Framework for Detection of Phishing Social Engineering Attacks

Balaji S [1] Punitha K [2]

[1 3] Dhanraj Baid Jain College (Autonomous),Chennai,India

[1] *newsrisaibalaji@gmail.com*, [2] *punithacs1996@gmail.com*

Abstract

Social enginnering refers to the act in which hackers gain unauthorized access into the computers by non technical means.Social Engineeringis also termed as the group of activities that gain access in to users credential,confidential and personal information illegally. Phishing is the most common type of social engineering attack The type of illicit endeavor that adventures both social building and specialized misdirection to obtain touchy secret information in emails is Phishing.It includes spam messages camouflaged as authentic with a subject or message intended to trap the casualities.This paper presents the framework for detecting attacks on Social enginnering with specific to phishing emails.The different parts of the email is extracted using feature selection process. The classifier approach is applied on data after feature selection process.The features extracted are classified using Bayes Net,J48 and Naive Bayes Classification algorithms .Training,testing and validation applied with Ten-fold cross-validation for experimentation. The performance of the technique is observed with the metrics .The method J48 Classifier provides the best accuracy to detect Phishing on the dataset.

Keywords:-Social engineering, Phishing ,Classifier, Bayes net, Navie Bayes and J48

1.INTRODUCTION

Human psychology is the principle behind most of the Social Enginnering attacks.According to Gartner research group, Social engineering(SE) is defined as the breach of the security systems with the manipulation of people instead of machines[1].

According to tripwire work on security in SE, Phishing attacks are the most common type of attacks leveraging social engineering techniques. Attackers use emails, social media and instant messaging, and SMS to trick victims into providing sensitive information or visiting malicious URL in the attempt to compromise their systems[2].

At a high level, most phishing scams endeavor to accomplish three things:

a. Obtain personal information such as names, addresses and Social Security Numbers.

b. Use shortened or misleading links that redirect users to suspicious websites that host phishing landing pages.

c. Incorporate threats, fear and a sense of urgency in an attempt to manipulate the user into responding quickly[3].

Between March 1 and March 23, 2020 Barracuda researchers detected 467,825 spear phishing email attacks. 9,116 of those detections were related to COVID-19, representing about 2 per cent of attacks.In comparison, a total of 1,188 coronavirus-related email attacks were detected in February, and just 137 were detected in January.Of the coronavirus-related attacks detected by Barracuda Sentinel through March 23, 54 per cent were scams, 34 per cent were brand impersonation attacks, 11 per cent were blackmail, and 1 per cent are business email compromise..

In the month of April 2020, Google reported that it was detecting about 18 million pandemic-themed malware or phishing messages per day and some 240 million Covid-linked spam messages. "Hackers frequently look at crises as an opportunity, and Covid-19 is no different," Shane Huntley of Google Threat Analysis Group said in a blog post.Hence there is a growing need to protect sensitive information in our email by protecting with the strategy.

2.LITERATURE SUPPORT

The works of Tsinganos, N.,, Sakellariou, G., Fouliras, P., Mavridis, I. covers an extensive overview of the existing systems and provide a comprehensive recognition of subsys-tems for their detection architecture; in influence, deception, personality, speech act and past experiences[4].

The work of Bhakta, R, Harris, I.G. provides a semantic based approach of diaglogues to detect social engineering attacks[10].

The work of Heartfield, R., Loukas G. show that the human factor is the weakest link in social engineering attacks and based on a human study the prove that[11].

Kumar et al. work utilized TANAGRA information mining apparatus on a tested spam dataset to assess the productivity of the messages classifier where a few calculations were applied on that informational index. Toward the end, the highlights determinations by Fisher spam channels and separating accomplished better characterizations. After fisher sifting has accomplished over 99% precision in recognizing spam, the tree Yuancheng et al. proposed an AI approach for the discovery of phishing site pages.The work accentuations on highlights of the site page, web picture and report item model to advance the highlights that are separated from the site page they have utilized quantum motivated transformative calculation[16].

Liu P et al. work attempted to locate a viable answer for sifting spam messages in their work. Theethodology for the investigating of the utilized content is about the E-mail as a watchword just to execute multiplex word handling. The investigation led, 4327messages in the CSDMC2010 SPAM preparing informational collection were assessed. The outcome of the demonstrated models shows an exactness of 92.8 per cent[17].

3.CLASSIFIER BASED APPROACH

Classification technique is proposed to differentiate spam email from ham email.

First step in the process is Data acquisition which is the email dataset .The email dataset consists of spam and ham email items.

After collecting emails ,it has to be converted into Electronic mail format.

In the stage of preprocessing,the process of Tokenization is applied to filter the words in EML format of emails.

characterization calculation was applied on pertinent highlights[12].

Chhikara et al. presented brief data about phishing, its assaults, steps that clients can take to defend their secret data. The paper additionally demonstrates an overview led by net craft on phishing[13].

M. Edwards. etc.,work analyzed how Human instinct can be a security hole that can be used by hackers to expose psychological and cognitive aspect to manipulate and obtain important information[14].

Zhijun Yan etl.,work highlighted phishing attempt on chinese Online shopping sites.URL and internet highlights are utilized for the location of Phishing[15].

The maximum efficiency is obtained after elimination of unnecessary words in email.The framework proposed for the detection of Phishing is given in figure 3.1.

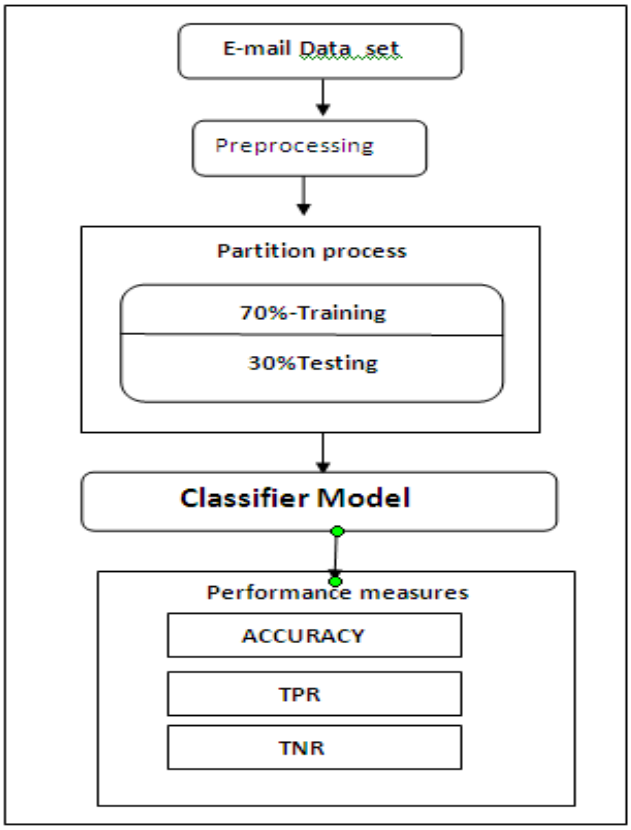

Figure 3.1 Framework for phishing detection using Classifier approach

The process of feature selection deals with extraction of feature attributes needed for processing.This has to be done with all the emails in the dataset. This stage extracts feature elements such as body of the email, header of the email and URL. to address field, from address filed, cc and bcc fields.After feature selection ,resultant file is converted into ARFF format .The ARFF file format is taken for applying classification procedure.The classifier procedure applied by dividing the dataset into training dataset and testing data.The rule based classification is applied on features selected. This Ordered model shows the email as spam one or ham one based on precision values.

4.PERFORMANCE OBSERVATION

EDRM dataset is used for validation purpose.The dataset of 1000 records is taken after preprocessing.The performance of the dataset is validated with three classifier techniques as Navie Byes ,J48 and Bayes Net ans J48.

The statistical measures such as Classifier accuracy,True Positive rate,True Negative rate,F-measure and Precision are taken in to account. J.Han and M.Kamber (2006) defined these measures in the form of confusion matrix. The components of confusion matrix are True Positive,True Negative ,False Positive and False Negative.

The performance matrices used to measure the effectivess and strength of the proposed model as follows

The percentage of tuples correctly classified is referred as accuracy of the classifer.

The proportion of Positive tuple that are correctly classified as TPR(True Positive Rate).

The proportion of Negative tuple that are correctly classified as TNR(True Negative Rate).

The feature set is applied in such a manner that it identifies the irrelevant feature.This measure plays a vital role in the process of model development based on classification technique.The model is then applied for phishing email classification.

EXPERIMENT	BAYESNET (%)	J48 (%)	NAVIE-BAYES (%)
TPR	96.15	96.7	96.65
TNR	98.5	99.41	98.$65
FPR	0.4	0.53	0.25
FNR	2.75	3.25	1.35
ACCURACY	97.32	98.1	97.65
PRECISION	98.49	99.46	98.64
SENSITIVITY	96.15	96.75	96.65
F-MEASURE	97.31	98.08	97.64

Table 4.1-Confusion matrix showing

comparision of Classification technique results

The experiment is carried out using Waikato environment for Knowledge analysis data mining software using Intel i8 machine with windows-10 environment. The accuracy of the models with 70-30% training -testing data.TPR and TNR are calculated to check the robustness of the developed models using confusion matrix.

The results of the confusion matrix is presented in Table 4.1.

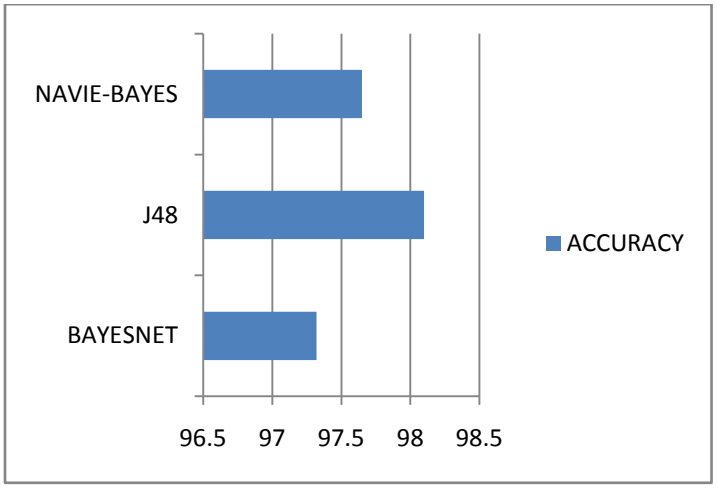

Figure 4.1 –Graphics describition of Accuracy of Models trained

False negative: The system fails to recognise a phishing attack.

• False positive: The system recognises a ham website as a phishing website.

The Experimental results shows accuracy of the bayesnet,J48 and Navie-bayes classifiers as 0.9732,0.981 and 0.9765.

The results of the model trained is presented in figure 4.1. The trained model with J48 classifier shows the best results compared to Bayes net and Naïve Bayes classifer.

5.CONCLUSION

The work in this paper focuses on identification of genuine email by separating spam messages with the intention to protect

Precision rates are 0.985,0.995,0.987 for bayesnet,J48 and Navie-bayes classifiers respectively.

The precision for a class is the number of true positives (i.e. the number of items correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class (i.e. the sum of true positives and false positives)[2].

vital data . The sensitive information of legitimate users is gained by Phishing attack.Phising email data with features is reduced upto 15 features using proposed model.The proposed model finds

reduced upto 15 features using proposed model. The proposed model finds J.48 Classifier with increased accuracy of 98.10 with only 10 features.This work can be extended by future scholars using other feature selection methods and deep learning.

6.REFERENCES

[1] Sawa, Y., Bhakta, R., Harris, I. G., Hadnagy, C. (2016) "Detection of social engineering attacks through natural language processing of conversations." In 2016 IEEE Tenth Inter¬national Conference on Semantic Computing (ICSC), pp. 262-265. IEEE, 2016.

[2]https://www.tripwire.com/state-of-security/security-awareness/5-social-engineering-attacks-to-watch-out-for/

[3] Mouton, F., Nottingham, A., Leenen, L., Venter, H. S. (2018) "Finite state machine for the social engineering attack detection model: SEADM." SAIEE Africa Research Journal 109, no. 2 : 133-148,2018.

[4] Tsinganos, N.,, Sakellariou, G., Fouliras, P., Mavridis, I. (2018) "Towards an Automated Recognition System for Chat-based Social Engineering Attacks in Enterprise Environ¬ments." In Proceedings of the 13th International Conference on Availability, Reliability and Security, p. 53. ACM, 2018.

[5] Cialdini, R. B. "The science of persuasion." Scientific American 284, no. 2 : 76-81,2001.

[6] Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D. (2014) "The Stanford CoreNLP natural language processing toolkit." In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, pp. 55-60. 2014.

[7] Hoeschele, M., Rogers, M. (2005) "Detecting social engineering." IFIP International Con¬ference on Digital Forensics. Springer, Boston, MA, 2005.

[8] Hoeschele, M. "CERIAS Tech Report 2006-15 DETECTING SOCIAL ENGINEERING." (2006).

[9] Abid, J., Asif, K., Ghulam, Z., Nazir, M.K., Alam, S. M., Ashraf, R. "MPMPA: A Mitigation and Prevention Model for Social Engineering Based Phishing attacks on Face-book." In 2018 IEEE International Conference on Big Data (Big Data), pp. 5040-5048. IEEE, 2018.

[10] Bhakta, R, Harris, I.G. "Semantic analysis of dialogs to detect social engineering attacks." Proceedings of the 2015 IEEE 9th International

Conference on Semantic Compu¬ting (IEEE ICSC 2015). IEEE, 2015.

[11] Heartfield, R., Loukas, G. "Detecting semantic social engineering attacks with the weakest link: Implementation and empirical evaluation of a human-as-a-security-sensor framework." Computers & Security 76 : 101-127,2018.

[12] Bezuidenhout, M., Mouton, F., Venter, H.S. "Social engineering attack detection model: Seadm."Information Security for South Africa. IEEE, 2010.

[13] Mouton, F., Leenen, L., Venter, H. S. "Social engineering attack detection model: Seadmv2." 2015 International Conference on Cyberworlds (CW),2015.

[14] Nicholson, J., Coventry, L., Briggs, P. "Can we fight social engineering attacks by social means? Assessing social salience as a means to improve phish detection." Thirteenth Symposium on Usable Privacy and Security {SOUPS} 2017.

[15] Krombholz, K., Hobel, H., Huber, M., Weippl, E. . "Advanced social engineering attacks." Journal of Information Security and applications 22 : 113-122,2015.

[16] Mouton, F., Leenen, Λ., Venter, H. S. "Social engineering attack examples, tem¬plates and scenarios." Computers & Security 59 : 186-209,2016.

[17] Thanh, N. N., Nguyen, V. D., Hwang, D. "An influence analysis of the number of members on the quality of knowledge in a collective." Journal of Intelligent & Fuzzy Sys¬tems 32.2 pp. 1217-1228,2017.

[18] Javad, S., Moallem, P., Koofigar, H. "Training echo state neural network using harmony search algorithm." Int. J. Artif. Intell 15.1 (2017): 163-179,2017.

[19] Precup, R. E., and Radu-Codrut D. Nature-Inspired Optimization Algorithms for Fuzzy Controlled Servo Systems. Butterworth-Heinemann,2019.

[20] Bilal H., Abedalguni. "Island-based Cuckoo Search with Highly Disruptive Poly¬nomial Mutation". Int. J. Artif. Intell 17.1 : 57-82,2019.

[21] Marchal, S., Franc Ãßois, J. State, R. , Engel, T. , "Phishstorm: Detecting phishing with streaming analytics," IEEE Transactions on Network and Service Management, vol. 11, no. 4, pp. 458-471,2014.

[22]M. Edwards, R. Larson, B. Green, A. Rashid, and A.Baron, "Panning for gold : Automatically analysing online social engineering attack surfaces," Comput. Secur., vol.69, pp. 18–34, 2017.

[23]Peng, T., Harris, I., Sawa, Y. (2018) "Detecting phishing attacks using natural language processing and machine learning." 2018 IEEE 12th International Conference on Semantic Computing (ICSC). IEEE, 2018.

[24] Zhijun Yan, Su Liu, Tianmei Wang, Baowen Sun, Hansi Jiang, Hangzhou Yang, "A Genetic Algorithm Based Model for Chinese Phishing E-commerce Websites Detection in HCI in Business", Government, and Organizations: eCommerce and Innovation, Springer International Publishing, 2016.

[25] Yuancheng Li, Rui Xiao, Jingang Feng, Liujun Zhao, "A semi-supervised learning approach for detection of phishing webpages", Optik-International Journal for Light and Electron Optics, vol.124, Issue 23, December 2013.

[26] P. Liu and T. S. Moh, "Content Based Spam E-mail Filtering", 2016International Conference on Collaboration Technologies and Systems(CTS), Orlando, FL, pp. 218-224, 2016.

[27] Jakobsson, Markus. "Displaying and counteracting phishing assaults." In Financial Cryptography, vol. 5. 2005.

[28] Chhikara , Jyoti, RituDahiya, Neha Garg , and Monika Rani. "Phishing and hostile to phishing methods: Case ponder." International Journal of Advanced Research in Computer Science and Software Engineering 3, no. 5, 2013.

[29] Kumar, R. K., Poonkuzhali, G., and Sudhakar, P. Similar investigation on email spam classifier utilizing information mining procedures. In Proceedings of the International Multi Conference of Engineers and Computer Scientist Vol. 1, pp. 14-16,march-2012.