Survey on Anomaly Detection Techniques in Social Networking

Bhagyashree Patil
Department of Computer Engineering
MIT College of Engineering
Pune, India

Prof. R. K. Bedi
Department of Computer Engineering
MIT College of Engineering
Pune, India

***Abstract*—Combine success of social networking sites and smart phone devices have change the way people communicate with each other. Leakage of personal information causes to enter the threat from cyber world to our real life which causes issues like profile cloning, social phishing, neighborhood attack, physical threat. Ultimately it disturbs the social, psychological, physical state of mind of user. Hence it is necessary to detect anomaly in social networking site to provide safe social networking and information security. An anomaly is a set of activities that deviate from the normal behavior of the user. In a given paper we have proposed the approach of anomaly detection in social networking site and overview of research on anomaly detection. In our proposed method we are going to use integration of two approaches like link anomaly detection and text anomaly detection. Model of anomaly detection will use data collected from user's profile, text, words, URLS shared by him. Model also requires monitoring of time series behavior of user. Bayesian probability model will be used for classification of unknown event into anomalous or non-anomalous event.**

Keywords*—: Anomaly detection, Bayesian probability model, outlier periodic pattern, time series, and social networks.

I. INTRODUCTION

Now-a-days social networking sites are becoming the main communication media among individuals and organizations. Social network sites are web-based services that allow individuals to construct a public or semi-public profiles, articulate list of other users with whom they share a connection, and traverse and view their list of connections and those made by others in the system[4].

People share their personal information, photos, videos, URLs, ideas on these sites. People live in contact with their family, friends and colleague. However leakage of personal information creates security problem, cyber bullying, spreading the hatred messages etc. Malicious users may cause many severe issues like De-Anonymization attack, neighborhood attack, profile cloning, social phishing, spam attacks and many more. Hence development of reliable anomaly detection in social networking sites is extremely important.

An anomaly is a set of activities that deviate from normal behavior of the user [9]. Anomalies are also called as outliers, abnormities, deviants, discordant, exceptions. If users on SNS had spread hate message in a group which harms the society or organizations, then behavior of such users in SNS deviate from normal. This can be used as a clue for tracking criminals.

This paper focuses on a detail introduction about several anomaly detection schemas for identifying anomalous and non-anomalous behavior of user as follows:

a) Detection of User Cluster with Suspicious Activity

Group of users with suspicious activities are identified based on sentiments in online messages and comments exchange over SNS over time [5].

b) Approach to detect suspicious profiles on social platforms

Aim of a dynamic approach is to alert the users of smartphone users about suspicious profiles located in his or her close circle of contacts on a given social network [6].

c) Detection of Random Link Attacks

Malicious users create false identities and used it to communicate with innocent users. While detecting Random Link Attack mining social networking graph which is extracted from user interaction in communication network is important [8].

d) Threat Detection through Graph Learning and Psychological Context.

Structural Anomaly (SA) detection will uses graph analysis to detect threat. Psychological Profiling (PP) will use behavioral pattern of a user to identify threat. Whatever outcomes o module SA and PP are fused and ranking of most probable threat is done [7].

e) Detection of Emerging Topics via Link-Anomaly Detection in Social Streams

Main focus is on detecting emerging topics from social network streams based on monitoring the mentioning behavior of users [1], [2].

II. RESEARCH METHODOLOGY

A. Detection of User Cluster with Suspicious Activity

Proposed system will detect the user cluster with suspicious activities in following steps:

a) Online data monitoring system and Database

Communication between users on online social networking is monitor by online data monitoring system. Data contains sender, receiver of message, actual message, date, timestamp when message send. Database is used to access the information of suspicious user.

b) Suspicious message identification using NLP/Keyword

Based on sentiment score, sentiment count, training based sentiment, dictionary based topic identification system suspicious message is identified and user profile is updated.

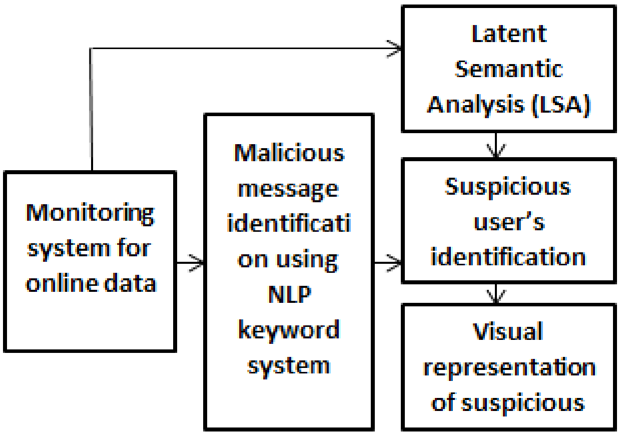
Latent Semantic Analysis (LSA)
Monitoring system for online data
Malicious message identification using NLP keyword system
Suspicious user's identification
Visual representation of suspicious

Fig.1.Proposed sytem to detect cluster of suspicious activity of a users

c) Latent semantic analysis (LSA) system

LSA identifies the messages which are having similar meaning but words used are different.

d) Suspicious users identification system

Suspicious activities of user are stored in database called 'user history'. Current and previous information is used to update user information which alerts about suspicious user.

e) Visual representation of suspicious users

Graphical representation of user network is obtained using visualization tool. This helps to identify suspicious node or link in SNS.

B. Approach to detect suspicious profiles on social platforms

Smartphone users are alerted about possible suspicious profiles located in his or her close circle of contacts on a given social network based on five indicators as follows:

a) Activity (A): Activity is no. of actions performed by a user in a given time period. Actions are no. of messages sent, no. of reply, no. of likes, no. of friend request send, no. of groups created.

*b) Visibility (V):*Visibility is amount of techniques users can perform in the aim to increase its audience in a given time period. Techniques are nothing but no. of keywords and references used in message. Keywords will attract the people of interested community. References indicate that message is received by targeted audience.

*c) Balance (V/A):*Balance indicator will give insight into suspicious behavior. Balance should be present between no. of messages sent with visibility. Anomaly user will try to participate actively in the network without showing its visibility or vice versa.

*d) Energy:*Energy is Euclidian distance from visibility and activity co-ordinated to the origin.

*e) Anomaly score:*Suspicious users will give unusual activity and visibility pair than normal user. Naïve Bayesian classifier is used to calculate anomaly score dynamically.

C. Detection of Random Link Attacks

In an RLA, the malicious user creates a set of false identities and uses them to communicate with a large, random set of innocent users. A is a (non-empty) subset of nodes in the social network graph G and V be a subset of nodes in G-A that share an edge with some node(s) in A. Random Link Attack can be identified if it satisfies following conditions

- $|A| \leq k$
- $|V| \geq \alpha\, |A|$
- Distinct external triangles $\leq \theta$

To launch a successful attack, the set of victims has to be much bigger in size than the attack set itself by a constant factor α of attack set. External triangles are the triangles formed by the attackers with the rest of the graph. These triangles contain single attack node and two non-attacker node (victims). Two external triangles are called distinct if they contain two different pairs of victim nodes. Since edges which are present in victims must be a part of these triangles, and it is observed that these edges are found to be few in number. So the number of distinct external triangles will also be very small.

To identify malicious node the count of external is minimum as possible. If no. of distinct external triangles is less than threshold value then it is considered as malicious node.

D. Threat Detection through Graph Learning and Psychological Context.

Threats are detected through graph learning as follows:

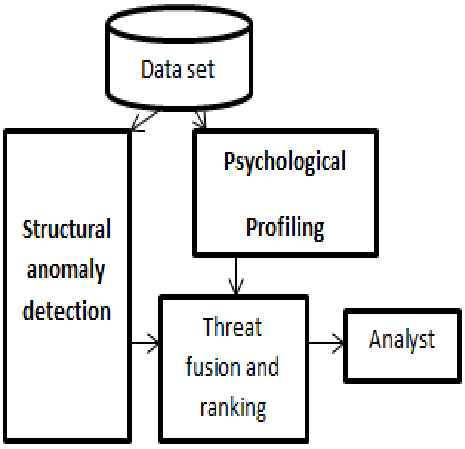

Fig.2.Componants of threat detection system

Fig.2. Anomaly detection using Structural anomaly detection and Psychological profiling

Emerging topics are detected using link anomaly detection as follows:

Structural Anomaly Detection (SA) from social networks and Psychological Profiling (PP) of individuals are used to find out graph anomaly and behavioral anomaly respectively. In SA Graph Structure Analysis gives network characteristics, Anomaly Detection finds unusual pattern in graph data.

In PP Human Psychology Model is built from human behavioral pattern. Behavioral, text, social predictor will predict the overall behavior of a user and then it will find the unusual behavior of a user. Statistical inference model will rank threat based on their probability and uncertainty.

E. Detection of Emerging Topics via Link-Anomaly Detection in Social Streams

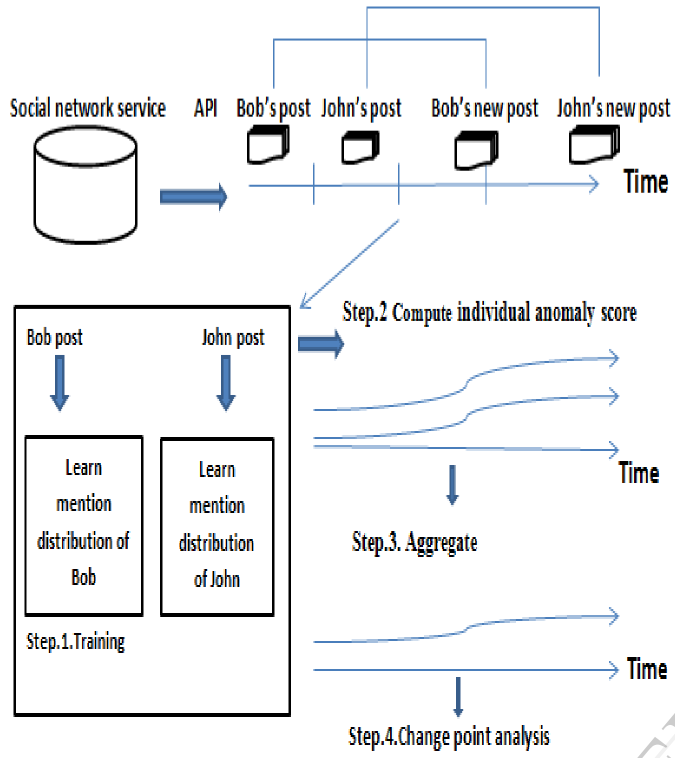

Fig.3. Emerging topic detection using link anomaly detection

Main focus is on detecting emerging topics from social network streams based on monitoring the mentioning behavior of users. Emerging topic is that people feel like to discuss, give comments, and forward to their friends.

Probability model consists of normal mentioning behavior of a user. It consists of number of mentions per post and occurrence of frequency of user in the mentions. Anomaly of future user behavior is measured by the model. Anomaly score is aggregated by change point detection technique. This technique is based on the sequentially dis-counting normalized maximum-likelihood (SDNML) coding. This technique can detect a change in the statistical dependence structure in the time series of aggregated anomaly scores, and notify where the topic emergence is.

III. PROPOSED SYSTEM

Proposed system of anomaly detection in social networking site consists of integration of two approaches first is text anomaly detection and second is link anomaly detection. Each module is explained in detail as follows:

A. *Text anomaly detection*

Dataset of social networking site like Facebook, tweeter is given to module of text anomaly detection. Content preprocessing is next step which consists of many other processes as follows:

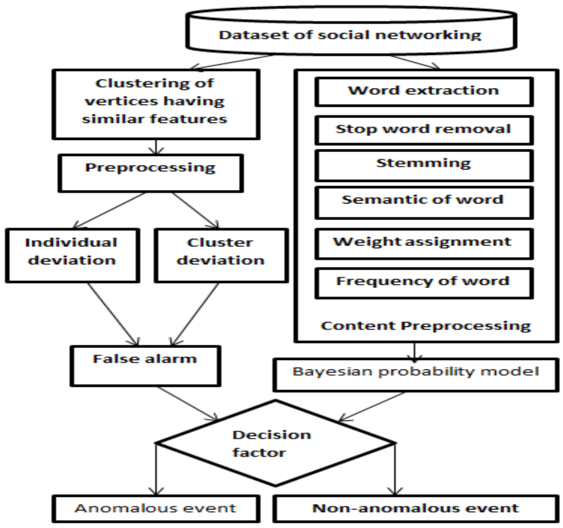

Fig.4. Proposed system

a) Word extraction: Words are extracted from text shared by user over social networking site.

b) Stop word removal: In some cases stop words can causes problems when searching for phrase that include them. Most commonly removed words are the, is, which, at and so on.

c) Stemming: Variant forms of a word are reduced to a common form. Stemming is the process of retrieving root or stem of word.

d) Semantic of word: Semantics is the study of meaning that is used for understanding human expression through language.

e) Weight assignment to word: Whatever words extracted from previous steps are assigned weight to them depending on prediction made from word.

f) Frequency of words: how many times particular words appear in a given time period is calculated.

Bayesian probability model for classification-Bayesian probability model will predict the probability of message being an anomalous or not[14]. Result of it forwarded to decision factor module.

B. *Link anomaly detection*

Dataset of social networking site is also given to link anomaly detection module. A step performed in this module is as follows:

a)Clustering of vertices having same features: We can do clustering of vertices depending on same communication behavior and build profile for each cluster. Individual vertex profiles are also built depending on the communication behavior of a vertex [10].

b) Preprocessing: For dynamic graph time span is divided into disjoint time interval. For particular time period static graph is built to summarize dynamic graph. For each vertex link based features are extracted and feature vector is generated. Cluster profiles and individual profiles are building based on these feature vectors.

c) Individual deviation: Under normal circumstances vertices should show close behavior to its cluster center and some variations are allowed its own individual center. If vertex will show significant deviation from cluster center or individual deviation then it introduce false alarm.

d) Cluster deviation: Cluster deviation of a vertex in a given time period is distance between current feature vector and cluster center. If distance is maximum then vertex will show cluster deviation and it introduce false alarm.

e) False alarm: False alarm introduces by individual and cluster deviations are taken into consideration and final false alarm is identified and possible anomaly score is forwarded to decision factor.

C. Decision factor: Result obtained from link anomaly module and text anomaly module is compared in decision factor and final anomaly is predicted.

TABLE 1. COMPARISION OF ANOMALY DETECTION TECHNIQUES

Sr.no	**Comparison of anomaly detection technique**		
	Title	***Pros***	***Cons***
1	Detection of User Cluster with Suspicious Activity	Visualization of nodes makes easier to identify suspicious users	Number of clusters required to be mention prior of the cluster formation
2	Approach to detect suspicious profiles on social platforms	Framework is applied to any ubiquitous device, low calculation cost	Applied to the users which are present in the contact list of smartphones user
3	Detection of Random Link Attacks	RLA is based on simple interconnection properties of individuals in any communication network.	If attackers employee collaboration strategy then harder to detect RLA by counting the triangles
4	Threat Detection through Graph Learning and Psychological Context.	Behavioral, text, social proctor together generate more accurate result than individual predictor	Approach will generate poor results when dataset is small
5	Detection of Emerging Topics via Link-Anomaly Detection in Social Streams	Faster than text based approach, preprocessing time is lesser than text based approach	Proposed link anomaly model does not immediately tell what the anomaly is.

IV. CONCLUSION

Thus we have discussed different techniques of anomaly detection.

We have proposed the model of anomaly detection in social networking site by integrating two approaches first is link anomaly detection and second is text anomaly detection which will generate more accurate results.

V. ACKNOWLEDGMENT

We would like to thanks mentors for valuable discussions, contributing the ideas and helping us to solve problems.

REFERENCES

[1] Faraz Rasheed, Reda Alhajj, “A Framework for Periodic Outlier Pattern Detection in Time-Series Sequences,” IEEE TRANSACTIONS ON CYBERNETICS, VOL. 44, NO. 5, MAY 2014

[2] Toshimitsu Takahashi, Ryota Tomioka, and Kenji Yamanishi, “Discovering Emerging Topics in Social Streams via Link-Anomaly Detection,” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 1, JANUARY 2014

[3] Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han, and Yong Yu, “Mining Social Emotions from Affective Text,” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING VOL. 24, NO. 9, SEPTEMBER 2012

[4] D. Boyd and N. B. Ellison, “Social Network Sites: Definition, History, and Scholarship,” Journal Computer-Mediated Communication, vol. 13, no. 1-2, Nov. 2007

[5] Sharath Kumar A and Sanjay Singh, “Detection of User Cluster with Suspicious Activity in Online Social Networking Sites,” Second International Conference on Advanced Computing, Networking and Security, 15-17 Dec. 2013

[6] Charles PEREZ, Marc LEMERCIER, Babiga BIRREGAH, “A dynamic approach to detecting suspicious profiles on social platform,” IEEE International Conference on communications Workshops (ICC), 9-13 June 2013

[7] Oliver Brdiczka, Juan Liu, Bob Price, Jianqiang Shen, Akshay Patil, Richard Chow, Eugene Bart, Nicolas Ducheneaut, “Proactive Insider Threat Detection through Graph Learning and Psychological Context,” IEEE CS Security and Privacy Workshops,24-25 May 2012

[8] Nisheeth Shrivastava, Anirban Majumder, Rajeev Rastogi, “Mining (Social) Network Graphs to Detect Random Link Attacks,” IEEE 24th International Conference on Data Engineering,pp.486-495,7-12 April 2008

[9] Sang Hyun Oh and Won Suk Lee, “An anomaly intrusion detection method by clustering normal user behavior,” ELSEVIER, Computer & Security, Vol 22, No 7, 2003

[10] Xiaomeng Wan, Evangelos Milios, Nauzer Kalyaniwalla and Jeannette Janssen, “Link-based Anomaly Detection in Communication Networks,” IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Volume 3,pp.402-405,9-12 Dec. 2008

[11] K. Hanumantha Rao, G. Srinivas, Ankam Damodhar, M. Vikas Krishna, “Implementation of Anomaly Detection Technique Using Machine Learning Algorithms,” International Journal of Computer Science and Telecommunications, Volume 2, Issue 3, June 2011

[12] Ryan Layfield, Bhavani Thuraisingham, Latifur Khan, Murat Kantarcioglu, Jyothsna Rachapalli, “Design and Implementation of a Secure Social Network System,” IEEE International Conference on Intelligence and Security Informatics, pp.236-247, 8-11 June, 2009

[13] Jiong Zhang, Mohammad Zulkernine, and Anwar Haque, “Random-Forests-Based Network Intrusion Detection Systems,” IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETIC, VOL. 38, NO. 5, SEPTEMBER 2008

[14] Kush R. Varshney, “Bounded Confidence Opinion Dynamics in a Social Network of Bayesian Decision Makers,” IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 8, NO. 4, AUGUST 2014

[15] Yang Li, Bin-Xing Fang, “A Lightweight Online Network Anomaly Detection Scheme Based on Data Mining Methods,” International Conference on Network Protocols, pp.340-341, 16-19 Oct. 2007

[16] Alexander Y. Liu and Dung N. Lam, “Using Consensus Clustering for Multi-view Anomaly Detection,” IEEE CS Security and Privacy Work, pp.117-124, 24-25 May 2012

[17] Gabriel Weimann, “Terror on Facebook, Twitter, and Youtube,” The Brown Journal of World Affairs, volume 16, issue 2, 2010

[18] Mr. A. A. Sattikar,Dr. R. V. Kulkarni, “Natural Language Processing For Content Analysis in Social Networking,” International Journal of Engineering Inventions, Volume 1, Issue 4, pp.06-09, September 2012