How Machine Learning and Big Data Analytics Can Aid in Social Engineering (Phishing)

Chandra Kanta Mallick, Hiren Kumar Praharaj

College of Engineering Bhubaneswar, Biju pattnaik University of Technology, Odisha, India

ABSTRACT:
The crucial area for cyber security is social engineering attacks. Social engineering is a technique whereby people use technology on other people. It is currently one of the most popular and simple ways to gain access to personal accounts and obtain items. Due to the automatic syncing that occurs when your devices connect to WiFi, it is currently the least secure technology for this purpose. This aims to clarify why social engineering assaults pose a significant risk to enterprises.

Social engineering exploits familiarity through emails and advertisements. Information gathering and use, pretesting, and baiting. driving too fast.

KEYWORDS: Social Engineering, Phishing, Cyber threats, Bigdata, Fraud, Hacking

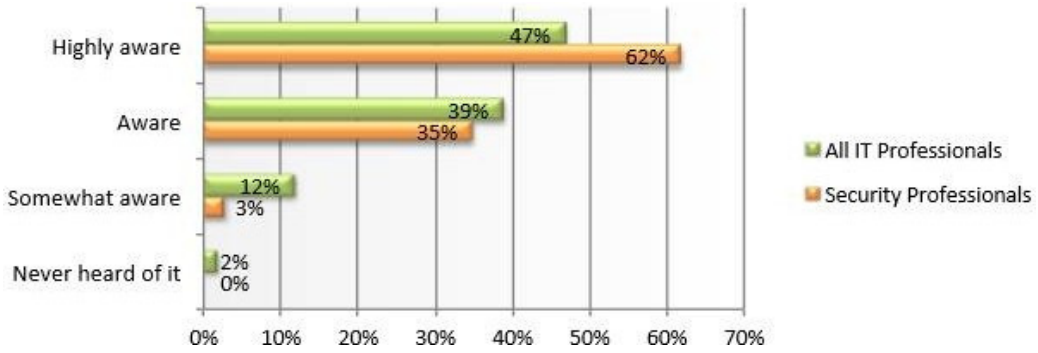
I. INTRODUCTION

This essay provides guidance on social engineering attacks in the online environment. As a result of the increased expansion in business, an increasing number of companies are implementing technology to expand their market reach. Internet users have benefited greatly from technological advancements and adaption, but they may also be exploited for illicit purposes. Email and SMS can be used to fall victim to phishing, which is the greatest threat to businesses and organizations. One sort of social engineering is phishing, in which a criminal mind, also known as a phisher, tries to falsely get user credentials through an automated process that imitates correspondence from a reliable source or open organization. Phishing emails also contain links that victims must click in order to be taken to another website or page that requests personal information or login credentials. Twenty-five thousand phishing campaigns are initiated each month, according to Ant Phishing Working People. Phishing emails aim to obtain as much personal information as they can, including usernames, passwords, credit card numbers, and personal identities. Phishers always target high-class profiles in an attempt to obtain confidential data, including social security numbers. Phishing poses a threat to the entire online business sector. More than only financial trust is harmed; their elements are degrading. may result in a loss of money in terms of time and resources. However, organizations rarely report on phishing occurrences, or only a very tiny portion of the real event is published, as the public disclosure of bad information can harm an organization's reputation and hurt investors. High danger is predicted by the statistical data over Middle-sized or small-sized businesses, based on (DBIR Verizon, 218) Only 4% of recipients of phishing emails click on them, compared to 21% of professionals, 41% of educators, and 32% of public sector workers. Small

businesses lack the security policies, knowledge, and resources necessary to prevent these attacks, as well as an efficient gateway solution. Because of this, small businesses are frequently the focus of email attacks. Individuals who are aware of the phishing process and its traits can better assist employees in spotting these emails and help them make better decisions. For this reason, the investigation will uncover a few areas of expanded understanding that can aid in improving decision-making and safeguarding individuals or organizations.

I. RELATED WORK

According to research survey 76% of target attacks being with a spear-phishing email containing a malicious attack or link using techniques which aredifficult to detect standard email of endpoint security (Trend Micro). Social engineering attacks can simply understand as it's all about finding gaps and use those against human, hack the human by the humanis more suitable in this case. Mostly this is a category of hacking personal information and use that again them way of using technology by human Wi-Fi is most hackable system as it connects automatically by that gabs can be identified easily secondly phishing/email is the most useableattach, spear phishing an email that appears to be from an individual or business that is known to you. Social Engineering is costly especially for larger organizations 48% of large companies and 32% of companies of all sizes have experienced 25 or more social engineering,48% of all participants cite an average per-incident cost of over $25,000,30% of large companies cite a per-incident cost of over $100,000 97% of security professionals and 86% of all IT professionals are aware or highly aware of this potential



Social engineer attack has different forms which impact personal information of human.

Phishing: Scams might be the most common types of social engineering attacks used today.

FamiliarityExploit: This is one of the best and it's a cornerstone of social engineering,In a nutshell, you are trying to make it appear perfectly normal to everyone that you should be there making yourself familiar to those that you want toexploit helps to lower their guard.

Gathering&Usinginformation: When it comes right down to it the key to being a successful social engineering in information gathering the more information you have about your mark the more likely you are to get what you want from him or her obviously.

Pretexting: is another form of social engineering where attackers focus on creating a good pretext or a fabricated scenario. That they can use to try and steal their victim's personal information.

Baiting: is in many ways like phishing attacks. However, what distinguishes them from other types of social engineering is the promise of an item or good hackers use to entice victims.

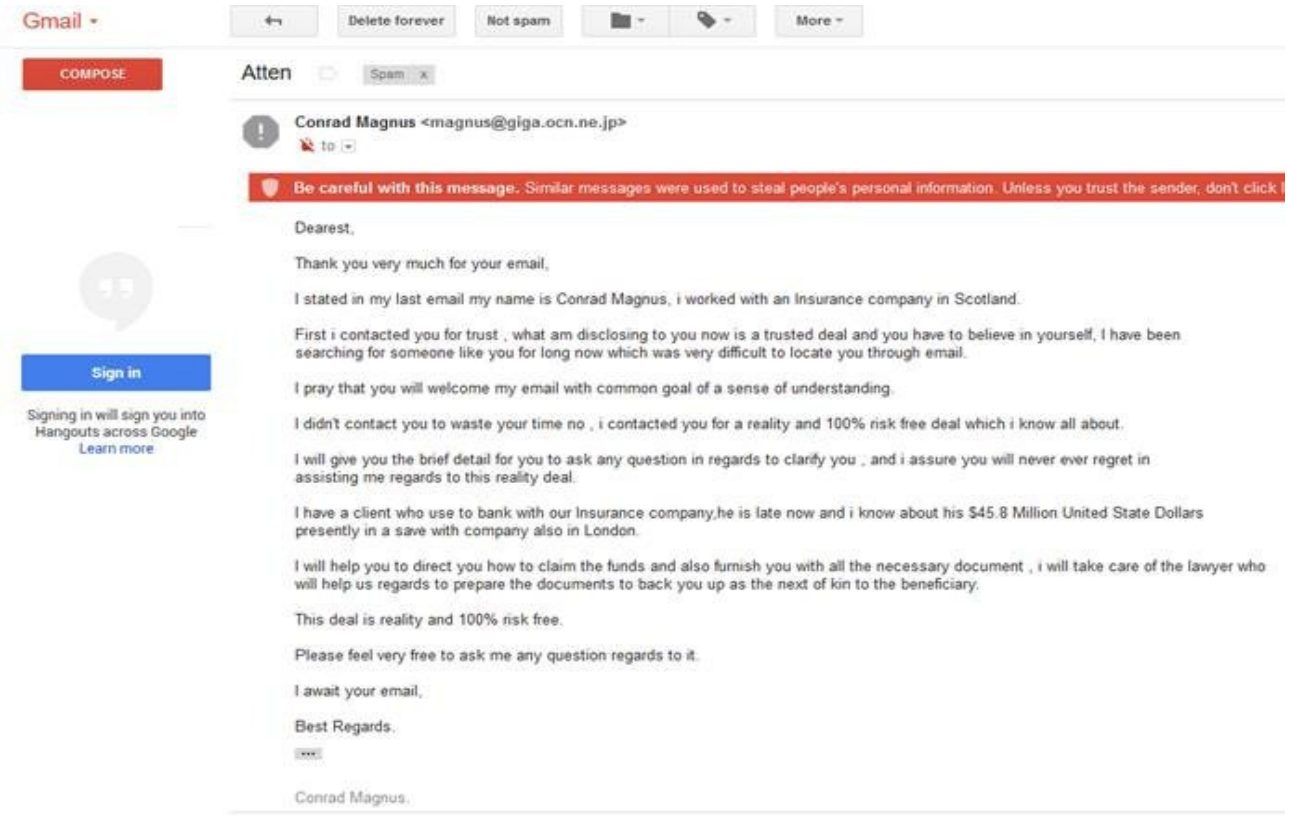
Tailgating: another social engineering all bock type is known as tailgating or "piggybacking" these types of attacks involve someone who lacks the proper authentication following an employee into a restricted area.

Phishing is a leading social engineering attack, it has no limits this fraud method has been growing rapidly approximately 8 million daily phishing attempts worldwide, Chinese phishers are more responsible for the full attack of phishing. There is a shared virtual server hacking attack, in this attack phisher hacks into webserver that hosts a larger number of domain websites /web application, Phisher places fake website pages to divert all information this way phisher can have thousands of websites live in few mints.

The APWG identified 263,538 attacks that used this strategy in 1Q 2018 this is 46% more by the Q4 2017.

Spammers taking advantage of occasion /event and holidays offerings, phisher create the spammers for these events,There where a larger number of phishing campaign run over on Japanese earthquake about Olympic and Xmas.

It can be any big event. The way criminal /Phisher executes operation there is much great opportunity to get success. Also popular scam of economical fears these scams include phishing emails those are coming from some financialinstitution or some economical upcoming are there.



III.PROPOSED ALGORITHM

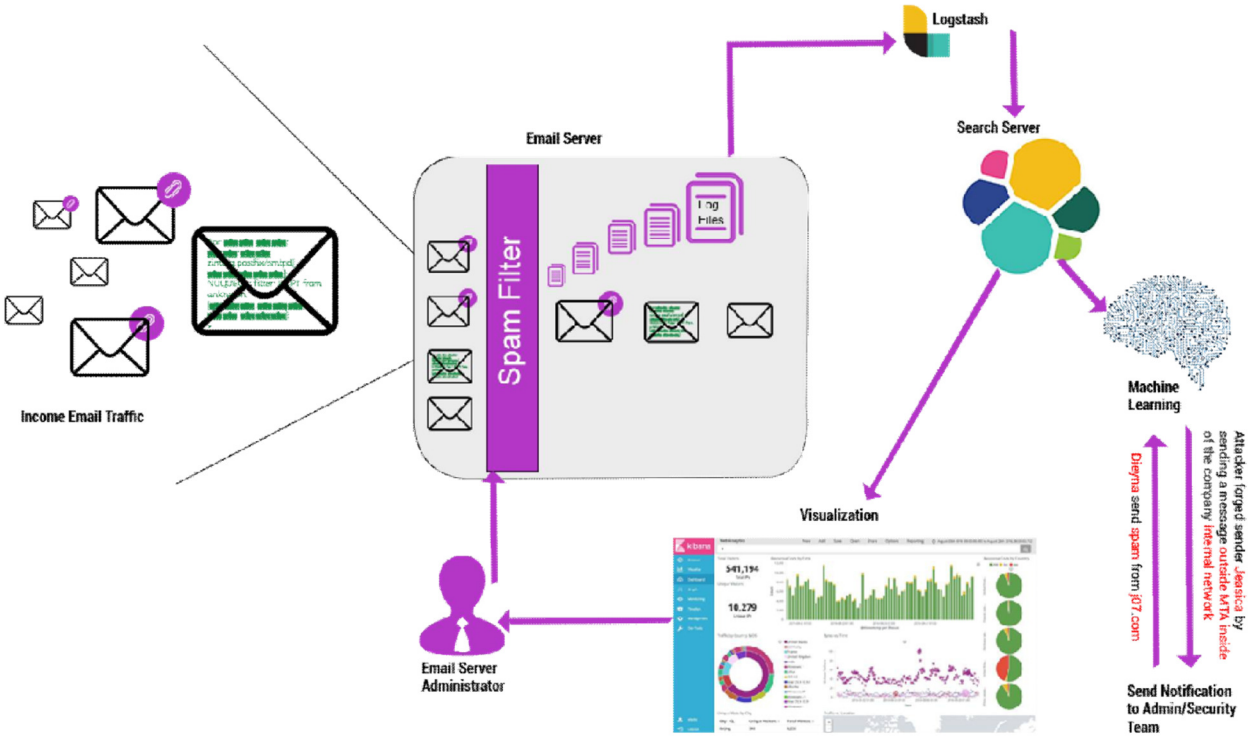
There is a need foran advance email filter technique to detect social engineering attacks. Mail server message log looks as follows

Apr 01 06:43:39 zimbra postfix/smtpd[31272]: NOQUEUE: filter: RCPT from unknown[115.159.87.234]:

<anna@1g77.com>: Sender address triggers FILTER smtp-amavis:[127.0.0.1]:10024; from=<anna@1g77.com> to=<removed> proto=ESMTP helo=<anna.1g77.com>

It perfectly shows this domain does not exist, need to identify these type of domain has block those a sap, by using bigdata analytics we can see how many domains is newly born and what are those.

Trail show how we can minimize the risk of social engineering “Phishing” in the organization,



Social Engineering email attack protection using 5 components. Email server spam filter which will be first componentinteracting with income email traffic it stops unwanted emails also logs are written.The second component read logfile and send to the search server, It will be continuously working with search server to send updated information to the server. One information is inside the search server, four and five component will start working on there given task four component is machine learning component will be taking data and training machine model by working on following questions

Who? Where? What? When? How? Who: Sends the emails ? to whom? Where? Does the email come from? What? Is the intention of this email? When? Is the email begin sent?
How? Is the email being relayed?

Once the model is trained and it will detect spam and scam emails e.g. “Dieyna send spam from j07.com” or “Attacker forged sender Jessica by sending a message outside MTA inside of the company internal network”, notification is sentto organization cyber security team as well as email server Admin.

Next component is visualization system which will get data to form search server and shows statistics when, where and who send emails, it will list down all new domains as well as a count of old domains which will give a statistical view of spam emails.

Social Engineering Attack Protection detects targeted attack emails and prevents them from reaching endpoints.

IV.SIMULATION RESULTS

The system for protecting phishing emails/ emails attack have detected spam and provided good protection this research includes big data analytics with machine learning to protect email attacks. Trail figure shows how much traffic is coming from which domain all the others domain are declared spam emails which are notified to cyber security cell.

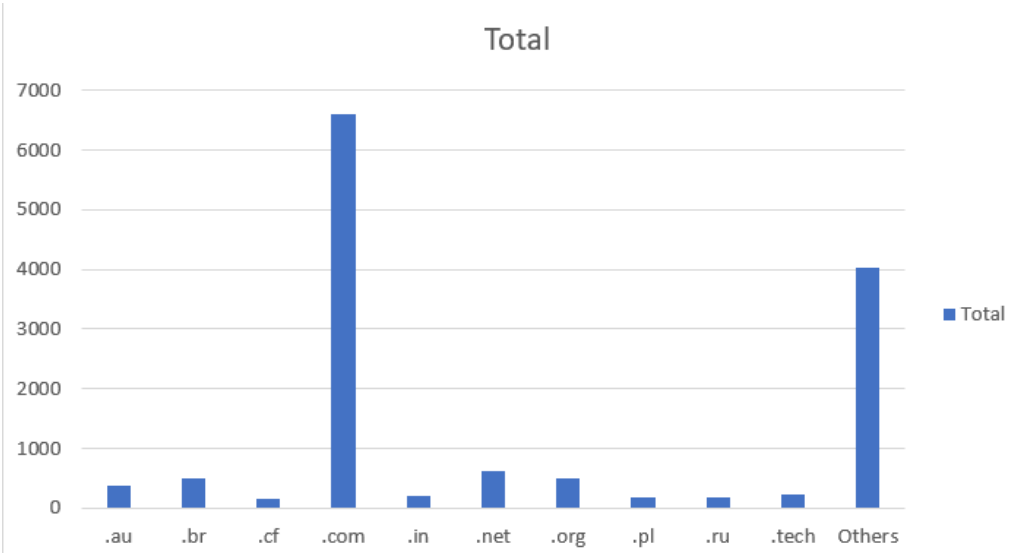


Figure 3: Proposed 5 Component Structure to Stop Social Engineering Attack.

V.CONCLUSION AND FUTURE WORK

Additionally, a suitable awareness program for staff members is necessary so that they are informed about attacks. One of the finest strategies to establish trust and allay worries about phishing is through education. Educating individuals on the types of emails that are damaging, Phishing always changes, utilizing human behavior to its advantage and even going so far as to hack people.

Subsequent research endeavors will involve the automatic adjustment of spam filters based on machine learning findings. such that there is no need for human interaction. It's also important to investigate the potential contributions of big data technology and robotic process automation.

REFERENCES

1. Ullah, F., and Babar, M.A (2018), "Architectural Tactics for Big Data Cybersecurity Analytics Systems: AReview"
2. Nicholas Carr. "The Limits of Social Engineering". https://www.technologyreview.com/s/526561/the-limits-of-social-engineering/
3. Neil DuPaul,"Hacking the Mind: How & Why Social Engineering Works"https://www.veracode.com/blog/2013/03/hacking-the-mind-how-why-social-engineering-works.
4. Davide Andreoletti, SUPSI and Enrico Frumento, CEFRIEL "https://www.dogana-project.eu/index.php/social-engineering-blog/11- social-engineering/92-cambridge-analytica "
5. David Kebo "Gartner:Social engineering, big data top security priorities for 2013" https://www.us-analytics.com/hyperionblog/qlikview/2012/11/gartner-social-engineering-big-data-top-security-priorities-for-2013"
6. Cárdenas, A.A., P.K. Manadhata,and S. Rajan, Big data analytics for security intelligence. *"https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Analytics_for_Security_Intelligence.pdf".*
7. https://www.us-analytics.com/hyperionblog/qlikview/2012/11/gartner-social-engineering-big-data-top-security-priorities-for-2013
8. Joseph A. Cazier, Christopher M. Botelho, "Social Engineering's Threat to Public Privacy" Appalachian State University.
9. Trend Micro, "https://www.trendmicro.tw/cloud-content/us/pdfs/business/datasheets/ds_social-engineering-attack-protection.pdf"*Trend Micro*.
10. Dimensional Research, THE RISK OF SOCIAL ENGINEERING ON INFORMATION SECURITY "https://www.stamx.net/files/The- Risk-of-Social-Engineering-on-Information-Security.pdf"
11. SymantecResearch, Fraud Alert:Phishing White Paper"http://www.symantec.com/content/en/us/enterprise/white_papers/b-fraud-alert-phishing-wp.pdf"
12. APWG, Phishing Activity Trends Report 1Q -2018 "https://docs.apwg.org/reports/apwg_trends_report_q1_2018.pdf"
13. DigiCert,inc,Phishing:A Primer on what phishing is and how it works"https://www.digicert.com/news/DigiCert_Phishing_White_Paper.pdf"