Analysis and Detection of Suspicious Chats in Terrorism Using Word Substitution

Shital Londhe[1], Vishakha Devare[2], Asmita Shinde[3], Madhuri Gujarathi[4], Prof P.S.Tambe[5]
[1,2,3,4]Student of Computer Engineering,
[5]Assistant Prof. Computer Department, SVIT Chincholi
Email ID
londheshital98@gmail.com[1],vishakhadevare22@gmail.com[2],asmitashinde017@gmail.com[3],
madhuri9765@gmail.com[4],prachi_gaikwad28@rediffmail.com[5].

***Abstract*—** Terrorist attacks are the major problem for the society. To know their intentions early and avoid the attacks is very necessary. In our technique we are going to allow automatic detection of suspicious messages. We are developing a desktop application where the admin who is monitoring all chats can crack the chat between two person whose chat seems to be suspicious. By performing Natural Language Processing Algorithm on the chat admin will able to know the exact meaning of the conversation by substituting the irrelevant words by the point of view of attacker. For this we have used Hadoop, Map-reduce, Word substitution like techniques.

Keywords: - ***Counter terrorism ,Data Mining, Document Analysis ,Map Reduce , Sentence Substitution , Natural Language Processing***.

I. INTRODUCTION

In our system admin and user roles, Admin is able to view the live chat and monitor the same. User can chat with any other user which he cannot monitor. He is only able to send and receive the messages to and from the other user. While chatting if there is any occurrence of obfuscated word then the algorithm is being able to detect the same and is able convert the obfuscated word with a normal one. Database is being used for analysis of the same where all the chat will be stored in a text file, the same text file will be stored into database as a storage purpose. Regarding the various propose models this system is motivated to find the obfuscated chat in terrorism, this techniques focuses on collection, observation and preservation of suspected words and information occurs during conversation, so as to analyze the suspected words. Our paper explains the tools in the digital media (storage), digital forensic methods and cyber crime. This system proposes a new tool which uses the word Substitution technique. The proposed system is designed for finding the suspected words occurs in conversation and replaces these words with normal ones which are stored in our hadoop database and the message is received in the form of normal text.

II.LITURATURE SURVEY

Factored language models and generalized parallel back: The problem of detecting a word that is somehow out of context occurs in a number of settings. For example, speech recognition algorithms model the expected next word, and back up to a direct interpretation when the next word becomes suciently unlikely. These problem diers from the problem addressed here because of the strong left context that is used to decide on how unlikely the next word is, and the limited amount of resources that can be applied to detection because of the near real time performance requirement Application for counter terrorism using word substitution

Winnow-based approach to context-sensitive spelling correction: Detecting words out ofcontext can also be used to detect (and correct) misspellings. These problem dyers from the problem addressed here because the misspelled words are nonsense, and often nonsense predictably transformed from the correctly spelled word, for example by letter reversal. Using common sense knowledge base for detection word obfuscate in adversarial communication future information security (Swati Agrawal , Ashish Sureka) : use of concept net to compute conceptual similarity between two term solving problems of word conduct empirical analysis on large and real word data set Conflict occur in substitution of words due to the complex design.

Detecting threats of violence in online discussion using bi-diagrams of important words (Hugo Lewi Hammer): Use of Classification method within text mining classification performed using LASSO logistic regression .But the drawback of this system is rate of wrongly classification of word is about 20 percent so cannot used for real time substitution.

Detecting word substitution in text. IEEE Transaction on knowledge and data (SW. Fong, D.Roussinovei and D.B.Skillicorn): design the system of measures that applied to sentence, positive detection rate is 90 percent and false rate is around 10 percent. Difficult to understand which design is used and conflict data structure.Surveillance issue for security over computer communication and legal implication. proceeding (Mr Shilpa Mehta, Dr.U.Eranna, Dr.k.Soundatatajan)

III. PROPOSED SYSTEM

In our system both user as well as the admin logs into the system. The user can chat with the other users. The admin can monitor the chat between the users. The admin can crack the chat between them. In our proposed system we are going to use natural language processing, word substitution, map-reduce like techniques. The input to our system will be in the form of simple text message, or posts from the social Medias or emails. The admin will provide the input for word substitution. Our system uses hadoop database to store data. To extract meaningful word or substitute the relevant word for the suspictitious word from the hadoop database we are using map-reduce function. Our proposed system also inculcates some algorithms for word substitution. After performing all the operations the final result is provided to the admin in the form of normal text message again.

IV. SYSTEM ARCHITECTURE

The above architecture of the proposed system shows, there are two suspicious persons i.e., user A and user B. Both the users communicate with each other. The admin monitors the chat between the two suspicious persons , cracks the chat between them. The admin then provides the chat in the form of normal text message for natural language processing . after performing natural language processing on the data the relevant word for the suspicious word is substituted from the Hadoop database . Then the output is provided back to the admin in the form of normal text file.

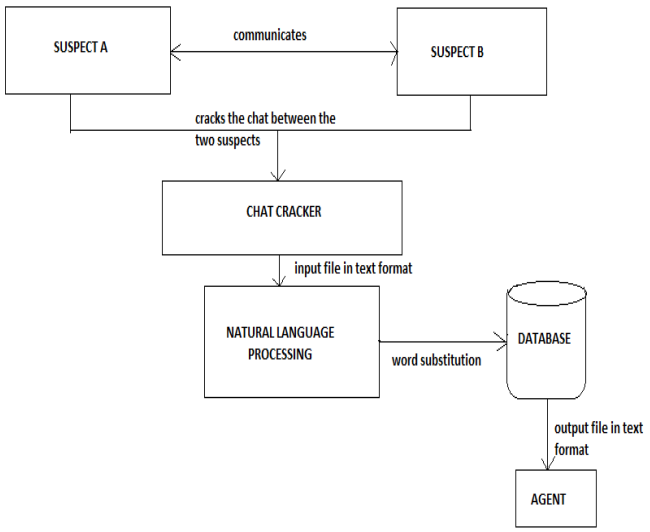

Fig 1.System Architecture

V. ALGORITHM AND MATHEMATICAL MODULE

Algorithm:

Mathematical Formulation:

S=I, F, O

Where,

I=Set of inputs

i1=Normal text messages

i2=Images.

i3=Audio conversation

F=Set of functions

f1=Hadoop

f2=Map reduce

f3=word substitution

O=Set of outputs

o1=Messages in normal form

o2=Identity of suspected person

o3=Location of suspected person

Success Condition=Gets the relevant information from the communication of suspected person.

Failure condition=Connection is loss system does not work, internet connection is compulsory.

Mathematical Model:

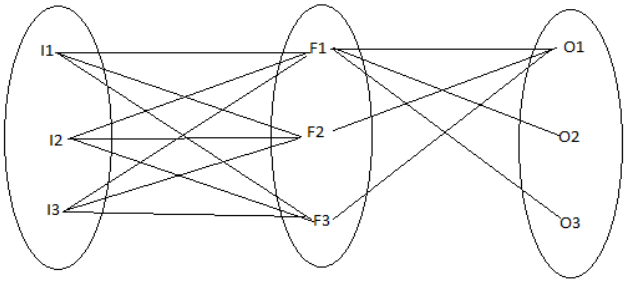

Fig 3.Mathematical module

USE CASE VIEW

To model a system the most important aspect is to capture the dynamic behavior rather than the static behavior. To clarify a bit in details, dynamic behavior means the behavior of the system when it is running operating .So only static behavior is not sufficient to model a system rather dynamic behaviorism more important than static behavior. In UML there are five diagrams available to model dynamic nature and use case diagram is one of them Now as we have to discuss that the use case diagram is dynamic in nature there should be some internal or external factors for making the interaction The use case is applicable to model the system or subsystem of an application. A single model captures a particular functionality or property of a system. And this use case shows the flow of our program. Initially when people communicates with each other their conversation is monitor by the admin, and if he finds any suspected word in conversation then admin will send it to system for further processing. For our system there are two actors: one is user and second is admin. The user is communicate with other person through calls, messages, online chats, or through any social media application .Admin view and monitor the chat and if he find any suspected words in their chatting then he browse the chat, analyze application. A single use case diagram captures a particular functionality of a system. It and send other uses cases for further operations.

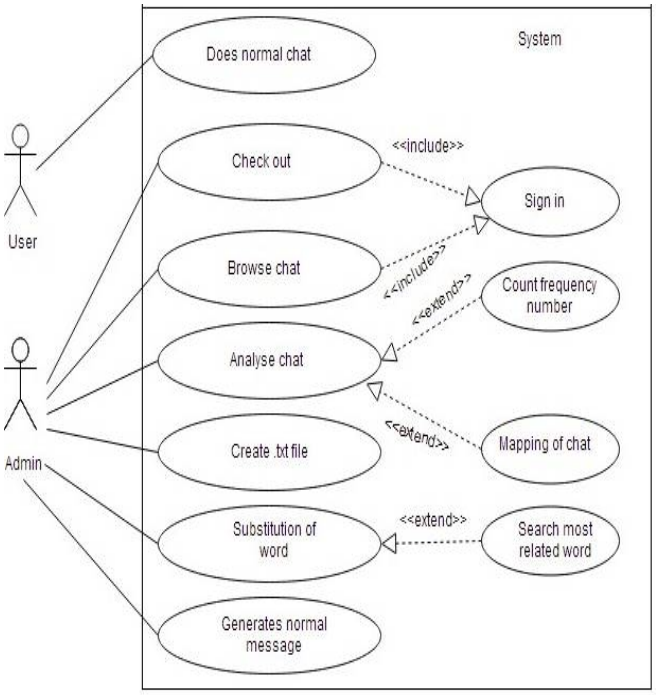

Fig 4.Usecase Diagram

VI.FUTURE SCOPE

In future, using this application we will able to give security alert to government agencies or secret agencies. By detecting the chat we will able to find the suspicious information or doubt full data. We can track the location of attacks and the main thing we can save the lives. In addition we can use this application in banks also to detect frauds.

REFERENCES

[1] SW. Fong, D. Roussinov, and D.B. Skillicorn. Detecting word substitutions in text. IEEE Transactions on Knowledge and Data Engineering, 20(8):10671076, 2008.

[2] J. Bartlett, J. Birdwell, and M. Littler, The rise of populism in Europe can be traced through online behaviour..., 21-January-2014.

[3].I. M. Sunde, Preventing radicalization and violent extremism on the Internet (Norwegian), The Norwegian Police University College 2013:1, 2013.

[4].D Mitri r Oussinov, S Zewang F Ong, D Avid S Killcorn, Detecting Word Substitution: PMI VS HMM, Sigir 2007, Amsterdam Proceedings.

[5]. Transcript: Senator Dick Durbin on Surveillance, Terrorism and the Constitution Newstin Oct 5 2009.

[6]. Arlen Specter, The Need to Roll Back Presidential Power Grabs The New York Review of Books, Volume 56, and Number 8 May 14, 2009.

[7]. S. Fong, D. Skillicorn, and D. Roussinov, Measures to Detect Word Substitution in Intercepted Communication, Proc. IEEE Intl Conf. Intelligence and Security Informatics (ISI 06), May 2006.

[8].Vaibhav Shinde, Komal Salunke, Punam Teli Deducing and conclusive analy-sis of declassified obfuscate chat in terrorisom using Word substitution, IERJ, March 2017.

[10].Swati Agarwal, Ashish Sureka using Commonsense Knowledge-base for detecting word obfuscation in adversarial communication. Future information security workshop, IEEE 2015.