Received 25 April 2021; Accepted 08 May 2021.

Detecting E Banking Phishing Websites Using Associative Classification PHP

Arelli Rithvik [1], Nadella Shyam Sudheer [2], Dr. V. Anbarasu[3]

[1,2]UG Student, Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur

[3]Associate Professor, Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur
aa1843@srmist.edu.in [1], nn8590@srmist.edu.in [2], anbarasv2@srmist.edu.in[3]

Abstract

The term “phishing” is an old type network attack where the phisher creates a duplicate page having similar structure to the original verified page , the idea of this is to trick the users believing the page is their service provider to give out their sensitive information like bank details and password of their personal things they not only steal the bank information things to take out money but also might steal data to use them for their next step. This concept used in this is based on an end-host, this anti-phishing algorithm is also known as link-guard algorithm, when the phisher sends the email or message to the user they can input the link in this algorithm ,by evaluating the characteristics of the URL sent in that email it ranks how safe the site is and has various criteria that might put the threat at different levels. The Link Guard is based on URL characteristics so it can detect and forestall unknown or new ones. This project uses the PHP, WAMP Server as Front-End and Back-End respectively.

Keywords: Phishing; Associative Classification; link guard; WAMP.

1. Introduction

The word “Phishing” is derives from the word 'fishing' as they are similar for fish that mistakes the bait from the attacker for a worm and goes into a trap, similarly in online the attacker sends a link through a mail or message similar to legit but actually is not. It is a trap to take the information from the user. Here the user is attracted by the attacker as they are calling from their service provider like bank and ask to share the sensitive information which may be data or the transaction details. The attacker collects the information if we enter it in the link

Though it is not a new concept. In the last 2 years there has been an increase in these cases

In our experiment we can say that phishing links might following characteristics:

1. The visual link sent in mail and the actual link opened are not the same.
2. The phishers might use a kind of dotted decimal IP address.
3. Fake DNS names that are similar but not the same with the target Web site.
4. tricks are used to encode the hyperlinks.

2.Significance Of The Study

Online shopping has become a trend now and of course it is really great because it is very easy to shop and very time saving. But the problem here is the transaction in the websites. The transactions seem to be safe but they may not be. The whole process seems to be legit but your personnel information like card numbers, CVV might be stolen by the hackers. The hackers make mimic banking websites which seems to be like your original banking websites, that’s the whole point. The user thinks this is a legitimate site and enters his/her personnel details and the details will directly be delivered to the hacker. These mimic websites are technically called as Phishing websites. Here we are using a classification algorithm to find out if a particular website is phishing or not. Their URL , domain address are used in examining the website. This web application can be used by any e-commerce website so that the users can safely shop from their sites and this software can also be embedded as an extension. So that the user can directly detect a website by just clicking on a button and the extension gives us the result using data mining which is better than the traditional system.

3.Review of Related Studies

Throughout the research process we have been able to find some good research papers as the foundation of our project, to learn and improvise from those phenomenal works.

[1] This research paper named Phishing detection based on search engine-based techniques. The authors of this paper are Adida, Hohenberge. This technique tracks all the images, texts present in the actual website and it

compares with the malicious site. It also checks its popularity in the search engine i.e., where will it be placed in the list of websites when searched for it. It is very mobile and lightweight. It can be added as an extension. It does not require much storage too.

[2] The next research paper is based on the machine learning techniques whose author is Sharif. In this technique the machine is learned to track all the details such as the content present in the page, the URL addresses. The use of machine learning is fully accurate and makes the scheme able to adapt. Under this technique there are algorithms used to identify best and unique things found in the page and assign weights to it.

[3] The next research paper is titled as Phishing detection using blacklist and whitelisted databases whose author is Dhamija and Tygar. Here the author used the concept of blacklist and whitelist where blacklist holds all the phishing URL links and white list has the legit sources stored as a database in a server, now we can compare the link inputted by user with blacklist and white list and may find a match, this is a lightweight system and can run in a browser smoothly.

[4] The fourth research paper is based on the visual similarity techniques done by Wu, Miller, and little. In this paper the author uses the visual similarity between the legit and the fake page by detecting the any visual differences between them, we have to extract the visual features and compare for these differences between them to know if the inputted website is legit or not this can even be done at client side with some software.

[5] The last paper is based on the most important thing, the DNS based techniques. [11][12]The author of this paper is Bridges and Vaughn. This is a very efficient way as this require a very less extraction needed from the source and is possible to apply on client side as more storage is possible this used the IP address; this uses DNS information to verify the authenticity of a website.

All the of the mentioned research papers were thoroughly studied and we came out of the best solution to the problem. Different limitations were observed over the various research papers and was contrasted in respect with our project to bring up the best possible result.

4. Objectives of the Study

People often purchase products online and make payment through e-banking. There are many E-banking phishing websites. To detect the e-banking phishing website our system uses an effective classification data mining algorithm. 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.

- 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.
- This application can be used by many E-commerce enterprises in order to make the whole transaction process secure.
- Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms.
- System uses machine learning technique to add new keywords into database.

5. Proposed Methodology

Online shopping has become a trend now and of course it is really great because it is very easy to shop and very time saving. But the problem here is the transaction in the websites. The transactions seem to be safe but they may not be. The whole process seems to be legit but your personnel information like card numbers, CVV might be stolen by the hackers. The hackers make mimic banking websites which seems to be like your original banking websites, that’s the whole point. The user thinks this is a legitimate site and enters his/her personnel details and the details will directly be delivered to the hacker. These mimic websites are technically called as Phishing websites. Here we are using a classification algorithm to find out if a particular website is phishing or not. Their URL , domain address are used in examining the website. This web application can be used by any e-commerce website so that the users can safely shop from their sites and this software can also be embedded as an extension. So that the user can directly detect a website by just clicking on a button and the extension gives us the result using data mining which is better than the traditional system

Pattern matching is a concept to handle unknown sites, These are totally new and we don’t know whether to trust this sites or not by liking at black-list or the white-list, all we have is a actual link as visual link does not contain any DNS or IP

We collect the DNS names from the mail sent which might be false and the other from the user , and second step is they take input from user and save names in seed-set

The DNS names of the visual and actual are compared if not identical then we call another method:

The similarity index is calculated by the most minimum number of change needed to change the original link to the fake link

If two strings are 0 then we give it equal and identical

If two strings has small changes then they are almost similar but not identical

And if the strings have a big changes then they are not too similar

Number of changes is inversely proportional to similarity

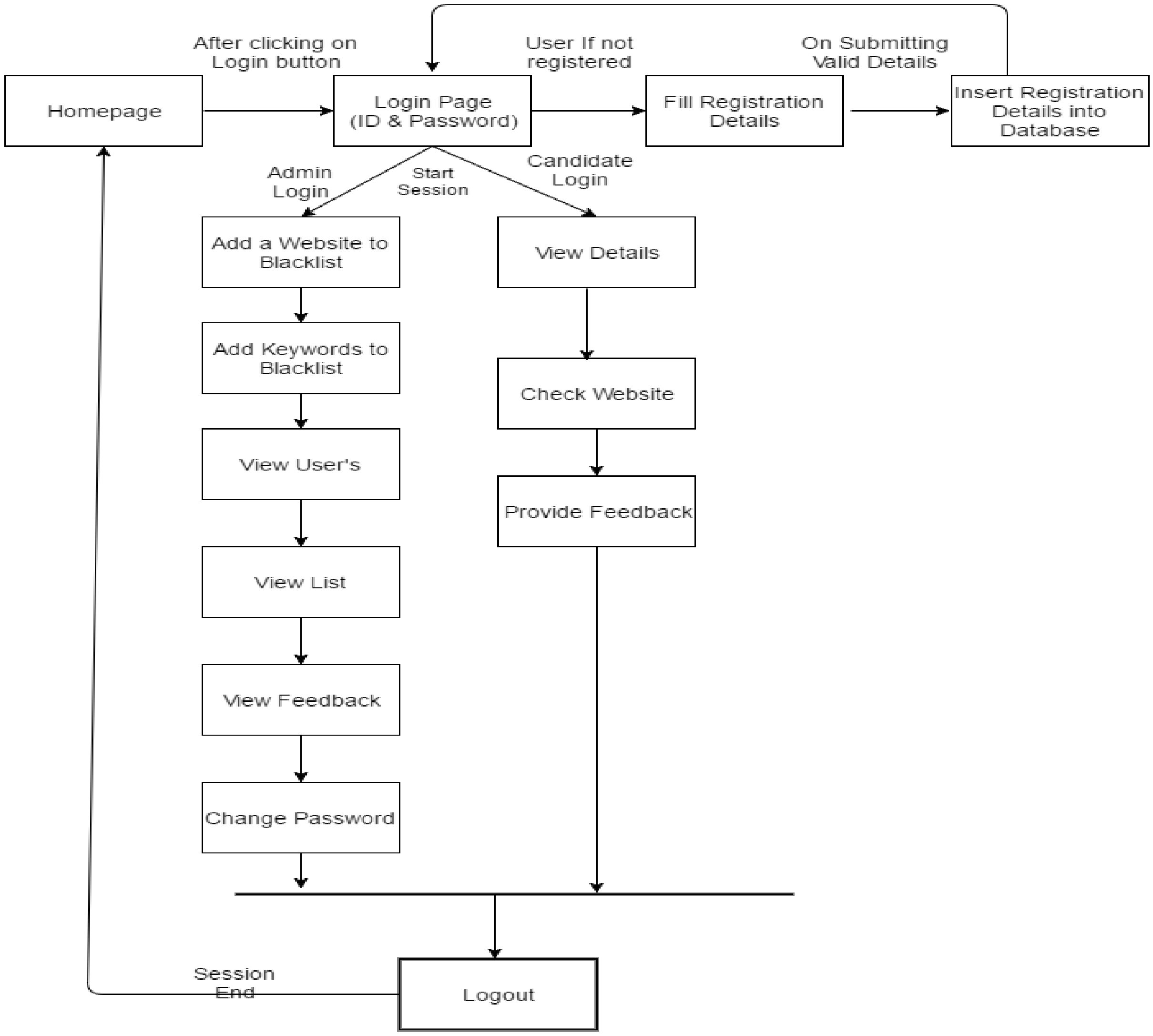
Here is aexample :

Shyam and Shuom has 3/5(2 characters are meant to replace)

Rithvik and rithvikkkk has 7/10(3 characters has to be removed)

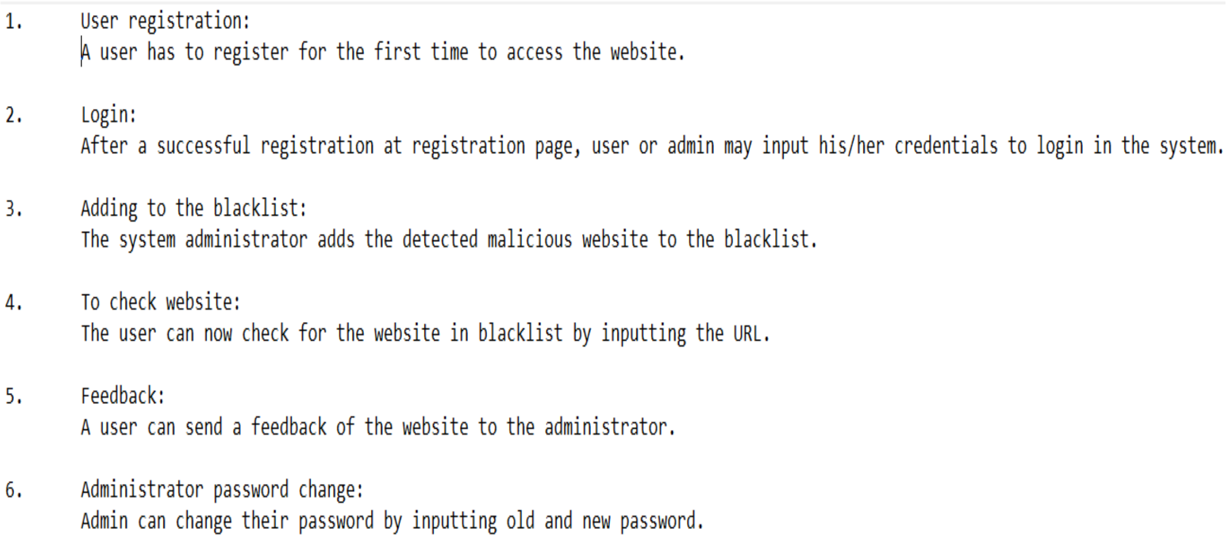
144449 and 1444449 is 6/7 (we need to insert a 4 to match it).

System Architecture of Proposed System:


After clicking on Login button
User If not registered
On Submitting Valid Details
Homepage
Login Page (ID & Password)
Fill Registration Details
Insert Registration Details into Database
Admin Login
Start Session
Candidate Login
Add a Website to Blacklist
Add Keywords to Blacklist
View User's
View List
View Feedback
Change Password
View Details
Check Website
Provide Feedback
Session End
Logout

Received 25 April 2021; Accepted 08 May 2021.

6.Result and Discussion



1. User registration:
A user has to register for the first time to access the website.

2. Login:
After a successful registration at registration page, user or admin may input his/her credentials to login in the system.

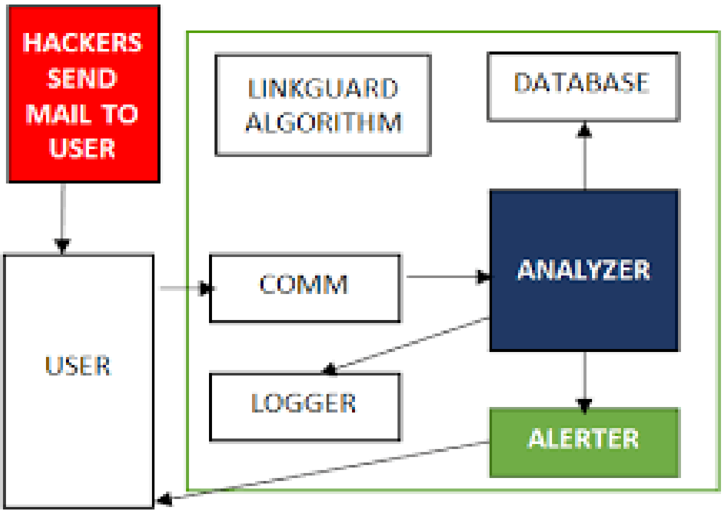
3. Adding to the blacklist:
The system administrator adds the detected malicious website to the blacklist.

4. To check website:
The user can now check for the website in blacklist by inputting the URL.

5. Feedback:
A user can send a feedback of the website to the administrator.

6. Administrator password change:
Admin can change their password by inputting old and new password.

The Proposed Algorithm:
This Link Guard examines the differences between the actual URL link and visual URL link.

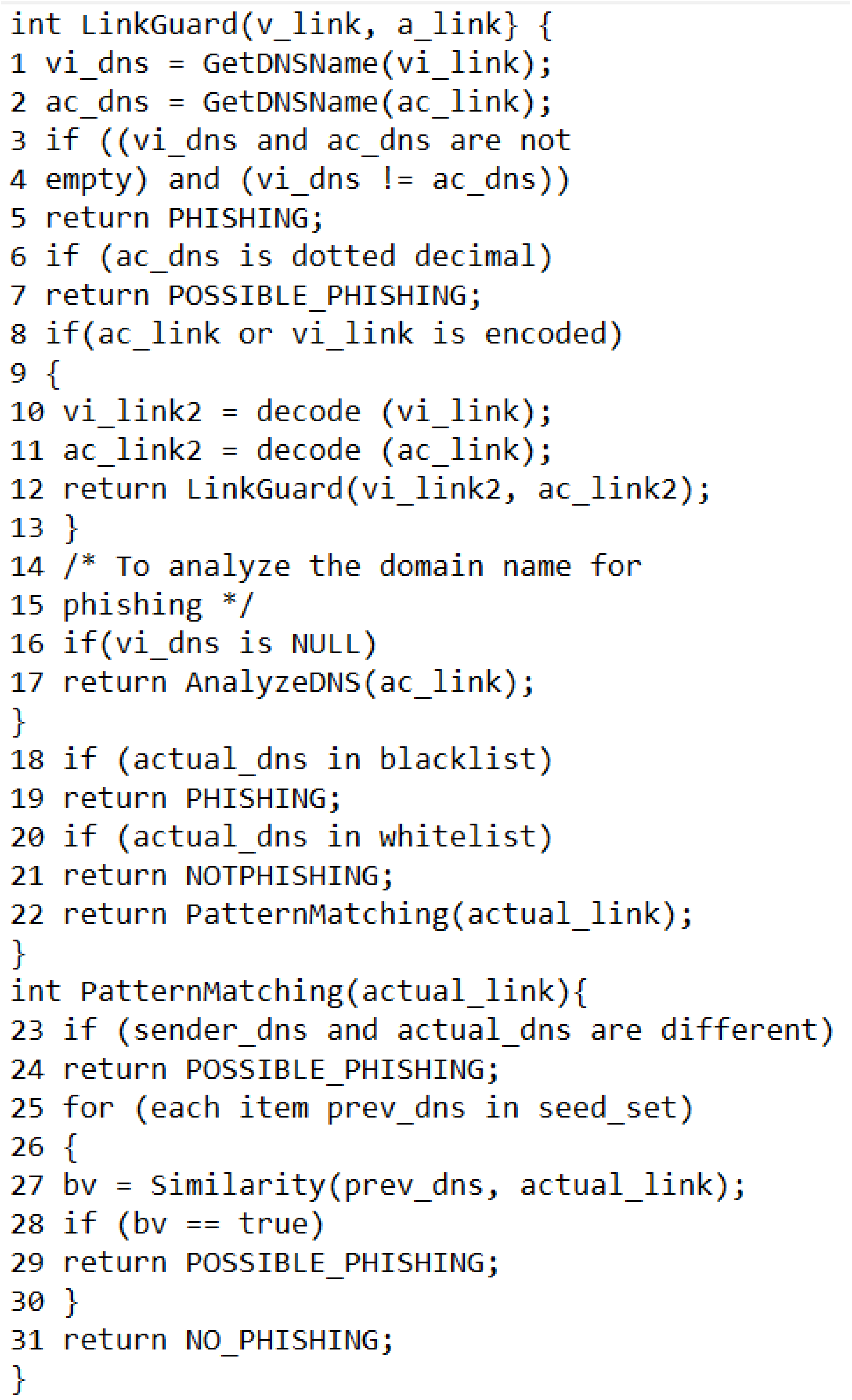


Operations of Link Guard algorithm:

1) Comm: All the information from the user is gathered by comm and it is bought to the analyser.
2) Database: The Url’s inputted by the user is preserved in the database
3) Analyser: Is applied on link guard algorithm. Certain data is collected by comm and the database which is utilized by the analyzer to examine and it is sent to the log modules.
4) Alerter: If any malicious act is detected it alerts the user. It takes the help of the Alerter.
5) Logger: All the details or information required for later use is archived by the logger.

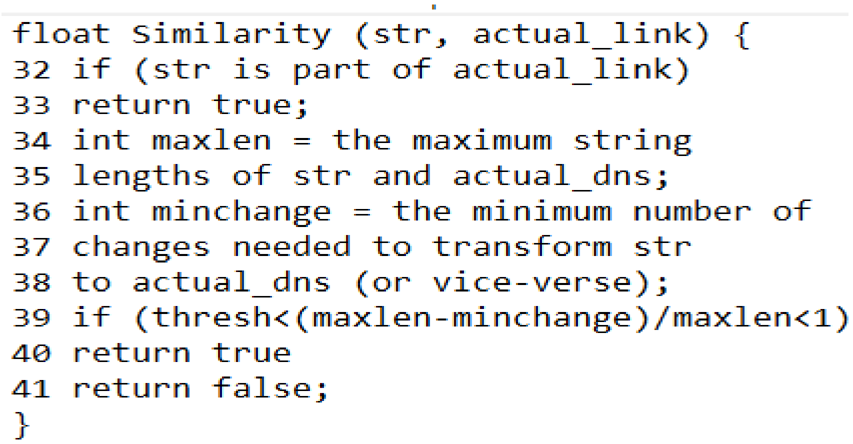
The following terms are used in the algorithm.
vi_link: visual link;
ac_link: actual_link;
vi_dns: visual DNS name;
ac_dns: actual DNS name;
sender_dns: sender’s DNS name.

Algorithm:



```
int LinkGuard(v_link, a_link} {
vi_dns = GetDNSName(vi_link);
ac_dns = GetDNSName(ac_link);
if ((vi_dns and ac_dns are not
empty) and (vi_dns != ac_dns))
return PHISHING;
if (ac_dns is dotted decimal)
return POSSIBLE_PHISHING;
if(ac_link or vi_link is encoded)
{
vi_link2 = decode (vi_link);
ac_link2 = decode (ac_link);
return LinkGuard(vi_link2, ac_link2);
}
/* To analyze the domain name for
phishing */
if(vi_dns is NULL)
return AnalyzeDNS(ac_link);
}
if (actual_dns in blacklist)
return PHISHING;
if (actual_dns in whitelist)
return NOTPHISHING;
return PatternMatching(actual_link);
}
int PatternMatching(actual_link){
if (sender_dns and actual_dns are different)
return POSSIBLE_PHISHING;
for (each item prev_dns in seed_set)
{
bv = Similarity(prev_dns, actual_link);
if (bv == true)
return POSSIBLE_PHISHING;
}
return NO_PHISHING;
}
```

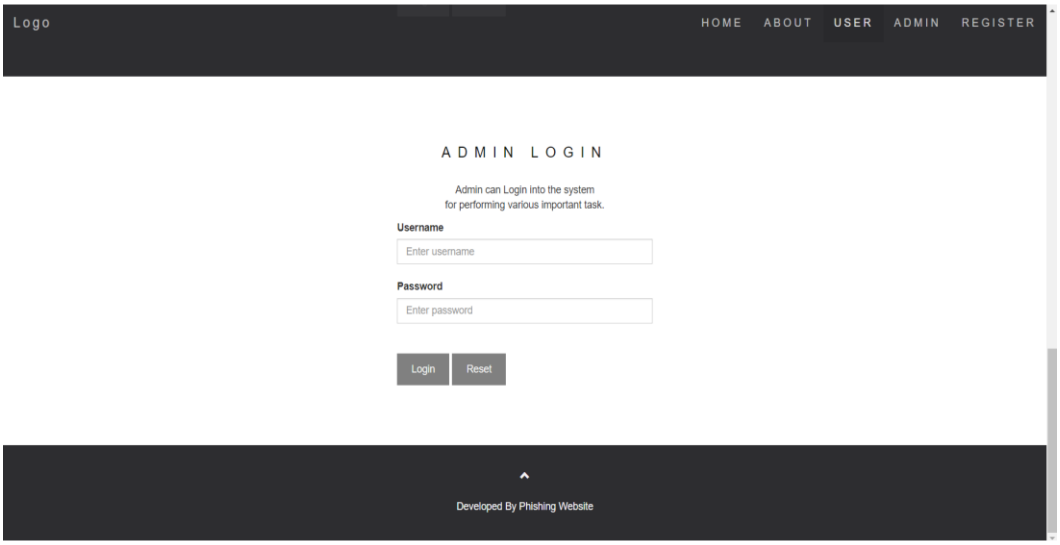
Received 25 April 2021; Accepted 08 May 2021.



```
float Similarity (str, actual_link) {
if (str is part of actual_link)
return true;
int maxlen = the maximum string
lengths of str and actual_dns;
int minchange = the minimum number of
changes needed to transform str
to actual_dns (or vice-verse);
if (thresh<(maxlen-minchange)/maxlen<1)
return true
return false;
}
```

This shows the skeletal structure of the implemented algorithm.

Screenshots of the result from the Project:

[1] 

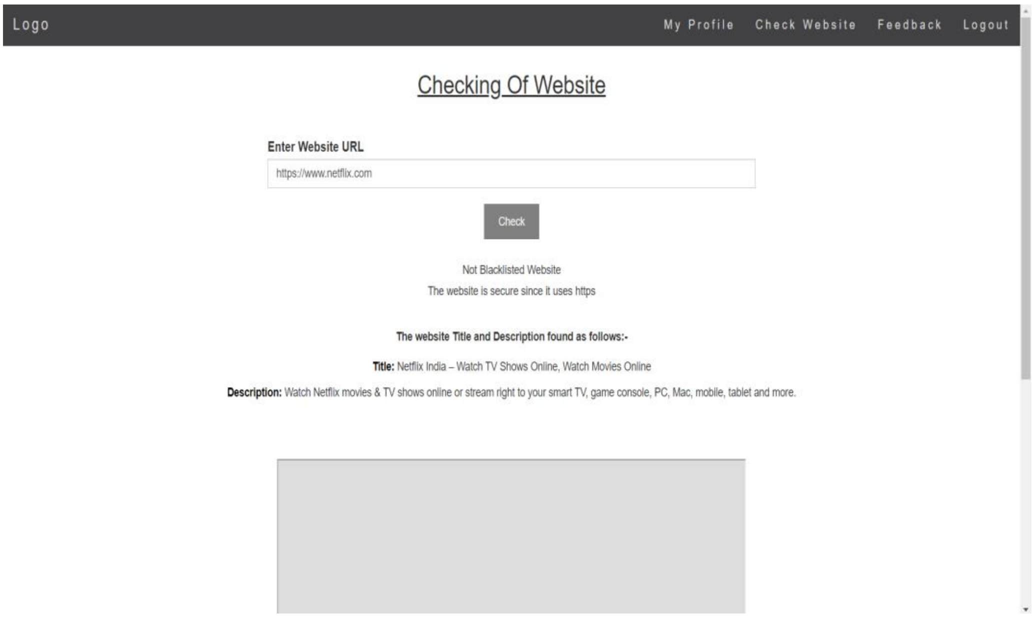
ADMIN LOGIN

Admin can Login into the system
for performing various important task.

Username: Enter username

Password: Enter password

Developed By Phishing Website

[2] 

Checking Of Website

Enter Website URL: https://www.netflix.com

Not Blacklisted Website

The website is secure since it uses https

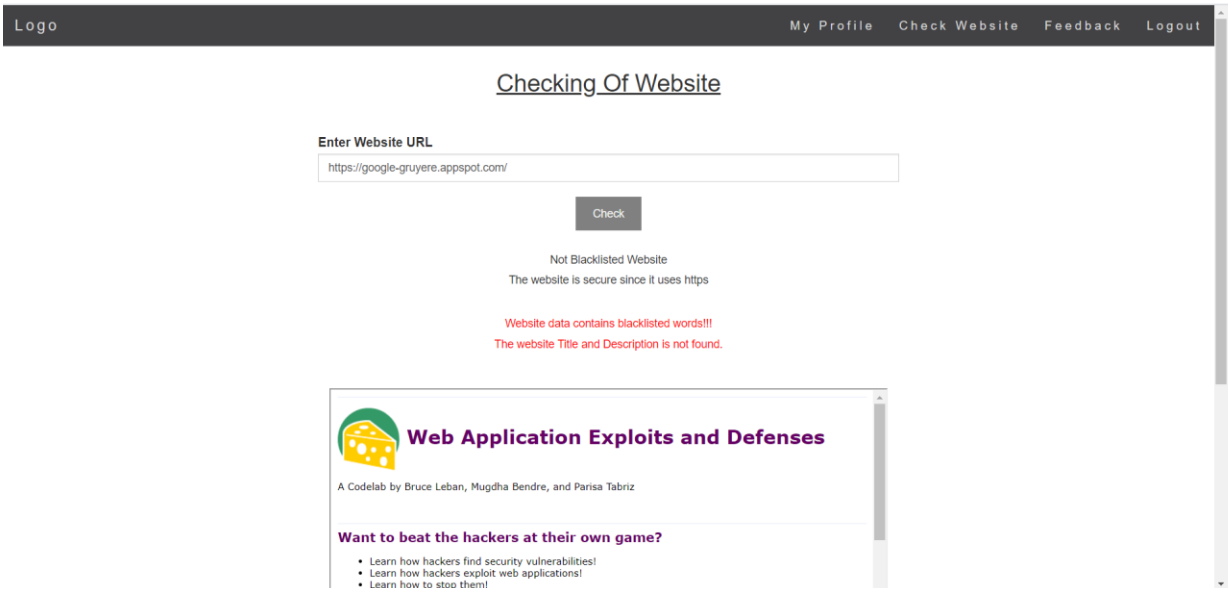
The website Title and Description found as follows:-

Title: Netflix India – Watch TV Shows Online, Watch Movies Online

Description: Watch Netflix movies & TV shows online or stream right to your smart TV, game console, PC, Mac, mobile, tablet and more.

Received 25 April 2021; Accepted 08 May 2021.

[3]



Checking Of Website

Enter Website URL

https://google-gruyere.appspot.com/

Not Blacklisted Website

The website is secure since it uses https

Website data contains blacklisted words!!!

The website Title and Description is not found.

Web Application Exploits and Defenses

A Codelab by Bruce Leban, Mugdha Bendre, and Parisa Tabriz

Want to beat the hackers at their own game?

- Learn how hackers find security vulnerabilities!
- Learn how hackers exploit web applications!
- Learn how to stop them!

7.Conclusion

We can know from the data and cases around us that "Phishing" has become a serious network security problem and is still growing, causing finical loss to both consumers and the e-commerce websites making online money transaction or who exchange sensitive online to provide services like banks. Phishing has made e-commerce less popular in normal consumers as they are scared to become the victims of phishing losing their money while they transact the money. Here, we have studied the characteristics of the phishing hyperlinks.

We then designed a anti-phishing algorithm by studying characteristics of the URL. Since E Banking Phishing Website is characteristic based, we can detect unknown phishing links too. Experiment in our system testing in real life showed that Link Guard can detect up to 96 percentage of unknown phishing we conclude that the algorithm can shield our users from malicious or unverified links in any form like e-mails/messages/popups/webpages , They also can be added to e-commerce sites to validate the site before they provide information..

8. References

[1] M. Aburrous and K. Dahal, "Intelligent phishing detection system for e- banking using fuzzy data mining." Jan. 2016.

[2] T. Moore and R. Clayton, "An empirical analysis of the current state of phishing attack and defence," vol. 1, Jan. 2017.

[3] B. Adida, S. Hohenberger, and R. Rivest, "Lightweight encryption for email" USENIX Steps to Reducing Unwanted Traffic on the Internet (SRUTI) (ICTer), Mar. 2015.

[4] SadiaAfroz, Rachel Greenstadt, "Phishzoo: Detecting phishing websites by looking at them", 2011 IEEE fifth international conference onSemantic computing, 368-375, 2011

[5] Moitrayee Chatterjee, Akbar-SiamiNamin, "Detecting phishing websites through deep reinforcement learning", 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC) 2, 227-232, 2019

[6] Mehdi Babagoli, Mohammad PourmahmoodAghababa, VahidSolouk, "Heuristic nonlinear regression strategy for detecting phishing websites", 227-232, 2019

[7] RouthuSrinivasa Rao, Alwyn R Pais, "Detecting phishing websites using automation of human behaviour", Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security, 33-42, 2017

[8] Luong Anh Tuan Nguyen, Ba Lam To, HuuKhuong Nguyen, Minh Hoang Nguyen, "Detecting phishing web sites: A heuristic URL-based approach", 2013 International Conference on Advanced Technologies for Communications (ATC 2013), 597-602, 2013

[9] VaibhavPatil, Pritesh Thakkar, Chirag Shah, TusharBhat, SP Godse, "Detection and prevention of phishing websites using machine learning approach", 2018 Fourth international conference on computing communication control and automation (ICCUBEA), 1-5,2018

[10] Shraddha Parekh, Dhwanil Parikh, SrushtiKotak, SmitaSankhe, 'A new method for detection of phishing websites: URL detection, 2018 Second international conference on inventive communication andcomputational technologies (ICICCT), 949-952, 2018.

[11] Ramkumar J, M. Baskar, M Viswak, M D Ashish, "Smart Shopping with Integrated Secure System based on

IoT ", International Journal of Advanced Science and Technology, Vol. 29, No. 5, pp: 301-312, ISSN: 2005-4238, April 2020.

[12] Ramkumar J, M. Baskar, K. Ravishankar, Venkateswara Reddy Yakkanti,“ Health monitoring through pills dispenser for Alzhiemer disease based on IoT”, International Journal of Advanced Science and Technology , Vol. 29, No. 4,pp: 1810-1818, ISSN: 2005-4238, April 2020