ISSN: 2321-7782 (Online)
Impact Factor: 6.047

Volume 5, Issue 1, January 2017

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Research Perspectives in Security Threat Detection in Social Media Networks

Dr. Savita Kumari Sheoran[1]
Associate Professor & Chairperson
Department of Computer Science & Applications
Chaudhary Ranbir Singh University,
Jind (Haryana) – India

Pratibha Yadav[2]
Research Scholar
Department of Computer Science & Applications
Indira Gandhi University Meerpur,
Rewari (Haryana) – India

Abstract: ***Now days the use of social media is widespread in all aspect of society. Owing to its features of fast updates, semantic richness and great relationship between wide public it imprints great impact on the society. Undoubtedly this novice media application provides a robust platform for social interaction with fast response by convoying a stream of real-time social media posts. These information and responses over network contains the public and private information about user, whose security and privacy is a serious concern. Since the social media data differ from their traditional counterpart in all aspects of size, texture and character etc, therefore their threats towards security are also stringent. The perusal to literature reveals that the major issue in this area is to detect the actual security threats in social media networks. This paper study about the security threats encountered in social media network and based upon it formulates an open research challenges which need immediate attention of research and scientific community engaged in this regime. It further proposed a theoretical model to address these research challenges in present scenario. This study may be helpful towards developing a practicable solution to security threats in social media networks.***

Keywords: Social media research, threat detection, social media networks, data security.

I. INTRODUCTION

Owing to its immense applications and ease to use, the social media has become a buzz word in recent years, which provides platform to connect people across the world and share their interests. With the proliferation of the Internet, cyber security is becoming an important concern. The ubiquitous use of social media has generated unparalleled amounts of social data. Data may be in the form of text, audio, video, images, numbers or facts that are computable by a computer. A particular data is absolutely useless until and unless it converted into some useful information. Therefore it is necessary to analyse this massive amount of data and extracting useful information from it. Information can be spread across social networks quickly and effectively, therefore become susceptible to different types of undesired and malicious spammer and hacker actions. Social Networking Sites are providing opportunities for cybercrime activities; hence there is an essential need for security in social media networks and industry [1].

Data are now woven into every sector of industry and function in the worldwide economy. These are generated from emails, online transactions, search queries, audios, videos, images, click streams, logs, health records, posts, social networking communications, sensors and science data, mobile phones and their applications. Portable and held hand devices also generated data, generally a workstation holds about 500 gigabytes, so it would be require about 20 billion PCs to collect all of the world's information. There are 6 billion mobile subscriptions in the world and 10 billion text messages are sent in every day. By the end of year 2020, 50 billion hosts will be connected to networks and the internet. With the fast development of information digitization, huge amounts of unstructured, semi-structured and structured data are generated quickly. This data is known as "Big Data" due to its volume, velocity and the variety.

There are three characteristics of Big Data, called "3V":

- **Volume** (the data volumes are huge which cannot be processed by traditional methods),
- **Velocity** (the data is created with great velocity and must be captured and processed quickly) and
- **Variety** (variety of data types: structured, semi-structured, and unstructured).
- Based on data quality, IBM has added a fourth V called: **Veracity.**
- However, Oracle has added a fifth V called: **Value**, highlighting the added value of Big Data.

For instance a total 5 exabytes of data were produced by human until 2003. Today this amount of information is generated and collected in just two days. In year 2012, digital data was extended to 2.72 zettabytes. It is predicted that this data will be double in every two years, and reaching about 8 zettabytes of by 2015. According to IBM, 2.5 exabytes of data generated daily and also 90% of the data produced in last two years [2].

By collecting, storing, analyzing, and mining these data, an enterprise can obtain large amounts of individual users' sensitive data. It is a big challenge for research community and scientist working in this area [3].

Public and private databases contain more threats and vulnerabilities, volunteered and unexpected leakage of data. Therefore, lacking of powerful security policies makes hackers to collect the confidential information. When data is moving from homogeneous to heterogeneous, it is more difficult to secure the data because of untrusted people [4]. At present no secure tools and technologies are available, therefore hackers attack on the database by sending threats such as denial of service, spoofing attack and brute force attack.

1.1. Threats

A threat is the opponent's goal, or what an opponent might try to do to a system. In computer security, a threat is a possible danger that might exploit a vulnerability to breach security and cause possible harm. Big Data is not useful without sharing data between the users [5]. But, sharing of data faces the challenge of data privacy and security. These issues have little attention till now. Only researchers pointed out that due to its big volumes Big Data creates new threats. Traditional security policies are less adequate to protect the Big Data from the threat to secure sensitive data. Big Data involve both the width of sources as well as depth of the information needed for programs to specify risks correctly, to defend against illegitimate activity and advanced cyber threats [6].

Financial institution and healthcare provider are more effected if attacker attacks on their data repository because the data volume is high and government regulations are exists. Still we are lacking the proper policy to secure the data therefore hackers can call any time. It is also a big challenge by research community. Threats can occur from outside or from the inside of an organization [7]. A threat can be internal, external or both external and internal entities. Outside threats are the attacks on system by someone from outside the organization. Hackers are outside attackers who break into computer systems and cause destruction within an organization. Ecological threats can be either internal, due to natural processes or external, due to natural process that begin outside the system boundaries. There are hackers who start diffusion viruses, and these viruses cause enormous harm to the files in various computer systems. But a more threatening issue is the insider threat. There are the insider threats which are information-related attacks. There are people inside an organization who have study the business practices and expand schemes to cripple the organization's information resources. These people could be regular staff or even those working at computer centers and causing all kinds of damage. By nature of action a threat can be malicious or non-malicious. The goal of attackers on a system can be malicious or non-malicious environmental threats are natural and so they are introduced with no malicious goals and committed mistakes are due to unintended actions.

1.1.1. Types of threats

With the increase the Communication in social media, organizations become susceptible to various types of threats such as cyber-attacks. Vulnerabilities consist of security weakness in a system which can be explored by the attackers that may lead to dangerous affects. When vulnerabilities continue living in a system, a threat may be manifested via a threat agent using a particular penetration technique to cause undesired effects [8].

Online Threats: There are four types of online threats,

a) **Classic Threats:** These threats have been originated with the Internet widespread usage. These threats can use the user's personal information published in a social network and their friends also i.e. Malware, Phishing attacks, Spammers, Cross site scripting and Internet Fraud.

b) **Modern Threats:** These threats are unique to Online Social Network environments. These threats target users' personal information as well as the personal information of their friends i.e. click-jacking, fake profile, de-anonymization attack, face recognition and information leakage.

c) **Combination Threats:** Attackers can combine classic threats along with modern threats to create more complex threats. i.e. the attacker may use the phishing attacks to gather a targeted user's social networks password and then forward a message containing a click-jacking attack on the user's timeline, thus attracting the user's Facebook friends to click on the posted message and install a hidden virus onto their own computers.

d) **Threats Targeting Children:** Children of any age group are affected by these types of threats i.e. Cyber bullying, Online predators and Risky.

1.1.2. Threat Impacts

A threat can cause one or several destructive impacts to the systems. Impact of threats can be divided into seven types:

- **Corruption of Information:** It is an unauthorized access which can corrupt many files stored on a host computer or transit data across a network. It is also called as tampering of information i.e. add, delete or change target system's memory, hard drives and other part like file virus i.e. spoof, falsification, malicious logic and repudiation.
- **Destruction of information:** Intentional destruction of a system component to perturb system operation e.g worm, denial of service attack.
- **Disclosure of Information:** The distribution of information to anyone who is not allowed to access that information. These threat actions can cause unauthorized disclosure i.e. Interception, exposure, intrusion and inference.
- **Theft of service:** The illegal use of computer or network services without humiliating the service to other users. i.e. theft of functionality, theft of service, theft of data, data misuse, software or/ and hardware misuse.
- **Denial of service:** The intended degradation or blocking of computer or network assets.
- **Elevation of privilege:** The Use of weaknesses in the system get authorization to access the target system. Such as guessing passwords,
- **Illegal usage:** Use the normal function of the system to attain the attacker's activities for other purposes. For example, an attacker uses the normal network connection to attack other systems, using vulnerabilities throughout the normal system services to attain attacker's aims.

1.2. Data Mining

Data mining has attracted more research attention in past few years because of the fame of Big Data. Data mining is the process of finding interesting patterns and knowledge from huge amounts of data. The outcomes of data mining include classification, clustering, forming associations, as well as detecting anomalies. With increase communication traffic in social media networks, malware technologies are being developed at a very rapid speed. These include worms, viruses, and trojan horses. To adapt to the multisource, enormous, dynamic Data [9], researchers have extended existing data mining methods for Big Data in many ways, including the improvement of single-source knowledge invention methods, designing a data mining mechanism from a multisource point of view [10-11], as well as the study of dynamic data mining methods and the analysis of stream data. In typical data mining systems, the mining process requires rigorous computing units for data analysis and comparisons. Among data mining methods, Naive Bayes algorithm is easy to implement and is an efficient and effective inductive learning algorithm for machine learning. Machine learning framework is efficient as it uses filter approaches to be able to successfully detect malware with a smaller feature set.

Big Data mining gives opportunities to go beyond the conventional relational databases to rely on less structured data: i.e. weblogs, e-mail, social media, sensors, and photographs. This stored data can be mined for useful information. A recent study revealed that users' "data access patterns" can also have severe data security issues and lead to disclosures of geographically co-located users or users with common interests (e.g., two users searching for the same map locations are likely to be geographically co-located). For mining, data processing framework rely on clustered computers along with a high-performance computing platform, where a data mining task being deployed by running some parallel programming tools, such as MapReduce on a large number of computing nodes (i.e. clusters).

The rest of this paper is organized as follow. Section 2 gives review of available literature in the area. Based upon this literature study, an open research challenge is formulated in section 3 while the section 4 constructs a theoretical model to way out a solution for the proposed challenge. Finally, section 5 concludes the paper.

II. Review of Literature

During the last decade, many researchers have been contributed in the areas of big data analytics. However a little works has done in the area of security and privacy. As systems become more complex, there are always exploitable weaknesses as a result of design and programming errors, or through the use of various "socially engineered" penetration techniques. Here we are presenting the works of various authors on big data analytics, information security and privacy.

Avita Katal et al (2013), introduced the new concept of Big Data, its importance and the existing projects. Hadoop tool for Big data is described in detail focusing on the areas where it needs to be improved so that in future Big data can have technology as well as skills to work with. Omar El-Gayar et al (2014) described the steps involved in evidence based medicine and identified current needs and potential for business intelligence and Big Data analytics. Study provides a research agenda for health informatics researchers and data scientists to address issues of pressing needs, reducing the cost and improving the cost of healthcare by broadening the practice of evidence based medicine through the applications of business intelligence big data analytics.

Adil Fahad et al (2014) proposed a categorizing framework to classify a number of clustering algorithms. The categorizing framework is developed from a theoretical viewpoint that would automatically recommend the most suitable algorithms to network experts while hiding all technical details irrelevant to an application. Fan Zhang et al (2014) introduced ordinal optimization using rough models and fast simulation to obtain suboptimal solutions in a much shorter timeframe. Ordinal optimization can be processed fast in an iterative and evolutionary way to capture the details of big-data workload dynamism. Leman Akoglu et al (2015) proposed a comprehensive overview of graph-based techniques for anomaly, event, and fraud detection, as well as their use for post-analysis and sense-making in explaining the detected abnormalities.

III. OPEN RESEARCH CHALLENGE

Social media data creates new opportunity for the world economy in the field of communication, brand development, attracting customers, product promotion, instant feedback from customers, news, and marketing. Attackers try to intrude into networks and computers to steal sensitive information from the organisation by using sophisticated malware. Social media websites permit users freely distribute and share information to their friends. Information can spread out easily and very fast using the social media networks. Online social networking websites are open to various types of unwanted and malicious spammer or hacker actions. There is a strong need in the society and industry for detection of these malicious activities and provide a security solution at organizational and social media networks levels.

Table 1 : Strengths and Limitations of Existing Techniques

Name of Existing Technique in Security Threat Detection	Strengths of Existing Techniques	Limitations of Exiting Techniques and Future Work
Graph based anomaly detection and description: a survey.	i. Graph-based approaches to anomaly detection are Inter-dependent nature of the data. ii. Powerful representation: Graphs represent the inter-dependencies by the links between the related objects. iii. Relational nature of problem domains: The nature of anomalies could exhibit themselves as relational. iv. Robust machinery: Graphs serve as more adversarial robust tools.	i. Anomaly detection on attributed dynamic graphs: While static attributed graphs have been exploited in abnormality detection, there exists only a few works on spotting anomalies by exploiting dynamic attributed graphs. ii. The history trace of dynamic updates: While most techniques for dynamic graph consider and work with edge/node updates, there exists no work that exploits the history of the updates.
A survey of data mining and social network analysis based anomaly detections technique.	i. Cluster based approaches is unsupervised in nature where no predefined set of labeled classes of data objects is required. ii. These methods involve fast comparison process as once clusters are constructed: It is faster to compare objects to clusters because a number of clusters available are comparatively less than number of objects.	i. Computational complexity is highest for all data mining methods applied. ii. Clustering approaches are a costly procedure for large data sets iii.Sometimes clustering process involves anomalous objects depicting similar behavior and hence forming the clusters.
Detecting spammers on social networks.	i.Through analyzing content feature and user behavior, it is capable to distinguish abnormal behavior from legitimate ones. ii. SVM classifier is capable to achieve best accuracy.	i. Feature extraction is based on statistical analysis and manual selection. So it is low adaptive and costive.
Zero-day Malware Detection based on Supervised Learning Algorithms of API call Signatures.	i. Machine learning framework has resulted in high accuracies in malware detection. ii. The system is signature-free and does not require knowledge about the API sequence of execution to classify a malware.	i. Effeciency is not improved by single machine learning algorithm rather uses combination of algorithms.

IV. Proposed Solution Model

For a solution to be viable, it must be highly scalable and support multiple heterogeneous data sources. Current state-of-the-art solutions do not scale well and preserve accuracy. Furthermore, by utilizing the data acquisition techniques, we are able to easily integrate and align heterogeneous data. Thus, proposed approach will create a scalable solution in a dynamic environment. Presently, no existing threat detection tools offer this level of scalability and interoperability. We have included the novel data mining techniques to create an efficient threat detection solution.

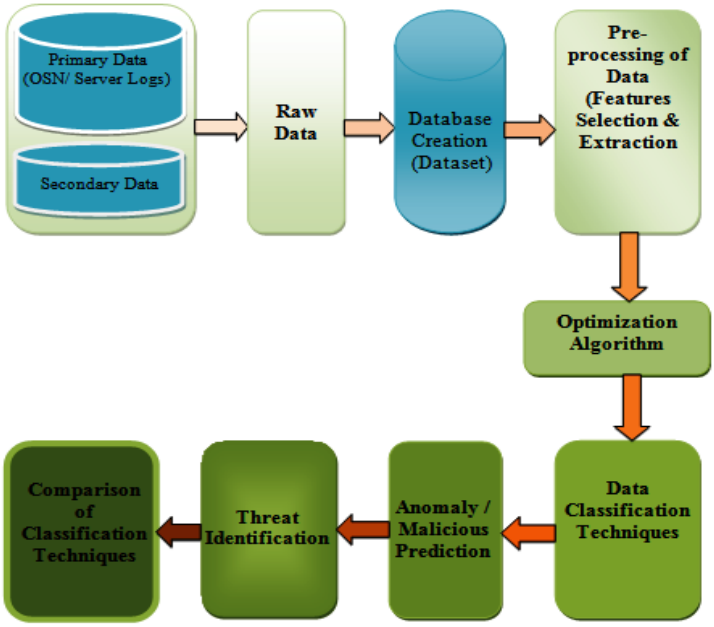

Figure 1: Proposed Model

Particularly, this approach will use primary data from social media data and/or server logs or secondary data as a dataset. After loading, data will be refined to reduce the redundancy and stop words. Database will be created from raw data collected from social networks or server logs. This solution will pull data from multiple sources and then data feature selection and extraction techniques will be applied to the database. Database may be large to process, so after feature selection, reduction of data to relevant data will be done. An optimization algorithm will ab applied for this reduction of data. After that, data classification techniques will be used to classify the dataset into malware or legitimate data. On the basis of classification techniques, we will be able to predict the anomalies or threats found on social media data and servers logs. Finally a comparison of classification techniques may be carried out on the basis of result obtained in classification step. Hence security can be enhanced by detecting the anomalies on server's logs and malicious activities on social media.

V. Conclusion

This research paper studied the security threat issues arises over novice socio-technological area of social media network. We have carried out an in-depth study of characteristics of data communicated over such networks, various types of possible threats, helpful data mining techniques and social information which may be at risk during transmissions. Since the major issue in this area is to detect and tract the proper threat. The open research challenge formulated in section 3 of this paper contain unaddressed question, which need immediate solution in the savior of mankind. The theoretical model proposed in section 4; provide necessary steps for security threat mitigation in area.

Since it is theoretical model, feasible to implement using synthetic or real time data through suitable computation platform compatible with Big Data. Such implementation will generate practicable solution to solve these issues.

References

1. S. Multani Harshal, Sinh Marod Amrita, Pillai Vinita, Gaware Vishal , "Spam Detection in Social Media Networks: A Data Mining Approach", International Journal of Computer Applications , Volume 115 – No. 9, pp.1-4, 2015.
2. Singh S. and Singh N., "Big Data Analytics", IEEE International Conference on Communication, Information & Computing Technology Mumbai India, pp.1-4, October 2012.
3. Dong Xinhua, Li Ruixuan, He Heng, Zhou Wanwan, Xue Zhengyuan, and Wu Hao, "Secure Sensitive Data Sharing on a Big Data Platform", Tsinghua Science And Technology, pp.72-80, Volume 20, Number 1, 2015.
4. Sarala Devi B, Pazhaniraja N, Victer Paul P, Saleem Basha M.S., Dhavachelvan P , " Big Data and Hadoop-A Study in Security Perspective", International Symposium on Big Data and Cloud Computing, Elsevier, pp. 596 – 601, 2015.
5. Beckwith Richard and Mainwaring Scott, "IEEE Conference on Privacy: Personal Information, Threats, and Technologies", pp.9-16, 2005.
6. Nikolaos E. Petroulakis, Ioannis G. Askoxylakis, Theo Tryfonas, " Life-logging in Smart Environments", IEEE Conference on Challenges and Security Threats", pp.5680-5684, 2012.
7. Bhatt Parth, Yano Edgar Toshiro, Jose Sao, M.Gustavsson Per, "Towards a Framework to Detect Multi-Stage Advanced Persistent Threats Attacks", IEEE International Symposium on service Oriented System Engineering, pp.390-395, 2014.
8. Chen R., Sivakumar K., and Kargupta H., "Collective Mining of Bayesian Networks from Distributed Heterogeneous Data", Knowledge and Information Systems, vol. 6, no. 2, Springer, pp. 164-187, 2004.
9. C. Douglas Craig, "An Open Framework for Dynamic Big-Data-Driven Application Systems (DBDDAS) Development", International Conference on Computational Science, Elsevier, Volume 29, pp. 1246–1255, 2014.
10. Wu X. and Zhang S., "Synthesizing High-Frequency Rules from Different Data Sources," IEEE Transaction. Knowledge and data engineering, volume 15, Issue. 2, pp. 353-367, 2003.
11. Wua Xindong, Zhangqi Chengqi , Zhang Shichao, "Database Classification for Multi-Database Mining," Elsevier , vol. 30, no. 1, pp. 71-88, 2003.

AUTHOR(S) PROFILE

Dr. Savita Kumari Sheoran presently working as Chairperson, Department of Computer Science & Applications and Dean, Faculty of Physical Sciences at CRS University Jind (India), is an academician and researcher of high tempore. She has graduated her Ph.D. from Banasthali University, Rajasthan (India) in 2011 and have about 12 years of teaching and research experience. She has more than 40 national / international papers, 03 books and 01 patent in her credit and has attended about 30 national/international events in India and abroad. She is on the panel of reviewer of various research journals and conferences and has organized national conference in capacity of organizing secretary. She has actively been engaged in research in the area of mobile computing, cloud computing, big data, social networks and learning platforms.

Ms. Pratibha Yadav presently pursuing towards her Ph.D. in Computer Science at Indira Gandhi University Meerpur, Rewari (India) in the area of Security Threat Detection in Social Media Networks. She holds M.Tech. in Computer Science and poses teaching experience in reputed colleges/university. She have about a dozen of published research papers/attended conferences.

