**ASSIGNMENT 2.**

**NAME**: STANLEY NGUGI CHEGE **REG NO**: SCT212-0065/2017

**Task**:

1. Literature Review
2. *Identify at least one book, one conference proceeding and two research articles. Copy the abstracts into a word document.*

***Part a)***

# Literature Review

Phishing is a type of social engineering attack often used to trick users into revealing private or sensitive information by masquerading as a trusted entity. These attacks occur in large numbers and have caused billions of dollars in losses (R. Verma et al., 2018).

**How Phishing is conducted**

An attacker conducts a successful phishing in 5 stages (Anjum & Shabut et al., 2016):

**Stage 1**: *Planning and Setup*: the attackers identify the target, digs out the essential details regarding their target, then set up the attacks to redirect the victim to the fraudulent URL.

**Stage 2**: *Phishing*: The attackers disguise themselves as some reputable organization, attract victim(s) and request confidential information from them.

**Stage 3**: *Break-in/Infiltration*: The victim clicks on the malicious link and either a malware that allows the attacker to access the device automatically installs on his device or the victim is redirected to a URL.

**Stage 4**: *Data Collection*: As soon as the attackers gain access to the victim’s system, they extract the required data.

**Stage 5**: *Break-out/Ex-filtration*: Once the attacker has access and gained the required information, they remove all the evidence then track the degree of success of their attack to refine their future attacks

**How Phishing is detected**

Two popular approaches are used to detect the phishing websites (Ali, Waleed, et al., 2017):

**Blacklist and whitelist-based approach**: Blacklists are essentially a database of URLs that have been confirmed to be malicious in the past. This database is compiled over time (often through crowd-sourcing solutions, e.g., PhishTank (LLC OpenDNS. et al., 2016). The main drawback of the blacklist and whitelist-based approach is that it cannot distinguish the newly created phishing websites from legitimate websites

**Intelligent heuristics-based approach**: In this approach, some features of websites are collected and evaluated to select the most influential website features, which play an important role in detecting the phishing websites. The selected significant features of many websites can be utilized as training dataset. Then, the machine learning techniques are trained based on the prepared training dataset in order to effectively classify the websites as either phishing or legitimate.

**SUPERVISED MACHINE LEARNING**

Machine learning concentrates on developing the computational algorithms that reason and induce patterns and rules from externally supplied instances and prior data in order to produce general models, which are able to make predictions about future instances. The machine learning is called supervised if known labels are given with instances in the training phase, whereas instances are unlabeled in unsupervised machine learning. (D. Sahoo, C. Li S.C.H. Hoi, et. al 2019.)

The 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. According to APWG (et al. 2021), Phishers continue to use certain domain name registrars to obtain domains for their schemes. In a URL, lexical features can be extracted such as the URL string, information about the host, and sometimes even HTML and JavaScript content. Host-based features obtained from the hostname properties of the URL allow us to know the location, identity, and the management style and properties of malicious hosts. An underlying assumption is that there is an array of features to differentiate malicious and benign URLs. Based on this information, a prediction model can be built, which can make predictions on new URLs. This can be formalized as a binary classification task of a machine learning algorithm. (D. Sahoo, C. Li S.C.H. Hoi, et. al 2019.)

***Part b)***

# Book 1

**Author**: Tuli Krishan, Dr. Juneja Neenu.

**Publication Title**: Cyber security challenges & online frauds on internet

**Abstract**

The fast evolution of on-line and mobile channels has etched out new markets and brought large opportunities for aborting and established organizations alike. However, sadly the past decade has additionally witnessed important disruption to ecommerce payment processes and systems. The interconnected, anonymous and fast nature of those channels has inevitably diode to the event of malicious threats targeting ecommerce and retail services corporations, their individuals and their customers.

These e-crime and digital fraud threats still evolve apace, with attackers utilizing progressively refined techniques to focus on vulnerabilities in individuals, processes and technologies. The e-crime threats, if with success completed, will undermine essential digital services, cause important injury to complete reputations, and end in wide money and operational pain for organizations and their customers.

Cybercrime is rising as a significant threat. Worldwide governments, police departments and intelligence units have begun to react. Initiatives to curb cross border cyber threats are taking form. Indian police have initiated special cyber cells across the country and have started educating the personnel. This text is a trial to supply a glimpse on cybercrime in Asian country. This text is predicated on numerous reports from journalism and news portal.

# Conference Proceeding

**Title**: Detecting Phishing Attacks Using Natural Language Processing and Machine Learning

**Published in**: 2018 IEEE 12th International Conference on Semantic Computing (ICSC)

**Date of Conference**: 31 Jan.-2 Feb. 2018

**Abstract**:

Phishing attacks are one of the most common and least defended security threats today. We present an approach which uses natural language processing techniques to analyze text and detect inappropriate statements which are indicative of phishing attacks. Our approach is novel compared to previous work because it focuses on the natural language text contained in the attack, performing semantic analysis of the text to detect malicious intent. To demonstrate the effectiveness of our approach, we have evaluated it using a large benchmark set of phishing emails.

# Research article 1

**Article Title:** Why phishing still works: User strategies for combating phishing attacks

**Year Published**: 2015

**Author**: Mohamed Alsharnouby

**Abstract:**

We have conducted a user study to assess whether improved browser security indicators and increased awareness of phishing have led to users׳ improved ability to protect themselves against such attacks. Participants were shown a series of websites and asked to identify the phishing websites. We use eye tracking to obtain objective quantitative data on which visual cues draw users׳ attention as they determine the legitimacy of websites. Our results show that users successfully detected only 53% of phishing websites even when primed to identify them and that they generally spend very little time gazing at security indicators compared to website content when making assessments. However, we found that gaze time on browser chrome elements does correlate to increased ability to detect phishing. Interestingly, users׳ general technical proficiency does not correlate with improved detection scores.

Research article 2

**Article Title**: A survey of phishing attacks: Their types, vectors and technical approaches

**Year Published**: 2018, Volume 106

**Authors**: Kang Leng Chiew, Kelvin Sheng Chek Yong & Choon Lin Tan

**Abstract:**

Phishing was a threat in the cyber world a couple of decades ago and still is today. It has grown and evolved over the years as phishers are getting creative in planning and executing the attacks. Thus, there is a need for a review of the past and current phishing approaches. A systematic, comprehensive and easy-to-follow review of these approaches is presented here. The relevant mediums and vectors of these approaches are identified for each approach. The medium is the platform which the approaches reside and the vector is the means of propagation utilized by the phisher to deploy the attack. The paper focuses primarily on the detailed discussion of these approaches. The combination of these approaches that the phishers utilized in conducting their phishing attacks is also discussed. This review will give a better understanding of the characteristics of the existing phishing techniques which then acts as a stepping stone to the development of a holistic anti-phishing system. This review creates awareness of these phishing techniques and encourages the practice of phishing prevention among the readers. Furthermore, this review will gear the research direction through the types of phishing, while also allowing the identification of areas where the anti-phishing effort is lacking. This review will benefit not only the developers of anti-phishing techniques but the policy makers as well.

# References

"Cyber security challenges & online frauds on internet", International Journal of Advanced Research in IT and Engineering, Year 2016, Volume 5, Issue 2First page : ( 1) Last page : ( 12) Online ISSN : 2278-6244. (https://www.indianjournals.com/ijor.aspx?target=ijor:ijarie&volume=5&issue=2&article=001)

T. Peng, I. Harris and Y. Sawa, "Detecting Phishing Attacks Using Natural Language Processing and Machine Learning," 2018 IEEE 12th International Conference on Semantic Computing (ICSC), 2018, pp. 300-301, doi: 10.1109/ICSC.2018.00056. (https://ieeexplore.ieee.org/abstract/document/8334479)

Mohamed Alsharnouby, Furkan Alaca, Sonia Chiasson, "Why phishing still works: User strategies for combating phishing attacks", International Journal of Human-Computer Studies,Volume 82, 2015, Pages 69-82, ISSN 1071-5819, (<https://doi.org/10.1016/j.ijhcs.2015.05.005>.) (https://www.sciencedirect.com/science/article/pii/S1071581915000993)

Kang Leng Chiew, Kelvin Sheng Chek Yong, Choon Lin Tan, "A survey of phishing attacks: Their types, vectors and technical approaches", Expert Systems with Applications, Volume 106, 2018, Pages 1-20, ISSN 0957-4174, (<https://doi.org/10.1016/j.eswa.2018.03.050>.) (<https://www.sciencedirect.com/science/article/pii/S0957417418302070>)

Rakesh Verma, Devin Crane, and Omprakash Gnawali, (2018). Phishing During and After Disaster: Hurricane Harvey. In 2018 Resilience Week (RWS). IEEE, 88–94

Shaikh, Anjum & Shabut, Antesar & Hossain, Alamgir. (2016). A literature review on phishing crime, prevention review and investigation of gaps. 9-15. 10.1109/SKIMA.2016.7916190.

LLC OpenDNS. 2016. PhishTank: An anti-phishing site. Online: https://www. phishtank. com (2016).

Ali, Waleed. (2017). Phishing Website Detection based on Supervised Machine Learning with Wrapper Features Selection. International Journal of Advanced Computer Science and Applications. 8. 10.14569/IJACSA.2017.080910.

Shaikh, Anjum & Shabut, Antesar & Hossain, Alamgir. (2016). A literature review on phishing crime, prevention review and investigation of gaps. 9-15. 10.1109/SKIMA.2016.7916190.

Phishing Activity Trends Report, 2021. 1 st Quarter 2021 Report, APWG. Activity January-March 2021

Ali, W. (2017). Phishing website detection based on supervised machine learning with wrapper features selection. International Journal of Advanced Computer Science and Applications, 8(9), 72-78.

Sarno, D. M., Lewis, J. E., Bohil, C. J., & Neider, M. B. (2019). Which Phish Is on the Hook? Phishing Vulnerability for Older Versus Younger Adults. Human factors. doi: 10.1177/0018720819855570

Mohammad R., Thabtah F., McCluskey L. Tutorial and critical analysis of phishing websites methods. Computer Science Review Journal. Volume 17, August 2015, Pages 1–24 Elsevier. 2015

Doyen Sahoo and Chenghao Liu and Steven C. H. Hoi, 2019. “Malicious URL Detection using Machine Learning: A Survey”. Eprint 1701.07179, arXiv, cs.LG