Data Security Approach Analysis on Cyber Crime with Web Vulnerability

Pusuluri Venkata Naga Raghavendra[1, *], Mr. Y. Venkata Narayana[2]

Department of Master of Computer Applications[1] Department of Information Technology[2] Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, AP, India

---***---

Abstract - Internet is the one of the major source in performing cyber crimes like terrorism activities in the form of speech, content as well as video formats. Many terrorist organizations like ISIS uses internet sources especially social networks to manipulate human individuals and promote terrorist activities through web pages and mails that inspire helpless and innocent people to join terrorist organizations. Especially, they have targeted people to inspire and make them join in their organizations through by spam emails. Therefore, cyber security is a major concern at such type of situations. In order to detect flags to such kind of malicious activities, we proposed a web-data mining system with the help of Naive Bayes method. We designed E-mail System for detecting the unwanted messages related suspicious terrorism activities. The proposed methodology successfully detected spam messages and separated them to spam folder directly to the recipient who is using the system.

Key Words: Cyber Crime, Cyber Security, Data Mining, Naive Bayes, Spam Emails, Spam Detection

1. INTRODUCTION

After the typical behaviour done by some of terrorists, by applying a data- mining algorithms to the textual content of terror-related harmful Web sites. The result profile was used by the system admin to perform detection of users suspected of being engaged in terrorist activities. And this algorithm is based on the content of previous existing terrorist sites and known terrorist traffic on the Web. Data mining is a technique which is used to mine out useful data from large data sets. Web mining also contains text mining methodologies that allows to scan and extract the useful content from unstructured data. This system will check the messages send by the sender and whether the message is promoting terrorism. using web mining as well as data mining together at same time causes efficient system development. System will find the messages that are unwanted and more susceptible to terrorism and will send directly to the receiver's spam account. It will give more awareness to the users.

2. LITERATURE SURVEY

In the literature survey there are some papers that i have found on detecting spam mails based on some techniques which avoid unnecessary mails to spam. Some use different techniques and different methods like Case Base Spam Filtering Method, technique which is called as Rule Based Spam Filtering Technique, Previous Based Spam Filtering Technique, Adaptive Spam Filtering Technique [1]. Spamming emails are necessary, though them harmful messages like terrorist suspicious activities are filtered and also they include viruses and spy wares, so there is an emergency need for detecting spam emails to prevent cyber attacks from internet. Several methods are there for detecting spam emails which are based on the methods of machine learning, they were submitted to reduce suspicious emails and get results of high priority for spam email classification. In this work, a new predictive method is submitted based on Naive Bayes algorithm. In an digital investigations that every investigator typically has to deal with thousands of digital artifacts. Among them, email is one of the main focus that potentially can generate useful information. However, i notice downplay the importance of analyzing spam emails as they are generally assumed to be irrelevant junk emails. In this article, we have illustrated on how these irrelevant messages might play a crucial role in digital investigations. Five scenarios are introduced in which the investigator likes to overlook crucial crime information that has been disguised as spam. The methods used by criminals in these cases are discussed. In light of these covert criminal communications, we call for more attention from the digital forensics community to realize how email spam may help in criminal activities [2]. Electronic mail (email) is significant for many kinds of group connection, which has become widely used by many people individuals and organizations. At the same time, email is one of the fast rising and costly problems linked with the internet today, in which this case it is called spam email. Spam emails are mainly idealistic or have attractive links to famous websites but they lead to sites that are interfering [3]. As a result, spam emails causes minimize in privacy, spreading viruses, occupying space in the email box, and destroying email servers. Therefore, the user wastes a lot of time in filtering email imports and cancelling the unwanted email. The discovery of unwanted emails categorizes the emails as spam or non-spam (harm), so this process is related to the classification problem [4].

3 EXISTING SYSTEM

In an Existing system, detection of terrorism was presented by using audit information on Web traffic content. After that the behaviour of terrorists activities by applying a data mining algorithm to the textual content of terror-related Web sites. The resulting profile was used by the system to perform detection of users suspected of being engaged in terrorist activities. And this algorithm is be based on the content of existing terrorist sites and known terrorist traffic on the Web.

4 PROPOSED SYSTEM

There are two features used in this system that is data mining and web mining. Data mining is a technique that is used to dig out useful data from large data sets. Here web mining as well as text mining both consists some methodologies that allows to scan and extract useful data content from the unstructured data.

This system will check the messages send by the sender and whether the message is promoting terrorism. Both Data mining as well as web mining are used together at times for efficient system development. System will find the unwanted or unnecessary messages that are more suspicious to terrorism and will send directly to the receiver's spam account. It will give more awareness to the users.

4.1 BAYES THEOREM

Bayes' theorem is one of those key mathematical concepts that can be used to answer anything from a simple problem such as whether to go on a picnic to answering complex questions on if a person has cancer given the result of a medical test and more importantly it is a powerhouse behind some day-to-day ML algorithms such as the Naives Bayes Classifier (commonly used for spam detection). Hence it becomes important for Data Scientists to understand the Bayes Rule and is also an often touched upon topic during interviews.In our proposed system, we have used Naive Bayes classifier to classify whether an email is spam or not. The probability of an event may depend on the occurrence or non-occurrence of another event. This dependency is written in terms of conditional probability.

4.1.1 CONDITIONAL PROBABILITY

$$P(A|B) = P(A\quad B) / P(B)$$

$$P(B|A) = P(A\quad B) / P(A)$$

$$P(A\quad B) = P(B|A)\, P(A) = P(A|B)\, P(B)$$

An event A is INDEPENDENT from event B if the conditional probability is the same as the marginal probability. P(B|A) = P(B) P(A|B) = P(A) From the formulas the Bayes Theorem States the Prior probability: Unconditional probabilities of our hypothesis before we get any data or any NEW evidence. Simply speaking, it is the state of our knowledge before the data is observed. Also stated is the posterior probability: A conditional probability about our hypothesis (our state of knowledge) after we revised based on the new data. Likelihood is the conditional probability based on our observation data given that our hypothesis holds.

$$P(A|B) = P(B|A)\, P(A)/ P(B)$$

$$P(B|A) = P(B|A)\, P(B)/ P(A)$$

Where P (A|B) is the posterior probability, P(B|A) is the likelihood and P(A) prior probability.

Very little is known about Thomas Bayes even though his name has been lent to an entire branch of statistics. As a matter of fact it is not even clear if anybody has seen Bayes.

4.2 SYSTEM ARCHITECTURE

System Architecture Design sometimes simply known as System Design is a conceptual representation of the components and subcomponents that reflects the behaviour of a system.

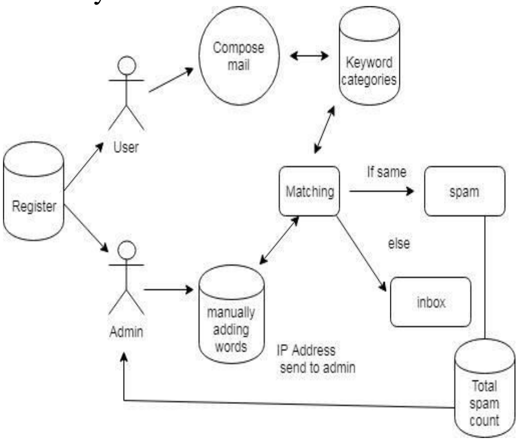


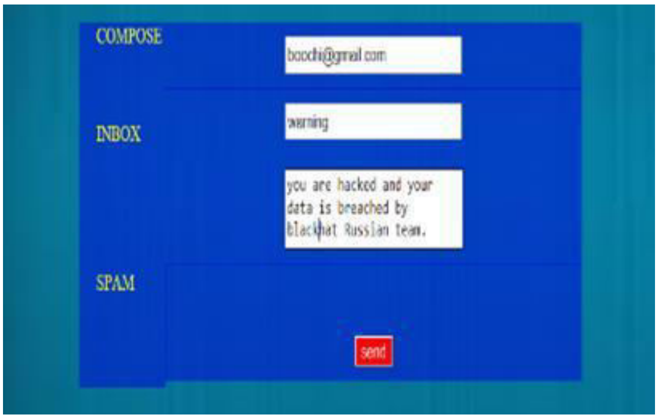
Figure 1. System Architecture

4.2.1 MODULE DESCRIPTION

This system comprises of 4 modules named as Mailing, Filtering, Spam Detection, Pre-Processing.

4.2.1.1 MODULE 1: MAILING

First, User should Register with their basic details through create an account link. By using that details they need to Login for enter into the system. Then they will receive the message of "success". Here, we are using the system like E-mail. Hence, it contain the features of inbox, sent mail, spam, recent histories, etc., user can compose the mail with whom to sent. It may be related to terrorism or may something related to common things. Here, the recent history denotes the person who is doing mail recently. To check the performance of proposed system, an email has been composed as shown in figure 2.



boochi@gmail com

warning

you are hacked and your data is breached by blackhat Russian team.

Figure 2. Mail Composition

4.2.1.2 MODULE 2: FILTERING

In this Module, I have a few data's in my Data-set. With that, I will check whether the sent message have contain the filtration words about terrorism or not? I have using Data mining technique to filter out text data from the large data sets for making the most useful of obtained results. Web mining consists of text mining methodologies. Through that text mining, we can extract the text or content what are all related to terrorism. If the filtration words are match with the sent message means, the receiver receives the mail in his/her spam box or else inbox. These data has been depicted in table 1.

Table 1 Filtering data

First Name	Last Name	User Id	Mobile Number	Email Id
Santhosh	Kumar	Santhosh	9789672188	Chennaisunday.cs0216@gmail.com
Suresh	kumar	Suresh	9789672189	Chennaisunday.cs0209@gmail.com
Sabari	nathan	Sabari	9789672189	Sabarinathan1350@gmail.com
Sanjai	kumar	Sanjai	9785372180	Sanjai12@gmail.com
Siva	krishna	Siva	9589452187	Siva12@gmail.com
Ram	Dev	Ram	6776776778	ram@gmail.com
James	james	James	7417417414	james@gmail.com
Raghava	raghava	Raghava	8798654878	8465009441naaga@gmail.com
Venkat	venkat	Venkat	8798654878	venkat@gmail.com
Hello	My name	Hellomyname	9987878789	boochi@gmail.com

4.2.1.3 MODULE 3: SPAM DETECTION :

In this Module, Admin should login first. It will contain the pre-defined user name and password. Admin side, it has the features of keywords, spam, analyze chart. By using Mining concepts Administrator can add few terrorism related words manually in few parameters/ categories. That keywords will also be going to add with the existing data-set. In spam, we can see what are all spam messages from starting. In analysis, It contains a mail having how many words in those keyword categories and their total count per each mail.

4.2.1.4 MODULE 4: PRE-PROCESSING :

In this Module, Admin can see all the spam mail sent and receive in this system, whereas, Spam Detection will contain pre-processing which means it will remove all the common words/stop words such as the, and, or, here, there, etc., Here, I have used the Naive Bayes algorithm. After pre-processing, I have highlighted the filtration words in mails. Then it contains every categories count as total spam Detection count. Finally by make use of the total spam Detection count, did the chart.

5 IMPLEMENTATION

To justify the proposed system's efficiency, a data security approach on cyber crime with web vulnerability was implemented, in which the online experiment was conducted to study users satisfaction about the sites recommendation system. this project was implemented using python and MySQL, and the front-end interface using Html, CSS, JavaScript.

6 ANALYSIS

In processing, we decided to code a data. We coded a bar graph, a data visualization of grapes. Data talks in grapes what types of data security crime related to cyber world. The analysis of crime have been depicted in figure 3 and figure 4.

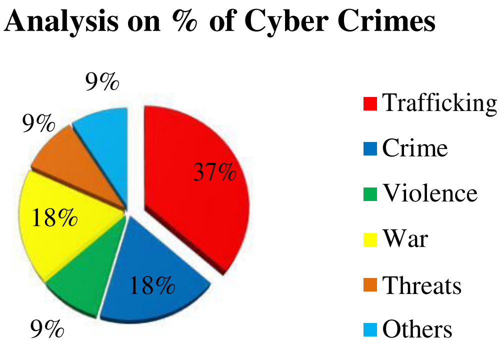


Figure 3. Crime Analysis

We pulled the datasets from online and uploaded in processing to analyze data security through web. If we observe figure 4, 40% of crimes are trafficking. This can be happened through dark web which is illegal and required advanced knowledge. In case of any email related to any cyber crime mentioned in figures 3 and 4 happened through email, our proposed architecture can easily detect.

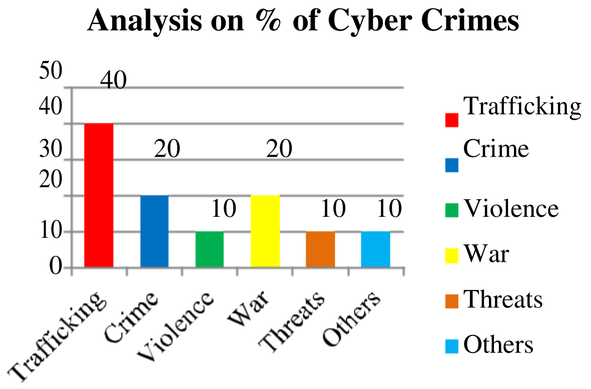
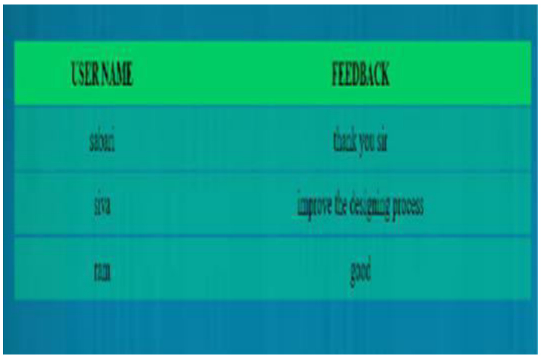


Figure 4. Crime Analysis

7 RESULTS AND DISCUSSION

After our experimental investigations, we got feedback from the clients who used our proposed system like good, bad, excellent etc. This feedback has been depicted in figure 5.



USER NAME	FEEDBACK
saloni	thank you sir
siva	improve the designing process
ram	good

Figure 5 Feedback from Clients

8 CONCLUSION

To curb and destroy the terrorism and spreading of their activities through online social media like mails through unwanted messages and images to cover the helpless people, we need to use the powerful method or system. Our proposed system can be useful to the cops to spread awareness to common people and also to find the person who are spreading the harmful words as well as who are all involved in terrorism.

References:

[1] Khorsi, Ahmed. "An overview of content-based spam filtering techniques." Informatica 31.3 (2007). https://doi.org/10.1016/j.fcij.2018.11.006.

[2] Wei, Chun, et al. "Mining spam email to identify common origins for forensic application." Proceedings of the 2008 ACM symposium on Applied computing. 2008.

[3] Pandey, Mayank, and Vadlamani Ravi. "Text and data mining to detect phishing websites and spam emails." International Conference on Swarm, Evolutionary, and Memetic Computing. Springer, Cham, 2013.

[4] Teli, Savita Pundalik, and Santosh Kumar Biradar. "Effective email classification for spam and non-spam." International Journal of Advanced Research in Computer and software Engineering 4.2014 (2014).

BIOGRAPHIES

Pusuluri Venkata Naga Raghavendra is post-graduate scholar in the Department of Master of Computer Applications, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, AP, India. He has Bachelor's Degree in Computer Science. He is a passionate Cyber Security Researcher and he mentors students of several universities at a reputed organization. His area of interest includes web application penetration testing, advanced ethical hacking and digital forensics.

Mr. Y. Venkata Narayana is an Assistant Professor in the department of Information Technology,Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, AP, India. He has published more 12 research papers in various reputed peer reviewed journals, international conferences and book chapters. He has more than five years of teaching experience in the field of Information Technology . His research interest includes big data, smart computing.