M.KUMARASAMY
COLLEGE OF ENGINEERING
NAAC Accredited Autonomous Institution
Approved by AICTE & Affiliated to Anna University
ISO 9001:2015 & ISO 14001:2015 Certified Institution
Thalavapalayam, Karur – 639 113.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

18CSE040L

PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

TITLE : **WEB PHISHING DETECTION**

DOMAIN : **APPLIED DATA SCIENCE**

FACULTY MENTOR : **Mrs K. MAKANYA DEVI M.E/AP-CSE**

INDUSTRY MENTOR : **Prof., SANDESH P**

PROJECT MEMBERS :

REGISTER NO	NAME	ROLE
19BCS4120	VARSHAA K K	LEADER
19BCS4119	TRINAYA S	MEMBER
19BCS4115	SUDHARSAN R	MEMBER
19BCS4097	RANJITH S	MEMBER

WEB PHISHING DETECTION

ABSTRACT:

Web phishing aims to steal private information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate entity. It will lead to information disclosure and property damage. Large organizations may get trapped in different kinds of scams. This Guided Project mainly focuses on applying a machine-learning algorithm to detect Phishing websites. In order to detect and predict e-banking phishing websites, we proposed an intelligent, flexible and effective system that is based on using classification algorithms. We implemented classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy. The e-banking 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.

LITERATURE SURVEY

[1] Author Name: Solomon Ogbomon Uwagbole

Phishing is a form of social engineering or website forgery whereby attackers mimic a trusted website or public organization or sending e-mails in an automated manner in order to steal sensitive information or credentials of online users. This is done in a way the user does not realize he is in a phishing environment and in turn reveals his sensitive information such as credit card information, employment details, online shopping account passwords and bank information. Phishers are still having their ways to succeed in their various nefarious activities and attacks. Different anti-phishing schemes however have emerged but phishers still find their ways around by breaking through various existing techniques. Against this backdrop, this project aims at developing a web enabled anti-phishing technique using enhanced heuristic approach.

[2] Author Name: skr aaa

Phishing has become a substantial threat for internet users and a major cause of financial losses. In these attacks the cybercriminals carry out user credential information and users can fall victim. The current solution against phishing attacks are not sufficient to detect and work against novel phishes. This paper presents a systematic review of the previous and current research waves done on Internet phishing mitigation in different areas of expertise and highlighted phishing attacks types and some existing anti-phishing approaches. Further the discussion about novel phishes and identify the elements of issues highlighted. The review can be valuable source of information to find and identify recent gap and challenges to fulfil the security flaws.

[3] Author Name: Jason Hong

Phishing websites, fraudulent sites that impersonate a trusted third party to gain access to private data, continue to cost Internet users over a billion dollars each year. In this paper, we describe the design and performance characteristics of a scalable machine learning classifier we developed to detect phishing websites. We use this classifier to maintain Google's phishing blacklist automatically. Our classifier analyses millions of pages a day, examining the URL and the contents of a page to determine whether or not a page is phishing. Unlike previous work in this field, we train the classifier on a noisy dataset consisting of millions of samples from previously collected live classification data.

[4] Author Name: Solomon Ogbomon Uwagbole

Authors designed a system with a detection technique involving a fresh approach for phishing website detection named PhishLimiter. The proposed system used Deep Packet Inspection (DPI) along with Software-Defined Networking (SDN) through web communications and emails for identifying malicious activities. The real-time DPI and phishing signature classification based on SDN programmability provided PhishLimiter, the flexibility to address phishing attacks in real-time. This also helped in better network traffic management and evaluated attacks in real-world environments proving an effective solution to identify phishing attacks.

[5] Author Name: L. MacHado and J. Gadge

The authors proposed a system to detect phishing using heuristic-based methods and feature extraction. The c4.5 decision tree algorithm was used for analysis and computing the heuristic values to determine whether a website is legitimate or phishing. The Dataset was imported from Phishtank and Google, proposed using the trained classifier, and used for detection purposes. The model achieved an accuracy of 89.40%.

TABLE OF ARTICLES

S.No	ARTICLE NAME	AUTHOR NAME	PUBLISHED YEAR	DRAWBACKS
1	Website forgery whereby attackers mimic a trusted website	**Solomon Ogbomon Uwagbole**	2021	Phishers are still having their ways to succeed in their various nefarious activities and attacks.
2	Phishing using heuristic-based methods and feature extraction	**L. MacHado and J. Gadge**	2014	In total 10 URL features including the host-based and lexical features were extracted from the obtained dataset. The testing phase of this model could not achieve an accuracy of 100 percent for the Random Forest classifier using the labeled dataset.
3	The proposed system used Deep Packet Inspection (DPI) along with Software-Defined Networking (SDN)	**Solomon Ogbomon Uwagbole**	2018	The real-time DPI and phishing signature classification based on SDN programmability provided PhishLimiter,without flexibility
4	Fraudulent sites that impersonate a trusted third party to gain access to private data	**Jason Hong**	2017	Classifier analyses millions of pages a day, examining the URL and the contents of a page to determine whether or

				not a page is phishing.
5	Internet phishing mitigation in different areas of expertise and highlighted phishing attacks types and some existing anti-phishing approaches	**skr aaa**	2019	The current solution against phishing attacks are not sufficient to detect and work against novel phishes

REFERENCES

1. (2014, June) Digicert phishing white paper: A primer on what phishing is and how it works. DigiCert, Inc. [Online].
   Available: http://www.digicert.com/news/DigiCert Phishing White Paper.pdf

2. Phishing Detection using Machine Learning based URL Analysis: A Survey.
   Available:https://www.markmonitor.com/download/report/Fraud Report-Q3 2012.pdf

3. (2014, June) The evolution of phishing attacks: 2011-2013. Kaspersky Lab ZAO. [Online].
   Available: http://media.kaspersky.com/pdf/Kaspersky Lab KSN report The Evolution of Phishing Attacks 2011-2013.pdf

4. (2014, June) Phishing activity trends report, 4th quarter 2012. Anti-Phishing Working Group. [Online].
   Available: http://docs.apwg. org/reports/apwg trends report Q4 2012.pdf

5. (2014, June) Fraud alert: New phishing tactics-and how they impact your business. Thawte, Inc. [Online].
   Available:https://community.thawte.com/system/files/download-attachments/Phishing%20WP D2.pdf