An Approach to Detect Terror Related Activities on Net

[1]Sonali Vighne, [2]Priyanka Trimbake, [3]Anjali Musmade, [4]Ashwini Merukar, [5]Sandip Pandit

[1] Sonali Vighne, BE, Dept. of Computer Engineering, Amrutvahini COE, Sangamner
[2] Priyanka Trimbake, BE, Dept. of Computer Engineering, Amrutvahini COE, Sangamner
[3] Anjali Musmade, BE, Dept. of Computer Engineering, Amrutvahini COE, Sangamner
[4] Ashwini Merukar, BE, Dept. of Computer Engineering, Amrutvahini COE, Sangamner
[5] Sandip Pandit, Prof. Dept. of Computer Engineering, Amrutvahini COE, Sangamner

ABSTRACT

In present scenario use of internet is in boom. But every coin has two sides, likewise the use of internet is beneficial as well as harmful to human being. Recent days many terror attacks were there on net. Such terror related activities are hazardous for peoples, organization and countries. Terrorist are using internet to spread terror and form terrorist groups. By using internet they easily do the same. To exchange information Internet infrastructure is used by different Terrorist cells and they recruit new members and supporters. To handle such situation we propose a system called ATDS.

.

Keyword: - *Terrorist trend detection, data mining, user modeling, anomaly detection, log mining, vector generator, threshold, clustering. ATDS (Advanced Terror Detecting System)*

.

1. INTRODUCTION

Day to day the use of internet is going on increasing drastically, relatively the huge number of techniques are there emerging everyday on various dynamic platforms. Use of internet has increased but along with these destructive minded peoples are using internet for making harm to the society and peoples. To exchange information Internet infrastructure is used by Terrorist cells and they recruit new members and supporters [2]. For example, high-speed Internet connections were used intensively by members of the infamous 'Hamburg Cell' that was largely responsible for the preparation of the September 11 attacks against the United States. This is one reason for the major effort made by law enforcement agencies around the world in collecting the information from the Web about terror-related activities.

It is believed that the detection of terrorist's activities on the Web might prevent further terrorist attacks. One way to detect terrorist activity on the Web is to eavesdrop on all traffic of Web sites associated with terrorist groups, organizations in order to detect the accessing users based on their IP address. Unfortunately it is difficult to detect terrorist sites since they do not use fixed.

Many terrorists or terrorist groups are used to make use of such techniques, applications, and internet to spread the terror. They used to attract the youths to be involved in such activities. It is necessary to detect such attempts in order to prevent such hazardous things. Terrorists groups are using social sites. It is the big challenge to detect such hazardous terrorist attacks. It is similar to eavesdrop. Terrorists are using different IP addresses changing them frequently such a that it will become more hard to identify them. To detect them is so difficult.

Many researchers and scientists looked into the matter to solve such problem. National security board also implemented best techniques to be aware of the terror activities and media because those are mostly visited by peoples frequently and peoples are addicted for those. Terrorists found it easy to scourge people, they can easily spread the terror among an organism, Society and some particular country.

Security is main aim behind this to protect our country from such hazards. Cyber security is one of the streams concentrating on security over net and detecting the crime scenes. Here we are going to express our proposed technique for detecting terror related activities on net. How such activities are to be detected and notified so that one come to know about future threats. Data mining deals with it since long period to more strengthen the security essentials.

2. LITERATURE SURVEY

Web based Text data is the most common content type on the net when it comes to author's opinion. Recently, following the progress of wireless internet and smartphone devices, iPhones the amount of data on the web is dramatically increasing with no constrain to time or location. In this paper we suggested the method for extracting the words from document names as WordNet Hierarchy [1]. This method was tested with the sampled New York Times articles by querying four distinct words from four different areas. Experimental results show our proposed method effectively extracts context words from the text and identifies terrorism-related documents. Text analysis is used to discover unknown, valid patterns and relationships in large data sets. Even text analysis has a great potential to identifying unknown text documents, there is a limitation that human written language is still complicated for machine to understand semantic meanings of it [1].

The learning Typical–Terrorist-Behavior part of the methodology defines and represents the typical behavior of terrorist users based on the content of their Web activities. It is assumed that it is possible to collect Web pages from terror-related sites. The content of the collected pages is the input to the Vector Generator module that converts the pages into vectors of weighted terms (each page is converted to one vector). The vectors are stored for future processing in the Vector of Terrorists Transactions DB. The Clustering module accesses the collected vectors and performs unsupervised clustering resulting in n clusters representing the typical topics viewed by terrorist users [2].

One major issue of today is the representation of textual content of Web pages. More specifically, there is a need to represent the content of terror-related pages as against the content of a currently accessed page in order to efficiently compute the similarity between them. This study will use the vector-space model commonly used in Information Retrieval applications for representing terrorists' interests and each accessed Web page.

3. PROPOSED SYSTEM

Terrorist groups use the Web as their infrastructure for various purposes. One example is the forming of new local cells that may later become active and perform acts of terror. The Advanced Terrorist Detection System (ATDS), is aimed at tracking down online access to abnormal content, which may include terrorist-generated sites, by analyzing the content of information accessed by the Web users. ATDS operates in two modes: the training mode and the detection mode. In the training mode, ATDS determines the typical interests of a pre specified group of users by processing the Web pages accessed by these users over time.

In the detection mode, ATDS performs real-time monitoring of the Web traffic generated by the monitored group, analyzes the content of the accessed Web pages, and issues an alarm if the accessed information is not within the typical interests of that group and similar to the terrorist interests. An experimental version of ATDS was implemented and evaluated in a local network environment. The results suggest that when optimally tuned the system can reach high detection rates of up to 100% in case of continuous access to a series of terrorist Web pages.

Training Mode: It is first module of project where we will design terrorist transaction database acknowledge their behavior from their web activities. As our database is prepared we will connect with our next module.

Detection Mode: In detection mode Terrorist behavior is given to detector module as reference data library. In this mode we calculate one threshold value.and content based detection. If we find such activity on web our system will make alarming reporting.

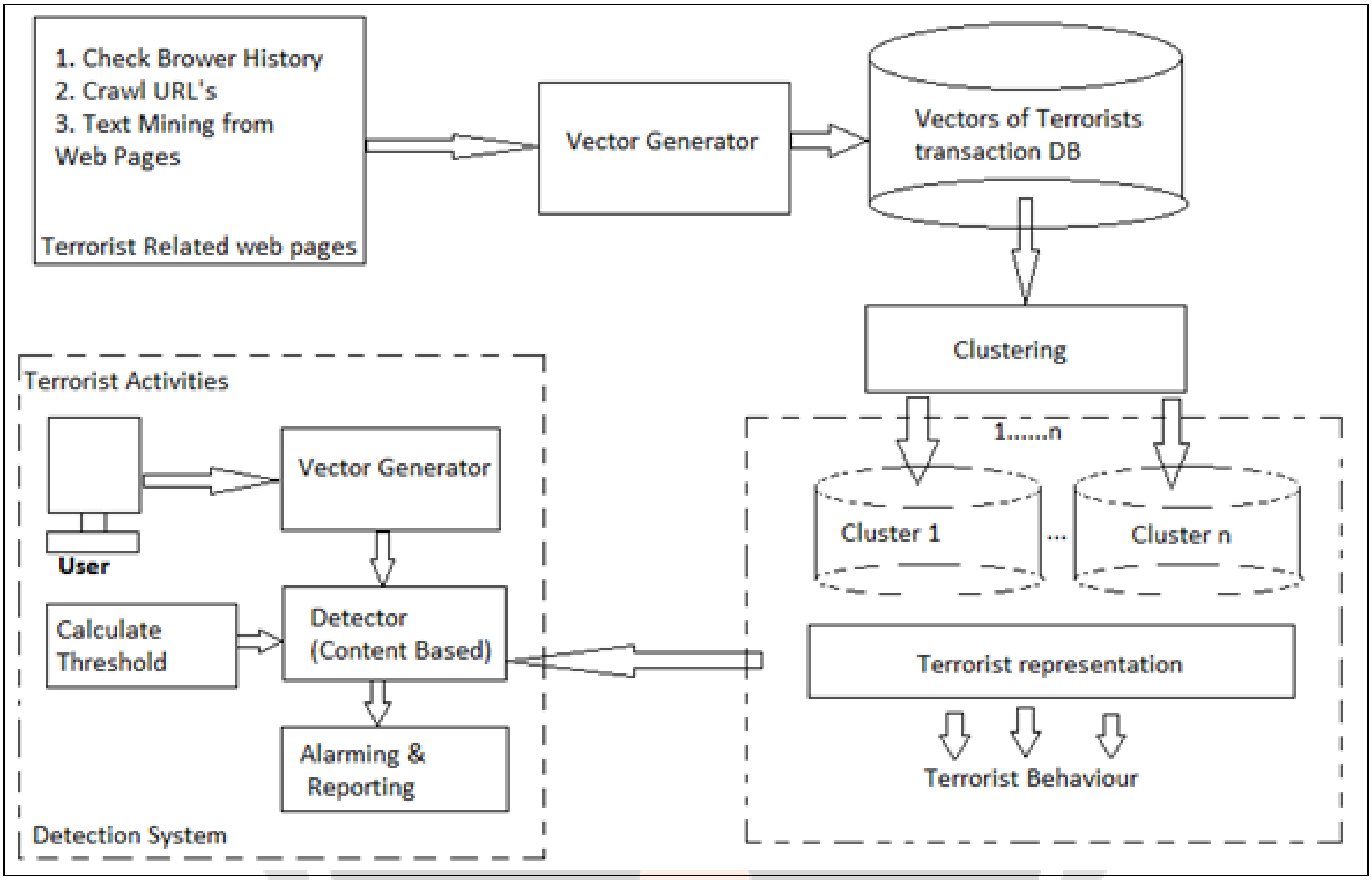

Fig 1. Proposed System Architecture diagram

1. Terrorist related web pages:

Terrorists are using web/net to communicate with each other and also to spread the terror. We can recognize that usage of net by using data mining techniques. We are extracting the log files over the net and check against some standard data base, to decide whether those URL's are suspicious or not [3]. Following are ways-

(a) Check Browser History
(b) Crawl URL's
(c) Text Mining from Web Pages

2. Vector Generator:

There is a need to represent the content of terror-related pages as against the content of a currently accessed page in order to efficiently compute the similarity between them. This study will use the vector-space model commonly used in Information Retrieval applications for representing terrorist's interests and each accessed Web page. In the vector-space model, a document d is represented by an n-dimensional vector d = (w1,w2,..,wn), where wi represents the frequency-based weight of term i in document d [5]. The similarity between two documents represented as vectors may be computed by using one of the known vector distance measuring methods such as Euclidian distance or Cosine [2].

3. Transaction Database:

We are storing vectors generated by Vector Generator in database called transaction database. This database is further used for the clustering.

4. Clustering:

Cluster analysis is the process of partitioning data objects (records, documents, etc.) into meaningful groups or clusters so that objects within a cluster have similar characteristics but are dissimilar to objects in other clusters. Clustering can be viewed as unsupervised classification of un-labelled patterns (observations, data items or feature vectors), since no pre-defined category labels are associated with the objects in the training set. Clustering results in a compact representation of large data sets (e.g., collections of visited Web pages) by a small number of cluster centroids.

Applications of clustering include data mining, document retrieval, image segmentation, and pattern classification. Thus, clustering of Web documents viewed by Internet users can reveal collections of documents belonging to the same topic. Clustering can also be used for anomaly detection [4].

5. Terrorist representation:

After applying clustering algorithm from transaction database n clusters are generated this is used to represent terrorist behavior which is further provided as input to the detector.

6. Detection system:

It is a new type of knowledge-based detection methodology that uses the content of Web pages browsed by terrorists and their supporters as an input to a detection process. In this study, refers only to the textual content of Web pages, excluding images, music, video clips, and other complex data types. It is assumed that terror-related content usually viewed by terrorists and their supporters can be used as training data for a learning process to obtain a Typical-Terrorist-Behavior. This typical behavior will be used to detect further terrorists and their supporters. A Terrorist-Typical-Behavior is defined as an access to information relevant to terrorists and their supporters. Following are components of detection system,

(a) User
(b) Vector Generator
(c) Detector(Content Based)
(d)Alarming and Reporting

4. PROCEDURE

Algorithm used

WORD EXTRACTION ALGORITHM

The word extraction algorithm is given below.

```
Input: Web URL's i.e. log file
Output: Extracted meaningful word
Begin
Step 1: Take the browser history
Step 2: Using Vector extract the word from that document
Step 3: Match that word with standard database
Step 4: Use clustering algorithm to do the cluster of similar words
Step 5: If match found and shows terror behavior then goto step 6
        Else goto step 7
Step 6: Report the problem using alarm or beep sound
Step 7: Close the connection
Exit
```

The Clustering module accesses the collected vectors and performs unsupervised clustering resulting in n clusters representing the typical topics viewed by terrorist users. For each cluster, the Terrorist-Represent or module computes the centroid vector (denoted by Cvi) which represents a topic typically accessed by terrorists. As a result, a set of centroid vectors represent a set of terrorists interests referred to as the 'Typical-Terrorist-Behavior.

Text analysis is used to discover unknown, valid patterns and relationships in large data sets. Even text analysis has a great potential to identifying unknown text documents, there is a limitation that human written language is still complicated for machine to understand semantic meanings of it. Over the years, many studies have been made by using statistical methods to represent documents into meaningful sequences, such as TF-iDF. This is the most basic method for determining which words are significant in the given text data set. However, performance state based on TF-iDF is not acceptable if the amount of data is too small. Moreover, this approach depends on the bag of words to calculate TF-iDF value.

$$\text{TF (t)} = \left(\frac{n}{N}\right)$$

$$\text{IDF (t)} = \log_e\left(\frac{T}{t}\right)$$

n = Number of times term t appear in a document
N = Total number of terms in document
T = Total number of document
t = Number of documents with term t in it

The Typical-Terrorist-Behavior is based on a set of Web pages that were downloaded from terrorist related sites and is the main input of the detection algorithm. In order to make the detection algorithm more accurate, the process of generating the Typical-Terrorist-Behavior has to be repeated periodically due to changes in the content of terrorist related site. Typical-Terrorist-Behavior depends on the number of clusters. When the number of clusters is higher, the Typical-Terrorist-Behavior includes more topics of interest by terrorists where each topic is based on fewer pages. It is hard to hypothesize what the optimal number of clusters is. In the case study presented in the next section detection performance for two settings of the number of clusters are presented. The detector issues an alarm when the similarity between the access vector and the nearest centroid is higher than the predefined threshold denoted by tr:

$$Max\left(\frac{\sum_{i=1}^{m}(tCv_{i1}.tAv_i)}{\sqrt{\sum_{i=1}^{m} tCv_{i1}^2.\sum_{i=1}^{m} tA_i^2}}, \ldots, \frac{\sum_{i=1}^{m}(tCv_{in}.tAv_i)}{\sqrt{\sum_{i=1}^{m} tCv_{in}^2.\sum_{i=1}^{m} tA_i^2}}\right) > tr$$

where ${}_iCv$ is the ith centroid vector, Av - the access vector, ${}_{i1}tCv$ - the i_{th} term in the vector ${}_iCv$, ${}_itAv$ -the i_{th} term in the vector Av , and m - the number of unique terms in each vector.

ADVANTAGES

i. It provides better marketing intelligence.
ii. Web logs or browser history provide an exciting new way of collecting information on visitors.
iii. We can easily differentiate the normal user and user which uses net for harming the society.

5. RESULTS

Overall an innovative knowledge-based methodology for terrorist detection by using Web traffic content and using data mining techniques as the audit information is presented here. The proposed methodology called ATDS, learns the typical behavior (profile) of terrorists by applying a data mining algorithm to the textual content of terror-related Web sites. The resulting profile consisting of logs is used by the system to perform real-time detection of users suspected of being engaged in terrorist activities. The Receiver-Operator Characteristic (ROC) analysis shows that this methodology can outperform a command based intrusion detection system.

In this way terror related activities can be detected using data mining techniques and log access and processing.

6. CONCLUSIONS

As per the goal of this project an innovative, knowledge-based methodology for terrorist activity detection on the Web is presented. The results of an initial case study suggest that the methodology can be useful for detecting terrorists and their supporters using a legitimate ways of Internet access to view terror related content at a series of evasive web sites. Present system just recognize the the words of terrorist's language. By developing such system, relationship between human and computer becomes much closer and secure. Thus it helps in overcoming the problem of Terrorism on net.

The proposed approach is efficient one to detect terror related activities.

7. REFERENCES

[1] Dongjin Choi , Byeongkyu Ko, Heesun Kim, Pankoo Kim, "Text analysis for detecting terrorism-related articles on the web", *Journal of Network and Computer Applications* 38 (2014) 1621.

[2] Mohammad Javad Hosseinpour,Mohammad Nabi Omidvar, "Detecting Terror-Related Activities on the Web with Using Data Mining Techniques", 2009 *Second International Conference on Computer and Electrical Engineering.*

[3] Ramesh Yevale, Mayuri Dhage, Tejali Nalawade,.Trupti Kaule, "Unauthorized Terror Attack Tracking Using Web Usage Mining", *(IJCSIT) International Journal of Computer Science and Information Technologies*,ISSN: 0975-9646, Vol. 5 (2) , 2014, 1210-1212.

[4] Y.Elovici, A.Kandel, M.Last, B.Shapira, O. Zaafrany, "Using Data Mining Techniques
for Detecting Terror-Related Activities on the Web" University of South Florida,4202 E. Fowler Ave. ENB 118 Tampa, FL, 33620, USA.

[5] Nisha Chaurasia1, Mradul Dhakar1, Akhilesh Tiwari2 and R. K. Gupta2, "A Survey on Terrorist Network Mining: Current Trends and Opportunities", *International Journal of Computer Science Engineering Survey (IJCSES)* Vol.3, No.4, August 2012.