing further as net mining to observe patterns and mine out matter info on websites. Here, we have a tendency to are victimization E-mail System to observe the unwanted messages that are additional prone to terrorist act and can send to the spam on to the recipient who is victimization the system.
Despite the speedy increase of cyber threats, there has still been very little analysis into the foundations of the topic or methodologies that would serve to guide info systems researchers and practitioners UN agency touch upon cybersecurity. additionally, very little is thought regarding crime-as-a-service (CaaS), a criminal business model that underpins the crime underground. This analysis gap and also the sensible cybercrime issues we have a tendency to face have impelled North American nation to research the crime underground economy by taking an information analytics approach from a style science perspective. to realize this goal, we: (1) propose a data analysis framework for analyzing the crime underground; (2) propose CaaS and crimeware definitions; (3) propose Associate in Nursing associated classification model, Associate in Nursingd (4) develop an example application to demonstrate however the planned framework and classification model might be enforced in apply. We then use this application to research the crime underground economy by analyzing an oversized knowledge set obtained from the net hacking community. By taking a style science analysis approach, this paper contributes to the planning of the email phishing in the area of cybercrime and the area is to fulfill the data security to the email messages consisting the data and providing them the security using the naïve bayes to planning the

industries will harden attacks by the crime underground. Among these, one of the attack is the email phishing.

Keywords:
CAAS (Crimeware-as-a-Service), ICMP (Internet Control Message Protocol), IRC (Internet Relay Chat), MIB (Management Information Base), CVE (Common Vulnerabilities and Exposures).

1. INTRODUCTION:

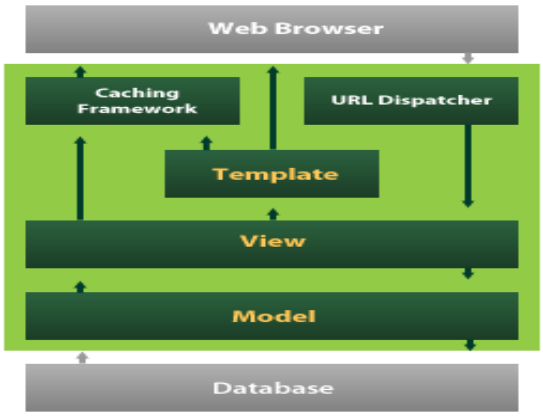
The rapid development of Internet technologies has immensely changed on-line users' experience, while security issues are also getting more overwhelming. The current situation is that new threats may not solely cause severe injury to customers' computers however conjointly aim to steal their cash and identity. Among these threats, phishing may be a noteworthy one and may be a criminal activity that uses social engineering and technology to steal a victim's identity knowledge and account info. Most social engineering attacks area unit initiated and administrated by the attackers in person. By means that of in person handled social engineering attacks area unit particularly those that use the manner of impersonation principally deception to be in distress, a troublesome state of affairs, or urgency. Social engineering attacks area unit initiated usually in 2 ways: By the attackers one by one and by creating use of computers. the opposite ways that of handling social engineering attacks area unit by exploitation computers or automatic means that. a way of assaultive is thru faux websites, that area unit simply created. Websites that appear as if the legitimate sites can also be created therefore simply. One very talked-about style of social engineering attack is finished by giving free downloads or terribly high discounts and inspiring them to use their official ids. The persons could also be attracted and supply substantial details within the method. In the scope of social engineering, attackers use some necessary approaches that may be place into physical, social, and technical class. in an exceedingly physical approach, because the name implies, the offender performs some physical actions so as to urge data regarding the victim like looking through Associate in Nursing organization's trash, that is termed Dumpster diving. A Dumpster is a valuable supply {of data |of data |of knowledge} like personal information regarding staff, manuals, memos of sensitive data. in a very social approach, attackers deem socio psychological techniques like Cialdini's principles of persuasion to govern their victims. samples of persuasion strategies embody the employment of authority. Attackers typically use search engines to assemble personal data regarding future victims. There are tools that may gather and mixture data from completely different internet resources.

The user has to create an account on the e-mail server by clicking the creating account on the login page. The user can send or receive the e-mail once he created the account. The system will check the mail data and will make the large data into smaller parts, classify the data and will crosscheck the server for the keywords which are spam. Then, it will classify whether the mail is spam or ham.

The overview of the cybercrime data mining is to mine out the patterns in the email to prevent the crime anticipate criminal activity. The Naïve Bayes and the K-means algorithm are used to classify the datasets and to mine the patterns. So, that it can record and form the prediction or the output about the spamming mails.

Data mining is a technique that are capable to scan the accuracy and performance in cybercrime. Web mining is also a technique of text mining technique which can mine out the patterns in the large datasets. The both are techniques to mine out the data present in the mail. After, finding the data in the dataset it is going to classify it and verify it with the keywords which are already in the given dataset. Web mining improves the functionality of an online software by classifying content and identifying web sites. It's utilized for internet searching (e.g., Google, Yahoo), as well as vertical searching (e.g., Fat Lens, Become, and so on). To forecast user behavior, web mining is used. Web mining is particularly beneficial to a certain website and e-service, such as landing page optimization. Website mining is separated into three categories of mining techniques: website mining, internet structure mining, and internet usage mining and data mining using methods.

Cybercrime has undergone a revolutionary modification, going from being product-oriented to service-oriented as a result of the fact it operates within the virtual world, with totally different abstraction and temporal constraints, differentiates it from different crime taking place within the physical world [11]. As a part of this alteration, the crime underground has emerged as a secret crime marketplace as a result of rising technological changes have provided organized cybercriminal teams with unexampled opportunities for exploitation [12]. The crime underground includes a extremely skilled business model that supports its own underground economy [5]. This business model, referred to as CaaS, is ‘‘a business model employed in the underground market wherever illegal services are provided to assist underground consumers conduct cybercrimes, like attacks, infections, and concealing in an automatic manner,’’ [3]. Thus, CaaS is referred to as a do-it-for-me service, not like crimeware that may be a do-it-yourself product.



2. PROBLEM STATEMENT:

The phishing email is one in all the numerous threats within the world nowadays and has caused tremendous monetary losses. Though the strategies of confrontation are frequently being updated, the results of these strategies aren't terribly satisfactory at this time.

Moreover, phishing emails are growing at associate degree dismaying rate in recent years.

Therefore, more practical phishing detection technology is required to curb the threat of phishing emails. during this project, we tend to 1st analyzed the e-mail structure.

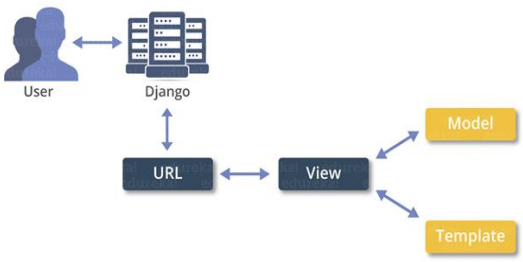
Then supported associate degree improved repeated Convolutional Neural Networks (RCNN) model with structure vectors and a focus mechanism, we tend to planned a replacement phishing email detection model named Titaness, that is employed to model emails at the e-mail header, the e-mail body, the character level.

3. PROPOSED WORK:

The proposed system uses machine learning algorithms to get implemented.

- There are some features used in the system they are data mining and web mining. Data mining could be a technique want to mine out patterns of helpful data from large data sets. Web mining conjointly consists of text mining methodologies that takes us to scan and extract and mine useful content from unstructured data.
- This system will check the sender messages and whether the message is promoting terrorism. Data mining and the web mining are used together at tough times for efficient system development. System will find the unnecessary messages that are more suspected to terrorism and will send directly to the receiver’s spam account.
- In this, the advantage is that the data can be easily taken into smaller parts for easy consideration. From, the small datasets the phishing can be easily done, and the process will be fast.

4. BLOCK DIAGRAM:



5. SOFTWARE REQUIREMENTS:

- Python
- NumPy
- MySQL
- Jupiter notebook

PYTHON

Python is a computer programming language frequently used to build websites and software, automate responsibilities, and conduct statistical analysis. Python is a preferred cause language, meaning it may be used to create a spread of different programs and is not specialized for any unique problems. Characteristics of Python Following are critical traits of Python Programming

- It helps practical and based programming strategies in addition to OOP.
- It can be used as a scripting language or maybe compiled to byte-code for constructing large packages.
- It provides very high-level dynamic data types and helps dynamic type checking.

Jupyter notebook:

The Jupyter Notebook is an unimaginably amazing asset for intelligently creating and Introducing information science projects. This article will walk you through how to set up Jupyter Journals on your nearby machine and how to begin utilizing them to do information science projects.
To capitalize on this instructional exercise, you ought to be acquainted with programming, explicitly Python and pandas. All things considered, assuming you have insight with another dialect, the Python in this article shouldn't be excessively mysterious, will in any case assist you with getting Jupyter Notebooks set up locally. Jupyter Notebooks can likewise go about as an adaptable stage for having the opportunity to grasp with pandas furthermore even Python, as will become clear in this article.
On Windows, you can run Jupyter using the alternate route Anaconda adds to your beginning menu, which will open another tab in your default internet browser. This isn't a journal presently, yet don't freeze! There's very little to it. This is the Notebook Dashboard, explicitly intended for making due to your Jupyter Notebooks. Consider it the Launchpad for investigating, altering, and making your scratchpad.

NumPy:

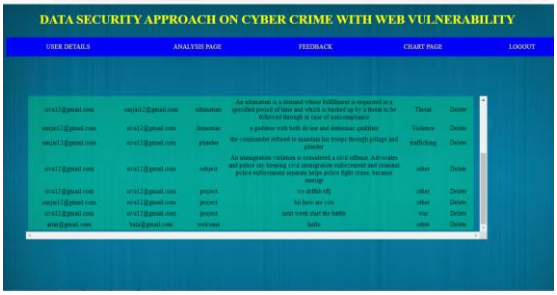
- Arrays of NumPy provide the introduction of modern statistics with a large amount of data. NumPy makes making these projects very easy and hassle-free.
- NumPy provides hidden lists and frequent items for similar members. It also comes with functions like logical transitions, Fourier variations, standard line algebra, and much more.
- While changing the layout of an N-dimensional array, NumPy will create new arrays of that and remove the old ones.
- This python package provides useful integration tools. You can easily integrate NumPy with programming languages such as C, C ++, and Fortran code.
- NumPy offers such services as MATLAB. Both allow users to speed up the process.

MySQL:

- MySQL is a relational database.
- MySQL offers a most powerful transactional database such as robust transactional support for the system.

Algorithm:

A naïve Bayes classifier is an algorithm that classifies things using Bayes' theorem. Naive Bayes classifiers are based on the assumption of robust (or naive) independence between data point properties. Spam filters, text analysis, and medical diagnosis are all examples of common applications for naive Bayes classifiers. Because they are simple to implement, these classifiers are commonly employed in machine learning. Simple Bayes or independent Bayes are other names for naive Bayes. Here, we have used the naïve Bayes classifier to mine out the common words or to stop words from the mail what's the user send. The common words like, and, then, the, there, or, therefore, hereafter, these, it, is, it's, this, he, she, her, him, etc., This are called preprocessing. In this process we have to eliminate these kind of words. From those messages, we are having a tendency to getting to light the filtration words. It is an useful algorithm for most of the machine learning side projects which are helpful to the society and the people. It keeps us to give more security to our objects.



Result and discussion:

In this section, we show the results obtained for the spamming. By using classification algorithms, we developed this code. By using this code, we can easily find the spam or ham using our code and the design.In this code mainly, Naïve Baye's classification takes place, everyone knows about the classification. For example, if we send a mail if it contains any war related means it goes under category war, or if it contains any violence words it goes under category of violence.

Conclusion:

In this paper, we developed a page to tell the email is spam or ham. To curb and destroy the terrorism and spreading of their activities through online social media through unwanted messages and images to cover the helpless people, we need to use the powerful method or

system. That system should be useful to the cops for easily give awareness to common people and find the person who are spreading the harmful words as well as who are all involved in terrorism. The system will destroy the terrorism found in the emails using their message content, sender mail and the receiver mail. So, finally the project is to destroy the ham mails in the emails and all. The system will easily get the spam mails and the ham mails using the algorithms and all.

REFERENCES

1. David Karig and Ruby Lee, "Remote Denial of Service Attacks and Countermeasures," Princeton University Department of Electrical Engineering Technical Report CE-L2001-002, October 2001.

2. Lincoln Stein and John N. Stuart. "The World Wide Web Security FAQ", Version 3.1.2, February 4, 2002. http://www.w3.org/security/faq/ (8 April 2003).

3. Paul J. Criscuolo. "Distributed Denial of Service Trin00, Tribe Flood Network, Tribe Flood Network 2000, And Stacheldraht CIAC-2319". Department of Energy Computer Incident Advisory Capability (CIAC), UCRL-ID-136939, Rev. 1., Lawrence Livermore National Laboratory, February 14, 2000.

4. "Yahoo on Trail of Site Hackers", Wired.com, February 8, 2000. http://www.wired.com/news/business/0,1367,34221,00.html (15 May 2003).

5. "Powerful Attack Cripples Internet". Associated Press for Fox News 23 October 2002. http://www.foxnews.com/story/0,2933,66438,00.html. (9 April 2003)

6. Joseph Lo and Others. "An IRC Tutorial", irchelp.com. 1997. http://www.irchelp.org/irchelp/irctutorial.html#part1. (8 April 2003).

7. Nicolas Pioch. "A Short IRC Primer". Edition 1.2, January 1997. http://www.irchelp.org/irchelp/ircprimer.html#DDC. (21 April 2003).

8. Kleinpaste, Karl, Mauri Haikola, and Carlo Kid. "The Original IRC Manual". March 18, 1997. http://www.user-com.undernet.org/documents/irc-manual.html#seen (21 April 2003).

9. Kevin J. Houle. "Trends in Denial of Service Attack Technology". CERT Coordination Center, Carnegie Mellon Software Engineering Institute. October 2001. www.nanog.org/mtg-0110/ppt/houle.ppt. (14 March 2003).

10. Federal Computer Incident Response Center (FedCIRC), "Defense Tactics for Distributed Denial of Service Attacks". Federal Computer Incident Response Center. Washington, DC, 2000.