ANALYZING AND PREDICTING SPEAR-PHISHING USING MACHINE LEARNING METHODS

Aadil Gani Ganie
PhD Student, Institute of Information Sciences, University of Miskolc
3515 Miskolc, Miskolc-Egyetemváros, e-mail: aadilganiganie@gmail.com

Samad Dadvandipour
Associate Professor, Institute of Information Sciences, University of Miskolc
3515 Miskolc, Miskolc-Egyetemváros, e-mail: aitsamad@uni-miskolc.hu

Abstract

Phishing implies misdirecting the client by masking himself/herself as a reliable individual, to take the Critical material, for example, bank account number, credit card numbers, and so on; one of the noticeably utilized Phishing these days is spear phishing, and it is one of the effective phishing assaults given its social, mental boundaries. In this paper, we will mitigate the impact of spear phishing by utilizing the multi-layer approach. The multi-layer approach is the best method of managing the web interruption, as the intruder needs to experience shift levels. Practically all the scientists are dealing with the content of the email; however, this paper picks a novel method to counter the phishing messages by utilizing both the attachment and content of an email. We applied sentimental analysis on emails, including both content of the email and the attachment, to check whether they are spam or not using SVM classifier and Randomforest Classifier; the former showed 96 percent accuracy while, as later offers 97.66 percent accuracy. SVM showed false-positive 0 percent and false-negative 4 percent, while RandomForest showed 0 percent false-positive and 2.33 percent false-negative ratios. We also performed topic modeling using LDA(Latent Dirichlet Allocation)) from Gensim package to get the dominant topics in our dataset. We visualized the results of our topic model using pyLDvis. The perplexity and coherence score of our topic model is -12.897670565510511 and 0.44700287476452394, respectively.

Keywords: *Phishing, Machine learning, Hashing, Email Phishing, Topic modeling, RandomForest*

1. Introduction

Spear phishing is a scam of email or electronic communications aimed at a particular person, organization, or company. The main difference between spear phishing and another type of Phishing is in the former, the intruder attacks the person, not the system, by using psychological, social engineering, and reverse social engineering techniques. While often meant for malicious reasons to steal information, cybercriminals may even aim to install malware on the computer of a targeted consumer. Security flaws are some type of software or hardware failure. Cybercriminals seek to steal it after acquiring knowledge of a flaw. An exploit is a concept used to characterize a program written to take advantage of the known vulnerabilities. An attack is called the act of using an exploit against a bug. The purpose of the attack is to obtain access to a device, to the information it hosts, or to a particular resource.

A major flaw, called SYNful Knock, was found in Cisco IOS in 2015. This security breach enabled hackers to gain control of enterprise-grade routers, such as the legacy Cisco 1841, 2811, and 3825 routers. The hackers could then control all network contact and corrupt other network components. Spear phishing is mass-mail. According to industry statistics, spear phishing has a success rate of 19% compared to just 5% for standard Phishing and less than 1% for spam [1]. Mobile phones are more prone to phishing attacks because of the following reasons. Since mobile phones are always with us compared to desktop, .we regularly check the emails on our phones. We will more likely click on a fresh attack that has not been discovered by IT security personnel or Security Company. Ponemon Institute has found that 29% of data breaches were linked to mobile phone usage.

The average cost of a cyber-attack in the UK was 1.9m. In 2010 HSBC was fined 3.2m for losing confidential customer information. Antivirus vendors estimate that around 60,000 new pieces of malware are created daily. It takes 11.6 days to recognize new malware [2]. We are expiring digital warfare nowadays, significant economies of the world are more exhausting more resources on cybersecurity. Spear phishing is a personal attack; we can say that it is a psychological attack .psychologists have demonstrated that personality traits are a stronger predictor than an economic factor of certain risk-based decisions [3]. Spear-phishing is a far more focused type of Phishing. Whereas generic Phishing includes spam email sent to any arbitrary email address, spear-phishing emails are designed to look to be from someone the receiver knows and trusts.—such as a colleague, business manager, or human resources department—and can include a subject line or content that is specifically tailored to the victim's known interests or industry. The RSA Security firm was attacked in 2011 is one of the most famous examples of a spear-phishing attack that succeeded despite its suspicious nature. The hackers sent two separate targeted phishing emails; for useful targets, the hackers would search their Facebook, LinkedIn, and other social networking profiles to gain knowledge about the target and choose the names of trustworthy friends in their circle to embody or the subject of concern to attract the victim and gain their confidence. Training the user is the most effective way to deal against this social engineering and reverse social engineering [17-21]

2. Results and discussion

Gartner suggests that 3.6 million users in the USA lose money each year due to phishing emails. He also suggests that this problem leads to losing 3.2 billion US dollars per year in the USA alone; in 2006, 2.3 million people fell victim to phishing emails, and in 2007, the number rose to 6 million people [1]. In this model, the features of phishing emails are extracted based on a weighing of message content and message header and select features according to priority. In this model, every email is passed as a text file to identify each header element to distinguish them from the body [4]. According to a Message lab intelligence report, spam now compromises approximately 88% of email traffic [5]. A surveillance report of October 2008 estimated that for every 5 00000 phishing emails sent to people, 2500 people were successfully seamed [5]. Understandable, authoritative [22], accurate [19], and active [23] warnings are effective. Passive warnings are easily ignored or clicked away [23-24]. The context determines whether and when to warn users [25-26].

Spear phishing is an Advanced Persistent Threat (APT) attack. Recent research by Trend Micro shows just how critical a technique spear phishing is. Having analyzed many APT-related emails, the company estimated that 91% of APT attack begins with (SP) emails. The malware is most commonly Remote Access Trojans (RATs) delivered in zip files disguised as spreadsheet or word processor document (usually .xls or .rlf formats) [6]. The first widely reported Apt was publicized by Google in January 2010. However, it is believed to have begun some six months earlier known as Operation Aurora; it targets 34 organizations, including Yahoo, Symantec, Northrop, Grumman, Morgan Stanley, and Dow chemical, and Google itself [7]. Harry Sverdlove, Bit9 says that "why to try to break into a company when you can walk through the front door" [6]. Greaux says, "people's idiosyncrasies are targeted, and attacks are designed to exploit people's emotional responses of fear, curiosity, and greed" [6].

According to Tod Beardsley. Most spear phishing invites targets to open an attachment file (94%, according to a Micro study). The remaining 6% are emails that ask targets to click on a link [6]. Several anti-phishing techniques have been proposed to protect the users [1-3] from attenuating the phishing issue. Two methodologies are fundamentally adopted among them: link-based approach and word list-based approach [4]. In a link-based process, the hyperlinks are examined through blacklist [5] Google safe browsing [6] SiteAdvisor [7], whitelist [8-10], and heuristic-based methods [10-13] to decide whether the email is a phishing email or legitimate. On the other hand, the word list-based approach examines the frequent keywords. In most instances, phishers employ these keywords to manipulate the victims [14-16]. The proposed model will work according to tiers, one tier followed by another. The different levels are:

a) Tier_1: the total number of output emails Eout from the T1 ML algorithm can be represented mathematically as

$$Eout \rightarrow n\ (M1_1\ U\ Ms_1) \qquad (1)$$

b) Tier_2: for the T2 ML algorithm, the output can be categorized into three different sets

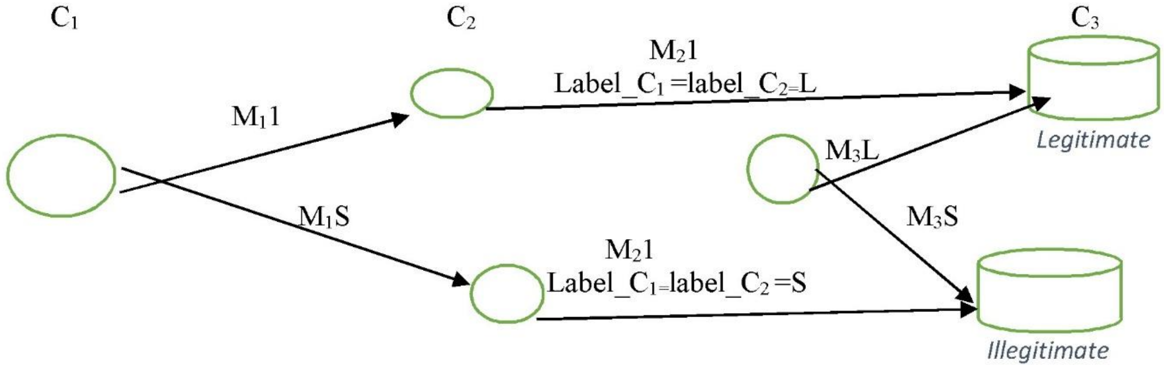

Figure 1. *Classifier*

In the first tier, we will analyze the contents of the attachment's mail and contents using the sentimental analysis to categorize the emails into spam and not spam; we used spam vs. not spam dataset available on Kaggle with more than 3000 emails. In the proposed model, we'll use a hashing function to check for the authenticity of the shared file. We will verify that the file has not been tampered with or compromised during the transfer by matching the hash values.

Table 1. *Hash value of uncorrupted file before transmission*

MD5	82aa432412486ce58f4d321c7250b54c
TIGER	5c67b08f2ee8de00606f2581d8fe157ea7bdd5750a3b57e8
SHA1	89e7ded069853ba61ef33e4d74e751fde6773665

Table 2. *Hash value of uncorrupted file after transmission*

MD5	82aa432412486ce58f4d321c7250b54c
TIGER	5c67b08f2ee8de00606f2581d8fe157ea7bdd5750a3b57e8
SHA1	89e7ded069853ba61ef33e4d74e751fde6773665

Table 3. *Hash value of corrupted file before transmission*

MD5	82aa432412486ce58f4d321c7250b54c
TIGER	5c67b08f2ee8de00606f2581d8fe157ea7bdd5750a3b57e8
SHA1	89e7ded069853ba61ef33e4d74e751fde6773665

Table 4. *Hash value of corrupted file after transmission*

MD5	5a9070eed14c65db25ea632d0f50717a
TIGER	f48d1cade17d4b627a223633482e64a362ab40752e13b454
SHA1	cf31889ad5db6352d45afd6e63788cac61a4ff17

From the above tables, it is evident that hash value remains the same for the uncorrupted file before and after transmission; however, the values change for the corrupted file for all the hashing functions before and after transmission, so this helps us to identify the emails with malicious attachments due to which accuracy touched 98 percent almost.

We categorized the mails into two categories: With attachment and Without attachment, Emails which carry attachments will be supplied to the machine learning model for detecting phishing content, and its attachment will be extracted first and then will be fed to hashing algorithms to calculate the hash value of the attachment before and after transmission. We have used seven hashing algorithms MD5, SHA1, and TIGER. We convert the attached file into a text file for sentiment analysis. After evaluating both the email contents and its attachment, it will be further categorized into VALID and MALICIOUS. If the Email is

VALID, it will be accepted and sent to the concerned user, and if there is malicious content, it will be categorized as spam and subsequently rejected. The Emails which are without attachments will be fed into a machine-learning algorithm trained for classifying emails into ham and spam. If it is a VALID mail, it will be accepted and sent to the concerned user, and if it contains malicious content, it will be discarded and categorized as spam.

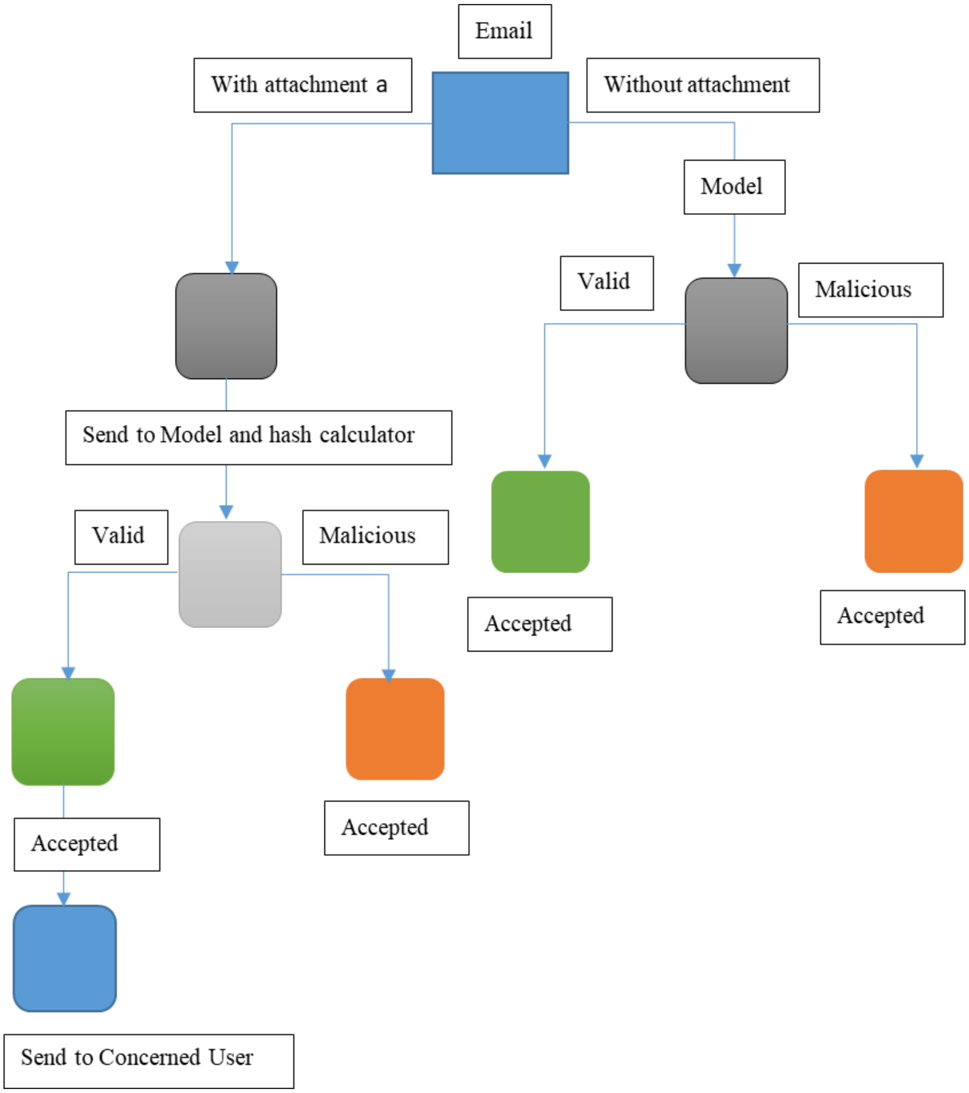

Figure 2. *Proposed model*

Sender or intruder will send a malicious email using the machine of the head of the organization which has been compromised; the receiver will receive the mail, here we will calculate the trust level of the email using machine learning algorithm if the trust level is suspicious, it will be redirected or forwarded to authentication. The authentication phase plays an important role, and It will send an OTP through phone number to the head of the institution to authenticate the email; if he/she has sent the mail, it will be given access and will send the mail to the receiver and if not then access will be denied, and an email will be categorized as spam. It an additional authentication which will protect the receiver from receiving the phishing emails.

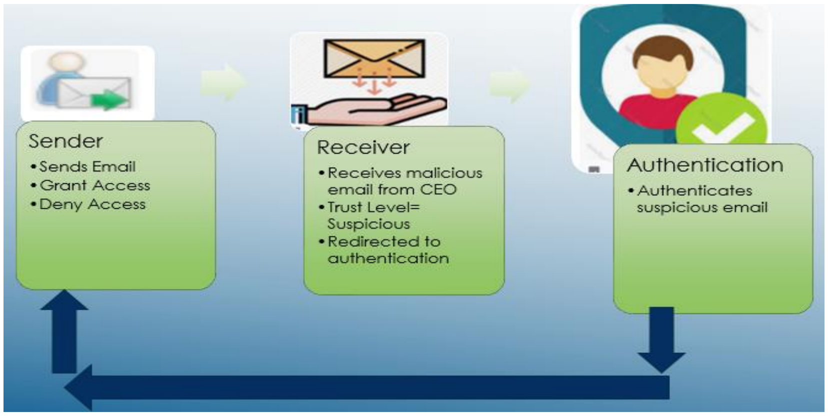
Sender
• Sends Email
• Grant Access
• Deny Access
Receiver
• Receives malicious email from CEO
• Trust Level= Suspicious
• Redirected to authentication
Authentication
• Authenticates suspicious email

Figure 3. *Additional authentication*

2.1 SVM results

This algorithm works on a simple strategy of separating hyperplanes. Given the training data, the algorithm categorizes the test data into an optimal hyperplane[27]. The simplest form of data classification, the goal is to find the hyperplane of the form$w^t x + b = 0$ **,** the distance between the hyperplanes is given by $\left|\frac{(\boldsymbol{b}-\mathbf{1})-(\boldsymbol{b}+\mathbf{1})}{||\boldsymbol{w}||}\right| = \frac{\mathbf{2}}{||\boldsymbol{w}||}$, the objective is to maximize $\frac{\mathbf{2}}{||\boldsymbol{w}||}$.

Table 5. *Support vector machine results*

	Precision	Recall	F1-Score	Support
0	0.95	1.00	0.98	501
1	1.00	0.76	0.86	99
accuracy			0.96	600
Micro avg.	0.98	0.88	0.92	600
Weighted avg.	0.96	0.96	0.96	600
Total accuracy =0.96				

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$, where TP= True Positive, TN= True Negative, FP= False Positive, FN= False Negative

Precision= $\frac{TP}{TP+FP}$Precision is the fraction of correctly classified position observations over all the observations classified as positive.

Recall= $\frac{TP}{TP+FN}$, the fraction of correctly classified positive observations over all the positive observations.

F1-score=2*$\frac{precision*recall}{precision+recall}$, it is the harmonic mean between precision and recall.

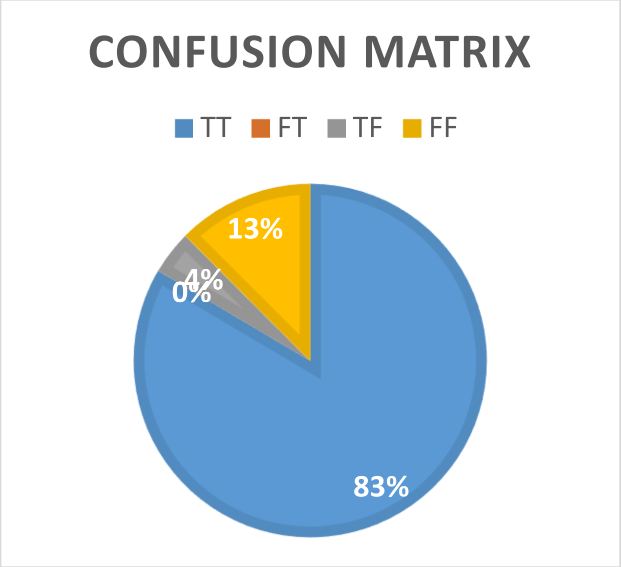

Figure 4. *Confusion matrix of SVM*

2.2 Random forest results

This algorithm is efficient in handling large datasets and thousands of input variables without their deletion. This model can deal with the over-fitting of data points [27]. Consists of multiple decision trees use mean square error (MSE) for solving regression problems, which are represented as

$$\mathbf{MSE} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{fi} - \boldsymbol{yi})^{\mathbf{2}} \qquad (2)$$

Table 6. *RandomForest results*

	Precision	Recall	F1-Score	Support
0	0.97	1.00	0.99	501
1	1.00	0.86	0.92	99
accuracy			0.98	600
Micro avg.	0.99	0.93	0.96	600
Weighted avg.	0.98	0.98	0.9	600
Total accuracy =0.98				

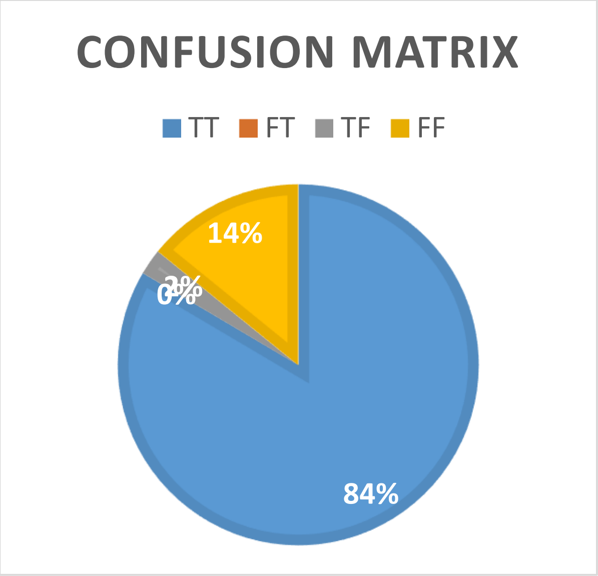

Figure 5. *Confusion matrix of Random forest*

2.3 Topic Modelling

Topic modeling is an unsupervised machine learning (ML) method that enables us to discover secret semantic structures in a text, allowing us to learn about text representations in a corpus. We convert all the mails into the significant corpus and find the dominant topic using our topic mode using LDA(Latent Dirichlet Allocation) from Gensim package. We visualize it using pyLDAvis. We build the topic model with several topics equal to 20, where each topic is a combination of keywords. Each keyword furnishes a certain weightage to the topic. To understand it in a better way, we will interpret first topic 0 from the dataset:

Topic 0 is represented as (0, '0.046*"prohibit" + 0.001*"reliance" + 0.000*"recommendation" + ' '0.000*"investor" + 0.000*"invest" + 0.000*"investment" + 0.000*"trading" + ' '0.000*"cbyi" + 0.000*"purchase" + 0.000*"cash"')

The interpretation of the above statement is that the top 10 words in topic 0 are 'prohibit,' 'reliance,' 'recommendation.' So, prohibiting having the highest weight; the weight reflects how important a keyword is essential to that topic. After observing the weightage, we can conclude that this topic is about purchasing, trading, etc. We calculated the Perplexity and Coherence of our topic model. Perplexity = -12.897670565510511 and Coherence score = 0.44700287476452394. Perplexity and coherence score are used to evaluate the model, the lower the perplexity best the model is, the higher the topic coherence, and the topic is more human interpretable.

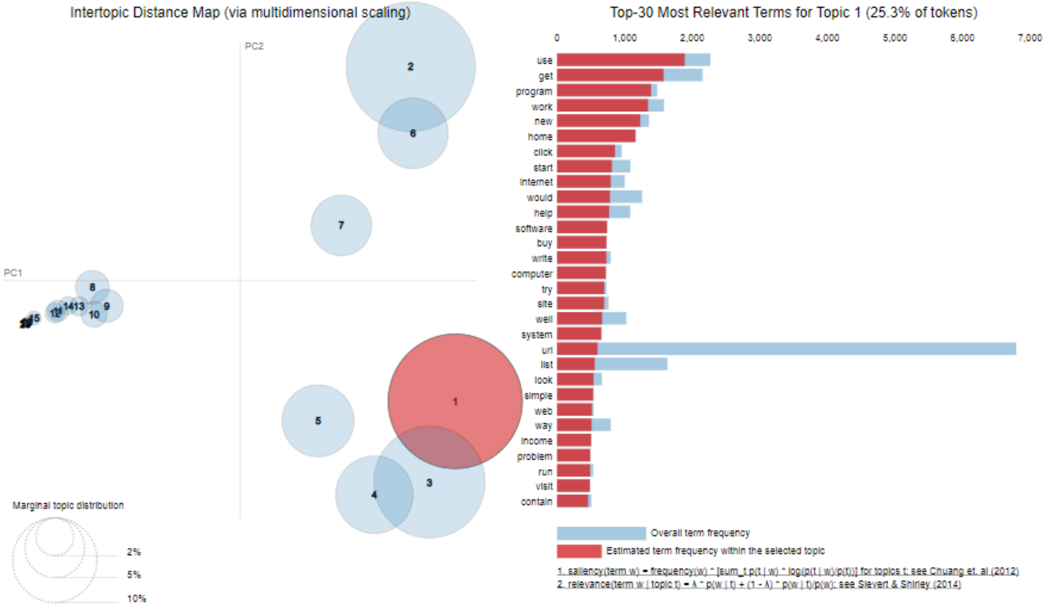

Figure 6. *Topic model results for Topic 1*

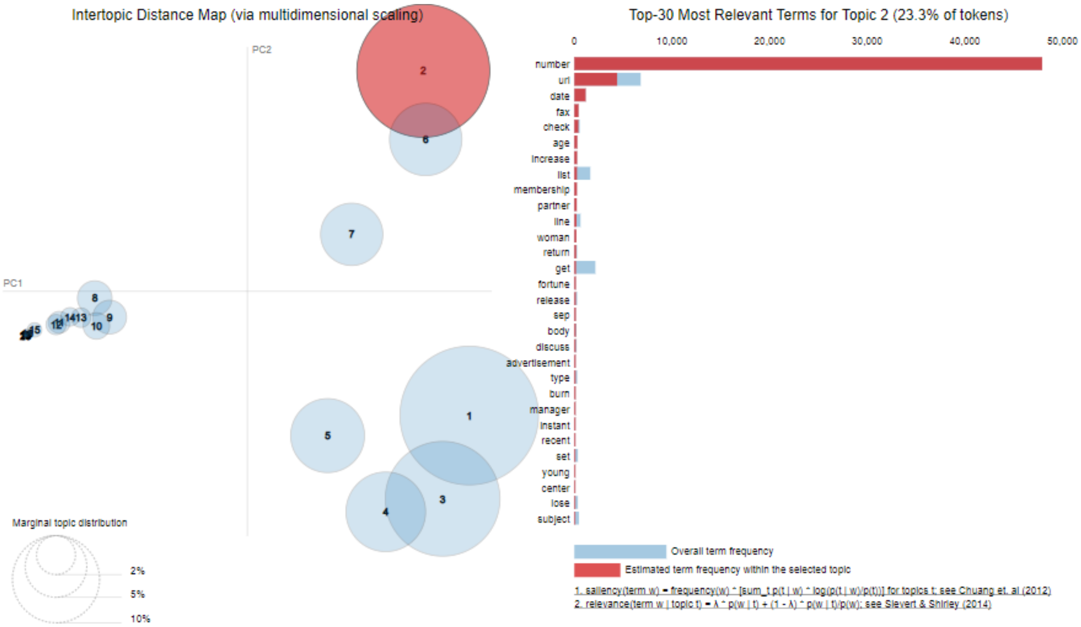

Figure 7. *Topic model results for Topic 2*

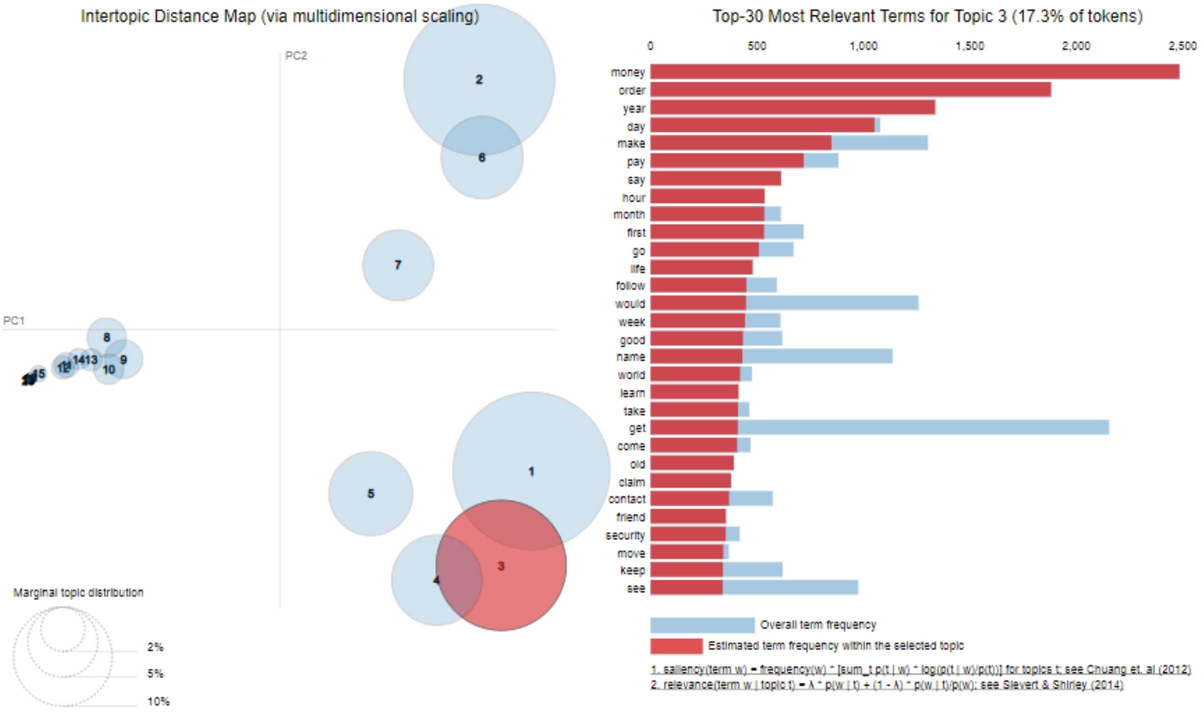

Figure 8. *Topic model results for Topic 3*

A topic is represented by each bubble on the left-hand side plot. The bigger the bubble, the more prevalent this topic is. A model with too many topics will typically have many overlaps, small-sized bubbles clustered in one region of the chart. It is evident from the above graphs how often topics like money, click, follow, site, etc., were used by the intruders for phishing purposes. The LDA approach to topic modeling considers each document as a set of topics and each topic as a keyword collection. Once you provide the algorithm with several topics, all it does is rearrange the distribution of topics within documents and distribute keywords within the topics to obtain a good composition of the distribution of topic-keywords. We created bigram and trigram models for topic modeling, bigram is the frequency of two words occurring together, and trigram is the frequency of three terms.

3. Conclusion

This paper analyzed the emails and categorized them into spam and non-spam using a multi-tier approach. In the first-tier, we did the sentimental analysis of the attachment's email content and content using machine learning algorithms such as SVM and Random-forest. SVM showed 96 percent accuracy with 0 percent false-positive and 4 percent false-negative ratios. Simultaneously, Random-forest proves to be more effective with 97.66 percent accuracy and 0 percent false-positive, and 2.33 percent false-negative ratios. In the second tier, to check the attachments' authenticity, we calculate the hash values both at the sender and receiver side; for uncorrupted files, hash values remain the same. At the same time, as it changes for corrupted files. Topic modeling with LDA helped get the dominant topics in our dataset, which intruders use for phishing purposes. The best part of this paper is the inclusion of email attachment, most of the work has been done on the contents of the mail, but significantly less work has been done on attachment. The paper introduced a new approach to mitigating the spear-phishing attack by introducing machine learning algorithms with hash functions.

References

[1] Chen, C., et al.: *Investigating the deceptive information in Twitter spam*, Futur. Gener. Comput. Syst., vol. 72, pp. 319–326, 2017, **https://doi.org/10.1016/j.future.2016.05.036**

[2] Shahriar, H., Zulkernine, M.: *Trustworthiness testing of phishing websites: A behavior model-based approach*, Futur. Gener. Comput. Syst., vol. 28, no. 8, pp. 1258–1271, 2012, **https://doi.org/10.1016/j.future.2011.02.001**

[3] Wenyin, L., Fang, N., Quan, X., Qiu, B., Liu, G.: *Discovering phishing target based on semantic link network*, Futur. Gener. Comput. Syst., vol. 26, no. 3, pp. 381–388, 2010, **https://doi.org/10.1016/j.future.2009.07.012**

[4] Almomani, A., Gupta, B. B., Atawneh, S., Meulenberg, A., Almomani, E.: *A survey of phishing email filtering techniques*, IEEE Commun. Surv. Tutorials, vol. 15, no. 4, pp. 2070–2090, 2013, **https://doi.org/10.1109/SURV.2013.030713.00020**

[5] Netcraft Protection Apps & Extensions | Netcraft. https://www.netcraft.com/apps/ (accessed Sep. 30, 2020).

[6] Google Code Archive - *Long-term storage for Google Code Project Hosting*, https://code.google.com/archive/p/google-safe-browsing/ (accessed Sep. 30, 2020).

[7] McAfee WebAdvisor.: Browse safely and steer clear of online dangers. https://www.mcafee.com/en-us/safe-browser/mcafee-webadvisor.html (accessed Sep. 30, 2020).

[8] Afroz, S., Greenstadt, R.: *PhishZoo: Detecting phishing websites by looking at them*, Proc. - 5th IEEE Int. Conf. Semant. Comput. ICSC 2011, pp. 368–375, 2011, **https://doi.org/10.1109/ICSC.2011.52**

[9] Cao, Y., Han, W., Le, Y.: *Anti-phishing based on automated individual whitelist*, Proc. ACM Conf. Comput. Commun. Secure., pp. 51–59, 2008, **https://doi.org/10.1145/1456424.1456434**

[10] Sonowal, G., Kuppusamy, K. S.: *PhiDMA – A phishing detection model with multi-filter approach*, J. King Saud Univ. - Comput. Inf. Sci., vol. 32, no. 1, pp. 99–112, 2020, **https://doi.org/10.1016/j.jksuci.2017.07.005**

[11] Mishra, S., Soni, D.: *Smishing Detector: A security model to detect smishing through SMS content analysis and URL behavior analysis*, Futur. Gener. Comput. Syst., vol. 108, pp. 803–815, 2020, **https://doi.org/10.1016/j.future.2020.03.021**

[12] Yearwood, J., Mammadov, M., Banerjee, A.: *Profiling phishing emails based on hyperlink information*, Proc. - 2010 Int. Conf. Adv. Soc. Netw. Anal. Mining, ASONAM 2010, no. May, pp. 120–127, 2010, **https://doi.org/10.1109/ASONAM.2010.56**

[13] Yu, W. D., Nargundkar, S., Tiruthani, N.: *PhishCatch-A phishing detection tool*, Proc. - Int. Comput. Softw. Appl. Conf., vol. 2, pp. 451–456, 2009, **https://doi.org/10.1109/COMPSAC.2009.175**

[14] Basnet, R., Mukkamala, S., Sung, A. H.: *Detection of phishing attacks: A machine learning approach,* Stud. Fuzziness Soft Comput., vol. 226, pp. 373–383, 2008, **https://doi.org/10.1007/978-3-540-77465-5_19**

[15] L'Huillier, G., Hevia, A., Weber, R., Ríos, S.: *Latent semantic analysis and keyword extraction for phishing classification,* ISI 2010 - 2010 IEEE Int. Conf. Intell. Secure. Informatics Public Saf. Secure., pp. 129–131, 2010, **https://doi.org/10.1109/ISI.2010.5484762**

[16] Pandey, M., Ravi, V.: *Detecting phishing emails using text and data mining,* 2012 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2012, 2012, **https://doi.org/10.1109/ICCIC.2012.6510259**

[17] Muniandy, L., Muniandy, B.: *Phishing: Educating the Internet users–a practical approach using email screenshots,* IOSR J. of Research & Method in Education, (IOSRJRME), 2013, *2*(3), 33-41. **https://doi.org/10.9790/7388-0233341**

[18] Darwish, A., El Zarka, A., Aloul, F.: *Towards understanding phishing victims' profile,* In 2012 International Conference on Computer Systems and Industrial Informatics, 2012, December, pp. 1-5. IEEE. **https://doi.org/10.1109/ICCSII.2012.6454454**

[19] Cranor, L. F.: Ca*n Phishing be foiled,* Scientific American, 2008, *299*(6), 104-111. **https://doi.org/10.1038/scientificamerican1208-104**

[20] Cone, B. D., Irvine, et al. l.: *A video game for cybersecurity training and awareness*, computers & security, 2007, 26(1), 63-72. **https://doi.org/10.1016/j.cose.2006.10.005**

[21] Dodge Jr, R. C., Carver, C., Ferguson, A. J.: *Phishing for user security awareness*, computers & security, 2007,*26*(1), 73-80. **https://doi.org/10.1016/j.cose.2006.10.009**

[22] Modic, D., Anderson, R.: *Reading this may harm your computer: The psychology of malware warnings,* Computers in Human Behavior, 2014, 41, 71-79. **https://doi.org/10.1016/j.chb.2014.09.014**

[23] Egelman, S., Cranor, L. F., Hong, J.: Y*ou've been warned: an empirical study of the effectiveness of web browser phishing warnings*, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2008, April, pp. 1065-1074. **https://doi.org/10.1145/1357054.1357219**

[24] Schechter, S. E., Dhamija, R., Ozment, A., Fischer, I.: *The emperor's new security indicators,* In 2007 IEEE Symposium on Security and Privacy (SP'07), 2007, May, pp. 51-65. IEEE. **https://doi.org/10.1109/SP.2007.35**

[25] Kern, N., Schiele, B.: *Context-aware notification for wearable computing,* In *Seventh IEEE* International Symposium on Wearable Computers, 2003. Proceedings. (pp. 223-230). IEEE.

[26] Avrahami, D., Fogarty, J., Hudson, S. E.: *Biases in human estimation of interruptibility: Effects and implications for practice,* In Proceedings of the SIGCHI conference on Human factors in computing systems, 2007, April, pp. 50-60. **https://doi.org/10.1145/1240624.1240632**

[27] Rane, A., Kumar, A.: *The sentiment classification system of Twitter data for US airline service analysis*, In 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC) 2018, July, Vol. 1, pp. 769-773). IEEE. **https://doi.org/10.1109/COMPSAC.2018.00114**