Scientific Journal Impact Factor (SJIF) : 4.364
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745 ISSN (PRINT) : 2393 - 8161

DETECTION AND BLOCKING SOCIAL MEDIA MALICIOUS POSTS

Miss. Sayali S. Karmode[1] and Prof. Vaishali B. Bhagat[2]
[1,2] CSE, P. R Patil College Of Engineering, Amravati

Abstract- Online Social Networks (OSNs) witness a rise in user activity whenever an event takes place. Malicious entities exploit this spur in user-engagement levels to spread malicious content that compromises system reputation and degrade user experience and has recently been reported to face much abuse through scams and other type of malicious content, especially during news making events. We have reached the era of social media networks represented by Facebook, Twitter, Flicker and YouTube. Internet users spend most of their time on social networks than search engines. Public figures and business entities set up social networking pages to promote direct interactions with the online users. Social media systems heavily depend on users for getting content and sharing. Information used is spread across the social networks in quick and effective manner. However, at the same time social media networks become vulnerable to different types of unwanted and malicious hacker or spammer actions. It has been observed that there is a greater participation in Facebook pages regarding malicious content generation. These contents will be in greater amount as compared to legitimate content. In this work we develop a detection mechanism to distinguish between malicious and genuine posts within seconds after the posts are uploaded by user. This work proposes an extensive way on the basis detection and blocking of malicious posts has been done.
Keyword- Social Network , malicious, blocking, etc.

I. INTRODUCTION

Cyber attacks primarily occur on social networks. Popular sites such as Facebook and Twitter currently have millions of active users. The popularity of social networks makes them exiting venues to for executing malicious activities. Due to the huge popularity of social media network these makes it easy for cybercriminal to misuse them. These can be in the form of media, thread or malicious post which does not belongs to a user. These posts upon clicking will take the user to some other pages created by malicious user. Cybercriminals create interesting posts that are actually baits which will be attracted by some users. Typical social engineering plans include the use of Interesting posts that ride on seasonal events, celebrity news and even disasters. So it needs to identify malicious post and blocking also by using techniques on social media [2]

We have reached the era of social media networks represented by Facebook, Twitter, Flicker and YouTube. Internet users spend most of their time on social networks than search engines. Public figures and business entities set up social networking pages to promote direct interactions with the online users. Social media systems heavily depend on users for getting content and sharing. Information used is spread across the social networks in quick and effective manner. However, at the same time social media networks become vulnerable to different types of unwanted and malicious hacker or spammer actions. It has been observed that there is a greater participation in Facebook pages regarding malicious content generation. In this work we develop a detection mechanism to distinguish between malicious and genuine posts within seconds after the posts are uploaded by user. This work proposes an extensive keyword set based on the textual content and URL features to identify malicious content on social media at zero time. The intent is to catch malicious or vulgar content that is currently evading social media detection mechanism

II. LITERATURE REVIEW

Sazzadur Rahman in 2007 has developed FraAppe, an accurate classifier for detecting malicious Facebook applications. Most interestingly, he highlighted the emergence of app-nets—large groups of tightly connected applications that promote each other.

Moreover, Lin in 2008 was interested in determining the events that are of interests to social networks' users based on their texts data. In this study they collected information from the internet, online communities, and social networks.

Justin Ma et al. have demonstrated the potential of a classifier based on suspicious URLs [4]. They train their dataset on properties such as host-name length, overall URL length, and the count of the sub domain separating character. Combining these lexical features with host information (e.g. DNS registry info), the researchers report an accuracy rate of over 95%.

Sakak in 2007 analyze the real-time interaction of micro blogging events especially on Twitter. In their opinion the user may be considered as a sensor to monitor tweets posted recently and to detect different events. Justin Ma et al. have demonstrated the potential of a classifier based on suspicious URLs [5]. They train their dataset on properties such as host-name length, overall URL length, and the count of the sub domain separating character. Combining these lexical features with host information (e.g. DNS registry info), the researchers report an accuracy rate of over 95%.

Gao et al. in 2009 presented an initial study to quantify and characterize spam campaigns launched using accounts on Facebook [6]. They studied a large anonymized dataset of 187 million asynchronous ―wall‖ messages between Facebook users, and used a set of automated techniques to detect and characterize coordinated spam campaigns. Authors detected roughly 200,000 malicious wall posts with embedded URLs, originating from more than 57,000 user accounts.

Following up their work, Gao et al. presented an online spam filtering system that could be deployed as a component of the OSN platform to inspect messages generated by users in real time [7]. Their approach focused on reconstructing spam messages into campaigns for classification rather than examining each post individually.

III. OBJECTIVES

- To detect malicious posts on social media
- To calculate the ratio of malicious content of social media post.
- To block malicious posts.
- To give more trustworthy, accuracy and efficiency while using social media.

IV. PROPOSED METHODOLOGY

It contains following steps:

1) For detection and blocking malicious contents, very firstly extraction of the data or posts uploaded by the user has to be done. Whatever the data in the form of numbers or texts has been extracted.
2) Then after extracting data from social media, splitter technique has been used here by considering space as separator [8].
3) After that, concept of k- means algorithm has been used where different clusters of malicious keywords are created according to category [9].
   a) For Violent;
      Violent= {v1, v2, v3, v4, v5, v6, v7, v8, v9…….vn}
   b) For Offensive;
      Offensive= {o1, o2, o3, o4, o5, o6, o7, o8, o9…….vn}
   c) For Hate;
      Hate= {v1, v2, v3, v4, v5, v6, v7, v8, v9…….vn}
   d) For Vulgar;
      Vulgar= {w1, w2, w3, w4, w5, w6, w7, w8, w9…… vn}

4) After that inputted post clusters i.e. C1 and malicious keywords clusters C(v), C(o), C(h), C(w) have been used for further purpose. Now here, brute force algorithm has been used for string matching. String matching is applied on clusters on both categories. Now cluster C1 is get compared with the malicious keywords cluster C(v), C(o), C(h), C(v) one by one. Earlier the counter has been set 0, as their no match found between both clusters. Now after finding the match between C1 and C(v) or C(o) or C(h) or C(w), the counter get increased by 1, and the matched string is selected for further process. Logic is as follows:

```
Statement stm=con.createStatement();
ResultSet rs=stm.executeQuery("select * from keywords");
while(rs.next()){
if(rs.getString(2).contentEquals(element)){
followed_word=""+list.get(i+1);
// out.println(""+followed_word);
wordmatch=wordmatch+","+element;
 if(rs.getString(4).contentEquals("Violant")){
 violantcount=violantcount+1;
}
if(rs.getString(4).contentEquals("Offensive")){
 offensivecount=offensivecount+1;
                    }
if(rs.getString(4).contentEquals("Vulgar")){
vulgurcount=vulgurcount+1;
                    }
 if(rs.getString(4).contentEquals("Sexual")){
 sexualcount=sexualcount+1;
                    }
```

5) Now the intersected matched string from both cluster is get compared with predefined followed keywords. After comparing if matching is done then the statement is declared as the malicious. Logic is as follows:

```
Statement selectword=con.createStatement();
ResultSet rscheck=selectword.executeQuery("select * from followedword");
while(rscheck.next()){
        if(followed_word.contentEquals(rscheck.getString(2))){
        malcontent=true;
        }
```

6) To decide the percentage of the malicious content of a particular post we used following formula which was previously used for detecting malicious contents using text analysis, 2016 and logic is as follows as follows:

$$\text{Malicious ratio} = \frac{\text{Total No.of a Malicious Word Match}}{\text{Total Number of Words in Post}} * 100$$

```
prst.setString(5,""+((float)violantcount/(udata.length()))*100);
prst.setString(6, ""+((float)vulgurcount/(udata.length()))*100);
prst.setString(7, ""+((float)offensivecount/(udata.length()))*100);
prst.setString(8, ""+(float)sexualcount/(udata.length())*100);
```

7) When the ratio of malicious contents of a particular user get cross the ratio 10%, then the user will blocked for posting on social media [10].

V. SYSTEM IMPLEMENTATION

Fig 1.1 User registrations and login panel

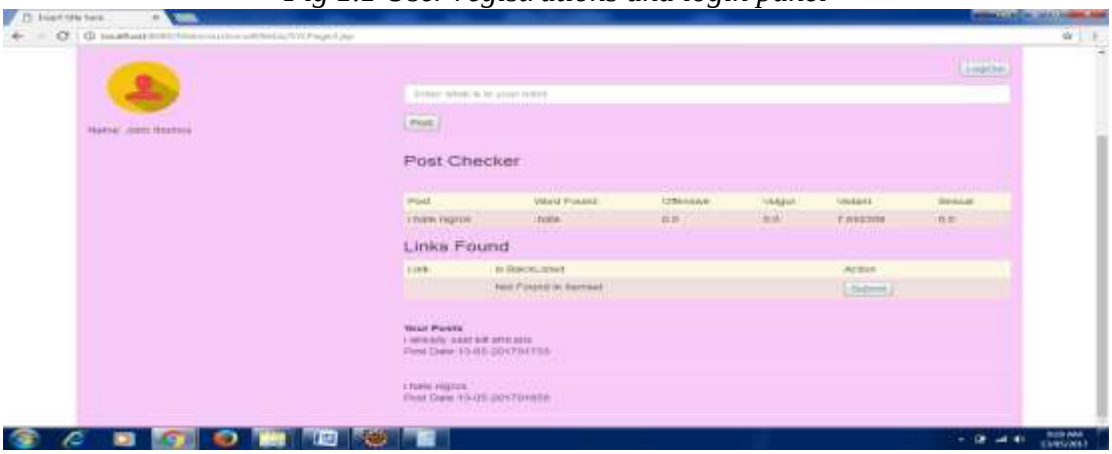

Fig 1.2 User timeline panel

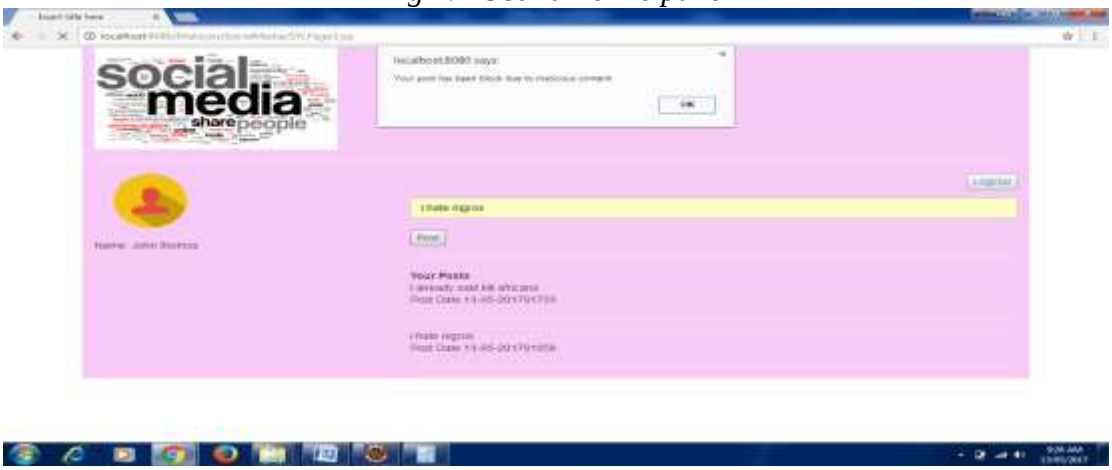

Fig 1.3 Blocking notification panel

Fig 1.4 Admin login panel

Fig 1.4 View all posts panel

Fig 1.5 View all users panel

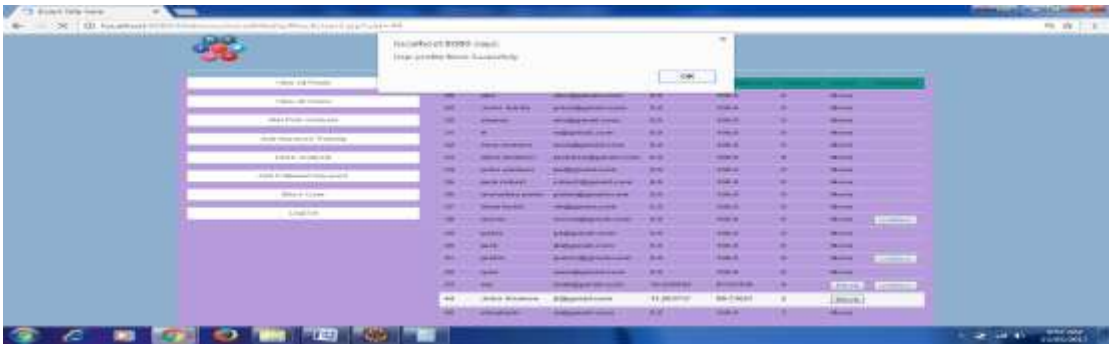

Fig 1.6Malicious analysis panel

Fig 1.7 Add keywords panel

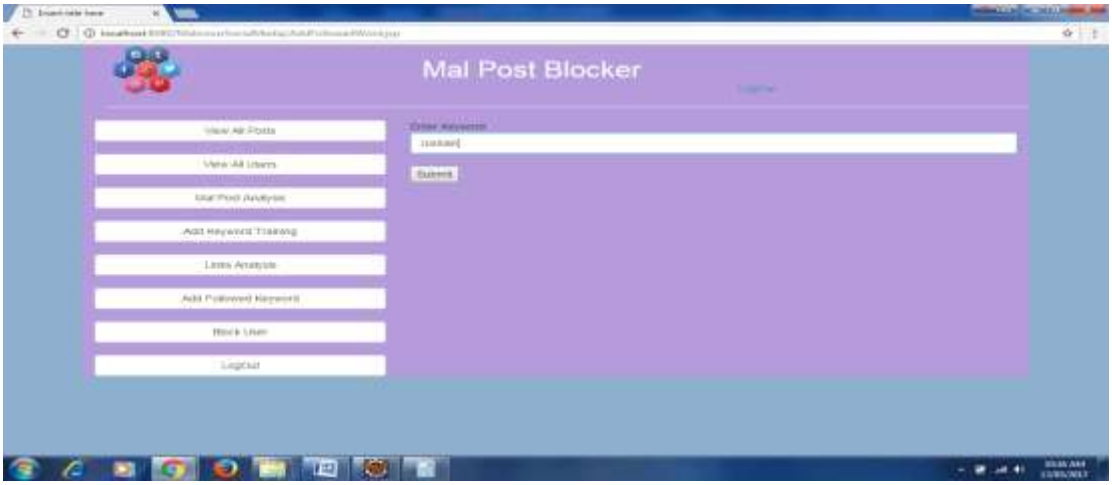

Fig 1.8 Add followed word panel

Fig 1.9 Block user list panel

Fig 1.10 Link analysis panel

VI. EXPERIMENTAL ANALYSIS

This chapter presents the results of a series of experiments designed to give insight into the performance, behavior, and scope of enzyme genetic programming. In particular, this chapter aims to document: The efficacy of functionality as an implementation of implicit context, the performance and behavior of recombination, the evolution of program size and structure, the role and evolution of redundancy, the evolution of compartmentalization, the complexity and consequences of development.

5.1 User Post Malicious percentage calculations

Here, we are going to studying and calculate the malicious and non-malicious percentage of the particular user's post of statements of keywords for study of results, here our system is analyzing the post which is posted by the user and calculating how much percentage it is malicious and non-malicious.

In table showing analysis of graph 5.1.1 and considered user posts P1, P2, P3, and P4. For every user of post we are calculating the malicious percentage with the help of keywords matches, here we get the results that are for user posts p1…p4, keywords are matched with the database blacklisted keywords and after applying the malicious formula we got results which is shows in graph.

Table 5.1.1 User post malicious percentage

User Post	Length of Statement	Keyword Matches	Malicious Percentage
P1	20	1	5
P2	18	3	16.66
P3	31	2	6.45
P4	15	1	6.66

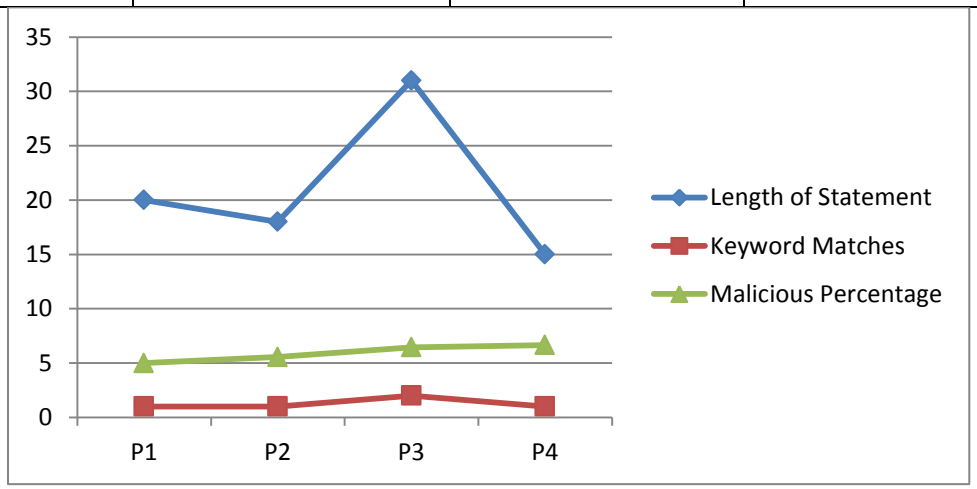

Figure 5.1.2 User post malicious percentage graph

5.2 Calculation of malicious link's percentage

Here we are going to studying and calculate the malicious and non-malicious percentage of the particular user's post of links for the study of the results, here our system is analyzing the post of link which is posted by the user how much percentage it is malicious and non-malicious. In table 7.2 showing analysis of graph 7.2 and considered user posts L1, L2, L3, L4. For every user of post we are calculating the malicious percentage with the help of keywords matches, here we get the results that is for user L1 56 keywords are matched with the database blacklisted keywords and after applying the malicious formula we got results which is shows in above table.

We consider for study blacklisted keywords in database is 120, on that basis we are calculating the malicious percentage of the link, here the blacklisted keywords in database is continuously updating so for study we are considering a fix value for database blacklisted keywords

that is 120. Out of 120 here for every users of malicious link will matches the pattern of keywords with database blacklisted keywords. Depending on the keywords matches the graph showing the results. No. of blacklisted keywords are matches more than malicious percentage of link also increase, so we can say that malicious percentage is directly depending on the blacklisted keywords of the database.

- Example of malicious link percentage analysis. Here consider no. of blacklisted keywords in database=120

Table 5.2.1 User link malicious percentage

User Link	Keyword Matches	Malicious Percentage
L1	56	46.66
L2	79	65.33
L3	67	53.33
L4	40	33.33

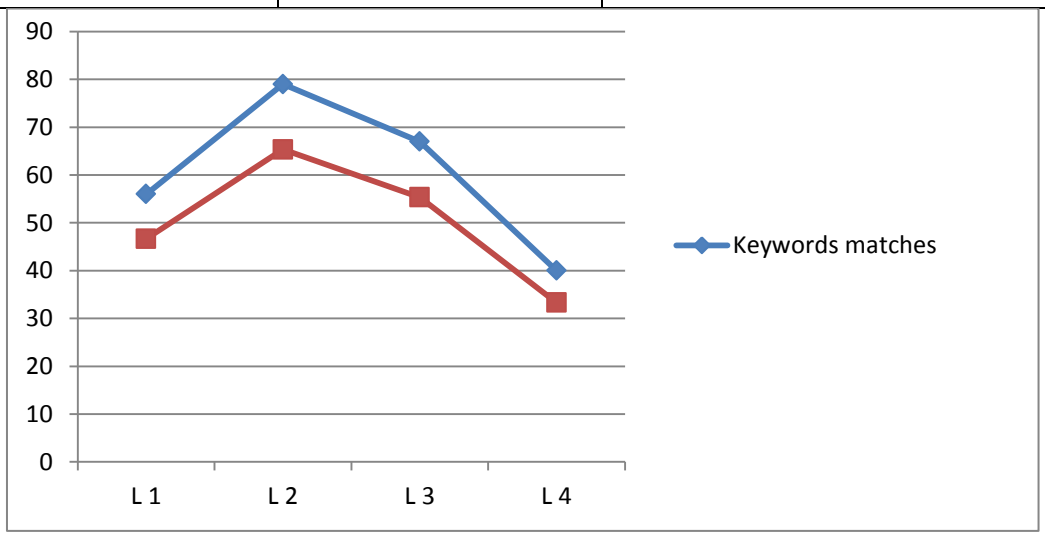

Figure 5.2.2 User link malicious percentage graph

The graph clearly shows that post of the user is malicious the red line in graph shows malicious percentage of post, and blue line shows that keywords matches with the database blacklisted keywords.

VII. COMPARATIVE ANALYSIS

Table 7.1 Comparative analysis

Sr. No	Method	Detection of malicious text posts	Detection of malicious urls	Blocking of malicious text posts/url posts
01	My PageKeeper	Yes	No	No
02	Frappe	No	No	No
03	FrappeLite	No	No	No
04	Proposed Approach	Yes	Yes	Yes

VIII. RESULT ANALYSIS

- Keyword Matching: In previous techniques for detection mechanism text processing is used, where only suspicious word get matched with predefined database. And after matching found

post get declared as malicious. But in our proposed work we are going to firstly compare the malicious word from the post and after getting match found, the suspicious or malicious word get checked with the followed word. If the meaningful statement of malicious activity found then the post will be blocked otherwise not.

- Malicious percentage calculation: In previous technique the post declared as malicious only on the basis of keywords matching. In our proposed work we are going to apply formula for calculating the percentage of malicious posts.
- Accuracy: Our proposed approach gives more accuracy because detection of malicious posts is not done on the basis of the keyword matching but by taking reference of followed keywords also which is not done in previously proposed work.
- Flexibility: Our proposed work gives more flexibility than other approaches. As we can modify the criteria of blocking the user just by increasing or decreasing limit of blocking percentage criteria, so the system is more flexible. In previous approaches no any blocking performed so there was no concern of flexibility.
- Accountability: The work and result achieved by the proposed work gives accuracy because of the system detection mechanism is not only depends on the keyword matching but also additional criteria get added here for detection and mechanism, so system is accountable

IX. CONCLUSION

It is the era where everyone gets addicted to use social media. Social media is the platform for sharing views of every individual. It's common nowadays most of the people post on social media by bad intention. So through our work we able to detect various post of social media and also blocking is possible. For future work, we plan to improve the system in term of execution time, developing automated classification and using other knowledge resources in order to improve the precision rates, the semantic of exchanged information will be used to identify more significant suspicious profiles. In this work, we are just going to work for detection and blocking of malicious posts and urls. In future by adding concept of OCR we can recognize malicious image also.

REFERENCES

[1] Bughin, M. Chui, and J. Manyika, "Clouds, Big Data, and Smart Assets: Ten Tech-Enabled Business Trends to Watch", McKinley Quarterly Conference, 2015.
[2] Miss. Sayali S. Karmode, Prof. V. B. Bhagat, "A review on Detection and Blocking Social Media Malicious Posts", IJMTER, 2016
[3] D.Centola, "The Spread of Behavior in an Online Social Network Experiment Science", in MARCTY Conference, 2015.
[4] C. Lin, B. Zhao, Q. Mei, and J. Han. Pet:"A statistical model for popular events tracking in social communities", In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data is mining ACM, 2014.
[5] T. Sakaki, M. Okazaki, and Y. Matsuo: "Realtime event detection by social sensors", In Proceedings of the 19th international conference on World wide web ACM, 2013.
[6] T. Sakaki, M. Okazaki, and Y. Matsuo: "Realtime event detection by social sensors", In Proceedings of the 19th international conference on World wide web ACM, 2013.
[7] T. Stein, E. Chen, and K. Mangla,"Facebook immune system "In Proceedings of the 4th Workshop on Social Network Systems", 2012.
[8] T. Stein, E. Chen, and K. Mangla,"Facebook immune system "In Proceedings of the 4th Workshop on Social Network Systems", 2011.
[9] L. Dari and F. Fusco, "Frappe for detecting malicious application", in 9th IEEE International Wireless Communications and Mobile Computing Conference (IWCMC), 2009.