ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Intelligent Classification Techniques for Effective Prevention of Inference Attacks in Online Social Networks

[1]P. Indira Priya, [1]A.Annapurni, [2] S. Indira Gandhi [3] S. Muthurajkumar, [3]M.Vijayalakshmi,[3]A.Kannan

[1]*Department of Computer Science and Engineering, Tagore Engineering College, Chennai -127, India.*
[2] *Department of Electronics,* [3]*Department of Information Science and Technology, Anna University, Chennai – 25, India.*

ARTICLE INFO

Article history:
Received 12 November 2014
Received in revised form 26 December 2014
Accepted 29 January 2015
Available online 10 February 2015

Keywords:
Social Network Analysis, Data Mining, Inference, machine learning methods, Ensemble Classification Algorithm , Sanitization, Encryption Techniques.

ABSTRACT

Network security mechanisms for congestion processing such as intrusion detection system (IDS) is the need of the hour today to keep up in pace with the high frequency networks. Security solutions in use today are insufficiently reliable as the congestion is quite high of network frequency. For combating the present insufficiency Sachet testing schemes are suggested to be used in the beginning of network screening systems. Presently used testing methods are unreliable because the current scenario in network trafficking is not able to get used to new situations. To meet this immediate requirement with minimum overheads Adaptive testing methods are being proposed on Predicted Sachet testing. The proposed testing method increases the ability of finding denial of Service attacks in Network IDS. This typical method does not reduce the size of the information that is to be scrutinized by an IDS and further more it also retains the essential similar attributes of the network congestion. IDS can use this method to detect DOS attacks if there is a small variation in the congestion similarity.

To Cite This Article: P. Indira Priya, A.Annapurni, S. Indira Gandhi, S. Muthurajkumar, M.Vijayalakshmi, A.Kannan., Intelligent Classification Techniques for Effective Prevention of Inference Attacks in Online Social Networks. ***Aust. J. Basic & Appl. Sci.,*** *9(6): 42-46, 2015*

INTRODUCTION

Social networking deals with maintaining connections with friends and relatives. Currently, all the people use the internet to communicate with friends available on social networks. Some important examples of popular social networks include LinkedIn, Facebook, and Twitter. Moreover, people use these services a daily basis and hence these communities are not only able to maintain friendship but also helps to generate income from advertising and additional paid services if it is necessary. Among the existing social networks, Facebook and Twitter, have a broad range of users. In addition, LinkedIn has positioned itself as a professional networking site in which profiles regarding resume information and stored. However, privacy issues are providing challenge to security in social networks.

Privacy of users in a social network can be classified into two categories namely privacy copying and leakage. When a person uses a website such as "theranking.com" for voting his/her favorites, while pressing the vote option it makes the user to connect their social network account. Therefore, intentionally or unknowingly, the user are forced to login into social network accounts. By performing this login operation, the corresponding website gathers the user's public information without the knowledge of the user. Moreover, some other famous games in social networks are also involved in gathering the user's public information when they use their system.

Private data leakage is another problem in social network where the details are based on the related attributes. In order to secure the public and private information of user it is necessary to propose new algorithm that can classify the attackers users effectively. In this proposed system we provide new techniques to predict leakage of private details for people. The proposed classification and sanitization approaches helps to prevent the data theft, while allowing the recipient of this data to perform inference on non-private details by using intelligent agents.

Related Work:

Raymond *et al* (2013) proposed new social network privacy categories and the classifiers for analysis. Recently, Johnson (2009) proposed new techniques to determine the user's social interest by collecting a small sub-graph from the facebook which contains the user's gender and opposite of this

Corresponding Author: P. Indira Priya., Department of Computer Science and Engineering, Tagore Engineering College, Chennai -127, India.
E-mail: 1indrapriyap@gmail.com

gender. Predicting an individual's character or some other personal detail helps the parents to monitor their children effectively. Michael Fire *et al*. (2013) proposed a recommendation system for social network users which assists them to enhance their security and privacy. Lars Backstrom *et al* (2007) studied about attacks patterns in social networks using graph based approach. According to them the goal of an attacker is any how to identify the people. Their communication graph consists of nodes (e-mail addresses), and a directed edges to represent communications. They considered the social network data including user's and as time-stamps. In such a scenario, provision of security is difficult. Jensen *et al*. (2007) proposed new ways anonymizing techniques for social networks. Moreover, the advances in internet technology made it possible for intruders to collect data about individuals and their connections. He *et al*. (2008) proposed new techniques to infer private information using a Bayesian network based classification. Jianming He (2006) developed new techniques which helps to choose the most effective details or links that are to be removed to protect the privacy of users. Sen and Getoor (2007) compared various of link-based classification techniques using different metrics such as robustness to noise, in attributes and links. Zheleva and Getoor (2008) proposed new methods to predict the private attributes of users in big data including Facebook. Talukder *et al*. (2010) proposed a newmethod for measuring the amount of information that a user reveals to the outside world which must be removed.

Macskassy and Provost [9] proposed a new Classification algorithm for effective classification of social network. In their model, the classes are linked to entities for which the class is to be estimated. They used intelligent classification algorithms for effective social network classification. However, all these systems do not provide a ensemble based classification technique which can handle both active and passive attacks. Hence, it is necessary to propose a new combined classification algorithm which can detect attacks and prevent them effectively through accurate classification. Reda M. Elbasiony *et al* (2013) proposed a data-mining-based network intrusion detection system. Two data-mining techniques are used in misuse, anomaly and hybrid detection. Gisung Kim *et al* (2014) proposed a new hybrid intrusion detection model which is used to integrate the misuse and anomaly detection models.

1. System Architecture:

Fig.1, Shows the architecture of the system proposed in this paper. The user interface module is used to connect to the search and web to make communication with friends. The Web data are stored in the online social network data base. The data collection module is responsible to collect the Web data for analysis which can use intelligent agents for getting relevant data. The data classification module applies rules intelligent agents to classify the users. If the user is an attacker, it sends the user details to the prevention module which prevents the user from further access to all the Web resources.

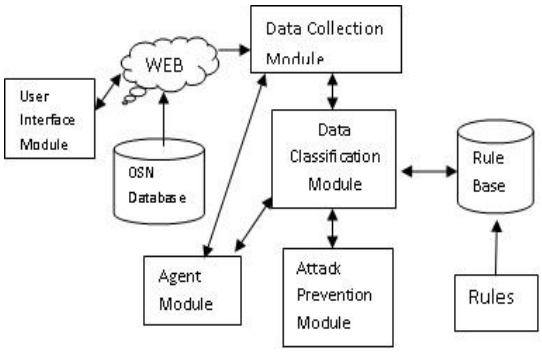

Fig. 1: System Architecture.

2. Proposed Work:

The proposed system solves the problem of private information leakage in social network. First, the social network data are collected and then converted into social graph. Three types of classifiers are applied on the graph to model inorder to identify the sensitive attributes of users and to classify them into a set of classes. Based on this classification, the system identifies what are all personal attributes to be removed or modified. The system manipulates the relationship among the users of network and decides whether the friendship links should be removed or not using rules and intelligent agents. Removing the links between users from network achieve only minimal securing when it is compared with other classification methods. Therefore, the ensemble based classification technique is effective in reducing the classification data on user details and the important data are classified as sensitive. Sanitization methods are used in this work to handle the utility of user details. For Securing Social network database, the generalization and suppress methods have been implemented and then the data has been encrypted before they are stored in the Web database.

4.1 Social Network Data Gathering:

In this experiment, the username and password details are collected for the users from social network sites, such as Facebook. Details regarding the number of logs to user accounts profiles of suspected users are monitored. Now HTML parser is used to parse the HTML files and to collect the attribute values of user profiles.

They are stored in the database. The records in database are exported into .csv format file for further classification using combined classifiers. Model the dataset file as network graph. In this work, the social network is represented using the graph data structure. This model consists of vertex and edges. Nodes are users and edges are communications.

4.2 Network Classification:

Intelligent ensemble based inference is the method proposed in this paper for classifying social network data. This ensemble has many classifiers namely local, relational, decision agent and the ensemble based inference algorithm. It uses Support Vector Machines (SVM) algorithm as a local classifier. The relational classifier is the second learning algorithm to consider edges and nodes.

Local classifiers consider only the details of the node it is classifying and its history. The relational classifiers consider only the link structure of a node and its neighbour. Ensemble inference uses both node and links and in addition in the network agent to improve the classifier accuracy. By using the SVM classifier in the first iteration, ensemble inference ensures that every node will have an initial uncertain classification. The algorithm then uses a relational classifier to reclassify nodes. The ensemble inference method also controls the length of time the algorithm runs. For ensemble inference, relaxation labeling is used since it is the best when there are few known labels. For relational classification, the link-based classifier clearly is used in this work preferable when many labels were known. The lower-variance methods (WVRN and CDRN) were applied when fewer labels were known. Finally, Relaxation Labeling is applied repeatedly to estimate class distributions on all unknowns, based on current estimates.

Steps involved in Ensemble classification:

Step 1: Assign initial label using SVM.

Step 2: In first iteration, the SVM classifies the data set into two categories namely normal and suspicious.

$V\ nb = \text{argmax}\ vj \in V\ P(vj)\ \pi\ P(ai \mid vj)$ (1)

Step 3: Validate the using voting.

Step 4: Find the combined probability by applying the score obtained for each node.

Step 5: Find the related nodes and group them.

Step 6: Apply relational classifier to each group iteratively and reclassify the labels.

Step 7: Apply relaxation labeling to form final groups.

Step 8: Perform Encryption.

Step 9: Store the data in the database.

Step 10: Store user queries in separate table.

Step 11: Analyze queries for inference attacks based on rules.

Step 12: Apply the sanitize technique frequently until it hiding or deleting the link.

Step 13: By removing both link and details reduce the private information leakage and efficiently enhanced the classifier accuracy.

RESULTS AND DISCUSSION

Instead of removing only node details or only friendship links, the system performs the removal of both node details and friendship links in order to reduce the classifier accuracy. By comparing the social network graph structure, only the lowest predictive accuracy is achieved. On the other hand when the system removes both details and links from the graph it provides more security by preventing the inference attacks happening through these links. This proposed algorithm is important since finding the solution for the inference problem is one of the most difficult problems in the security world. This is due to the fact that inference often comes from our brains making connections between certain datasets. There is no prefect way to defend against someone using their brain and own intelligence to make inferences. Preventing human intelligence from doing what it is supposed to do is a task that unlikely to be solved. Due to the abstractness of inference, many security professionals are unaware of this problem and/or dismiss it since there is no quick and easy fix. The proposed system provides best to make it harder for such events to occur by applying the intelligent classification methods and sanitization techniques.

The algorithm used to prevent inference attacks is as follows:

(i) Collect the details about nodes and links from the social network.
(ii) Find the communications made between different nodes.
(iii) Note the interest of communicators and their data access patterns.
(iv) Find the links used by them frequently.
(v) Out the nodes from the path used by anomous users.
(vi) Sanitize the path frequently and apply inference rules.
(vii) Redefine access control policies.
(viii) Use ensemble inference to make final decision.
(ix) Repeat until there are no more inference attacks.

The main reason for the improvement in classification accuracy for Ensemble Inference is due to many reasons. First, it uses only significant features. Second, it uses rules which are fired by the intelligent agents for effective decision making.

Table 1: Classifier accuracy for different user private data.

Accuracy		
Private Data	SVM	Ensemble Inference
Gender	95.21	95.89
Religion	85.10	86.2
Political Views	75.50	76.10
Character	79.60	80.2

Finally, the classification time is reduced due to due to the behavior of the different groups of users.

Fig. 2, 3, shows the accuracy analysis for gender, religion and political views and characters private data groups.

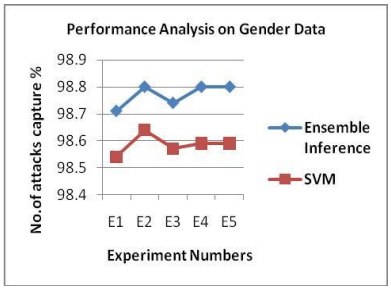

Fig. 2: Gender data analysis.

From these figures, it can be observed that the proposed Ensemble Inference classification algorithm provides higher accuracy than the existing SVM in detection all types of data groups. This is due to the fact that intelligent agents and rules applied in the proposed model enhanced the classification accuracy.

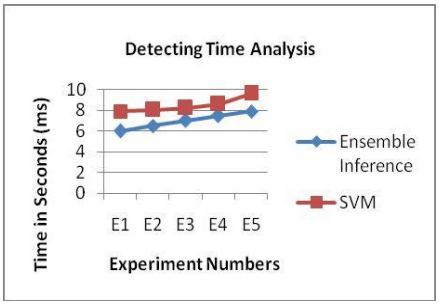

Fig. 3: Time analysis for ensemble inference.

From the experiments conducted using these classification algorithms, it has been shown that these algorithms help to predict and prevent inference attack effectively.

Conclusion And Future Work:

In this paper, an ensemble based classifier algorithm is proposed and implementation to detect and prevent private information leakage in social networks. This system considers friendship and relations. This system uses a graph model, where link deletion and addition are used to represent the cut in communication and the estimation of communication. The classification methods based on ensemble proposed in this paper enhance the classification accuracy by more than 5% to 98.9% accuracy. Further works in this direction can be the consideration of temporal constraints for effective prevention.

REFERENCES

Heatherly, R., M. Kantarcioglu, 2013. Thuraisingham, B. Preventing Private Information Inference Attacks on Social Networks. IEEE Trans. Knowledge And Data Engineering, 25(8) 1849-1861.

Johnson, C., 2009. Project Gaydar. The Boston

Globe.

Fire, M., R. Goldschmidt, Y. Elovici, 2013. Online Social Networks: Threads and Solutions. Telekom Innovation laboratories at Ben-Gurion university of the negev, The Knesset Research and Information centre, 1-32.

Backstrom, L., C. Dwork, K. Kleinberg, 2007. Wherefore Art Thou r3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography. Proceedings 16th International Conference World Wide Web (WWW '07), 181-190.

Hay, M., G. Miklau, D. Jensen, P. Weis, 2007. Srivastava, S. Anonymizing Social Networks. Technical Report 07-19, Univ. of Massachusetts Amherst.

Liu, K., E. Terzi, 2008. Towards Identity Anonymization on Graphs. Proceedings ACM SIGMOD International Conference Management of Data (SIGMOD '08), 93-106.

He, J., W. Chu, V. Liu, 2006. Inferring Privacy Information from Social Networks. Proceedings Intelligence and Security Informatics.

Sen, P. L. Getoor, 2007. Link-Based Classification. Technical Report CS-TR-4858, Univ. of Maryland.

Macskassy, S.A. F. Provost, 2007. Classification in Networked Data:A Toolkit and a Univariate Case Study. J. Machine Learning Research, 8: 935-983.

Talukder, N., M. Ouzzani, Elmagarmid, A.K., Elmeleegy, H. Yakout, 2010. Privometer: Privacy Protection in Social Networks. Proceedings IEEE 26th International Conference Data Eng. Workshops (ICDE '10), 266-269.

Zheleva, E., L. Getoor, 2008. To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private user Profiles. Technical Report CS-TR-4926, Univ. of Maryland, College Park.

Reda, M., Elbasiony, Elsayed, A. Sallam, Tarek, E. Eltobely, Mahmoud, M. Fahmy, 2013. A Hybrid Network Intrusion Detection Framework based on Random Forests and Weighted K-Means. Ain Shams Engineering Journal, 4: 753-762.

Gisung, Kim, Seungmin, Lee, Sehun, Kim, 2014. A Novel Hybrid Intrusion Detection Method Integrating Anomaly Detection with Misuse Detection. Expert Systems with Applications, 41: 1690-1700.