Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 9, Issue, 08, pp.56411-56414, August, 2017

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

REVIEW ARTICLE

DATA MINING OF SECURITY THREATS IN SOCIAL MEDIA AND WEBSITES

***Ambiga, N., Kanimozhi, V. and Jayasri, R.**

Department of Computer Science, Sri Akilandeswari women's college, Wandiwash, Tamilnadu, India

ARTICLE INFO

Article History:
Received 23rd May, 2017
Received in revised form
16th June, 2017
Accepted 26th July, 2017
Published online 31st August, 2017

Key words:

Online society,
Social media.

ABSTRACT

In today's online society, social media contains information to be connected with other people, to express themselves to a bigger audience or to gain all kind of relevant information from other users. Thereby, all these activities produce data, that is collected by the service providers. It is often not clear and it contain false information, noisy, unwanted information This is a big threat to the people. We don't know in what extent a service provider makes use of this data. On the one hand, personal data collections can serve a useful and necessary purpose, such as the improvement of the user experience or the facilitation of core functionality of the service. It can give also the opportunity for new businesses and research fields to emerge. On the other hand personal data collections has led to a widely applied practice of data analysis and personal information trading, which can highly compromise the users' privacy and lead to many further problems. In face of these contradictions, this course deals with the implications for the society from personal data collections. Therefore the main subjects are the discussion of business models based on personal data common practices to analyse personal data collections and alternatives to the current data driven social media landscape.

Citation: Ambiga, N., Kanimozhi, V. and Jayasri, R. 2017. "Data mining of security threats in social media and websites", *International Journal of Current Research*, 9, (08), 56411-56414.

INTRODUCTION

We discuss techniques for information extraction from text or images that is present on websites, and present two applications that use these techniques. We focus in particular on social media texts (Twitter messages), or any other images or videos which present challenges for the information extraction techniques because they are noisy and short. The first application is extracting the locations mentioned in websites and the second one is detecting the location of the users based on all the text or images posted or written by each user. The same techniques can be applied and used for extracting other kinds of information such as social media texts, with the purpose of monitoring the topics, events, emotions, or locations of interest to security and defence applications.

Social media services and revenue models

1. To study about the media services be familiar with the interest groups of social media services and know their revenue models.
2. be aware of threats which can appear with the collection of personal data in social media and implications of these threats.
3. be able to lead a discussion about different infrastructures for social media services and will know their strengths.
4. know methods of protecting users' privacy in social network services with a single service provider and their benefits and limitations.
5. have a basic understanding of technical aspects of data mining as well as machine learning and its usage
6. know ethical issues and problems, that arise with the usage of data mining for processing personal information

The Social Media Industry threats

Regarding today's social media landscape many popular social media services are provided by commercial companies. Those companies have got expenses for maintaining and extending their server infrastructure, implementing new features in their services, employing qualified staff and so forth. Even though many of these services remain free to use. This aspect often affects the design of the service, which tends to collect personal data for commercial purposes and threatens the users' privacy. The concentration must be on different options to generate revenue streams, their impacts on users as well as the explanation of basic terms and characteristics of a server based infrastructure. At first the term "social media" will be clarified and an understanding distinction between the terms service, service provider and technology will be developed. Afterwards

Corresponding author:* **Ambiga, N.
Department of Computer Science, Sri Akilandeswari women's college, Wandiwash, Tamilnadu, India.

business models of popular social media services and alternative approaches will be discussed. As a part of this the concepts of homophily, influencer and ad targeting based on user information will be introduced followed by a discussion about their impacts on users.

Threats, preventions and dilemmas in social media

Besides the known threats for privacy in social media services, additional threats arise with the involvement of adverse parties. This includes technology-based threats as well as organisational threats like impersonation attacks or social engineering. Users of these services might not be aware of these threats or are unable to protect themselves due to limited technical knowledge or insufficient prevention methods. On the other hand leaving a service can imply several consequences like the affection of social life. Therefore the analysis of social dilemmas like leaving a social media service and rules defined by a service provider is taking place in this course. Furthermore the course will deal with common risks and their prevention apart from privacy violations. Beyond that users' trust in service providers and the efficiency of recent developments like encryption in messengers to protect the users' privacy will be discussed.

Decentralization of Social Network Services (SNS)

After exploring possible threats and preventions this lecture will focus on approaches to avoid personal data collections by a single service provider and associated problems. This attempt can be addressed by decentralisation of SNS. Especially in the area of SNS, scientific approaches and production systems use peer-to-peer networks or decentralised servers to distribute the burden of single service provider across several participants. Those approaches often reduce economic incentives; offer more scalability, openness and protection of users' privacy. Though, services with a single service provider have a lot of advantages in relation to security, which cannot be archived by decentralised services in the same manner. Furthermore, decentralized services bring new challenges to researchers in order to make these services open, secure, reliable, scalable and easy to use. We will concentrate first and foremost on with the infrastructure of SNS. Therefore, possible decentralized infrastructures like peer-to-peer systems and decentralised servers will be introduced and compared with a centralised infrastructure in the context of SNS. After that it will be discussed which problems can be addressed by these solutions and how the data is protected. In addition the course will analyse decentralised SNS for limitations in protecting the users' privacy.

Data Mining and Machine Learning

The collecting of social media data by a service provider is just the first step because the vast amount of data has to be processed to gain useful information. This is where data mining comes into play. With this technology it is possible to discover patterns in the data. Therefore, data mining allows researchers to analyse a huge amount of data and to gain information for what individuals wouldn't be able to do or would need a long time to achieve. Nevertheless, further steps (like preparation, interpretation, selection and evaluation) are required to gain valuable information. This process (also known as Knowledge Discovery in Databases) is the topic of this workshop. Additionally machine learning and classifiers will be introduced and their relevance for social media data will be outlined. Afterwards ethical issues, which arise with the usage of this technology on social media data, will be explored with use cases by the participants. Recommendations, search engines people and documents, marketing, communications, advertising, and many others. Nowadays social network analysis (SNA) is used to study a variety of eco world Economic Forum ahead of its annual meeting, continues to receive attention from the business world and reaches a level of credibility in a crowded arena of forecasting. Cyber attacks on businesses will increase with a denial of service, data breach, cloud provider compromise and extortion being major concerns for IT departments. The sensitive geopolitical context, the rise of cyberattacks and major data breaches and hacks, as well as the global insurgency of violent extremism and radicaliist is always higher than to random people. On the one hand, this is good, since forming a loyal audience around the company, brand or person. But on the other hand, it is an opportunity for attackers.

3. Possibility of substitution of person or masquerade: for sure it is not clear exactly who hide their actions behind the name of friends or hiding behind photos friends in social network profile. It is possible by the IP-address of sender to gather at least some information about him in the correspondence by e-mail, that Is not work in social network.
   This masquerade is possible at the corporate level also. The result of such malicious script can be phishing, the organization of "black PR" or "Antipiar". There were many instances where it was not clear who created the site on behalf of any company – it is created a problem for the original brand.
4. Stealing passwords and phishing. As the identification of social networks uses passwords, it is sufficient To know the sequence of characters and can be possible to send advertising, some information on behalf of others, or to motivate recipients to any negative action, in particular to pass on the link and run the malicious code, and do other (often illegal) cases. Besides, some companies use social network to promote their own products, and the theft of administrator group password allows to steal the group itself. To obtain confidential information traditionally, phishing, dummy sites, social engineering, and more others are used. Protection against these attack methods are considered DLP-system (Data Loss Prevention) and reputation technologies that are integrated into a variety of anti-virus products.
5. URL shortening services usage. In recent years, URL shortening services allow to mask unwanted website address under the short link are especially popular. In fact, the domain redirects the visitor. Today there is an active struggle against these risks – URL shortening service began to use improved mechanisms for the detection of spam and other threats. However, for users of social networking this threat is keeping alluring messages and offers from familiar contacts that have been hacked, often lead to downloading malicious software or display unwanted web pages.
6. Using the same user names and passwords on the corporate network and external social resources. As a result, hacking profiles of social network users significantly increases the risk of penetration to corporate resources on behalf of one of the company's employees.

7. Web-attack. As social networks are web-based applications, they can be used by hackers to organize at-tacks on vulnerabilities in browsers. The tools for such attacks can be Trojan applications, fake antiviruses, social worms, which are used to spread own friends lists and other. Their main goal is to get into the information system of social network visitor and gain a foothold in it. Such traditional tools as anti-virus Software, that are able to work in real time and block the download of malicious code are used for protection.
8. Information leakage and compromising company employee's behavior. Social networks can be used to Organize leaks of important information for the company, as well as to undermine its reputation. Such attack can conduct internal employees who are dissatisfied with the leadership, or specially embedded insiders. In social networks persons often behave quite differently from the corporate communication environments, and it is possible that shocking publication and rough replicas can cause some damage to the reputation of their employers. DLP-systems and products for the analysis of publications on the Inter-net intended to protect against these threats.
9. The growth of traffic, especially viewing video sources.
10. Inducement of minors for sexual purposes (grooming).
11. Content with signs of incitement to racial, ethnic or religious hatred, propaganda of totalitarian sects.
12. Propaganda and public justification of terrorism [85].
13. Cyber humiliation and cyber bullying.
14. Promotion and distribution of drugs.

To protect from this threats, the information security services solve next problems:

- detection of information attacks: define the nodes from which the attack is made, the optimal placement of signal points;
- preventing information attacks: estimated cost of the attack on the object of attack and defense costs;
- formation and destruction of different networks: social and/or information;
- detection of intruders communities: such as terrorists, tracking malicious activity.

The following directions to counter information and psychology closure, traffic shaping;

- legal and regulatory practices – criminal responsibility of organizers and participants of the virtual com-munities;
- Internet censorship;
- monitoring and analysis of social networks.

Lets consider advantages and disadvantages of each method. The first two methods are effective in the short term, but they have some disadvantages: the lack of geographical boundaries and limitations for Pocessing and use of information beyond the scope of laws legal regulation of any government; anonymity; easily accessible variability of information in electronic form. Censorship works poorly in democratic states based on freedom of speech. Methods of monitoring and analysis of social networks are more effective in the long term, but require the Involvement of specialists in various fields of science. As virtual social groups have the ability to reorganize, the main task of monitoring and analysis of virtual communities that represent a threat to the national security of information is not their destruction, but management and control of their activities by a variety of methods.

Monitoring and analysis of social networks great number of special software to monitor and analyze the Internet environment were developed by this Time. Major functions of these systems are:

- monitoring: provides automated information search in the Internet environment, to determine and change the keywords to information search using information retrieval languages;

- analysis: automatic processing of information flows, revealing facts and events, visualization of analyti-cal data in the form of digests, charts, graphs, and other types of reports.
Monitoring refers to the process of continuous information collection from social networks in order to maintain further analysis. So, the search conduct in the scientific investigations is considered by a global search engine for social networks of commercial search engines for special applications not take into account the peculiarities of functioning discussion pages.
Some possible approaches to social networks analysisIn present time the four main approaches are allocated in social networks analysis structural, resource, regulatory and dynamic.
The network content serves as a source for a wide range of applications that focus on the extraction and data analysis. The use of network content helps to improve the quality of conclusions in social networks analysis significantly, for example in the problems of clustering and classification. Four types of network content analysis can be identified The methods of random walks are used in the analysis of general information with arbitrary data types. One of the most well-known algorithms by such methods is the reference ranking algorithm (PageRank).

This algorithm can also be used for search and classification of entities and participants in the social network, to assess the probability of visiting a particular vertex. It is natural that vertices are better located with structural point of view and have a higher weight, and, therefore, they are more important. The methods of random walk also could be useful to bring together participants in the group relative to the most influential members.

2. For sensory and flow analysis the use of data integration techniques coming from sensors and data available on the social networks. Modern mobile phones support users interaction with each other dynamically in real time, depending on their location and status. They are used to obtain information about a person or a combination of the properties of objects that are monitored.

3. Analysis of multimedia. There are many sites (Flickr, YouTube, and others) for the exchange and sharing of media: photo, video, audio. In the presence of tags or comments multimedia analysis can be reduced to the text information analysis in network.

4. Analysis of textual information. A lot of text information contains in various forms, for example, it is possible to add comments, links to posts, blogs or news articles in the social network. Sometimes, users can tag each other, which is also a form of text information in the form of links. The placement of

tags (labels or keywords) describing various objects: images, text, video is of particular interest. Under this approach, properties of tags flow, models tagging, semantic tagging, imaging tags, applications for their placement, etc. are studying. Normative approach studies the level of trust between the participants, and the rules, regulations and sanctions influencing on behavior of participants in the social network and processes of their interactions. In this case, the social roles of analysis are associated with the network edge, for example, the definition of organizers, managers and implementers of illegal actions; the relationship manager and a subordinate, friendships or family connections. Since social networking is based on the interaction between the various participants, it is natural to Assume that this interaction has influence on the participants in terms of their behavior. The issues for this Direction: how to simulate the influence on the basis of information about the participants; how to simulate Some tasks of social network analysis.

Conclusion

Communities in the network are characterized by the presence of a large number of connections between their participants and significantly fewer contacts with other participants. The simplest case is a community, where Each participant is associated with each other, and the other members cannot be included in this group, as they have no communication with members of the community (clique). Clique is the most complete subgraph of a given graph. The detection of communities is an important problem, including classification by network mem-bers, and as a result, the identification of homogeneous groups, groups of leaders or groups of critical connecection of community is actually analogue to clustering, traditional task of data mining in relation to various social networks. The approaches to the allocation of target groups by identifying communities allow their simulation, followed by the use models of information influence and manageme. The same time SNA investigates the structure of relationships between participants of various application areas by detecting the implicit links between them involving graph. The analysis of explicit and implicit communities allows to study the stability of social structures. To analyze the stability of a group structure over time the following technique is typically used. First three-dimensional matrix is constructed where rows represent the estimates of interactions of participant with all the other participants, submitted by the participants themselves. The columns are participant's own estimates of interaction. The time periods are located on the third axis. Further graph shows the structural changes of community over time. Thereafter, the techniques for dimensionality reduction are applied (for example, principal component analysis), i.e., the projection of the vertices into Euclidean space of reduced dimension to describe the relationships between the rows and columns of the matrix is considered. As a result, you can visualize the changes of network user status against the backdrop of changes in subgroups staus. Sustained projection can be clustered using a standard iterative clustering algorithms (for example, k-means) or the hierarchical ones. advantage of hierarchical methods is the possibility to represent the clustering result in a dendrogram. In this case we can obtain not only a partition of the graph into groups, but also.

REFERENCES

Aggarwal, C. 2011. Introduction to social network data analytics. Springer US. Retrieved from doi: 10.1007/978-1-4419-8462-3

Aggarwal, C., Karthik, S. 2014. Evolutionary Network Analysis: A Survey. ACM Computing Surveys, 47(1), Article 10.

Ajay Kumar Singh Kushwah, Amit Kumar Manjhvar, 2016. A Review on Link Prediction in Social Network. *International Journal of Grid and Distributed Computing*, 9(2), 43-50.

Ant colony optimization algorithms. Retrieved from https://en.wikipedia.org/wiki/Ant_colony_optimization_algorithms. Accessed 07 March 2017.

B through social network analysis: short survey Kirichenko Lyudmyla Doctor, Professor, Department of Applied Mathematics, Kharkiv National University of Radioelectronics, Ukraine Radivilova Tamara Ph.D., Associated Professor, Department of Info communication Engineering, Kharkiv National University of Radioelectronics, Ukraine Carlsson Anders Lecturer, Department of Computer Science and Engineering, Blekinge Institute of Technology, Sweden © The Authors, 2017. This article is published with open access at ARMG Publishing.

Batura, T.V. 2013. Modeli i metody analiza komp'yuternykh sotsial'nykh setey [Models and methods of analysis of computer social networks]. *International Journal Programmnye Produkty i Sistemy*, 3, 130-137.

Bonchi, F., Castillo, C., Gionis, A., Jaimes, A. 2011. Social Network Analysis and Mining for Business Applications, *ACM TIST*, 2(3), 22-58.

Bonchi, F., Castillo, C., Jaimes, A. 2011. Social network analysis and mining for business applications. *ACM Trans Intell. Syst. Technol*., 2(3), 1-37.

****