International Journal on Applications in Engineering and Technology
Volume 8 : Issue 3 : July 2022, pp 15 –18 www.aetsjournal.com ISSN (Online) : 2455 - 0523

MONITORING SUSPICIOUS DISCUSSIONS ON ONLINE FORUMS USING DATA MINING

JEEVITHA . M , A . SRINIVASAN

***Abstract*—** In today's world people are showing lot of interest towards internet as a discussion medium. As internet usage had been increasing more and more, this led to many legal and illegal activities. First the news has been presented in internet before they are published in regular media. Dissemination of copyrighted movies, threatening messages and online gambling are some of the illegal activities happening due to internet technology. The law implementation bodies are searching for solutions to tackle this problem by downloading suspected postings. To handle this problem we proposed an effective system. To monitor criminal activities and illegal postings data mining algorithm is used. From selected discussion forums this method downloads postings continuously and hot topics are identified by data mining techniques and by using word-based user profiles authors are clustered into different groups. For security purposes Internet news, blogs, etc are monitored and analyzed by this system. This is achieved with the help of text mining method. With the support of devising of patterns and trends high quality information is derived. Online plain text sources are analyzed by the system from selected discussion forums and different groups of text will be formed and legal and illegal postings are decided by the system. Illegal activities happening on internet reduced by this system.

***Keywords*—** Online gambling, Monitoring, Discussion Forums, Data Mining.

I. INTRODUCTION

Increasing digital media crimes give an alert to law enforcement systems to monitor online activities continuously. A system needs to be derived to monitor online suspicious postings. Information on internet changing constantly is proved by lot of surveys and facts, thus it is difficult to manage information. To overcome the above problem, data mining is the right choice for collection of data. Raw data is collected from a large text corpus by applying data mining methods and in pre-processing stage structured data will be created from the raw data.

Jeevitha M, Student, M.Sc Computer Science, Rathinam College of Arts and Science, Coimbatore, Tamil Nadu, India – 641021, (e-mail: jeevithakm2002@gmail.com).

Mr. A. Srinivasan, Assistant Professor, Department of Computer Science, Rathinam College of Arts and Science, Coimbatore, Tamil Nadu, India – 641021, (e-mail: srinivasanar91@gmail.com).

II. SYSTEM DEVELOPMENT

A. Existing System:

The problem is thus particularly tricky, on one hand, the development of the Internet allows complex and sophisticated services to be offered, and on the other hand, these services offer to the attacker many new weaknesses and vulnerabilities to exploit. Almost all traditional approaches for building secure systems only focus on preventing attacks to be successful. Such approaches are becoming insufficient when used in the context of open networks like the Internet, which are characterized by frequent appearance of new attacks. Current systems are so complex that it is impossible to identify and correct all their vulnerabilities before they are put in operation. The application of content-based filtering on messages posted on OSN user walls poses additional challenges given the short length of these messages other than the wide range of topics that can be discussed.

B. Proposed System:

The automatic filtering scheme has been applied in order to eliminate the unwanted messages in the social networks. The proposed system applies short text classifier (STC) in the extraction and selection of a set of characterizing and discriminate features.

The proposed system is a system used to detect unwanted messages in multitier web applications/ social networks. Our approach can create customized rule models of users. To achieve this, employ a lightweight virtualization technique to assign each user's filtering scheme to a dedicated container, an isolated virtual computing environment.

III. PROPOSED MODULES

1)Interface Creation

The interface creation module represented as the implementation media, where the proposed system co clusters and wall filtering has been implemented on. The system contains some basic process and procedure in this module.

2)Profile Creation

In this module the profile will be created with the user details like name, address, user id, password and other details.

3)Authentication

This module is to ensure that the correct user entering and access the resources. The owner can enter and alter the information provided by the user. The modules will ask for the password for the authentication.

This module contains the user and the administrator authentications. The admin will have permission to view the entire processes done by the user. The user can only view the authenticated page after getting registered to the approach. User can view their personal information and the data which sent by him. In the server module have the static and secure login to enter and starts the server to receive the data.

4)Rules And Policy Setting

Spam detection system controls traffic by matching outgoing packets against a set of rules grouped in policy files. Each rule is defines the parameters against which each connection are compared , resulting in a decision on what action to take for each connection. As soon as a network packet matches a rule , that rule is applied , and processing stops .

5) Filter User Messages

The core components of the proposed system are the customized rule based Messages Filtering and the Text Classifier modules. The later component aims to classify messages according to a set of categories.

The unwanted wall posting is a malicious activities or policy violations. Then there should be an attempt to stop an intrusion attempt. This module is for avoiding spam by analyzing the rule violation. The person of a company will not be allowed to send any message to the restricted user without the administrator permission. If the person sends the person mail will be blocked.

IV. TESTING METHODOLOGIES

The different types of testing are:-

1. Unit Testing
2. Integration Testing
3. Validation Testing
4. Output Testing
5. User Acceptance Testing

1)Unit Testing

Unit testing focuses verification efforts on the smallest unit of software design, the module. This is also known as “Module Testing” The modules are tested separately this testing is carried out during programming stage itself. In this step each module is found to be working satisfaction as regard to the expected output from the module.

2)Integration Testing

Integration testing focuses on the design and construction of the software architecture. Data can be lost across an interface, one module can have adverse effect on another sub functions and show on. Thus integration testing is a systematic technique for constructing test to uncover errors associated with in the interface. In this project, all the modules are companied and then the entire program is tested as a whole.

3)Validation Testing

Validation testing is the requirement established as a part of software requirement analysis is validated against the software that has been constructed. This test provides the final assurance whether the software needs all functional, behavioural and performance requirements. Thus the proposed system under consideration has been tested by using

validation testing and found to be working satisfactory.

4) Output Testing

After performing the validation testing, the next step is the output testing of the proposed system, since no system could be useful if it does not produce required output in the specific format. Tested asking the users about the format required by them, the output is considered into two ways: one is on the screen and the other is printed format. The output format on the screen is found to be correct as the format designed according to the user needs, for the hard copy also, the output comes as specified by the user. Hence output testing does not result in correction in the system.

5) Whitebox Testing

White box Testing is done with the project which drive test cases that do the following

- Exercise at least once.
- Exercise all Guarantee that all the independent paths with in modules have been logical decision on the true and false side.
- Execute all loops at the boundaries and within their operation bounds.
- Exercise internal data structures to ensure the validity

It is aimed at ensuring that the system works accurately and efficiently before live operation command.

6) Blackbox Testing

Black box System methods focus on the functional requirement of the software. Using the black box testing method the following errors are identified and rectified in the package.

- Incorrect or Missing functions
- Interface Errors
- Errors in data Structures or external database access.

7) User Acceptance Testing

User acceptance testing of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keep in touch with the prospective system user at time of developing and making changes wherever required

8) System Implementation

In this project, propose virtualizing Harvard architecture on top of the existing memory architecture of modern computers, including those without non-executable memory page support, so as to prevent the injection of malicious code entirely. Harvard architecture is simply one wherein code and data are stored separately. Data cannot be loaded as code and vice-versa. In essence, we create an environment where in any code injected by an attacker into a process' address space cannot even be addressed by the processor for execution.

In this way, we are attacking the code injection problem at its root by regarding the injected malicious code as data and making it un addressable to the processor during an instruction fetch. Split memory architecture produces an address space where data cannot be fetched by the processor for execution. For an attacker attempting a code injection, this will prevent him from fetching and executing any injected code.

V. CONCLUSION

Thereby, this project meets and satisfies all the end user's needs by accomplishing the inputs provided and the output delivered. This web application was successfully created and stored all the modules, testing details, and deployment details into the application using this application. The application was tested very well and the errors were properly debugged. Testing also concluded that the performance of the system is satisfactory. All the necessary output is generated. This system thus provides an easy way to automate all the functionalities of consumption. If this application is implemented in little consumption, it will be helpful. Further enhancements can be made to the project, so that the website functions in a very attractive and useful manner than the present one. It is concluded that the application works well and satisfy the needs. The application is tested very well

and errors are properly debugged. It also acts as the sharing of files to the valuable resources.

REFERENCES

[1] Murugesan, M. Sururthi, R. Pavitha Devi, S. Deepthi, V. Shri Lavanya, and Annie Princy. Automated Monitoring Suspicious Discussions on Online Formus Using Data Mining Statistical Corpus Based Approach.
[2] Imperial Journal of Interdisiplinary Research (IJIR) Vol- 2, Issue-5, 2016.
[3] Harika Upgaganlawar, Nilesh Sambhe. Surveillance of Suspicious Discussions on Online Forums Using Text Mining. International Journal of Advances in Electronics and Computer Science, Volume- 4, Issue-4, April-2017.
[4] Suhas Pandhe and Sahil Pawar. Algorithm to Monitor Suspicious on Social Networking Sites Using Data Mining Techniques. International Journal of Computer Applications. Volume 116 – No. 12, April 2015.
[5] Javad Hosseinkhani, Mohammad Koochakazei, Solmaaz Keikhaee and Yahaya Hamedi Amin. Detecting Suspicion Information on Web Crime Using Crime Data Mining Techniques. International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol.-3, No. 1, 2014, Page 32-41.
[6] G.Vinodhini, R.M Chandrasekran. Sentiment Analysis and Opinion Mining: A Survey. International Journal of Advanced Research in Computer Science and Software Engineering. Volume 2, Issue 6, June-2012.
[7] Fabio Calefato, Filippo Lanubile, Nicole Novielli. EmoTxt: A Toolkit for Emotion Recognition from Text, University of Bari Aldo Moro.
[8] M.F Portar. An Algorithm for Suffix Stripping Program. Vol. 14 Issue: 3, pp. 130-137.