Terror Activities Detection by Using Data Mining on Social Networking Sites

[1] SHADAB ADAM PATTEKARI, [2] DR. RAVINDRA NATH KATIYAR
[1][2] Department of Computer Science and Engineering
University Institute of Engineering & Technology, CSJMU, Kanpur

Abstract—*This paper represents application of data mining techniques to analyze fraud. The impact of the fraud on organizations is becoming increasing costly. The problem of fraud detection is concern with not only capturing the fraudulent activities, but also capturing them as quickly as possible. Now a days there are many terror attacks happened by using social network, such a terrorist attacks are hazardous for peoples, organizations and countries. Terrorist are using internet to spread terror and form terrorist group .This paper is to represents the data mining methods to detection of terror activities on social networking sites.*

Index Terms—. ***Intrusion Detection System (IDS), Vector Space Model, Data mining, Clustering, ATDS (Advance terror detection system)***

I. INTRODUCTION

In our day to day life use of social networking sites increasing very vastly. There are lots of several terrorist and terrorist groups are used to make use of such technology, websites, social networking sites, methods to spread the terror in all over the world. By using data mining fraud they used to attract the young generation to be involved in such activities. Terrorism is most hazardous things in our nation. By attacking on our computers as well as networks, databases and the Internet could be harmful to businesses and many organizations. It is projected that cyber-terrorism could cause loss of billions of dollars to businesses. Consider a banking information system. If terrorists attacks on such a system and make use of accounts of the funds, then that bank could lose their millions and perhaps billions of dollars.

The detection of a terrorist's activities on the networking sites may be prevent next terrorist attacks, for that purpose there are some new techniques are invented, By eavesdropping all the traffic of networking sites associated with terrorist and their organizations in order to detect access of user based on their IP address we can detect terrorist activities, How such activities are to be detected and notified so that one come to know about future threats, such a process is named as intrusion detection system. Data mining deals with it from long period to provide more the security essentials. Data mining is mostly used technique in the fraud detection.

Data mining is a process of analysis of data and sort out and gather them into the useful and meaningful data.

This paper is organized as follows. The first section is of a brief review of network analytics in fraud detection ,in the second section the intrusion detection systems is get explained in detail, then the third section is about data mining and clustering analysis.

II. LITERATURE REVIEW

Text data on the web is the best 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 increasing with no limit to time or location. This method of learning Typical–Terrorist-Behavior is represents the typical behavior of terrorist users based on the content of their web activities. It is believed that it is possible to collect web pages from terror-related sites, and it is possible to use them for their inhuman actions. The content of the collected pages is the input to the Vector Generator module that converts the pages in to the vectors. These vectors are stored in the use of future processing in the vector of terrorists transactions data base. The web pages are clustered by using unsupervised of clustering technique [2]. Clusters serve as data indicating the typical terrorist behavior or the profile of the terrorist or their supporters.

One major issue of today is the representation of textual content of Web pages. Specifically, there is a necessity to represent the data of terror-related pages as against the content of a currently accessed page in order to compute the similarity between them.

2.1 SOCIAL NETWORK ANALYSIS:

Social Network Analysis (SNA) is one of the most used technologies for studying criminal and terrorist networks. It is the one of the data mining methods in fraud analytics, is a technique which represents the entities as nodes and the relationships between the entities as links. The SNA technique represents the role between the actors within the social networks. In the fraud detection, the interaction and exchanges can be viewed as heterogeneous networks with multiple participants. The numbers of participants are generally huge, but the kind of interaction among the individuals is generally in limits only and known. Graph analysis techniques can used further to identify suspicious individuals, groups, relationships, unusual changes over time/geography, and anomalous networks within the overall graph structure.

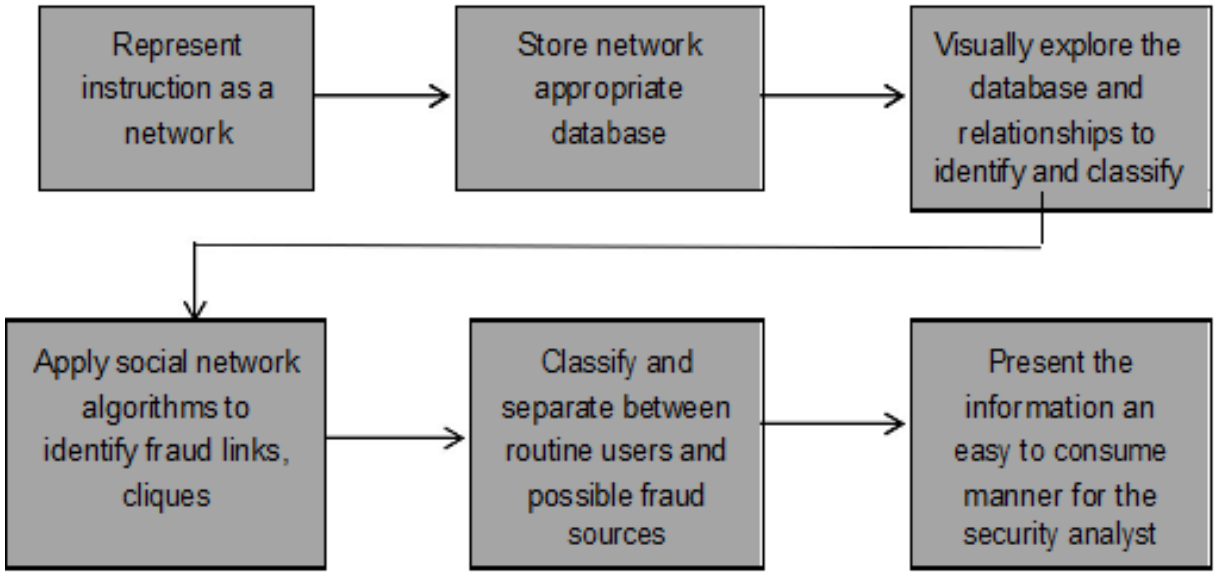

Figure 1: End to end fraud analytics approach using social network analysis methods

Social Network Analysis is having an ability to detect subgroups and discovering their patterns of the interaction, and identifying central individuals.

2.1.1 Subgroup Detection

Social Network Analysis uses the cluster analysis to partition the network into subgroups of individuals who interact with each other called clusters. they are not noticeable in data.

BACKGROUND OF THE RESEARCH:

This research integrates issues from the research fields of computer security (Intrusion Detection Systems), data mining (cluster analysis), Learning Typical Terrorist Behavior. The following subsections include a brief overview of these methods and their relation to the newly proposed methodology.

III. TERRORIST DETECTION METHODOLOGY

3.1 INTRUSION DETECTION SYSTEM:

The detection of the contents from the existing sites and the known terrorist traffic on the web is done by using intrusion detection system. An Intrusion Detection System is a network security technology to build for detection of frauds and threats. . An Intrusion Detection System (IDS) monitors all the activities in a definite environment and decides whether they are part of a possible hostile attack or a legal use of the environment. The environment may be several computers connected in a network or the network itself it may be social networking sites.

The IDS can investigate various types of information about actions from the environment and evaluates the probability that they are symptoms of intrusions. Such information includes, for example, configuration information about the current state of the system, audit information describing the events that occur in the system for e. g event log in Windows XP, or network traffic.

There are several measures for evaluating IDS have been suggested. In this it includes accuracy, performance, and efficiency and fault tolerance. The more common used measures are the True Positive rate it is nothing but the percentage of intrusive actions for an example terror related pages which is detected by the system, and False Positive rate which is the percentage of normal actions (e.g. Web pages viewed by normal users) the system incorrectly identifies as intrusive, and the real time accuracy which is the percentage of alarms found to represent abnormal behavior out of the total number of alarms. In this research True Positive rate, False Positive rate and Accuracy measures were adopted to evaluate the performance of the new methods.

3.2 Vector-Space Model:

The data which is evaluated by the IDS is represented in the textual content of Web pages. Consider the document D is represented by the n-dimensional vector V= (v1, v2, v3…vn) where vi represents the frequency-based weight of term i in D. The

Euclidian distance or cosine method these two vector distance measuring, methods are used to compute similarity between two documents which is represented as vectors. The cosine similarity measure is commonly used to estimate the similarity between an accessed Web page and a given set of terrorists' topic of interest [6].

3.3 Clustering technique:

Clustering is a technique to partitioning data objects into meaning groups and clusters, so that the data objects have the similar properties which are dissimilar than the objects or the data of other cluster .The unsupervised of clustering is performed for classification of patterns. The clustering is performed on the web content by clustering them into content of similar interest.

Cluster applications include data mining, image segmentation, data retrieval, and pattern classification (Jain et al. 1999). So that, clustering of web a document which gets viewed by Internet users can display the collections of documents within the same topic. Clustering is also used for anomaly detection, fraud detection, and threat detection.

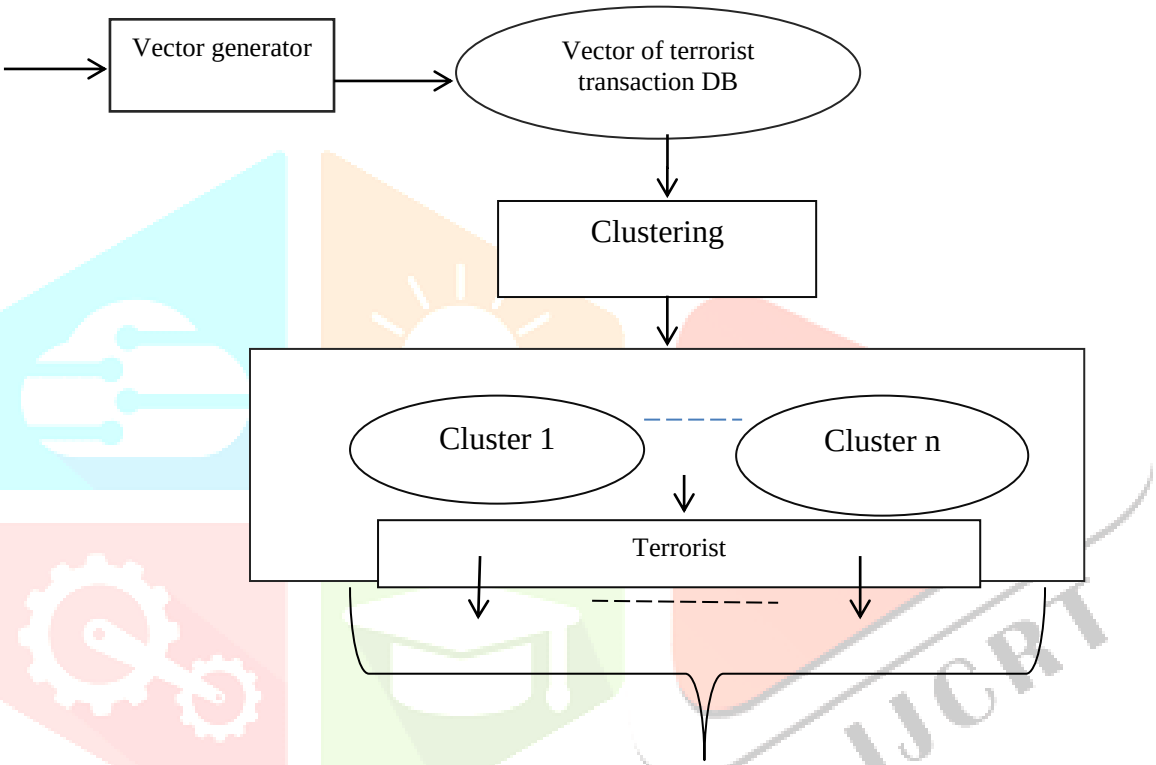

Figure 2: Learning the typical Terrorist Behaviour

IV DETECTION OF TERROR-RELATED ACTIVITIES

4.1. LEARNING TYPICAL TERRORIST BEHAVIOR:

In this the web pages are downloaded from terrorist sites. The web pages are clustered by using unsupervised of clustering technique. Clusters serve as data indicating the typical terrorist behavior or the profile of the terrorist or their supporters. The unsupervised clustering are performed on these vectors. A centroid vector Ci is computed for each cluster by the Terrorist Represent or module and thus representing typical terrorist behavior. The web pages downloaded are fed as the input to the vector generator module as shown in the fig.1

V PROPOSED SYSTEM

Terrorist organization uses web technologies for their inhuman purposes. For e.g. by forming of new local cellphones that may later become active and perform acts of terror. ATDS is formed to track down online access to anomalous content, which includes terrorist sites, by studying the content of information accessed by the web users.

ATDS operates in two modes: First is the training mode and second is the detection mode. In the first mode, ATDS verifies the typical interests of a pre specified group of users by

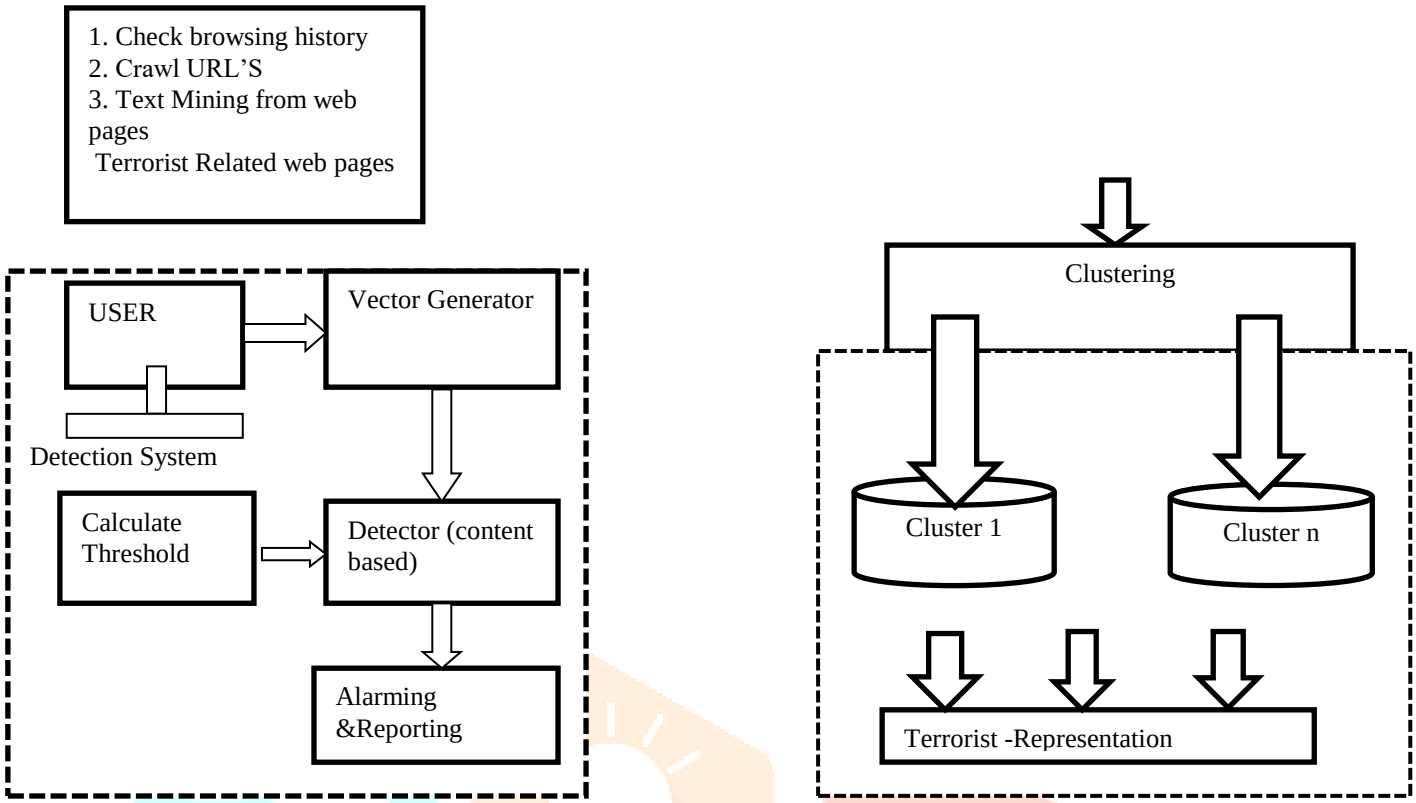

Figure 3. Proposed System Architecture diagram

Processing the web pages accesed by these users over time.

In the second mode, ATDS performs real –time monitoring of the Web traffic generated by the monitored group, analyzes the content of the accesed Web pages, and issues an alarrm 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 this mode we calculate one threshold value, and data based detection. If we find such activity on web our system will make alarming reporting.

VI Future Opportunities

It is analyzed that the use of Fuzzy Logic makes the detection process similar to the real world, defining the candidate set on the basis of uncertainty in support and confidence framework [7] while the Genetic Algorithm optimizes the detection process making the intrusion detection effective and optimizing the membership function. Further, these membership functions and patterns are stored for future use [8]

VII Conclusion

In this paper not only studies about what is the data mining is, But also explores the major developments to detect the terrorist networks. The purpose of this detection process is to powerfully detect and to stop the terrorist activities. In present paper SNA introduces the detection process through network analysis in the form of a graph and cluster analysis for subgroup detection; and also the detection using IDS, which monitors all the activities of terrorist and terrorist organization by eavesdrops traffic on net. After that the learning Typical Terrorist behavior model which monitors all the terrorist behavior by using unsupervised clustering technique. In this study case our main purpose was to consider the most relevant problems in social network analysis from fraud detection point of view and to restrict the terrorist activities.

By using this new techniques we can stop terrorism and to stop terrorist to fulfill their inhuman goals.

VIII REFERENCES

1. Sequeira, K., Zaki, M. (2002) ADMIT: Anomaly-based Data Mining for Intrusions, Proceedings of SIGKDD 02, pp. 386-395, ACM.

2. Mohammad Javad Hosseinpour, Mohammad Nabi Omidvar, "Detecting Terror-Related Activities on the Web with Using Data Mining Techniques", 2009 Second International

3. Jain, A.K., Murty, M.N., Flynn, P.J. (1999) Data Clustering: A Review, ACM Computing Surveys, 31, 3:264-323.

4. Debar, H., Dacier, H., Dacier, M., Wespi, A. (1999) Towards a taxonomy of intrusion-detection systems, Computer Networks, 31, pp. 805–822.

5. Kelley, J. (2002) Terror Groups behind Web, encryption, USA Today, URL:http://www.apfn.org/apfn/WTC_why.htm

6. Hosseinpour, M.J.; Omidvar, M.N., (2009) "Detecting Terror Related Activities on the Web with Using Data Mining Techniques", Proceedings of the Second International Conference on Computer and Electrical Engineering, 2009(ICCEE '09), Vol 2, 152-157.

7. German Florez, Susan M. Bridges, and Rayford B. Vaughn (2002) "An Improved Algorithm for Fuzzy Data Mining for Intrusion Detection", Proceedings of NAFIS2002 Annual Meeting of the North America, 457-462.

8. Shingo Mabu, Member, Ci Chen, Nannan Lu, Kaoru Shimada, and Kotaro Hirasawa,(2011) "An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming", Proceedings of the IEEE Transactions on Systems, Man, and Cybernetics—Part c: Applications and Reviews, Vol. 41, No. 1, 132-139.