IJAER/ Aug-Sep 2012 /Volume-1/Issue-2 ISSN: 2278-9677
International Journal of Arts & Education Research

"PREVENTION OF CYBER CRIME USING CLASSIFIER ACCURATELY AND EFFICIENTLYDETECTS SUSPICIOUS URLS"

Prachi Verma

Research scholar

Department of Computer Science

CMJ University, Shillong Meghalaya.

Dr. Rajesh Verma

Supervisor

Department of Computer Science

CMJ University, Shillong Meghalaya.

1.1Introduction

Online Social Networks enables to control the messages which were presented on evade the undesirable substance on their own private space showed by the users. This framework has direct full control on the messagesof user which were posted on their limits. Tragically, in inappropriate hands, there are additionally powerful devices for spam battles to be executed. It is an adaptable guideline based structure that it concurs of users that make so as to the filtering rules to be applied to their dividers. To remodel spam communication into crusade for arrangement decently than inspecting them into exclusively, machine Learning-based delicate classifier naturally discovers suspicious messages on the side of substance basedfiltering. For the most part in this section manages twitter and face book to order suspicious URLs and its related spam data so that we designed our framework as java as front end and My-SQL as back end.

1.1.2 TIPS TO GET PROTECTED FROM CYBER CRIME

Some easy tips to protect computers from the growing threats: **Terminate Online Session Completely:**

Shutting the program window or composing in another website address without logging out may give others a possibility of accessing your account data. Continuously end your online meeting by tapping on the "Log out or Sign Out" button. Abstain from utilizing the choice of "recollect" your username and password data.

Create Backup of Important Data:

Backup of all the significant records whether personal or expert ought to be made. Becoming acclimated to back up your records normally is the first step towards security of your personal computer.

Use Security Programs:

On the off chance that your system doesn't have data insurance programming toensure on the web, at that point by all methods purchase internet security program for your computer. Today, practically all new computer systems accompany a security programs introduced.

Protect Your Password:

Take a stab at making a password that comprises of a blend of letters (both capitalized and lower case), numbers and exceptional characters. Password ought to be changed routinely. Try not to impart your password to others.

Participation in Social Networking:

While taking an interest in most long range informal communication locales don't uncover the personal data to other people and these destinations have a specific power of authority over security issues. Use privacy settings to forestallpersonal data being communicated.

Use Your Own Computer:

It's commonly more secure to get to your financial accounts from your own computer as it were. On the off chance that you utilize some others computer, consistently erase all the "Temporary Internet Files", and clear all your "History" in the wake of logging off your account.

Update Your Software Package Regularly:

Frequent online updates are needed for all the Internet security software installed on your computer system.

Using Email:

A straightforward principle in utilizing this communication instrument isn't to open any connections in messages from individuals you don't have the foggiest idea. Hackers do utilize E-mail as the principle target looking to take personal data, financial data, security codes and other. Try not to utilize the connection sent to you. In the event that you need access to any website, visit the websiteby composing the location in your menu bar. Cyber crime, being a consuming issue far and wide, numerous nations is starting to execute laws and other administrative components trying to limit the occurrence of cybercrime. The laws in numerous nations on adequacy of the punishment and counteraction of computer crime requires a hearty number and extent of the guidelines, and even the procedures, which lingers a long ways behind the truth of interest for computer crime in legal practice.

1.2 WHAT IS URL REDIRECTION

URL redirection, which is additionally called URL sending, is a WorldWide Web methodology for making a web page existing under more than one URL manages its location. At the point when a web program attempts to open a URL and that must be diverted, a page with an alternate URL is discharged. The URL redirection can be utilized for URL limitation and to stop broken connections when web bring are moved and furthermore consent to such a numerous space names fit in to a similar proprietor to allude to a specific web website. It likewise aides such route into and somewhere else even to a portion of a website, for privacy wellbeing, and for less guiltless standard, for example, phishing attacks. A divert is a page which has no mollified itself, however dispatch the peruser to anotherinvestigate, section of a page, or article for the most part from an alternate substitute title. The content prearranged in the connection on a beset divert page should precisely coordinate the objective section heading or stay composition, including capitalization. URL diverts the got connections to anobsolete URL that can be sent to the honest area. It is a crime in the event that you apply these to any website without getting any authorization from the proprietor. In the event that crime is demonstrated, the person will be punished under Information Technology (Amendment) Act, 2008, Section 43(a) read with section 66 is applicable. The malicious users will be punished of imprisonment, which may extend to three years or fine with fivelakhs rupees under section 43(a).

1.3 SYSTEM ANALYSIS AND DESIGN

1.3.2 Problem Definition

Past suspicious URL discovering frameworks are disgraceful at security against limited redirection servers that separate examiners from standard

programs and divert them to delicate pages to shroud resentful points of arrival. Here new suspicious URL and spam data recognition framework has been discovered for Twitter and Face book that is ALERT SYSTEM Application device. Not at all like the past ALERT SYSTEM frameworks, is this application instrument hearty and cautious against provisional redirection, because it doesn't depend on the facial appearance of noxious foyer pages that may not be accessible and spam data. Rather, it checks the gathering point on the relationship of different divert shackles that share redirection spam data and its servers.

Existing System

Account include based configureuration utilize the characteristic highlights of spam account, for example, the proportion of tweets that hold the URLs. The analyst pages of character URLs in each tweet, which may not be viably gotten and which will be viewed as the relationship of URL divert chains are taken out from various tweets. The attacker's assets are commonly restricted and when required reuse their URL divert chains, which for the most part share similar URLs. The associated URL diverts chains and their tweet to find a few highlights and setting the data that can be utilized to group uneasy URLs and spam data. Conventional suspicious URL and spam location frameworks are inadequate in their security against limited redirection servers that separate specialists from typical programs and forward them to mercifully display pages to cover malicious arriving of pages.

Demerits of Existing System

1.3.2.1 It is ineffective against features fabrications

1.3.2.2 No user defined BL

1.3.2.3 Lack of BL Management

1.3.2.4 Spammers can easily change the shape of message

1.3.2.5 Plotting twitter graph is somewhat difficult

1.3.2.6 It consumes much time and resources

1.3.3 Proposed System

Twitter and Facebook users share a URL and data with friends through tweets or messages. So as to lessen the URL Length of messages in interpersonal organization, they use URL abbreviated services and furthermore their connection will concentrate on increasingly vigorous highlights where malicious suspicious user can't interfere in their network. Account and connection include based plan doesn't identify spam messages from traded off accounts, because the undermined accounts have kind highlights. A connection cultivating attacks for expanding spammers' social impacts have been directed. To adapt to malicious tweets, many Twitter and Face book Spam Detection Schemes being proposed. Here in this exploration, an application device - suspicious URL and spam identification framework for Twitter and Facebook has been proposed. By utilizing this device numerous twitter and Facebook user account can be included. At first, empower follow, unfollow, tweet, retweet and so on ought to befollowed.

By selecting tweet we can enter the message in the content field and it naturally refreshes or can send message in twitter and Face book open course of events. In the event that any suspicious URLs and data are available in that account which implies it can without much of a stretch be adjusted and distinguished by our Application instrument.

For the explanation that attackers' resources are restricted and their should be reused and a segment of their divert chains must be shared.

Merits of Proposed System

1.3.3.1 Investigate correlation of URLs

1.3.3.2 Effectively remove unwanted message by FR.

1.3.3.3 User defined block list is possible

1.3.3.4 Email provides the beneficiary a selection to respond instantaneously.

1.3.3.5 A business opportunity for email service providers can be set asnew data.

1.3.4 Very quick and easy to find relevant information.

System Architecture

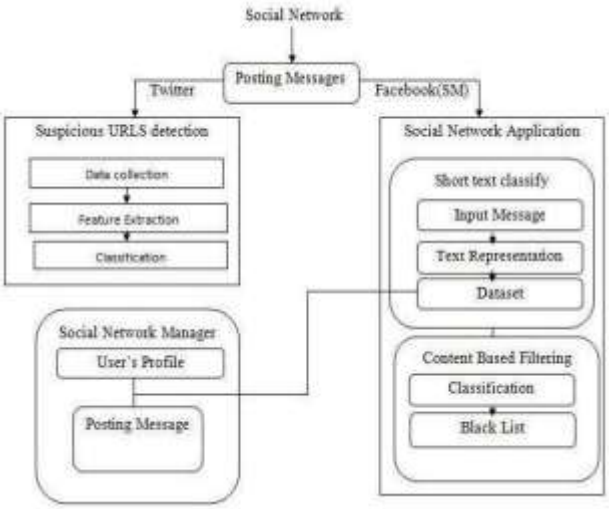

Figure 1.1: System Architecture

1.3.5 Data Flow Diagram

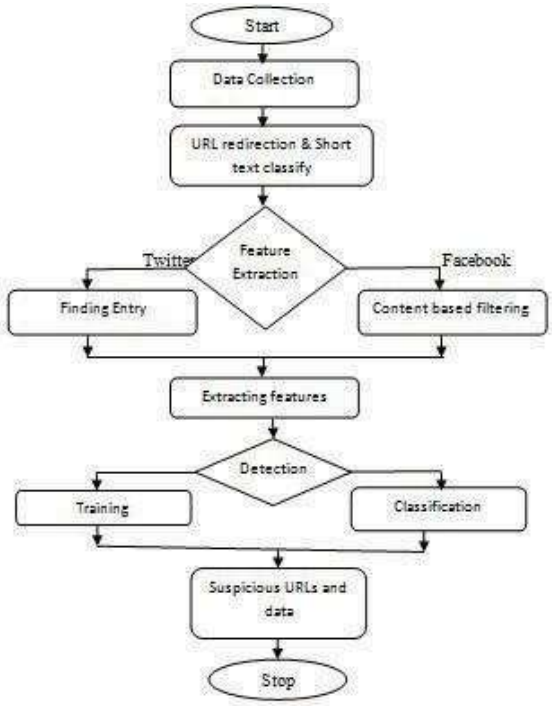

Figure 1.2: Data Flow Diagram

System Requirements

Hardware Requirements

System:	Pentium IV 2.4 GHz
Hard Disk:	160 GB
Monitor:	15 VGA color
Mouse:	Logitech.
Keyboard:	110 keys enhanced
Ram:	1GB

Software Requirements

O/S:	Windows XP.
Language:	Java.
IDE:	NetBeans 6.9.1

Data Base: MySql

1.3.6 System Specification

- **Java**

Java is an Object Oriented programming language which was initially evolved by James Gosling and his colleagues at Sun Microsystems (a subordinate of Oracle Corporation Pvt. Ltd) and discharged Java in 1995 as a center segment of Java Software from Sun Microsystems' Java stage. The language gets most its grammar from C and C++ and furthermore it has a less difficult item model with low-level offices in less. Java applications are regularly amass to byte code (class document) that can be utilized to run on any stage for Java Virtual Machine (JVM) which were paying little mind to its computer engineering and stage it has. Java is a class-based, object- situated language, simultaneous, broadly useful language which is explicitly designed to have as meager execution dependency as achievable. It is intended to let interface application designers to raise "compose once, and can run anyplace".

- ***Java Platform***

One characteristic of the Java is compactness, which implies the computer programs written in Java language should run correspondingly on any of the Hardware or in its particular Operating System Platform. This is accomplished by incorporating Java programming language code to a halfway language portrayal called Java byte code. These Java byte code guidelines are profoundly streamlined arrangement of guidance which are similar to machine code and are intended to be deciphered by a virtual machine (VM) composed deliberately for the host equipment System.

- ***Java a High-level Programming Language***

A high-level Object Oriented programming language was created by Sun Microsystems. Java was at first called OAK language which was additionally named in an edible oil partnership, and at first designed for handheld devices and furthermore for set-top boxes. Oak was fruitless so that in 1995 Sun Microsystems changed the name from OAK to Java and altered the programming language to exploit prospering World Wide Web.

- ***Net Beans and J2EE***

The Net Beans Platform is a reusable Integrated Development Environment system to improve the programming advancement of Java Swing work area application. The Net Beans IDE gives group to Java SE and every one of its bundles which contains records and organizes what is expected to begin to create Net Beans module and furthermore to its Net Beans Platform based applications. Here, no extra SDK is required in this product.

- ***WAMP Server***

WAMPs are packages of exclusively made arrangement of code which were introduced on computers that may utilize a Microsoft Windows operating system stage. WAMP is an abbreviation framed from the underlying of the Microsoft Windows operating system and the central segment of this package: Windows, Apache, MySQL and one of PHP or Perl or Python.

- **MySQL**

The MySQL improvement venture made source code realistic under the conditions of the GNU General Public License which is a freeware, just as under a selection of restrictive programming understandings. MySQL have ownership of and supported by a solitary organization revenue driven firm, the Swedish organization called MySQL AB, presently again got by Oracle Corporation.

- **Apache**

Apache is a web server which is prestigious as a Tomcat web server for freeware. MySQL is an open-source database which has given a commitment to all other open source programming. PHP is a scripting language which is utilized to control data held in a database and can likewise create web pages powerfully each time when it is mentioned by a program window. Different programs may likewise be remembered for this package, for example, phpMyAdmin which give a Graphical User Interface to MySQL database chief availability, or the option scripting dialects, for example, Python or Perl. Proportionate packages are LAMP (for the Linux operating system) and MAMP (for the Apple Mac).

1.4 SYSTEM IMPLEMENTATION

1.4.2 System Modules

1.4.2.1 Suspicious URLs Detection System

1.4.2.2 Social Network account Creation and Posting Messages

1.4.2.3 Short Text Classifier

1.4.2.4 Content Based Filtering

Module Description

Suspicious URLs Detection System

Detection is the mining of specific data from a greater progression of data without exact help from or synchronization with the dispatcher. The expression "detector" was first utilized for a device that detects the nearness or nonappearance of a straightforward radio communication signals. Duplication detector finds sections in which the content on the two pages is the equivalent. The detection system can't see suspicious URLs with active substance. These disclosure systems may even now not measure up to see far fetched destinations with confined practices. Attackers can build the highlights of their ambush to avoid our recognition system. This online report of ALERT SYSTEM returns dicey URLs that have become visible in the earlier hour close to actual time detection. It suggests that attackers had changed the characteristics of their accounts to stay away from detection.

Social Network account Creation and Posting Messages

- Interactions among individuals in which they make, offer, and trade data and thoughts in virtual communities and systems.
- The scope of apparatus strategies in work expands on indistinguishable qualities from

those for data mining.

- An Information filtering plan is a system that expel overflow or pointless data from a grouping stream utilizing programmed or computerized techniques before the course of action of a human user.
- This informal community viewpoint give a lot of strategy to examining the development of entire common elements just as a scope of hypotheses clarification to the example saw in these game plan.
- Social set-up and the investigation of them is a normally interdisciplinary scholarly field which rises up out of social brain science.
- One of the fundamental ideal models in current human science is social set-up examination. It is likewise working with various other social and authority sciences.
- The term is utilized to represent a social plan dictated by such relations.
- The informal community approach in circumspect social communication is that social wonder ought to be for the mostpart imagine and investigate through the properties of relationship between and inside these units, in its place the properties of the units inside themselves.

Short Text Classifier

- The significant exertion in building a hearty Short Text Classifier (STC) is amassed in the extraction and selection of a lot of characterizing and segregated highlights.
- The text characterization might be grouped by their subjects or as indicated by different traits.
- At designing and assessing different portrayal techniques are joined with a neural information plan to semantically group short messages.
- A rundown of evaluations is then utilized by the ensuing periods of the filtering procedure.

Content Based Filtering

- The substance of every passage is spoken to as a lot of terms or descriptors, normally the words that occur in a record.
- Same terms are utilized to speak to the user profile and they are developed by investigate the satisfied of issue which have been seen by the customer.
- The terms are alloted routinely a technique must be selected that can mine these terms from objects.
- The terms must be spoken to such an extent that both the user layout and the articles can be looking at in a significant manner.
- The user layout is spoken to with a similar jargon and developed by break down the placated of things which have been seen by thecustomer.
- Several issues must be estimated cautiously when execute a substance based filter system.
- The substance of every section is spoken to as a lot of articulation or descriptor, regularly the articulations that emerge in any record.
- Genetic Algorithms and Neural systems are generally much more slow contrasted with other information techniques as a few emphasess are alluring to choose whether or not an article isconnected.
- The capacity of an information technique is to adjust to the adjustments in the user's inclinations likewise assumes a significant job in content based filtering.

1.5 RESULT AND DISCUSSION

1.5.2 Algorithm used to find out suspicious URLs and its related dataStep 1:

At first before instating factors we need to store some suspicious URLs, ordinary URLs, spam data and typical data in database utilizing Wamp server.

Step 2:

1.5.2.1 In twitter, suspicious URLs are classified using Enhanced Offline Supervised Algorithm

1. Get the input from user
2. Compare with database
3. Provide output to user

- In Facebook, URLs and its related spam data are classified usingtwo techniques
  1. Short text Classify-Used to expand short text
  2. Content based filtering-Used to filter spam data

Step 3:

- In twitter, after enabling all task time has to be set to run aparticular task
  1. Follow amount per time.
  2. Interval time between 2 follow.
  3. Maximum follows per day.
- In case of facebook the following terms are calculated
  1. No of variables in posted message
  2. Calculate TF, DF, IDF, TFIDF value in postedmessage

Step 4:

- In twitter, subsequent to setting show time to utilizing Enhanced Offline Supervised Algorithm suspicious URLs can be effectivelycharacterized.
- Whereas, on account of Facebook messages it must be filtered utilizing content based filtering technique.

Step 5:

- After filtering suspicious URLs and its related spam data, particular message can be posted or unwanted message can be removed.

BIBLIOGRAPHY

1. Dr. Amita Verma, Cyber Crimes & Law (Central Law House Publications, Allahabad, 2009).
2. Justice A.K. Ganguly, "Legal framework inadequate to tackle cyber crime", The Hindu, 2008.
3. Dr. M. Das Gupta, Cyber Crime in India: A comparative Study (Eastern Rule Home Pvt. Ltd., Kolkata, 2009).
4. Dinniss Harrison, Heather, Cyber Warfare & the Laws of War, Cambridge University Press, U.K., 2012.
5. Farooq Ahmed, Cyber Law in India- Law on Internet (Fresh Age Law Publications, Delhi 2008).
6. Dr. V. Paranjape, Legal magnitude of Cyber Crimes and Defensive Law through individual Reference to India (Central commandment Organization Publication, 2010).
7. K. Mani, "A sensible budge toward to cyber laws", Kamal Pub., 2012.
8. Nandan K., Law connecting to Computers, Internet: A Guide to Cyber Laws & the IT Act, 2000 (Universal Law Pub. Co., 2009).
9. RK Chaubey, An foreword to Cyber Crime & Cyber Law, Kamal Law House Publication, (2009)
10. Vivek Sood, Cyber Law Simplified, (Tata Mcgraw- Hill Publishing Company Limited, New Delhi, 2008).
11. Jaishankar, K. (2011). Cyber Criminology: Explore Internet Crimes & Criminal Behavior. CRC Press: Taylor & Francis Collection, USA.
12. Kamini D, "Cyber offense in the Society: Problems and Preventions", Journal of option perspective in the Social Sci. (2011) Vol. 3, No 1, 240-259.
13. Yar, Majid (2006), Cyber Crime and Society, Sage Publications, London.
14. Clarke, Richard (2010),Cyber War: The Next Threat to National Security and What to Do About It. Harper Collins Publisher, USA.
15. Higgins, George (2010), Cybercrime: An Introduction to an Emerging Phenomenon, McGraw Hill Publishing, New York.
16. Holt, Thomas J (2011), Crime Online: Correlates Causes and Contexts. Durham, Caroline Academic Press, USA.
17. Brenner,W.Susan (2010), Cybercrime:Criminal threats from cyberspace. Greenwood Publishing group, Westport.
18. Brenner Susan, W., Cyber Threats, The Emerging Fault Lines of the Nation State, Oxford University Press, New York, 2009.
19. Rege, Aunshul (2009), What's love got to do with it? Exploring online dating scams and identity. International Journal of Cyber Criminology, Vol.3(2) , 494-512.
20. Das, Biswajet and Sahoo, Jyoti (2011),Social Networking Sites. A Critical Analysis of its impact on Personal and Social life. International Journal of Business and Social Science, Vol.2 (14)
21. Koh (2006), Click, Click, Who's Really There ? Charlotte North Carlifornia : LHK Publisher.
22. Duggal,Pavan (2009), Cyberlaw:The Indian Perspective, Saakshar Law Publications, New Delhi.
23. Vadhera, Sharad (2012), Fate of Social Networking Sites in India. Kan and Krishme, Global Advertising Lawyers Alliance.
24. NourMohammad "Internet Peculiarity and Territorial Traditionalism Converge on Cyberspace: A Study of Techno- legal Synchronization in the USA", Cyber Law Cybercrime Internet an Ecommerce, By Prof. VimlenduTayal, Bharat Law Publications (2011),P 444.
25. Dr. Gupta & Agarwal "Information Technology Law and Practice", Premier Publishing Company.(2009)
26. Ravindra Kumar Singh "Formation and Enforcement of E-Contracts in the Cyber World: An indian Perspective", Cyber Law CybercrimeInternet an E-commerce, By Prof. VimlenduTayal, Bharat Law Publications (2011),P 410
27. Dr.AshokMakkar in "Legislative Framework to Combat Cybercrimes in India: An Overview", Cyber

Law CybercrimeInternet an E-commerce, By Prof. VimlenduTayal, Bharat Law Publications (2011).

28. Dr. V.Tayal," Cyber Piracy in the Indian Information Technology Regime: Issues and Challenges" Cyber Law CybercrimeInternet an E-commerce, By Prof. Vimlendutayal, Bharat Law Publications (2011).