E-TGPS: Enhanced Terrorist Group Prediction System for Counter Terrorism

Abhishek Sachan
Ericsson India Global Services Pvt. Ltd. India

ABSTRACT
Terrorist group prediction using historical data of terrorist incidents has been less explored due to the lack of meticulous terrorist data which contains terrorist group's attacks and activities information. There are many reasons for less exploration like its confidentiality & sensitivity. This paper presents an enhanced system that helps to predict the terrorist groups involved in the attack under investigation named E-TGPS. This system initially learns similarities of terrorist activities from various past terrorist incidents to predict the responsible group. This system can be considered as a vital tool for security agencies and intelligence analysts, by providing more reliable and predictive solutions to take effective counter-terrorism measures. The system has been validated by experimental results. The overall performance of the system displays a fair degree of accuracy. This paper also lays emphases on the meticulous analysis of optimal parameters weight estimation, to improve the predictive accuracy of the system.

Keywords
Terrorist groups; pattern matching; prediction; terrorism; counter-terrorism; group detection; privacy; security.

1. INTRODUCTION
Today, the biggest problem of mankind is terrorism. The whole world is under constant threat from these well-planned, sophisticated and coordinated terrorist operations. Now every country is focusing on counter-terrorism. Counter-terrorism (CT) is the strategies, practices, tactics, and techniques that governments, militaries, police and security agencies use to counter terrorist threats. Some agencies have tried to solve this problem by using various advanced and emerging information techniques/technologies. There is a massive amount of data that has been collected by government and intelligence agencies. The information is sparse, distributed and difficult to analyse. Intelligence agencies have huge amount of data that is scattered in different places. They are continuously monitoring terrorist activities by analysing these data sources. But, the problem is shortage of trained officers to process the huge amount data, in that limited time span they have, for decision making about terrorist activities and attacks thus manual processing of bulk amount of data is not efficient.

The data that has been collected by agencies is huge to analyse, hence decision making from this data requires the automated approaches to solve the problem of terrorism. The problem of terrorism can be solved by two approaches namely- proactive and reactive approach. Both the approaches have their own benefits. In the proactive approach security agencies tries to stop terrorist activities before the incidents. This includes capturing, killing, or disabling suspected terrorists before they can mount an incident. In the reactive approach security agencies try to catch the culprits after the incident to mitigate the effects of the terrorist incident. In both the approaches privacy of the people should be preserve because it cannot be fully compromised for national security as security and privacy are inversely proportional to each other [1]. In counter terrorism the first step after any incident is to find the terrorist group name that might be involved in that particular incident and to make a planned strategy to catch them. This paper discusses a system that is based on the reactive approach of CT for the identification of responsible terrorist group for a particular incident based on the available historical information. Advanced and emerging information techniques and technologies play a major role for the development of automated analysis and decision making systems.

Automated approaches of CT are becoming more popular. Various research projects are still going on to enhance the automated systems. Automated approaches increase the efficiency and accuracy of the system that has been developed for CT, by reducing manual efforts many CT mechanisms are derived and developed for the sake of this purpose. Increasing the CT techniques is a challenging requirement. The goal of each mechanism is to develop an efficient CT system by the use of computational intelligence to save lives. The system that has been discussed in this paper aims to achieve the goal of creating automated system for counter terrorism that process the bulk data efficiently and help to make decisions rapidly. From analysis of experimental results it has been observed that automated approaches towards CT can achieve the goal of CT successfully. Security is an important aspect that has been given top priority by all political, government and intelligence agencies worldwide. They are aiming to reduce crime incidence [2, 3]. Intelligence analysis might be applied to any of the intelligence sources like Open Source Intelligence (OSINT), Signals Intelligence (SIGINT) and Imagery Intelligence (IMINT) based upon requirement [4].

Nowadays most researchers focus on using social network analysis (SNA) for structural and positional analysis of terrorist networks [5, 6, 7, 8]. In SNA required information is provided from non-crime data. Many researchers are working on Dark Web Analysis for the same purpose as these dark webs are hot spot for terrorist groups to exchange their ideas, spread propaganda, recruitment of new members, and plan attacks [9]. The prediction of terrorist group using historical data of attacks has very less explored; this is due to the lack of meticulous terrorist data which contains terrorist group's attacks and activities information [10]. The use of data mining technologies in counter terrorism has been flourishing since the U.S. Government encouraged the use of information technologies in security [11]. This paper is about a prediction system that has been developed as an enhancement of our previous terrorist group prediction model (TGPM), using historical data to predict the terrorist group involved in a given incident. The database includes terrorist attacks in India from year 1998-2008.

2. PREDICTING TERRORIST GROUP
In counter terrorism, prediction of terrorist group after an attack is the most important step. As quickly as the name of the involved group is sorted, the strategies to catch those culprits could be made. The general process followed to detect the terrorist group responsible is by using email, telephone signal information, terrorist web sites and social network analysis [8, 10]. Terrorist activities occurred in the past are available in

many criminal/extremist/terrorist databases [12, 13, 14]. These databases can be used to detect terrorist group responsible for an attack. Terrorist group prediction model (TGPM) has been developed which learns the pattern of terrorist attacks from the available historical data and make an association between terrorist group and previous incidents [15]. Every terrorist group can be differentiated based on their style of attack, common targets like police, private organizations, public property etc.; so by analysing these patterns TGPM predicts the group that may be involved in a given incident under investigation.

3. TERRORIST GROUP PREDICTION MODEL (TGPM)

TGPM [15] has been developed to detect the responsible terrorist group by using historical data. TGPM uses the concept of Crime Prediction Model [2, 10], Group Detection Model (GDM) [16] and Offender Group Detection Model (OGDM) [16, 17]. The predictive accuracy of the system is 80.41%. This model works well for big groups, but it fails to identify small groups or sub groups which are working with a big group (because subgroup's parameters will be similar to that of big group and hence they cannot be differentiated). The main limitation of TGPM is that it is unable to identify the terrorist groups which are involved in fewer incidents.

4. ENHANCED TERRORIST GROUP PREDICTION SYSTEM (E-TGPS)

E-TGPS has been developed as an enhancement of TGPM to detect the responsible terrorist group by using historical data. This is the improvement and refinement of our terrorist group prediction model (TGPM) [15]. The goal of this system is to help in the analysis, anticipation, and countering of terrorist/extremist threats. E-TGPS narrows down the shadow effect problem of TGPM. Shadow effect is the effect in which big group's effects are overlapping to the small group's effect due to their more involvement in the incidents/attacks. Due to shadow effect identification of small group is difficult. E-TGPS uses various parameters like attack type, location, target type, weapon type, hostage/kidnapping and suicide attack etc. E-TGPS uses terrorist corpus, parameter's value and parameters weight as input to the system. E-TGPS narrows down the shadow effect by making two changes in TGPM. Firstly, instead of directly using the percentage of attack of each group as group percentage use the natural logarithm of parentage of attack of each group. Secondly, change the association function.

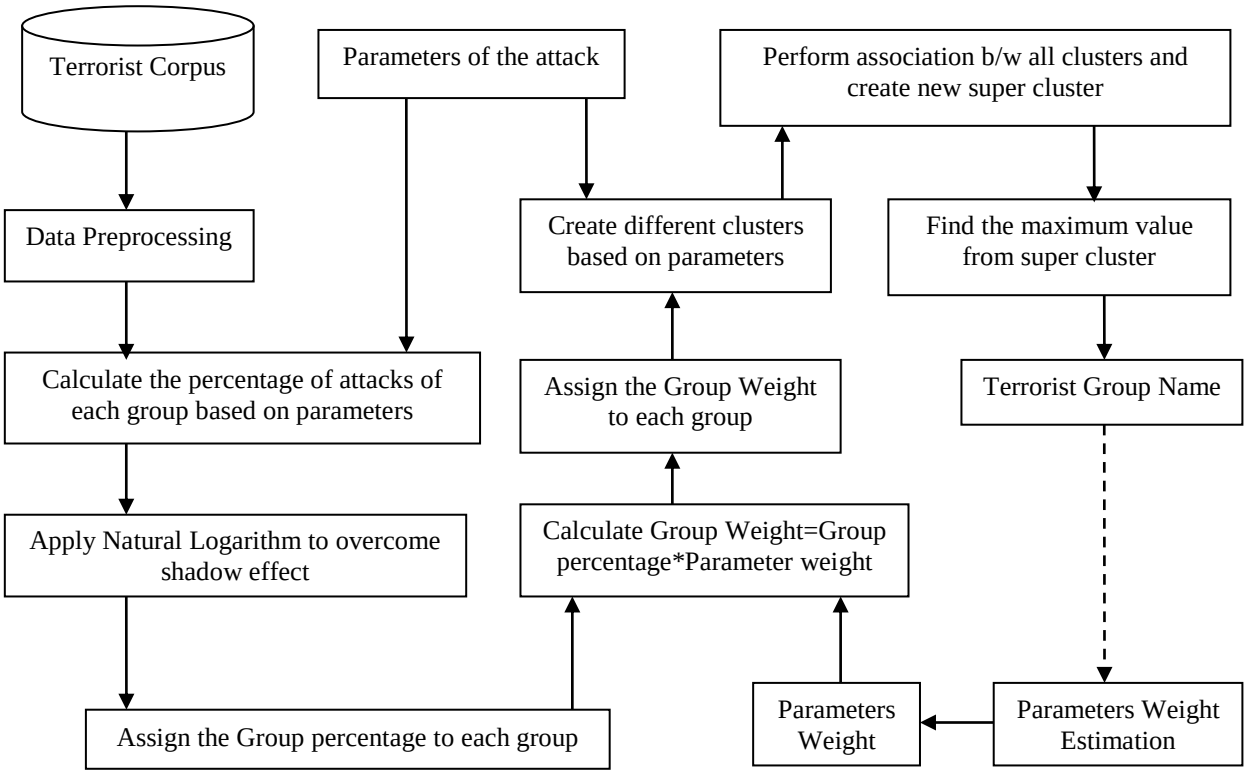

Fig 1: Enhanced Terrorist Group Prediction System (E-TGPS)

4.1 Methodology

E-TGPS uses same set of data as used in TGPM. First step of E-TGPS is data preprocessing. This is an important step in which missing values are filled up, redundancies are removed and filtering is performed to increase the efficiency and accuracy of the system. The missing values of terrorist corpus have been filled by using various terrorist databases available over internet for terrorism research [12, 13, 14]. After pre-processing of database, percentage of attacks of each group is calculated based on input parameters. Natural logarithm has been applied on it to narrow down the shadow effect. Now system assigns the group percentage to each group. Group percentage is the natural logarithm of parentage of attack of each group. Each parameter is assigned a weight based on its impact over the incident. The group weight is calculated by using group percentage of each group and the parameters weight. Different clusters are created. Association between these clusters is being performed and highest value from these associations is obtained. Group name corresponding to the highest value may be the most probable responsible terrorist group.

4.2 Mathematical Formulation

The mathematical equations to perform various major steps of E-TGPS have been discussed here.

The percentage (P) of the total attacks in which each group involved is estimated using Eq. (1).
Here,
G = No. of time a group involved in attacks.
n = no. of unique groups retrieved from dataset.

$$P_j = \frac{G_j}{\sum_{i=1}^{n} G_i} \quad \text{where } j = 1, 2, 3 \ldots n \qquad (1)$$

Equation (2) is used to estimate the percentage of the total attacks in which each group involved based on parameters.
Here value of n will depends on the retrieved records from dataset based on parameters.
m = no. of parameters used in the system.

$$P_{kj} = \frac{G_{kj}}{\sum_{i=1}^{n} G_{ki}} \quad \text{where } j = 1, 2, 3 \ldots n \text{ and } k = 1, 2, 3 \ldots m \qquad (2)$$

Equation (3) is used to narrow down the shadow effect.
α = Group percentage after applying natural logarithm to parentage of attack of each group.

$$\alpha_{kj} = \log_e P_{kj} \quad \text{where } j = 1, 2, 3 \ldots n \text{ and } k = 1, 2, 3 \ldots m \qquad (3)$$

Equation (4) is used to calculate the group weight based on each group probability and parameters weight.
Here, GW = Group Weight.
β= Parameter Weight.

$$GW_{kj} = \alpha_{kj} * \beta_k \quad \text{where } k = 1, 2, 3 \ldots m \qquad (4)$$

Equation (5) will associate the different clusters values based on parameters.
Here C_L is a super cluster created after association.
The value of n will be the total no. of unique groups in all clusters.

$$C_L = \sum_{k=1}^{m} GW_{kj}/m \quad \text{where } L = 1, 2, 3 \ldots n \qquad (5)$$

Equation (6) will determine the highest value from the super cluster that will give the group name as result.

$$R = \text{Max}\{C_L\} \qquad (6)$$

5. OPTIMAL PARAMETERS WEIGHT ESTIMATION

Estimation of the optimal parameters weight is an important step to increase the efficiency and accuracy of the system. E-TGPS uses optimal parameters weight estimation as a part of the system to improve the predictive accuracy of the system. The optimal parameters weight estimation has been performed over training data. The system uses various combinations of the parameters weights to check the performance of the system over training data. System runs itself until all the combinations of the parameters weights have been processed. The values of the parameters for which system gives most efficient result are optimal parameters weight. These optimal parameters weight has been used by the system to identify the terrorist group. E-TGPS performance has been evaluated on standard Global Terrorism Database (GTD) [12].

GTD database cannot be use directly as terrorist corpus. Firstly it should be process because some of the values of different fields of the database are missing and redundant. Database size is reduced according to the requirement for validating the system performance. Terrorist incident data from 1998-2008 have been used to validate the performance of the system. Data related to the Naxalites, Maoist Communist Center (MCC), Maoists, People's War Group (PWG), Communist Party of India- Marxist-Leninistare removed as these group got merged to form a new entity, the Communist Party of India-Maoist (CPI-Maoist) on September 21, 2004 [13, 18]. So there parameters will not be identical. Top 11 groups involved in 590 incidents have been used to evaluate the system and the concept of 3:1 ratio of training and testing have been followed. It means that 447 incidents have been use as training data and 143 incidents as testing data. The same sets of data as used in TGPM have been used to validate the performance and predictive accuracy of the system over TGPM [15].

The different parameters used in the system are attack type, location, target type, weapon type, hostage/kidnapping and suicide attack. When all the parameters are given equal weight 1, the overall performance of the system found to be 52.44%. This is because all parameters are never having same weight. Some parameters are more effective and some are less to distinguish the pattern of groups. To improve the performance, different parameters should have different weights. To determine the optimal values of the weight for different parameters, several experiments have been conducted. The graphs are plotted by varying the weight of only one parameter and keeping all others parameters weight equal to 1. The graphs in fig.2, fig.3, fig.4, fig.5, fig.6 and fig.7 are showing the effect of parameters weight over the result of system.

The performance of the system varies with the parameters weight. The overall performance 79.72% of the system is achieved when location parameter weight is more than 7.25. Although the overall performance of the system increases but the performance of the system to identify small groups decreases. This is because when location parameter is having much higher parameter's weight than other parameter's weight, system will identify the different groups which are working in different locations but it will not be able to distinguish between groups which are working on same location. To identify the groups which are working on same location system will require other parameters. So we can say that a proper combination of different parameters weight will give higher system performance.

In this experiment various parameters weight have been changed so that the performance of the system can be evaluated with different weight conditions. The results show that when the parameter weight of different parameters of system except location parameter weight increases, the performance decreases. This is due to the fact that the overall percentage weight of location reduces when other parameters weight increases. Based on the findings, it is the conclusion that the location is a major factor when discussing about the terrorist groups working in a country because each group is having its own working region. Parameters weight values ranging from 0-10 have been assigned to the system.

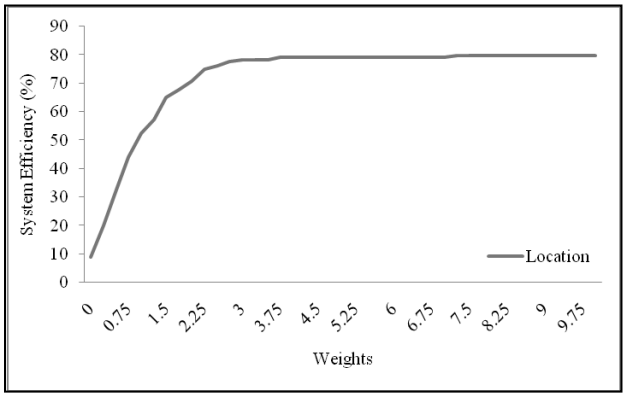

Fig 2: Location parameter weight graph

Fig.2 shows the variation of system accuracy with respect to weights of location keeping other parameter weights as 1. It can be observed that as the location parameter weights are increased, accuracy of the system increases because location parameter is an important factor to identify the groups. For the weights taken in the table, accuracy of the system ranges from 9.09 to 79.72. It has been identified that, when location parameter's weight becomes greater than 7.25, the performance of the system is constant because now system cannot detect more groups this is due to remaining groups are from same location as big groups are detected. To identify remaining small groups system requires proper parameter's weight of different parameters.

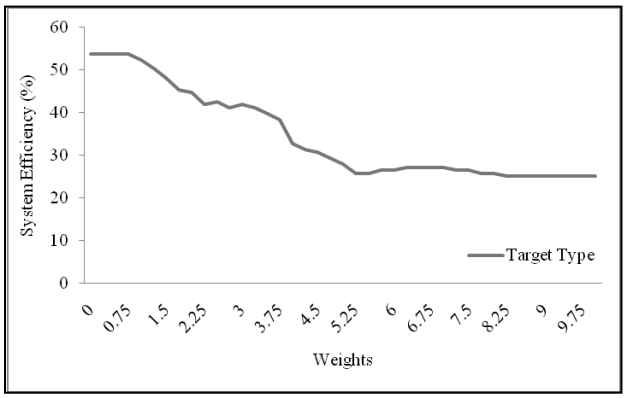

Fig 3: Target parameter weight graph

Fig.3 shows the variation of system predictive accuracy with respect to weights of target parameter keeping other parameter weights as 1. It can be observed that as the target parameter weights are increased, predictive accuracy of the system decreases. The predictive accuracy of the system varied in the range of 53.84 – 25.17.

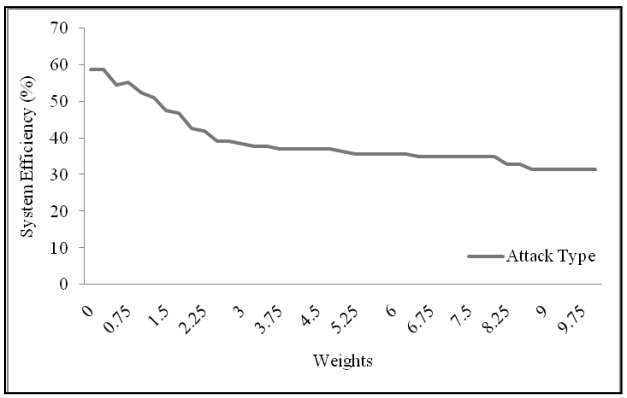

Fig 4: Attack parameter weight graph

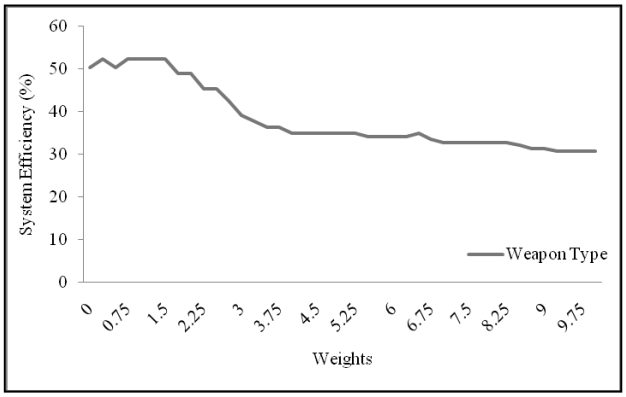

Fig 5: Weapon parameter weight graph

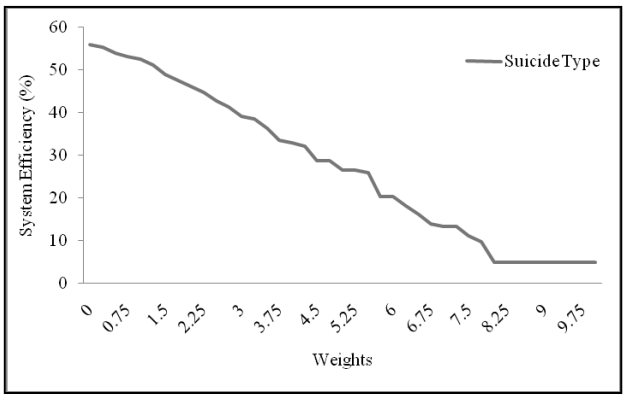

Fig 6: Suicide parameter weight graph

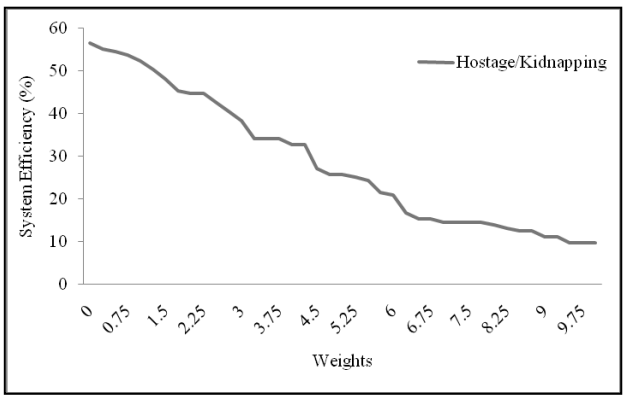

Fig 7: Hostage/Kidnapping parameter weight graph

It is the conclusion that, except location weight all other parameters weights are showing almost the same effects over the performance of the system.

6. EVALUATION AND RESULTS

The research has been performed to estimate the effect of parameters and the optimal parameter weights (β). Optimal parameter weights (β) have been used to calculate the system performance. After estimation of optimal parameters weight and analysing them, the conclusion is that optimal location type weight be 7.75, target type weight be 0.75, attack type weight be 0.75, weapon type weight be 0.50, suicide weight be 0.25, hostage/kidnapping weight be 0.25. By using these weights the performance of the system is found to be 81.12%. This system works well for both big groups and small groups, but some time it fails to identify small groups or sub groups which are

working with a big group (because subgroup's parameters will be similar to that of big group and hence they cannot be differentiated). Another problem of the system also associated with the system is that it fails to detect terrorist groups which uses different pattern each time or those which are new or unheard.

The predictive accuracy of E-TGPS is 81.12% which is greater than TGPM predictive accuracy of 80.41%. There is no big difference in predictive accuracy of these two systems but E-TGPS has directed a way toward narrow down the shadow effect and detection of small groups. The performance of the system is dependent on the input parameters. More the parameters, more accurate will be the results. The performance of the system is also associated with the availability of proper data and parameters for training the system so that specific small groups can also be detected.

7. CONCLUSION

This paper has presented the possibilities to predict the group involved in the given incident by using historical terrorist data. The system discussed in this paper is simple, efficient, easy to implement and user interactive as compare with other existing complex systems. Here it has been indicated that how to narrow down the shadow effect of big terrorist groups. However, further study can be carried out to narrow down the shadow effect by using more unique parameters of small groups. Mathematical equations along with optimization function and techniques can be used in the system for more realistic and accurate results. For further research point of view it is suggested to use different artificial intelligence and computational intelligence techniques for group detection and include more parameters like phone calls, email data etc. so that more accurate results can be obtained. This potential can only be achieved by providing more realistic and authentic information to the system.

8. REFERENCES

[1] Abhishek Sachan, Devshri Roy, Arun P.V., "An Analysis of Privacy Preservation Techniques in Data Mining", In: Proceedings of the Second International Conference on Advances in Computing and Information Technology (ACITY) July 13-15, 2012, Chennai, India, vol. 3, pp. 119-128, Springer Berlin Heidelberg, 2013.

[2] A. Malathi and Dr. S. Santhosh Baboo, "Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model", In International Journal of Scientific & Engineering Research, June 2011, Vol. 2, Issue 6.

[3] David, G., "Globalization and International Security: Have the Rules of the Game Changed?", In Annual meeting of the International Studies Association, California, USA, http://www.allacademic.com/meta/p98627_index.html, 2006.

[4] H. Chen, D. Denning et al., "The Dark Web Forum Portal: From multi-lingual to video", InIntelligence and Security Informatics (ISI), IEEE conference, 2011.

[5] Nooy, W.d., Mrvar, A., et al.: Exploratory Social Network Analysis with Pajek. Cambridge University Press, New York, 2005.

[6] Scott, J.: Social Network Analysis. SAGE Publications, London, 2005.

[7] Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications, 1994, pp. 266.

[8] Coffman, T.R., Marcus, S.E.: Pattern Classification in Social Network Analysis: A case study. In: 2004 IEEE Aerospace Conference, March 6-13, 2004.

[9] Abhishek Sachan, "Countering Terrorism through Dark Web Analysis", In: Proceedings of Third International Conference on Computing Communication and Networking Technologies (ICCCNT-12), Coimbatore an IEEE conference, 2012.

[10] Faith Ozgul, Zeki Erdem and Chris Bowerman, "Prediction of Unsolved Terrorist Attacks Using Group Detection Algorithms", In: LNCS, vol. 5477, pp. 25-30. Springer, Heidelberg, 2009.

[11] Taipale KA, "Data mining and domestic security: connecting the dots to make sense of data", In Columbia Sci Tech Law Rev 5, 2003, pp. 1–83.

[12] Global Terrorism Database, http://www.start.umd.edu/gtd/, Retreived on 05/02/2012.

[13] South asia terrorism portal , http://www.satp.org/satporgtp/countries/india/terroristoutfits/CPI_M.htm, Retreived on 05/02/2012.

[14] Dark web portal, "http://cri-portal.dyndns.org/portal/Home.action", Retrieved on 19/02/2012.

[15] Abhishek sachan, Devshri roy, "TGPM: Terrorist Group Prediction Model for Counter Terrorism," In: International Journal of Computer Applications, Vol 44 (10), 2012, pp. 49-52.

[16] Ozgul F., Bondy J., Aksoy H., "Mining for offender group detection and story of a police operation", In: Sixth Australasian Data Mining Conference (AusDM 2007). Australian Computer Society Conferences in Research and Practice in Information Technology (CRPIT), Gold Coast, Australia , 2007.

[17] Ozgul F., Erdem Z., Aksoy H., "Comparing Two Models for Terrorist Group Detection: GDM or OGDM?", In: ISI Workshops 2008. LNCS, vol. 5075, pp. 149–160. Springer, Heidelberg, 2008.

[18] Council on Foreign Relations, http://www.cfr.org/india/terror-groups-india/p12773, Retreived on 05/02/2012.