LITERATURE SURVEY

DOMAIN : APPLIED DATA SCIENCE

**TOPIC : ** WEB PHISHING DETECTION

1.A MACHINE LEARNING APPROACH TO PHISHING DETECTION AND DEFENSE:

AUTHORS : OA Akanbi, E Fazeldehkordi

ABSTRACT :

Phishing is one of the most widely-perpetrated forms of cyber attack, used to gather sensitive information such as credit card numbers, bank account numbers, and user logins and passwords, as well as other information entered via a web site. The authors of A Machine-Learning Approach to Phishing Detection and Defence have conducted research to demonstrate how a machine learning algorithm can be used as an effective and efficient tool in detecting phishing websites and designating them as information security threats.

REFER LINK : Amiri: A machine-learning approach to phishing detection... - Google Scholar

2.PHISHING

AUTHORS : Koceilah Rekouche

ABSTRACT :

Phishing is a type of social engineering where an attacker sends a fraudulent (e.g., spoofed, fake, or otherwise deceptive) message designed to trick a person into revealing sensitive information to the attacker or to deploy malicious software on the victim's infrastructure like ransomware. Phishing attacks have become increasingly sophisticated and often transparently mirror the site being targeted, allowing the attacker to observe everything while the victim is navigating the site, and transverse any additional security boundaries with the victim. As of 2020, phishing is by far the most common attack performed by cybercriminals, the FBI's Internet Crime Complaint Centre recording over twice as many incidents of phishing than any other type of computer crime.

REFER LINK : https://en.wikipedia.org/wiki/Phishing#History

3.‘ TEXAS SCHOOL DISTRIC lOSES $2.3 MILLION TO PHISHING SCAM ’, BEC, 2020. AVAILABLE AT :

AUTHORS : Trend Micro

ABSTRACT :

Manor Independent School District (MISD) in Texas is investigating an email phishing attack after a series of seemingly normal school-vendor transactions resulted in the loss of an estimated US$2.3 million. According to the statement posted on Twitter, the district is cooperating with the Manor Police Department and the Federal Bureau of Investigation (FBI), and encouraged the community to share Any information related to the incident.

REFER LINK :

https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/texas-school-district-loses-2-3-million-to-phishing-scam-bec

4.FACEBOOK AND GOOGLE WERE CONNED OUT OF $100M IN PHISHING SCHEME ’. AVAILABLE AT

AUTHORS : Gibbs S

ABSTRACT :

Google and Facebook were phished for over $100m, it has been reported, proving not even the biggest technology companies in the world are immune from the increasingly sophisticated attacks of online scammers. Last month it was reported that two major tech companies were tricked by a Lithuanian man into sending him over $100m (£77m). Evaldas Ramanauskas, 48, was charged with wire fraud, money laundering and aggravated identity theft for impersonating Quanta Computer – a Taiwanese electronics manufacturer that includes Google, Facebook and Apple as clients.

REFER LINK :
https://www.theguardian.com/technology/2017/apr/28/facebook-google-conned-100m-phishing-scheme

5. VADE SECURE DISCOVERS NEW PHISHING ATTACK TARGETING 550 MILLION EMAIL USERS GLOBALLY ’. AVAILABLE AT

AUTHORS : Hadley E

ABSTRACT :

Vade Secure has discovered a new phishing attack that represents more than 550 million emails sent since Q1 2018. First detected in early January, the phishing attack is targeting consumers around the world. Countries with high concentrations of impacted email users include the US, UK, France, Germany, and the Netherlands.

The phishing attack attempts to steal users’ bank account details by offering them a coupon or discount in exchange for participating in a quiz or online contest. The emails masquerade as popular brands, online streaming services, and telecom operators based on the country of the recipients. Examples include Canada Pharmacy in the US, as well as Orange and Carrefour in France. Moreover, the content of the messages is adapted according to the local language.

REFER LINK : https://www.vadesecure.com/en/phishing-attack-targets-550-million/

6.PHISHING ACTIVITY TRENDS REPORTS, 2020. AVAILABLE AT

AUTHORS : M.Wisecrackers

ABSTRACT :

The APWG Phishing Activity Trends Report analyzes phishing attacks reported to the APWG by its member companies, its Global Research Partners, through the organization’s website at http://www.apwg.org, and by e-mail submissions to reportphishing@antiphishing.org. APWG also measures the evolution, proliferation, and propagation of crimeware by drawing from the research of our member companies. Phishing is a criminal mechanism employing both social engineering and technical subterfuge to steal consumers’ personal identity data and financial account credentials. Technical subterfuge schemes plant crimeware onto PCs to steal credentials directly, often using systems to intercept consumers online account user names and passwords -- and to corrupt local navigational infrastructures to misdirect consumers to counterfeit Web sites (or authentic Web sites through phisher-controlled proxies used to monitor and intercept consumers’ keystrokes).

REFER LINK : https://docs.apwg.org//reports/apwg_trends_report_h1_2017.

7. GOOGLE SAFE BROWSING, 2020. AVAILABLE AT

AUTHORS : Tewari A

ABSTRACT :

Google Safe Browsing helps protect over four billion devices every day by showing warnings to users when they attempt to navigate to dangerous sites or download dangerous files. Safe Browsing also notifies webmasters when their websites are compromised by malicious actors and helps them diagnose and resolve the problem so that their visitors stay safer. Safe Browsing protections work across Google products and power safer browsing experiences across the Internet.

Our Transparency Report includes details on the threats that Safe Browsing identifies. The Transparency Report includes our Site Status diagnostic tool that you can use to see whether a site currently contains content that Safe Browsing has determined to be dangerous. Safe Browsing launched in 2007 to protect users across the web from phishing attacks, and has evolved to give users tools to help protect themselves from web-based threats like malware, unwanted software, and social engineering across desktop and mobile platforms. Our Safe Browsing engineering, product, and operations teams work at the forefront of security research and technology to build systems that help users protect themselves from harm. Check out our Research and the Google Security Blog for updates on Safe Browsing and other Google security technology.

REFER LINK : https://safebrowsing.google.com/

8. PHISHING PAGE DETECTION VIA LEARNING CLASSIFIERS FROM PAGE LAYOUT FEATURE

AUTHORS : Mao J. Bian J. Tian W.

ABSTRACT :

The web technology has become the cornerstone of a wide range of platforms, such as mobile services and smart Internet-of-things (IoT) systems. In such platforms, users' data are aggregated to a cloud-based platform, where web applications are used as a key interface to access and configure user data. Securing the web interface requires solutions to deal with threats from both technical vulnerabilities and social factors. Phishing attacks are one of the most commonly exploited vectors in social engineering attacks.

REFER LINK : Mao: Phishing page detection via learning classifiers... - Google Scholar

9. NEW RULE-BASED PHISHING DETECTION METHOD

AUTHORS : Moghimi M. Varjani A.

ABSTRACT :

a new rule-based method to detect phishing attacks in internet banking. Our rule-based method used two novel feature sets, which have been proposed to determine the webpage identity. Our proposed feature sets include four features to evaluate the page resources identity, and four features to identify the access protocol of page resource elements. We used approximate string matching algorithms to determine the relationship between the content and the URL of a page in our first proposed feature set.

REFER LINK : Moghimi: New rule-based phishing detection method - Google Scholar

10. A PHISH DETECTOR USING LIGHTWEIGHT SEARCH FEATURES

AUTHORS : Varshney G. Misra M

ABSTRACT :

Web phishing is a well-known cyber-attack which is used by attackers to obtain vital information such as username, password, credit card number, social security number, and/or other credentials from Internet users via deception. A number of web phishing detection solutions have been proposed and implemented in the recent years. These solutions include the use of phishing black list, search engine, heuristics and machine learning, visual similarity techniques, DNS, access list and proactive phishing

REFER LINK : Varshney: A phish detector using lightweight search features - Google Scholar

11. PHISHING WEBSITE DETECTION BASED ON MULTIDIMENSIONAL FEATURES DRIVEN BY DEEP LEARNING

AUTHORS : Yang P. Zhao G. Zeng P

ABSTRACT :

As a crime of employing technical means to steal sensitive information of users, phishing is currently a critical threat facing the Internet, and losses due to phishing are growing steadily. Feature engineering is important in phishing website detection solutions, but the accuracy of detection critically depends on prior knowledge of features. Moreover, although features extracted from different dimensions are more comprehensive, a drawback is that extracting these features requires a large amount of time.

REFER LINK : Yang: Phishing website detection based on multidimensional... - Google Scholar

12. A CONTENT-BASED APPROACH TO DETECTING PHISHING WEB SITES

AUTHORS : Zhang Y. Hong J.I. Cranor L.

ABSTRACT :

Phishing is a significant problem involving fraudulent email and web sites that trick unsuspecting users into revealing private information. In this paper, we present the design, implementation, and evaluation of CANTINA, a novel, content-based approach to detecting phishing web sites, based on the TF-IDF information retrieval algorithm. We also discuss the design and evaluation of several heuristics we developed to reduce false positives. Our experiments show that CANTINA is good at detecting phishing site.

REFER LINK : Zhang: Cantina: a content-based approach to detecting... - Google Scholar

13. PHISHING-ALARM: ROBUST AND EFFICIENT PHISHING DETECTION VIA PAGE COMPONENT SIMILARITY ’, IEEE ACCESS, 2017

AUTHORS : Mao J. Tian W.

ABSTRACT :

A Social networks have become one of the most popular platforms for users to interact with each other. Given the huge amount of sensitive data available in social network platforms, user privacy protection on social networks has become one of the most urgent research issues. As a traditional information stealing technique, phishing attacks still work in their way to cause a lot of privacy violation incidents. In a Web-based phishing attack, an attacker sets up scam Web pages (pretending to be an important Website such as a social network portal)

REFER LINK : Mao: Phishing-alarm: Robust and efficient phishing... - Google Scholar

14. PHISHING WEBPAGE DETECTION VIA IDENTITY KEYWORDS EXTRACTION AND TARGET DOMAIN NAME FINDER

AUTHORS : Tan C.L. Chiew K.L. Wong K

ABSTRACT :

This paper proposes a phishing detection technique based on the difference between the target and actual identities of a webpage. The proposed phishing detection approach, called PhishWHO, can be divided into three phases. The first phase extracts identity keywords from the textual contents of the website, where a novel weighted URL tokens system based on the N-gram model is proposed. The second phase finds the target domain name by using a search engine, and the target domain name is selected based on identity-relevant

REFER LINK : Tan: PhishWHO: Phishing webpage detection via identity... - Google Scholar

15. A FEATURE-RICH MACHINE LEARNING FRAMEWORK FOR DETECTING PHISHING WEB SITES

AUTHORS : Xiang G. Hong J. Rose C.

ABSTRACT :

Phishing is a plague in cyberspace. Typically, phish detection methods either use human- verified URL blacklists or exploit Web page features via machine learning techniques. However, the former is frail in terms of new phish, and the latter suffers from the scarcity of effective features and the high false positive rate (FP). To alleviate those problems, we propose a layered anti-phishing solution that aims at (1) exploiting the expressiveness of a rich set of features with machine learning to achieve a high true positive rate (TP)

REFER LINK : Xiang: Cantina+ a feature-rich machine learning framework... - Google Scholar

16. DETECTING PHISHING WEBSITES VIA AGGREGATION ANALYSIS OF PAGE LAYOUTS

AUTHORS : Mao J. Bian J. Tian W.

ABSTRACT :

Phishing websites are typical starting points of online social engineering attacks, including many recent online scams. The attackers develop web pages mimicking legitimate websites, and send the malicious URLs to victims to lure them to input their sensitive information. Existing phishing defense mechanisms are not sufficient to detect with new phishing attacks. In this paper, we aim to improve phishing detection

techniques using machine learning techniques. In particular, we propose a learning-based aggregation analysis mechanism

REFER LINK : Mao: Detecting phishing websites via aggregation... - Google Scholar

17. A NEW HYBRID ENSEMBLE FEATURE SELECTION FRAMEWORK FOR MACHINE LEARNING-BASED PHISHING DETECTION SYSTEM

AUTHORS : Chiew K.L. Tan C.L. Wong K.

ABSTRACT :

This paper proposes a new feature selection framework for machine learning-based phishing detection system, called the Hybrid Ensemble Feature Selection (HEFS). In the first phase of HEFS, a novel Cumulative Distribution Function gradient (CDF-g) algorithm is exploited to produce primary feature subsets, which are then fed into a data perturbation ensemble to yield secondary feature subsets. The second phase derives a set of baseline features from the secondary feature subsets by using a function perturbation ensemble.

REFER LINK : Chiew: A new hybrid ensemble feature selection framework... - Google Scholar

18. CHIEW: A NEW HYBRID ENSEMBLE FEATURE SELECTION FRAMEWORK... - GOOGLE SCHOLAR

AUTHORS : Gowtham R. Krishnamurthi

ABSTRACT :

Phishing is a web-based criminal act. Phishing sites lure sensitive information from naïve online users by camouflaging themselves as trustworthy entities. Phishing is considered an annoying threat in the field of electronic commerce. Due to the short lifespan of phishing webpages and the rapid advancement of phishing techniques, maintaining blacklists, white- lists or employing solely heuristics-based approaches are not particularly effective. The impact of phishing can be largely mitigated by adopting a suitable combination

REFER LINK : Gowtham: A comprehensive and efficacious architecture... - Google Scholar

19. TWO LEVEL FILTERING MECHANISM TO DETECT PHISHING SITES USING LIGHTWEIGHT VISUAL SIMILARITY APPROACH

AUTHORS : Rao R.S. Pais A.R.

ABSTRACT :

The visual similarity-based techniques detect the phishing sites based on the similarity between the suspicious site and the existing database of resources such as screenshots, styles, logos, favicons etc. These techniques fail to detect phishing sites which target non-whitelisted legitimate domain or when phishing site with manipulated whitelisted legitimate content is encountered. Also, these techniques are not well adaptable at the client-side due to their computation and space complexity. Thus there is a need for light weight visual.

REFER LINK : Rao: Two level filtering mechanism to detect phishing... - Google Scholar

20. ‘DETECTING PHISHING WEBSITES AND TARGETS BASED ON URLS AND WEBPAGE LINKS ’. 24TH INT. CONF. ON PATTERN RECOGNITION

AUTHORS : Yuan H. Chen X. Li Y.

ABSTRACT :

In this paper, we propose to extract features from URLs and webpage links to detect phishing websites and their targets. In addition to the basic features of a given URL, such as length, suspicious characters, number of dots, a feature matrix is also constructed from these basic features of the links in the given URL's webpage. Furthermore, certain statistical

features are extracted from each column of the feature matrix, such as mean, median, and variance.

REFER LINK : Yuan: Detecting phishing websites and targets based... - Google Scholar