International Journal of Information Technology and Management

Vol. VIII, Issue No. XII, May-2015, ISSN 2249-4510

DETECTING E-BANKING PHISHING USING ASSOCIATIVE CLASSIFICATION

www.ignited.in

Detecting E-Banking Phishing Using Associative Classification

Akshay Pandey[1]* Dr. M. K. Sharma[2]

[1]B. Tech. (CSE) R. D. Engineering College, Ghaziabad

[2]Associate Professor, Amrapali Institute, Haldwani

Abstract – There are number of clients who buy items on the web and make installment through e-saving money. There are e-managing an account sites who request that client give touchy information, for example, username, secret word or charge card subtle elements and so on regularly for noxious reasons. This kind of e-saving money sites is known as phishing site. With a specific end goal to recognize and foresee e-managing an account phishing site. We proposed a canny, adaptable and compelling framework that depends on utilizing grouping Data mining calculation. We executed order calculation and strategies to extricate the phishing informational collections criteria to arrange their authenticity. The e-saving money phishing site can be recognized in light of some critical qualities like URL and Domain Identity, and security and encryption criteria in the last phishing discovery rate. When client makes exchange through online when he makes installment through e-keeping money site our framework will utilize information mining calculation to distinguish whether the e-managing an account site is phishing site or not. This application can be utilized by numerous E-trade endeavors keeping in mind the end goal to influence the entire exchange to process secure. Information mining calculation utilized as a part of this framework gives better execution when contrasted with other conventional groupings calculations. With the assistance of this framework client can likewise buy items online with no dithering.

KEYWORDS: Banking Phishing, Associative Classification, Websites.

-------------- X --------------

INTRODUCTION

Phishing websites is the process of enticing people to visit fraudulent e-banking websites and persuading them to enter identity information such as usernames and passwords. The information is then used to impersonate victims in order to empty their bank accounts, run fraudulent auctions, launder money, and so on. The word phishing from the phrase “website phishing” is a variation on the word “fishing”. The idea is that bait is thrown out with the hopes that a user will grab it and bite into it just like the fish. The motivation behind this study is to create a resilient and effective method that uses Data Mining algorithms and tools to detect e-banking phishing websites in an Artificial Intelligent technique. Associative and classification algorithms can be very useful in predicting Phishing websites. It can give us answers about what are the most important e-banking phishing website characteristics and indicators and how they relate with each other.

There are number of users who purchase products online and make payment through e- banking. There are e- banking websites who ask user to provide sensitive data such as username, password or credit card details etc often for malicious reasons. This type of e-banking websites is known as phishing website. In order to detect and predict e-banking phishing website. We proposed an intelligent, flexible and effective system that is based on using classification Data mining algorithm. We implemented classification algorithm and techniques to extract the phishing data sets criteria to classify their legitimacy. The e-banking phishing website can be detected based on some important characteristics like URL and Domain Identity, and security and encryption criteria in the final phishing detection rate. Once user makes transaction through online when he makes payment through e-banking website our system will use data mining algorithm to detect whether the e-banking website is phishing website or not. This application can be used by many E-commerce enterprises in order to make the whole transaction process secure. Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms. With the help of this system user can also purchase products online without any hesitation.

Advantages

- This system can be used by many E-commerce Websites in order to have good customer relationship.
- User can make online payment securely.
- Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms.
- With the help of this system user can also purchase products online without any hesitation.

Disadvantages

- If Internet connection fails, this system won't work.
- All e-banking websites related data will be stored in one place.

FEATURES

A. Load Balancing

Since the system will be available only the admin logs in the amount of load on server will be limited to time period of admin access.

B. Easy Accessibility:

Records can be easily accessed and store and other information respectively.

C. User Friendly:

The Website will be giving a very user friendly approach for all users.

D. Efficient and Reliable:

Maintaining the all secured and database on the server which will be accessible according the use requirement

without any maintenance cost will be a very efficient as compared to storing all the customer data on the spreadsheet or in physically in the record books.

E. Easy Maintenance:

E-Banking Phishing Website is design as easy way. So maintenance is also easy

HOW IT WORKS?

- This system uses effective classification data mining algorithm to detect the e-banking phishing websites.
- The e-banking phishing website can be detected based on some important characteristics like URL and Domain Identity, and security and encryption criteria in the final phishing detection rate.
- This application can be used by many E-commerce enterprises in order to make the whole transaction process secure.
- Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms.
- By using this system user can make purchase products online securely.

Case Study: Website Phishing Experiment

We engineered a website for phishing practice and study. The website was an exact replica of the original Jordan Ahli Bank website www.ahlionline.com.jo designed to trap users and induce them by targeted phishing email to submit their credentials (username and password).

The specimen was inclusive of our colleagues at Jordan Ahli Bank after attaining the necessary authorizations from our management. We deliberately put lots of known phishing features and factors when creating the faked website in order to measure the user's awareness of these kinds of risk. For example, using IP address instead of domain name, http instead of https, poor design, spelling errors, absence of SSL padlock icon and phony security certificate. We targeted 120 employees with our deceiving phishing email, informing them that their e-banking accounts are at the risk of being hacked and requested them to log into their account through fake link attached to our email using their usual customer ID and password to verify their balance and then log out normally. As shown in table 2, The website successfully attracted 52 out of the 120-targeted employees representing 44%, who interacted positively by following the deceiving instructions and submitting their actual credentials (customer ID, Password).

Surprisingly IT department employees and IT auditors constituted 8 out of the 120 victims representing 7%, which shocked me, since we expected them to be more alert than others. From other departments 44 employees of the 120-targeted employee's victims representing 37%, fell into the trap and submitted their credentials without any hesitation. The remaining 68 out of 120 representing 56% were divided as follows: 28 employees supplied incorrect info, which seems to indicate a wary curiosity representing 23%; and 40 employees, received the email, but did not respond at all representing 33%. The results clearly indicate that target phishing factor is extremely dangerous since almost half of the employees who responded were victimized; particularly, trained employees such as those of IT Department and IT Auditors. Increasing the awareness of all users of e-banking regarding this risk factor is highly recommended.

PHISHING WEBSITE METHODOLOGY

1. Data Mining Techniques

We utilized data mining classification and association rule approaches in our new e-banking phishing website detection model to find the most important phishing features and significant patterns of phishing characteristic or factors in the e-banking phishing website archive data. Particularly, we used a number of different existing data mining association and classification techniques including JRip , PART, PRISM and C4.5, CBA, MCAR algorithms to learn and to compare the relationships of the different phishing classification features and rules. The experiments of C4.5, RIPPER, PART and PRISM algorithms were conducted using the WEKA software system (FDIC, 2004), which is an open java source code for the data mining community that includes implementations of different methods for several different data mining tasks such as classification, association rule and regression. CBA and MCAR experiments were conducted using an implementation version provided by the authors of. We have chosen these algorithms based on the different strategies they use to generate the rules and since their learnt classifiers are easily understood by human.

We used two web access archives, one from APWG archive (Gartner, 2006) and one from Phish tank archive. We managed to extract the whole 27 phishing security features and indicators and clustered them to its 6 corresponding criteria.

MINING E-BANKING PHISHING CHALLENGES

The age of the dataset is the most significant problem, which is particularly relevant with the phishing corpus. E-banking Phishing websites are short-lived, often lasting only in the order of 48 hours. Some of our features can therefore not be extracted from older websites, making our tests difficult. The average phishing site stays live for approximately 2.25 days (Fadi, et. al., 2005). Furthermore, the process of transforming the original e-banking phishing website archives into record feature row data sets is not without error. It requires the use of heuristics at several steps. Thus high accuracy from the data mining algorithms cannot be expected. However, the evidence supporting the golden nuggets comes from a number different algorithms and feature sets and we believe it is compelling (Weka, 2006)

CONCLUSION

Phishing has turning into a genuine system security issue, making finical loss of billions of dollars the two buyers furthermore, internet business organizations. Furthermore, maybe more essentially, phishing has made web based business questioned what's more, less appealing to typical customers. In this paper, we have considered the attributes of the hyperlinks that were implanted in phishing messages.

We at that point outlined a hostile to phishing calculation, in light of the inferred qualities. Since E Banking Phishing Website is trademark based, it cannot just distinguish known assaults, yet additionally is powerful to the obscure ones. We have actualized Link Guard for Windows XP. Our trial demonstrated that Link Guard is light-weighted what's more, can identify up to 96% obscure phishing assaults in real time. We trust that Link Guard is not just valuable for identifying phishing assaults, yet in addition can shield clients from vindictive or spontaneous connections in Web pages and Instant messages.

E-banking phishing website model based on classification data mining showed the significance importance of the phishing website two criteria’s (URL & Domain Identity) and (Security & Encryption) in the final phishing detection rate, and also showed the insignificant trivial influence of some other criteria like ‘Page Style & content’ and ‘Social Human Factor’ in the final phishing rate. The rules generated from the associative classification model showed the correlation and relationship between some of their characteristics which can help us in building phishing website detection system.

REFERENCES

"Putting an End to Account-Hijacking Identity Theft" 2004.

"Weka -Data Mining with Open Source Machine Learning Software in Java" in New Zealand EN:WEKA - University of Waikato 2006.

B. Adida S. Hohenberger R. Rivest (2005). "Lightweight Encryption for Email" <em>USENIX Steps to Reducing Unwanted Traffic on the Internet (SRUTI)</em>.

Bing Liu Wynne Hsu Yiming Ma "Integrating Classification and Association Rule Mining" <em>Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98 Plenary Presentation)</em> 1998.

Bing Liu, Wynne Hsu, Yiming Ma (1998). "Integrating Classification and Association Rule Mining." Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98, Plenary Presentation), New York, USA.

C. Jackson D. Simon D. Tan A. Barth (2007). "An evaluation of extended validation and picture-in-picture phishing attacks" <em>In Proceedings of the 2007 Usable Security</em>.

FDIC (2004). Tech. Rep., “Putting an end to account-hijacking identity theft”, [Online]. Available: http://www.fdic.gov/consumers/idtheftstudy/identity theft.pdf.

Gartner Says Number of Phishing E-Mails Sent to U.S. Adults Nearly Doubles in Just Two Years</em> November 2006.

I.H. Witten and E. Frank (2005). “Data Mining: Practical machine learning tools and techniques”, 2nd Edition, Morgan Kaufmann, San Francisco, CA.

I.H. Witten E. Frank (2005). "Data Mining: Practical machine learning tools and techniques" in San Francisco CA: Morgan Kaufmann 2005.

Ian Fette, Norman Sadeh and Anthony Tomasic (2006). “Learning to Detect Phishing Emails”, Institute for Software Research International, CMU-ISRI-06-112.

J. Cendrowska "PRISM: An algorithm for inducing modular rule" <em>International Journal of Man-Machine Studies</em> vol. 27 no. 4 pp. 349-370 1987.

J. Cendrowska (1987). "PRISM: An algorithm for inducing modular rule", International Journal of Man-Machine Studies (1987), Vol.27, No.4, pp. 349-370.

J.R. Quinlan "Improved use of continuous attributes in c4.5" <em>Journal of Artificial Intelligence Research</em> vol. 4 pp. 77-90 1996.

Kantardzic and Mehmed (2003). “Data Mining: Concepts, Models, Methods, and Algorithms.”, John Wiley & Sons. ISBN 0471228524. OCLC 50055336.

Kantardzic Mehmed (2003). "Data Mining: Concepts Models Methods and Algorithms" in John Wiley & Sons 2003 ISBN 0471228524.

R. Dhamija J.D. Tygar (2005). "The Battle against Phishing: Dynamic Security Skins" <em>Proc. Symp. Usable Privacy and Security</em> 2005.

Sebastian Misch (2006). “Content Negotiatio in Internet Mail”, Diploma Thesis, University of Applied Sciences Cologne, Mat.No.:7042524.

T. Fadi C. Peter Y. Peng (2005). "MCAR: Multi-class Classification based on Association Rule" <em>IEEE International Conference on Computer Systems and Applications</em> pp. 127-133 2005.

T. Moore R. Clayton (2007). "An empirical analysis of the current state of phishing attack and defence" <em>In Proceedings of the Workshop on the Economics of Information Security (WEIS2007).

T. Sharif "Phishing Filter in IE7" (2006).

U.M. Fayyad (1998). "Mining Databases: Towards Algorithms for Discovery" <em>Data Eng. Bull.</em> vol. 21 no. 1 pp. 39-48.

U.M. Fayyad (1998). “Mining Databases: Towards Algorithms for Discovery,” Data Eng. Bull., vol. 21, no. 1, pp. 39-48.

WEKA - University of Waikato, New Zealand, EN, 2006: “Weka - Data Mining with Open Source Machine Learning Software in Java”; http://www.cs.waikato.ac.nz/ ml/weka (2006/01/31).

Corresponding Author

Akshay Pandey*

B. Tech. (CSE) R. D. Engineering College, Ghaziabad

E-Mail –