ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Anti-Phishing Detection System to Detect and Prevent Deceptive Phishing in SoNet Sites

Aishwarya.D.Shetty[1], Dr.Niranjan.N.Chiplunkar[2]

M.Tech Student, Dept. of CSE, NMAMIT, Nitte, India

Principal, NMAMIT, Nitte, India

ABSTRACT: Violation in Cyber Security because of phishing messages is discovered from instant messages which are sent through SoNet site(Social Networking sites) , these Violation lead to disturbance in network communication and larceny of personal identifiable information (PII) that causes plenty of issues like identity theft and cyber scam. To solve these problems, a system is developed using Ontology based Information Extraction technique (OBIE) and Association rule mining (ARM) named as Anti Phishing Detection System that discover and then predict the phishing activity by maintaining continuously updated phishing database which consists of information obtained from previous attempts to violate security; thus, prevent the phishing activity to protect the user information.

KEYWORDS: Phishing, Ontology, Association Rule Mining ,Wordnet.

I. INTRODUCTION

The phishing in the context of Social Networking Sites (SNS) is defined as a deceptive mechanism of social engineering where in phisher impersonate as a reliable sources by exchange of trustworthy messages to gain confidence of the user to extract personal identifiable information (PII) which can lead to various cyber security issues. Now-a-days SNS had gained global recognition from desktop PC's to mobile phones, where users interact with one another to share information rapidly instead of using emails. Phisher's foresee SNS as an easy target to steal user's private information from Facebook, Twitter, LinkedIn, Viadeo, WhatsUp, Google+, QQ, Wechat, IMO and Weibo. Facebook stands as the top SNS which is mostly used with many people connected and at the same time are prone to various types of cyber threats.

A. Need of Proposed system

Proposed system combat phishing by considering the rule-based and Ontology techniques to successfully detect and predict phishing attack. When messages are found suspicious, then the details of culprits are traced and reported to the victim with predictable type of threat activity.

II. RELATED WORK

Phishing website is a recent problem, nevertheless due to its huge impact on the financial and on-line retailing sectors and since preventing such attacks is an important step towards defending against website phishing attacks, there are several promising approaches to this problem and a comprehensive collection of related works.

Julei Fu and Jian Chai[3] have proposed six-element analysis method for terrorist activities based on social network. However, this method analyzed on data obtained from previous year incidents, which is in the form of 420 web pages to get information of the terrorist events incited by East Turkistan.

Michael Robertson, Yin Pan and Bo Yuan[4] explained about the social approach to detect malicious web content for Facebook with security heuristics is limited to identify malicious URL links. Recently the Facebook static messages are scanned to identify criminal's behavior. Detection of suspicious emails from static messages using decision tree induction proposed which is purely dependent on highest information entropy that identifies the messages are deceptive or non-deceptive.

DOI: 10.15680/IJIRCCE.2015. 0405111

John Resig and AnkurTeredesai[5] detect suspicious messages from the data gathered by anomaly detection, topic detection and social network analysis, which will not disclose all suspicious messages. Hence new offenders will not be traced by this system.

Mohd Mahmood Ali and Lakshmi Rajamani[6] proposed framework with an idea of instant message secure system that identifies suspicious messages that leads to illegal activities by offenders. But it does not focus on securing messages by using encryption techniques and also does not concentrate on short form messages. This paper gives various ideas about stemming algorithm and apriori algorithm.

Sharath Kumar and Sanjay Singh[7] concentrates on cluster of users in SNS who perform illegal activity based on their messages with the help of past history of the user. But in present system, offenders are smarter than investigators. They are not using same way of writings.

Farkhund Iqbal, Benjamin C.M.Fung, MouradDebbabi [8] concentrate on entity such as name of a person and tries to find which group in social networks the person belongs to. It also focused on the messages sent by the same person in the group. But it never concentrates on suspicious words given by other offenders who are presently chatting with only one person. So it will focus on old group of offenders who are already in database. It is not providing full details to crime investigators.

Mohd Mahmood Ali, KhajaMoizuddinMohd and Lakshmi Rajamani[9] proposed framework for secure instant messaging system using ontology. This paper does not focus on code words and short form chat messages. Here, ontology construction means dividing the instant messages semantically with the help of Word Net database into various topics such as murder, robbery and so on. But ontology is not updated regularly with new code words that are found using data mining techniques.

All the papers mentioned above are concentrated on security in instant messaging in the form of simple chat logs. But nowadays offenders are too smart to use code words and short forms of messages. And none of the paper focuses on proper ontology updates. Proposed work focuses on this area of messages and proper ontology based information extraction system.

The Framework proposed in this paper will try to identify the type of cyber attack using Ontology based Information Extraction technique (OBIE), Association rule mining (ARM) a data mining technique with set of pre-defined Knowledge-based rules (logical), for decision making process that are learned from domain experts and past learning experiences of suspicious dataset like GTD (Global Terrorist Database).

III. PROPOSED ALGORITHM

The Proposed system initiates the steps for capturing the phishing words that are exchanged between the users and then stores them into database for identifying phishing words using pre-defined phishing rules. This system, identifies the culprit details and report to the victim client.

In OBIE, user messages are given as input to the preprocessor component that converts the text to pure textual format. The preprocessor, uses NLP (Natural Language Processing) tools. These tools perform functions such as Part-Of-Speech (POS) tagging, sentence splitting and identifying occurrences of regular expressions.

Filtering of unnecessary words from unstructured text is done using information extraction techniques. From each instant message, stop words i.e. words that are not significant are removed, i.e. prepositions, conjunctions, articles, adjectives, adverbs, etc. Stop word examples are: from, into, in, for, while, a, an, the, that, these, those, under, over, about, although, how, what, when, who, whom, etc.The removal of these words improves the efficiency of retrieval. Fig.1 shows working of the proposed system.

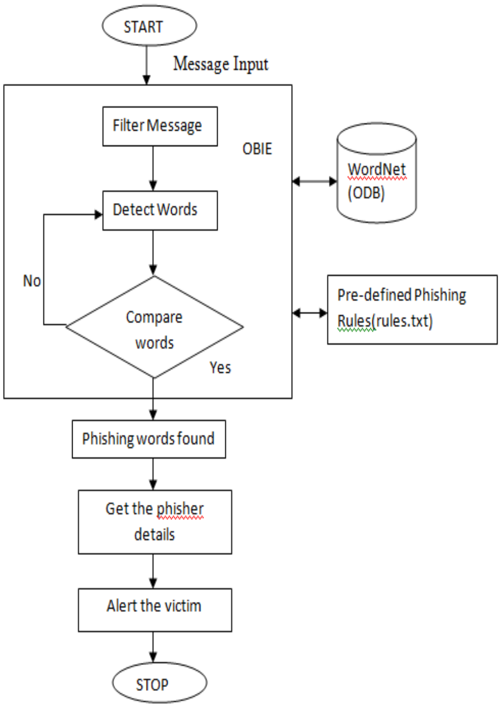

Fig. 1: Working of Proposed System

IV. PSEUDO CODE

Steps are:

1. OBIE Filter and Detect Words

- Take each message
- Apply pos tagging, sentence splitting
- Remove stopwords
- Take nouns and adjectives
- Apply stemming on nouns and adjectives
- Store the stemmed words in TPDB

2. Build Ontology Tree

Form the words in TPDB build a ontology tree by using the concept relations in SSPWDB.

3. Detect Phishing Attack

Check repetition of patterns and detect phishing attack.

4. Collect Profile

Once phishing attack detected, collect the profile of sender like his email-id, ip address, phone number, ISP-details, location.

5. Send Alert

Prepare a report with attacker profile and send to victim.

V. EXPERIMENTAL RESULTS

Fig.2 specifies the computation time taken to identify phishing words using Data mining and Wordnet Ontology. Proposed system identifies the phishing words faster than keyword based approach. This is denoted in the graph using red line.

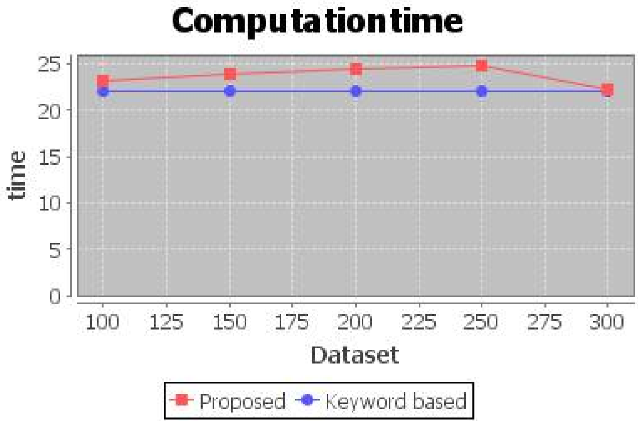

Fig.2 Comparison of Computation Time

Fig.3 compares the accuracy of proposed system with keyword based system. Red line indicates accuracy of the proposed system and blue line indicates keyword based. So, by this graph we can say that proposed system which uses Association Rule Mining and Ontology is more accurate than Keyword based system.

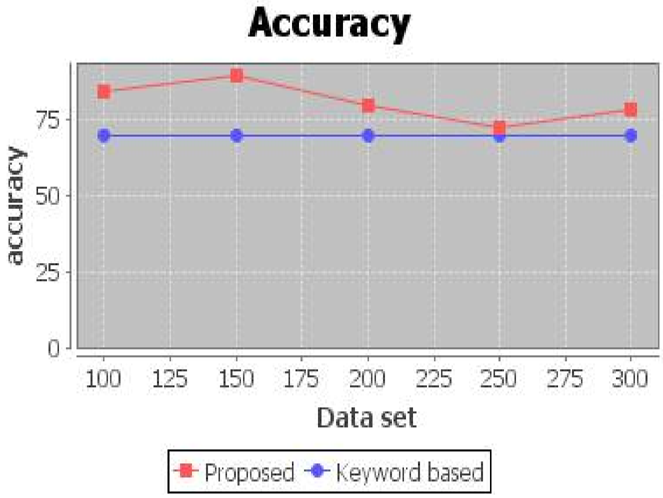

Fig.3 Comparison of Accuracy

VI. CONCLUSION AND FUTURE WORK

Now a days progress in the internet results in computer scandal. Scandalous will send distrustful messages via cell phones, IM and SoNet Sites, which is hard to follow their criminal exercises powerfully. After surveying different structural design of ‘Mobile Phones, Instant messengers and SoNet sites’, it helped to develop a new Framework, which battle phishing by using the rule-mining and Ontology methods to correctly identify and to guess phishing assault. When messages are discovered suspicious, then the details of offenders are traced and victim is alerted with predictable type of threat activity.

As a future work, Phishing messages can be identified in organizational level to build robust SoNet Sites and Detection should be done if misleading messages are sent using multimedia format.

REFERENCES

[1] Mohd Mahmood Ali, Owais A.W. Siddiqui, Mohd. Nayeemuddin and Lakshmi Rajamani, "An approach for Deceptive Phishing Detection and Prevention in Social Networking Sites Using Data Mining and WordNet Ontology", IEEE, DOI-10.1109/EESCO.2015.7253731,2015.

[2] F.J.Fu, J.Chai and S.Wangl., "Multi-factor analysis of terrorist activities based on social network", Business Intelligence and Financial Engineering (BIFE), 2012 5th International Conference, pp. 476-480, 2012.

[3] Michael Robertson, Yin Pan, and Bo Yuan, "A Social Approach to Security: Using Social Networks to help detect malicious web content," IEEE ,2010.

[4] John Resig and AnkurTeredesai, "A Framework for Mining Instant Messaging Services" , proceedings of the 2004 SIAM Lake Buena Vista - ejohn.org Date: 2011-04-19.

[5] Mohd Mahmood Ali and Lakshmi Rajamani," APD: ARM Deceptive Phishing Detector System Phishing Detection in Instant Messengers Using Data Mining Approach", Springer-Verlag Berlin Heidelberg 2012, Part I, CCIS 269, pp. 490–502, 2012.

[6] Sharath Kumar and Sanjay Singh, "Detection of user cluster with suspicious activity in online social networking sites" , IEEE, pp. 220-225, 2013.

[7] Farkhund Iqbal, Benjamin C.M.Fung, MouradDebbabi, "Mining Criminal Networks from Chat Log" , IEEE/WIC/ACM International Conference, pp.332-337, 2012.

[8] Mohd Mahmood Ali, Khaja MoizuddinMohd and Lakshmi Rajamani, "Framework for surveillance of Instant Messages in Instant Messengers and Social networking sites using Data Mining and Ontology" , IEEE Students' Technology Symposium, pp.297-302, 2014.

BIOGRAPHY

Aishwarya.D.Shetty is M.Tech Student in the Department of Computer Science & Engineering, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Nitte, India.

Dr. Niranjan N. Chiplunkar is Principal in Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Nitte, India.