Detection of Phishing Websites

Avaneesh C S[1]
Department of Computer Science and Engineering
Sri Ramakrishna Engineering College Coimbatore, India

Varun Ganapathy S[2]
Department of Computer Science and Engineering
Sri Ramakrishna Engineering College Coimbatore, India

Vasanth E[3]
Department of Computer Science andEngineering
Sri Ramakrishna Engineering College Coimbatore, India

Ranjeethapriya[4]
Department of Computer Science and Engineering
Sri Ramakrishna Engineering CollegeCoimbatore, India

Abstract:- Phishing is a cyber attack in which an attacker creates a copy of an existing web page to trick users into submitting personal, financial or password information, making them think that this is the real website that everyone uses. The strategy followed here is an edge server-based anti-phishing algorithm called “Link Guard” uses the property of hyperlinks in phishing attacks. The purpose of this Link Guard algorithm is to find phishing emails sent by phishers to obtain information about end users. Link Guard carefully analyzes the characteristics of phishing hyperlinks. That's why all end users use it using the Link Guard algorithm. By doing this, end users catch and don’t respond tp phishing emails. Because Link Guard is based not only on the detection and prevention of phishing attacks, but also on unknown attacks. This project uses PHP and MySQL server.

The program uses a link protection method that detects phishing content based on the characteristics of phishing hyperlinks. In the hyperlink distribution method, important information is collected from victims; Phishers often try to trick users into clicking on hyperlinks embedded in phishing emails. The link protection algorithm works by analyzing the difference between apparent links and real links. The Link Guard algorithm also evaluates similarity to established trustworthy sources. The Link Guard algorithm functions by initially extracting DNS names from both genuine and apparent DNS names, followed by a comparison between the two sets of DNS names.

Keywords:- *Phishing Detection, Link Guard Algorithm, Emailsecurity, Classification of Phishing Hyperlinks.*

I. INTRODUCTION

In these emails, the attacker will give some excuse, such as that your credit card password was entered incorrectly many times or that they are offering an upgrade service, and the hyperlink provided in the email will redirect you to their website for details or updates. . your password, account number and password. When you enter your username and password, the attacker collects the information on the server and can use this information to do further work. Phishing itself is not a new concept, but in recent years more and more phishers have been using it to steal user information and commit commercial crimes.

In existing systems, phishing websites can be detected and blocked in a timely manner. stoped. It is easy to determine (manually) whether a website is a phishing site, but the real challenge is to detect these phishing sites in time. DNS scanning puts additional load on the DNS system and can cause problems with normal DNS queries, and most crimes do not require a DNS name at all. Even if phishers get some of the victim's information, they cannot complete their mission. This technology requires additional equipment to authenticate users and websites; This increases costs and causes some inconvenience Proposed System.

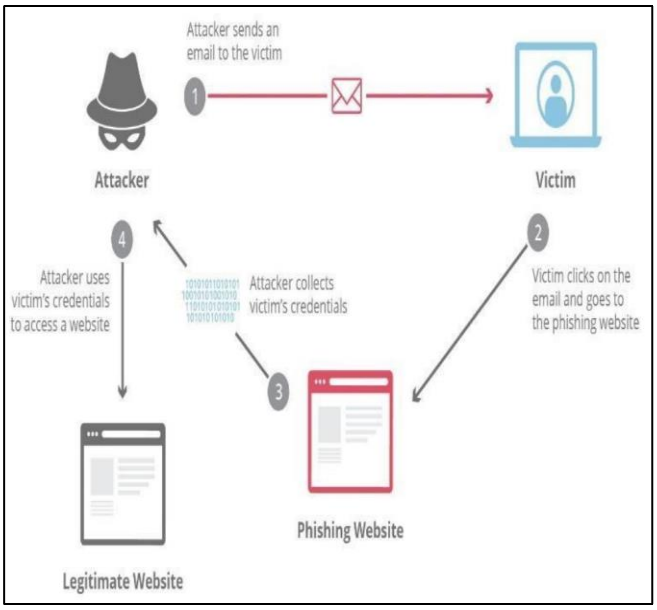

Attacker sends an email to the victim
1
Attacker
Victim
4
2
Attacker uses victim's credentials to access a website
Attacker collects victim's credentials
Victim clicks on the email and goes to the phishing website
3
Phishing Website
Legitimate Website

Fig 1 Phishing Process

The new system employs anti-phishing techniques to detect and capture phishing content by analysing the characteristics of phishing hyperlinks. When sharing hyperlinks, the system gathers crucial information from potential victims. Phishers often try to trick users into clicking on hyperlinks embedded in criminal emails. The structure of the bridge is as follows href="URI"> Link text<\a> Where - URI - (Universal Resource Identifier) Provides network resource information required for users to log in, "link text" is the text on the user's website is displayed in the browser. The link protection algorithm works by comparing the difference between the phished link and the original link. The Link Guard algorithm works as follows. In the main Link Guard functionality, it first extracts the real and apparent DNS names and then compares the real and apparent DNS names.

Detection of phishing content according to the characteristics of phishing hyperlinks has been completed. Connection Protection Algorithm It works by detecting the difference between the optical connection and the real connection.

II. MODULE DESCRIPTION

- *Mail Server Creation*

Mail Server Creation serves as the initial module of the project. These servers are specifically developed to facilitate communication with various individuals through email. The mail server allows recipients to send emails from various editing options and receive emails from any sender.

Additionally, it provides features to access incoming mail,spam, and outgoing mail.

- *User Registration*

In the user registration module, in order to send emails to many recipients, they must first create an account. After completing the registration process, individuals can sign in to their accounts to send and receive emails from various users. When registering, comprehensive details about each user are gathered and stored on the server. Each time a user attempts to log in, the system verifies if they are authorized to access the account.

- *Mail Composer*

A user abstraction is created in this module Users may transmit valuable information to customers via their email ID, message content and context, or message link. They can access a list of emails sent in the Sent Mail folder and any messages flagged as spam will be located in the spam folder. Once an email is sent, the sender will receive a notification confirming its successful delivery.

- *Phishing Checking*

It can check received emails for violations provided in the next model. Fill out the mail options, including the ID number option. Spoofing allows people to send emails using different addresses. This is to demonstrate the link protection algorithm.

- *Link Guard Algorithm*

This module includes the implementation of the link protection algorithm. Users can add new names under the site and classify them as whitelist or blacklist. If an email is found to be a software breach, the email's author will be blacklisted. The Link Guard algorithm checks whether the author falls into one of 5 categories of hyperlinks in the offending email. It also mentions information about blacklist and whitelist entries and determines the status of illegal or non-illegal words.

III. LANGUAGE DESCRIPTION

PHP is employed for server-side scripting, handling formsubmissions, and interacting with the MySQL database. MySQL serves as the database management system, storing user credentials and other relevant data. Detection involves input validation, URL analysis, user authentication, and monitoring user behaviour for suspicious activity. Security measures include SSL/TLS encryption, secure password storage using hashing and salting, and the use of parameterized queries to prevent SQL injection attacks. These components collectively form a robust system to detect and mitigate phishing attempts within a PHP and MySQL environment.

- *Er Diagram*

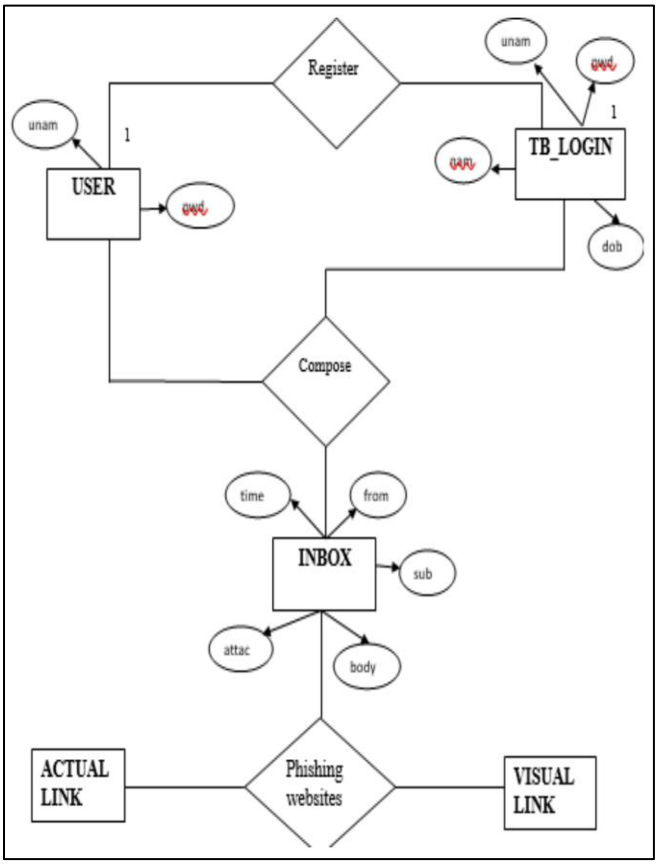


Fig 2 Er Diagram

IV. PROPOSED METHOD

The utilization of the Link Guard algorithm employing PHP and MySQL involves several key steps. First, a dataset of known phishing and legitimate websites iscollected. Next, features are extracted from hyperlinks within the webpage content, such as URL length, HTTPS presence, and domain reputation. These features, along with corresponding labels indicating the legitimacy of the websites, are stored in a MySQL database. A machine learning model, trained on this dataset, is deployed to predict the legitimacy of hyperlinks within a webpage. PHPscripts are developed to parse HTML content, extract hyperlinks, and pass them to the prediction model. A user interface is created using PHP to allow users to input URLs for analysis and view the results. A feedback mechanism enables users to report false positives or negatives, improving the model's accuracy over time. Continuous monitoring, regular updates to the dataset, and retraining of the model ensure adaptability to evolving phishing techniques. Security measures are implemented to protect the database and ensure user privacy. Performance evaluation metrics are utilized to assess the algorithm's effectiveness in detecting phishing hyperlinks, ensuring its reliability in combating phishing threat.

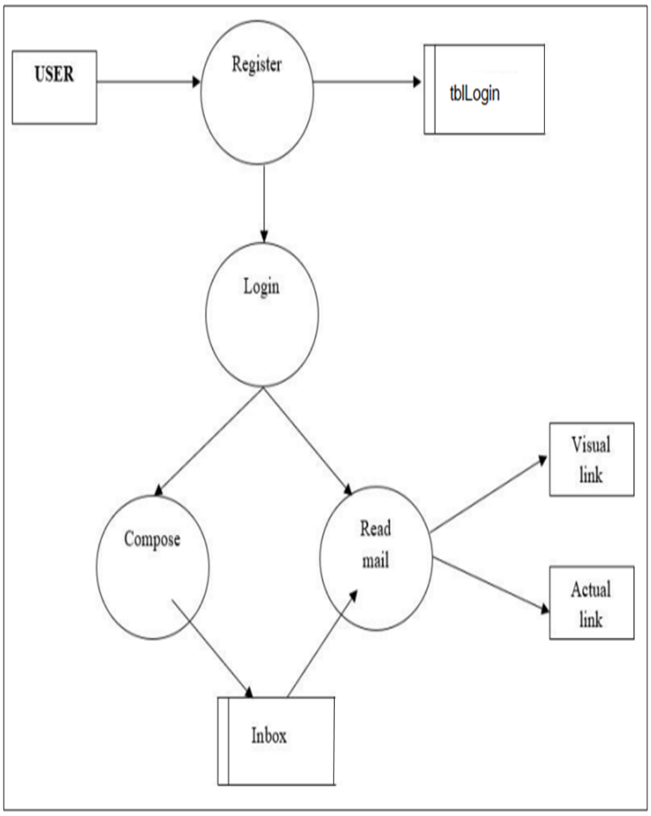
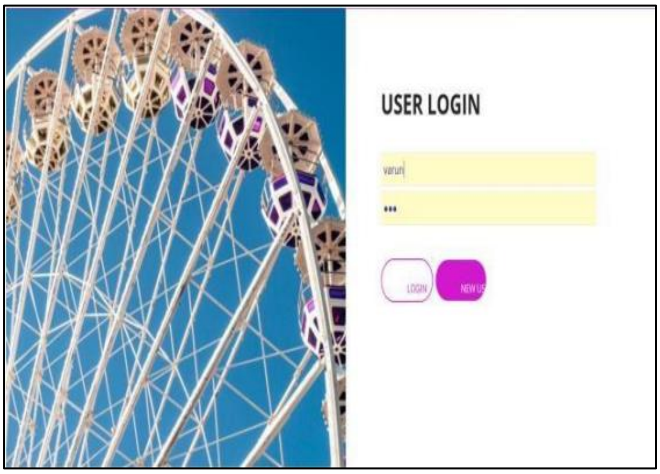


Fig 3 Flowchart

V. RESULT

In a scenario where a wrong link is sent to a user, potentially as part of a phishing attempt, the detection website equipped with the Link Guard algorithm serves as a vital defense mechanism. Upon receiving the suspicious link, the user can input it into the detection website's interface for analysis. The website then parses the URL, extracts relevant features, and feeds them into the trained machine learning model. If the link is indeed malicious, the detection website accurately identifies it as such, providing the user with detailed information about the actual vulnerable link, along with a warning about its potential risks. This empowers the user to recognize the threat and avoid clicking on the malicious link, thereby safeguarding their personal information and preventing potential harm such as data theft or malware infection. The detection website's ability to quickly and accurately identify malicious links enhances user awareness and strengthens their defenses against phishing attacks.

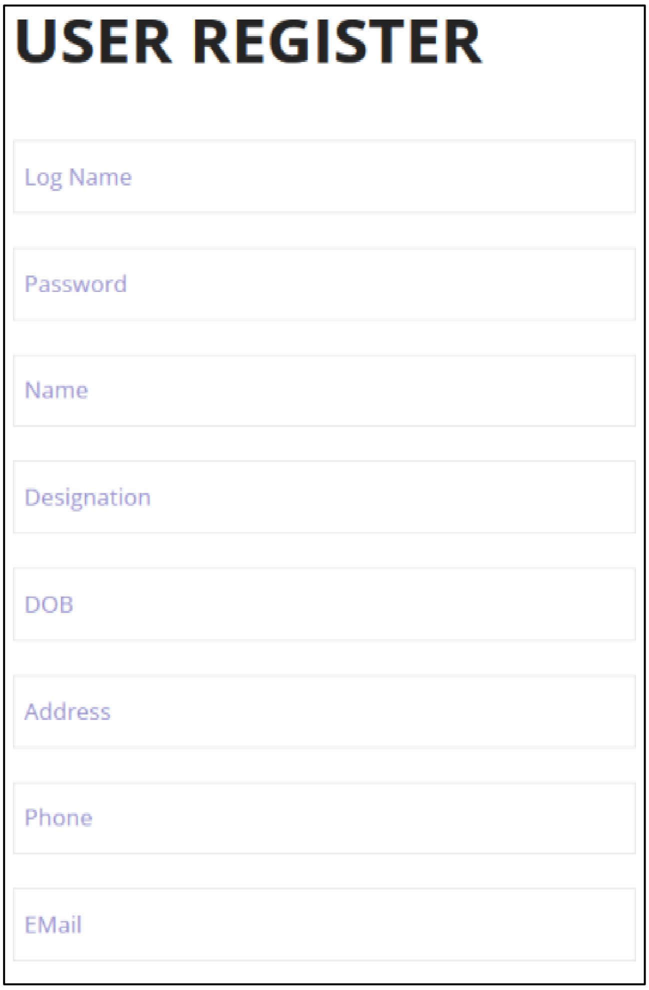


USER LOGIN

varun

•••

Fig 4 User Login



USER REGISTER

Log Name

Password

Name

Designation

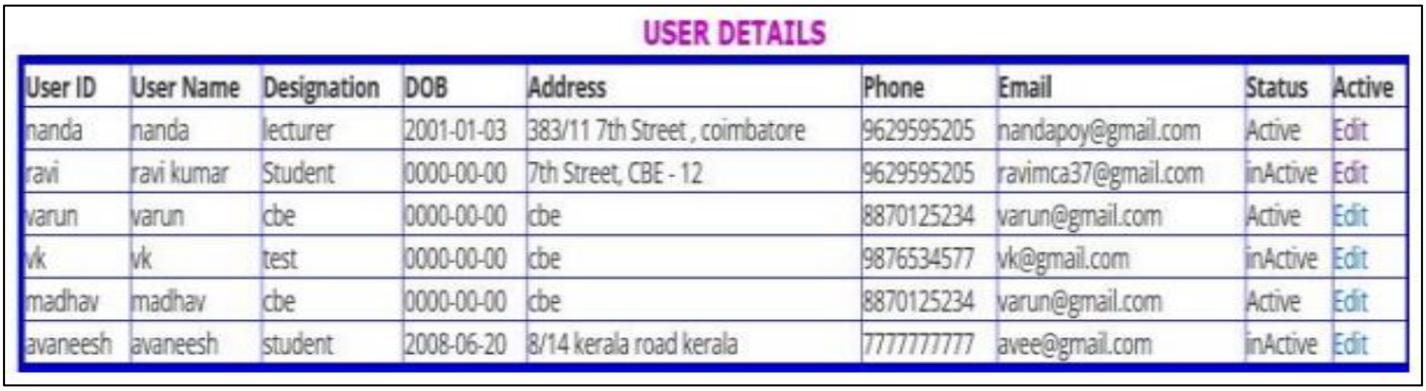
DOB

Address

Phone

EMail

Fig 5 Registration Page

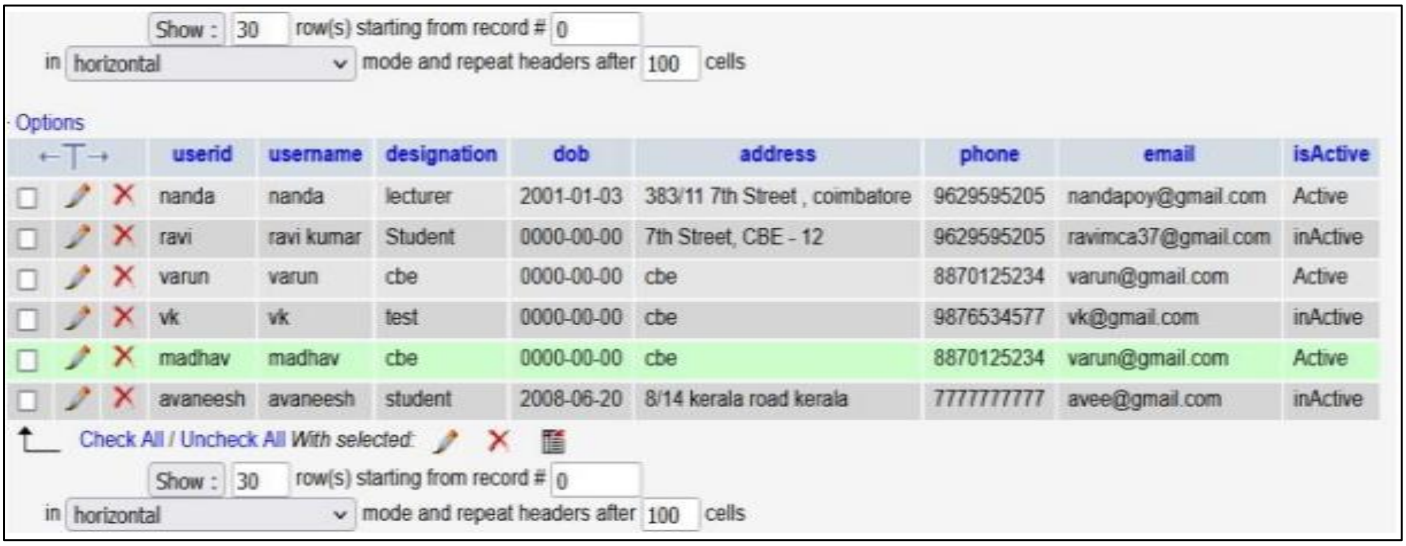


USER DETAILS

User ID	User Name	Designation	DOB	Address	Phone	Email	Status	Active
nanda	nanda	lecturer	2001-01-03	383/11 7th Street , coimbatore	9629595205	nandapoy@gmail.com	Active	Edit
ravi	ravi kumar	Student	0000-00-00	7th Street, CBE - 12	9629595205	ravimca37@gmail.com	inActive	Edit
varun	varun	cbe	0000-00-00	cbe	8870125234	varun@gmail.com	Active	Edit
vk	vk	test	0000-00-00	cbe	9876534577	vk@gmail.com	inActive	Edit
madhav	madhav	cbe	0000-00-00	cbe	8870125234	varun@gmail.com	Active	Edit
avaneesh	avaneesh	student	2008-06-20	8/14 kerala road kerala	7777777777	avee@gmail.com	inActive	Edit

Fig 6 User Details

Figure 6 shows admin can inspect the number of users presentand can edit the details of each person available respectivelyby making the users active or inactive.



Show : 30 row(s) starting from record # 0
in horizontal mode and repeat headers after 100 cells

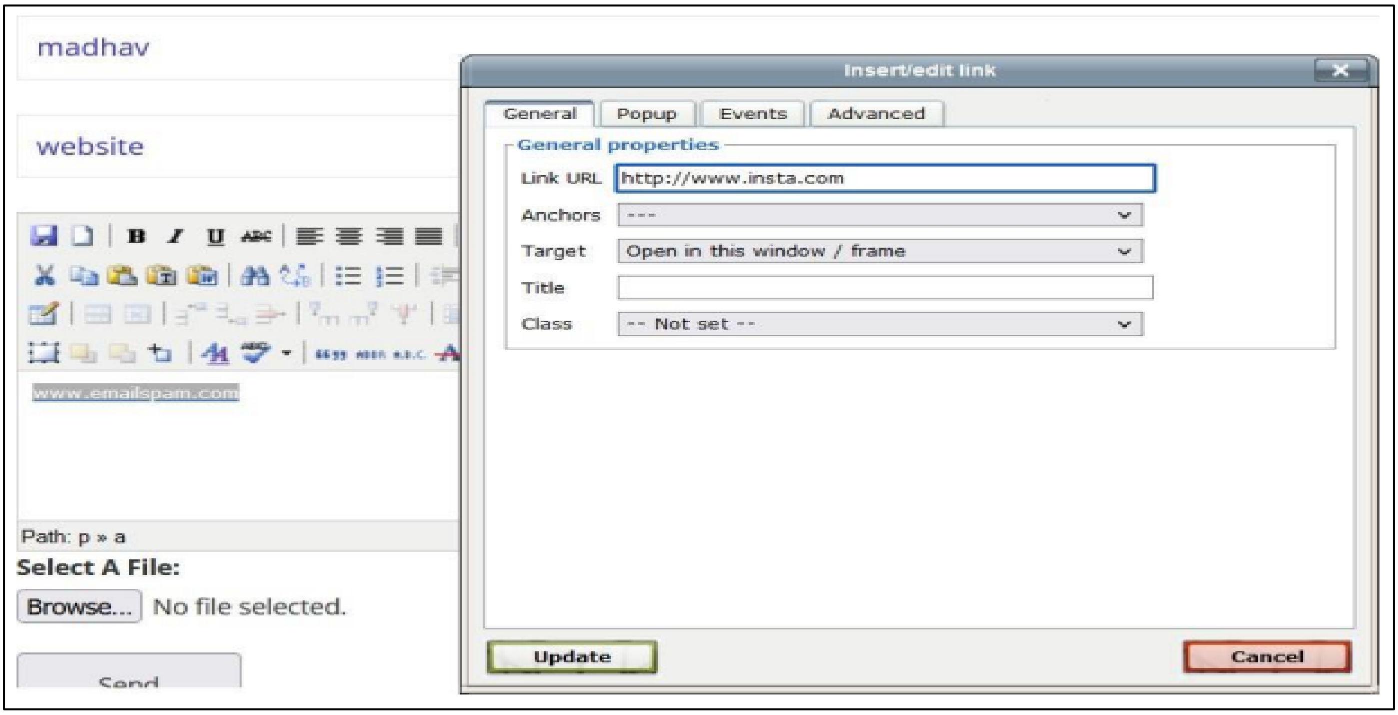
Options

←T→			userid	username	designation	dob	address	phone	email	isActive
□	✎	✕	nanda	nanda	lecturer	2001-01-03	383/11 7th Street , coimbatore	9629595205	nandapoy@gmail.com	Active
□	✎	✕	ravi	ravi kumar	Student	0000-00-00	7th Street, CBE - 12	9629595205	ravimca37@gmail.com	inActive
□	✎	✕	varun	varun	cbe	0000-00-00	cbe	8870125234	varun@gmail.com	Active
□	✎	✕	vk	vk	test	0000-00-00	cbe	9876534577	vk@gmail.com	inActive
□	✎	✕	madhav	madhav	cbe	0000-00-00	cbe	8870125234	varun@gmail.com	Active
□	✎	✕	avaneesh	avaneesh	student	2008-06-20	8/14 kerala road kerala	7777777777	avee@gmail.com	inActive

Check All / Uncheck All With selected: ✎ ✕

Show : 30 row(s) starting from record # 0
in horizontal mode and repeat headers after 100 cells

Fig 7 Database



madhav

website

www.emailspam.com

Path: p » a

Select A File:

Browse... No file selected.

Insert/edit link

General properties

Link URL: http://www.insta.com

Anchors: ---

Target: Open in this window / frame

Title:

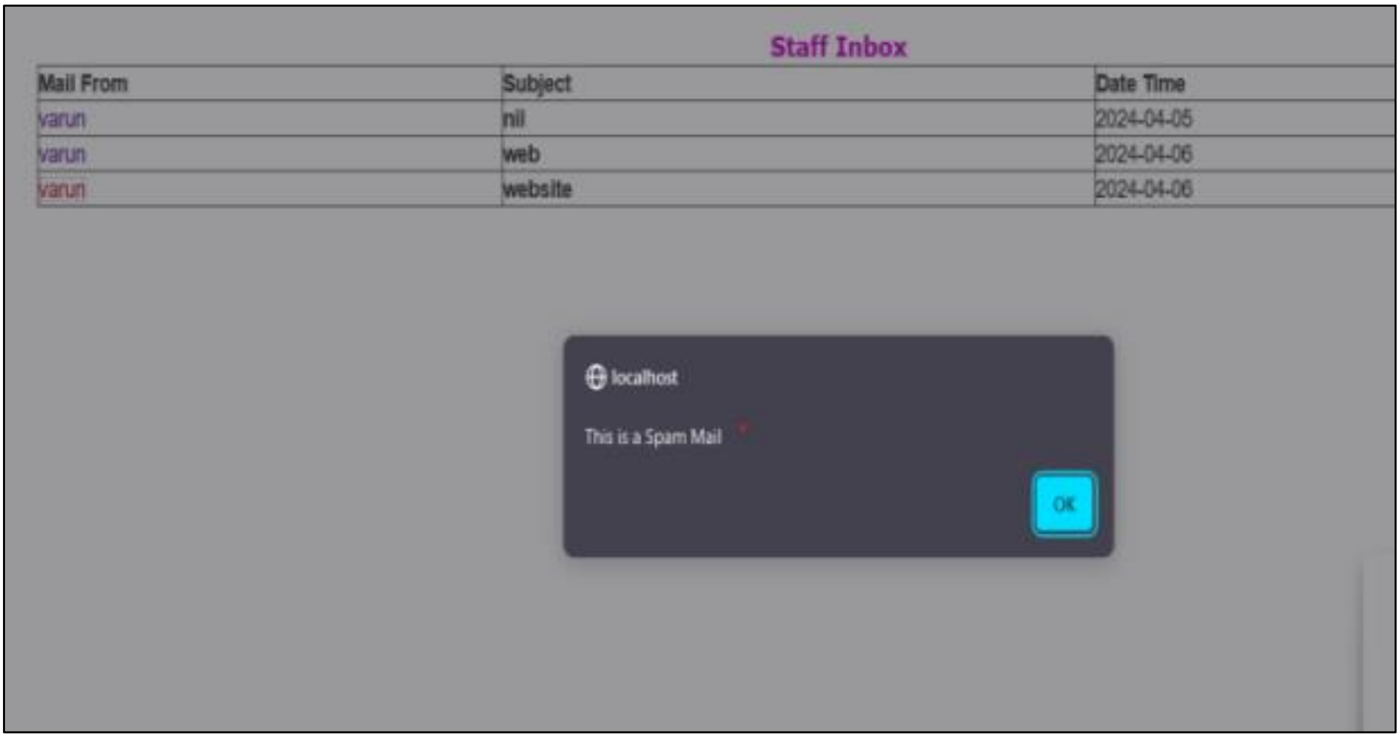
Class: -- Not set --

Fig 8 Converting Malicious Link into Normal Link

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2269

The above image shows the process of injecting a spam link which is converted into normal link and sent to the respective victim who thinks as genuine link and falls into trap.

So, Our project helps to identify these kind of spam just the process is to login as a user in the platform and automaticallythe site shows whether it is a genuine link or malicious link.



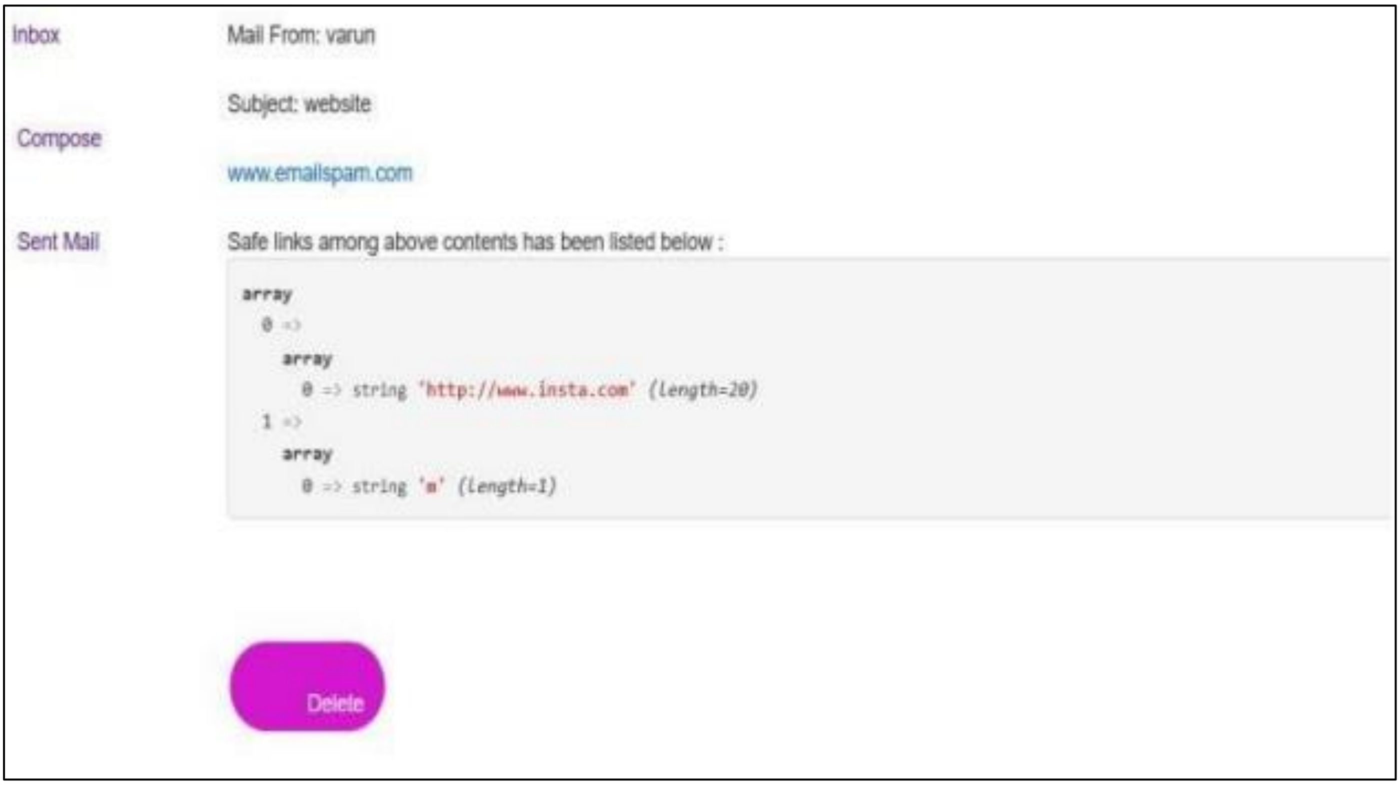
Staff Inbox

Mail From	Subject	Date Time
varun	nil	2024-04-05
varun	web	2024-04-06
varun	website	2024-04-06

localhost

This is a Spam Mail

Fig 9 Spam Alert



Mail From: varun

Subject: website

www.emailspam.com

Safe links among above contents has been listed below :

```
array
  0 =>
    array
      0 => string 'http://www.insta.com' (length=20)
  1 =>
    array
      0 => string 'm' (length=1)
```

Fig 10 Output

When the user clicks the button [Ok] in the above page its directs to the new site where the real name of the link is displayed to the user it can be used for further complaining process and proceeding to legal actions

VI. CONCLUSION

Phishing has emerged as a significant concern, resulting in network security issues and leading to substantial financial losses for both consumers and e-commerce companies, totalling billions of dollars. As a result, phishing has eroded trust in e-commerce and diminished its appeal to regular consumers. This paper thoroughly examines the characteristics of the hyperlinks contained in phishing emails. An anti-phishing algorithm is created from the given features. Since Link-Guard is signature-based, it is effective against unknown attacks as well as detecting attacks. Our tests show that Link Guard is powerful and can instantly detect 96% of unknown phishing attacks.

Since we use this method by taking the URL and registratio n model into consideration, different models need to be usedin the future.

REFERENCES

[1]. Mohamed Abdelkarim Remmide, FatimaBoumahdi, Narhimene Boustiaa, Chalabia Lilia Feknous, Romaissa Della DETECTION OF PHISHING URLS USING TEMPORAL CONVOLUTIONAL NETWORK, Universit´ eBLIDA1, Laboratoire LRDSI, Facult´ eSciences, B.P270, Route de Soumaa, BLIDA, ALGERIE

[2]. Abdulhamit Subasi, Emir Kremic, COMPARISON OF ADABOOST WITH MULTIBOOSTING FOR PHISHING WEBSITE DETECTION, Effat University, College of Engineering, Jeddah, 21478, Saudi Arabia, Federal Institute of Statistics, Sarajevo. 71000, Bosnia and Herzegovina

[3]. Asadullah Safi, Satwinder Singh, A SYSTEMATIC LITERATURE REVIEW ON PHISHING WEBSITE DETECTION TECHNIQUES, Nangarhar University, Ministry of Higher Education, Afghanistan, Dept. of Computer Science & Technology, Central University of Punjab, Bathinda, Punjab, India

[4]. Saad Al-Ahmadi, Yasser Alharbi, A DEEP LEARNING TECHNIQUE FOR WEB PHISHING DETECTION COMBINED URL FEATURES AND VISUAL SIMILARITY, College of Computer and Information Science, Computer Science Department, King Saud University, Riyadh, Saudi Arabia

[5]. Ayesha Arshad, Attique Ur Rehman, Sabeen Javaid, Javed Anjum Sheikh, Muhammad Azeem, Tahir Muhammad Ali, A SYSTEMATIC LITERATURE REVIEW ON PHISHING AND ANTI-PHISHING TECHNIQUES, Department of Software Engineering, University of Sialkot, Sialkot, Pakistan, Department of Computer Science, Gulf University of Science and Technology, Kuwait

[6]. Jian Mao, Jingdong Bian, Wenqian Tian, Shishi Zhu, Tao Weic, Aili Li, Zhenkai Liang, Detecting Phishing Websites via Aggregation Analysis of Page Layouts, "School of Electronic and Information Engineering, Beihang University, Beijing 100183, China Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai 200240, China Baidu USA LLC., Sunnyvale, CA 94089, USA, Information Technology Service Center, China National Petroleum Corporation, Beijing 100007, China, School of Computing. National University of Singapore, Singapore 117417, Singapore

[7]. Dong-Jie Liu, Guang-Gang Geng, Xiao-Bo Jin , Wei Wang ,An efficient multistage phishing website detection model based on the CASE feature framework: Aiming at the real web environment, Computer Network Information Center, Chinese Academy of Sciences, Beijing, China b University of Chinese Academy of Sciences, Beijing, China c College of Cyber Security, Jinan University, Guangzhou 510632, China

[8]. Mouad Zouina1 and Benaceur Outtaj A novel lightweight URL phishing detection system using SVM and similarity index An efficient multistage phishing website detection model based on the CASE feature framework: Aiming at the real web environment Mohammed V University of Rabat, Rabat, Morocco

[9]. ABB Corporate Research P. Cunha, J. Brandão, J. Vasconcelos, F. Soares and V. Carvalho, "Augmented reality for cognitive and social skills improvement in children with ASD," 20 13th International Conference on Remote Engineering and Virtual Instrumentation (REV), Madrid, 2020, pp. 334-335, doi: 10.1109/REV.2016.7444495.

[10]. Chandrasekaran, M., Narayanan, K., Upadhyaya, S.: Phishing email detection based on structural properties. In: NYS CyberSecurity Conf.

[11]. Garera, S., Provos, N., Chew, M., Rubin, A.: A framework for detection and measurement of phishing attacks. In: Proc. 2007 ACM Workshop on Recurring Malcode, pp. 1–8.

[12]. Abu-Nimeh, S., Nappa, D., Wang, X., Nair, S., 2007. A comparison of machine learning techniques for phishing detection. In: Anti-Phishing Working Groups Ecrime Researchers Summit. pp.

[13]. Bottazzi, G., Casalicchio, E., Cingolani, D., Marturana, F., Piu, M., 2015. MP-shield: A framework for phishing detection in mobile devices. Proceedings - 15th IEEE International Conference on Computer and Information Technology, CIT 2015, 14th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2015, 13th IEEE International Conference on Dependable, Autonomic and Se, 1977– 1983.

[14]. Lotter A., Futcher L.: A Framework to assist Email users in the identification of Phishing Mails. In: Proc. 8th Int‟l Symposium on Human Aspects of Information Security and Assurance.

[15]. Wardman, B., Stallings, T., Warner, G., Skjellum, A., Nov 2011. High- performance content-based phishing attack detection. In: eCrime Re- searchers Summit. IEEE, San Diego, CA,