Web Vulnerability Detection

V.G SaranyaVaishalini, P. Gayathri, V. Lavanya

[1] Assistant Professor, [2, 3] Student,
CSE, P.S.R Engineering College,
Sivakasi, TamilNadu, India.

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 12, Issue 1, (January-February 2024)

Abstract

Malicious URLs pose a significant cybersecurity threat due to their potential to distribute malware, steal personal information, and launch phishing attacks. Conventional methods of detecting malicious URLs, such as blacklists and heuristics, are becoming less effective as attackers develop new evasion techniques. This study introduces a novel approach using Multilayer Perceptron (MLP) to quantify and predict the behavior of Malicious Web Services. This approach not only measures but also predicts the response time of these services, allowing for a quantitative ranking rather than a qualitative assessment. The proposed methodology aims to automatically select the most reliable Malicious Web Service by considering metrics like system predictability and response time variability. Through the use of real-world data and experiments, the researchers demonstrate the feasibility and usefulness of their approach.

Keywords: Machine Learning, Malicious URL Detection, Adversarial Attacks, Malicious Web Services.

Introduction

Machine Learning

Machine learning, a branch of artificial intelligence (AI), enables computers to learn without explicit programming. Through training on data, computers can identify patterns and make predictions. Machine learning algorithms find applications in various domains, including spam filtering, fraud detection, product recommendation, and image recognition. These algorithms can be categorized into supervised learning, unsupervised learning, and reinforcement learning. Machine learning proves to be a powerful tool for solving diverse problems. However, it is crucial to acknowledge that the effectiveness of machine learning algorithms heavily relies on the quality and completeness of the training data. Biased or incomplete training data can lead to biased or inaccurate predictions.

Malicious URL Detection

Malicious URL detection involves the identification of URLs that direct users to malicious websites. These websites can distribute malware, steal personal information, or launch phishing attacks. Traditional methods of malicious URL detection, such as blacklists and heuristics, are progressively losing their effectiveness as attackers develop new evasion techniques. Machine learning presents a promising approach to address this issue. By training machine learning algorithms, patterns in malicious URLs that are difficult for humans to detect can be identified. These patterns may include the utilization of specific keywords or domains, the presence of suspicious characters in the URL, and the reputation of the hosting website.

Adversarial Attacks

Adversarial attacks involve carefully crafted inputs that aim to deceive machine learning algorithms, causing them to make errors. These attacks can target various types of machine learning algorithms, including those used for image classification, object detection, and natural language processing. Adversarial attacks are typically created by making subtle changes to the input data, which may be imperceptible to humans. For instance, an adversarial attack on an image classification algorithm might entail adding a small amount of noise to the image. Although this noise may go unnoticed by humans, it can be sufficient to trick the algorithm into misclassifying the image. Adversarial attacks pose a significant threat to the security of machine learning systems. If successfully executed, an attacker could potentially gain control of the system or manipulate it to make harmful decisions.

Malicious Web Services

Malicious web services refer to web services that are intentionally designed to carry out malicious activities, such as distributing malware, stealing personal information, or launching phishing attacks. These services often masquerade as legitimate ones, making them challenging to identify. Malicious web services can be utilized in various ways, including, Malware distribution: Malicious web services serve as a means to distribute malware, such as viruses, trojans, and worms. This can be achieved by embedding malware within web pages, scripts, or other downloadable files. Personal information theft: Malicious web services are also employed to steal personal information, including names, addresses, credit card numbers, and Social Security numbers.

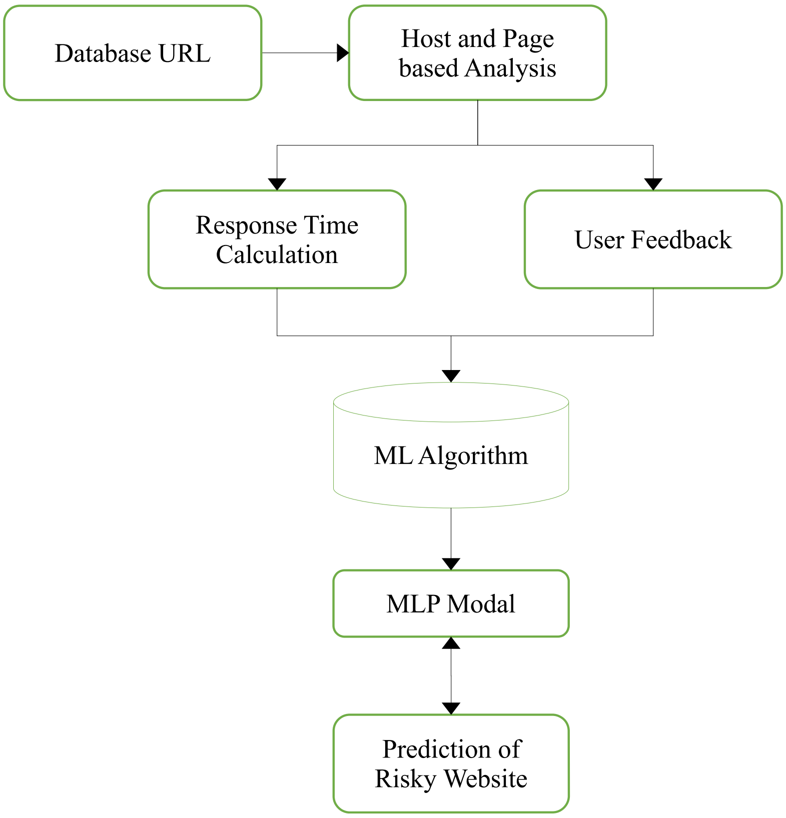
Existing System

In this article, a methodology is proposed to utilize Machine Learning (ML) in order to detect web application vulnerabilities. The analysis of web applications is particularly challenging due to their diversity and the widespread use of custom programming practices. Therefore, ML proves to be highly beneficial for enhancing web application security as it can incorporate the human understanding of web application semantics into automated analysis tools by leveraging manually labeled data. The proposed methodology is applied in the development of Mitch, which is the first ML solution designed for the black-box detection of Cross-Site Request Forgery (CSRF) vulnerabilities. Through the utilization of Mitch, a total of 35 new CSRFs were identified on 20 major websites, along with 3 new CSRFs on production software.

Proposed System

The proposed system utilizes a Multilayer Perceptron (MLP) to quantify and predict the behavior of Malicious Web Services (MWSs). Initially, a set of features is extracted from the behavior of MWSs, which can include factors such as response time, size, and content of the responses sent by MWSs. These extracted features are then inputted into the MLP model, which is trained to predict MWSs behavior specifically in terms of response time. Once the MLP model is trained, it can be utilized to rank MWSs in a quantitative manner. This ranking system aids in identifying the most reliable MWSs, which are those that are more likely to exhibit consistent and predictable behavior. The proposed system offers several benefits, including enhancing the accuracy of malicious URL detection systems, improving the performance of security applications reliant on MWSs, and reducing the cost and complexity associated with managing MWSs.

Architecture Diagram


Database URL
Host and Page based Analysis
Response Time Calculation
User Feedback
ML Algorithm
MLP Modal
Prediction of Risky Website

Response Time Calculation

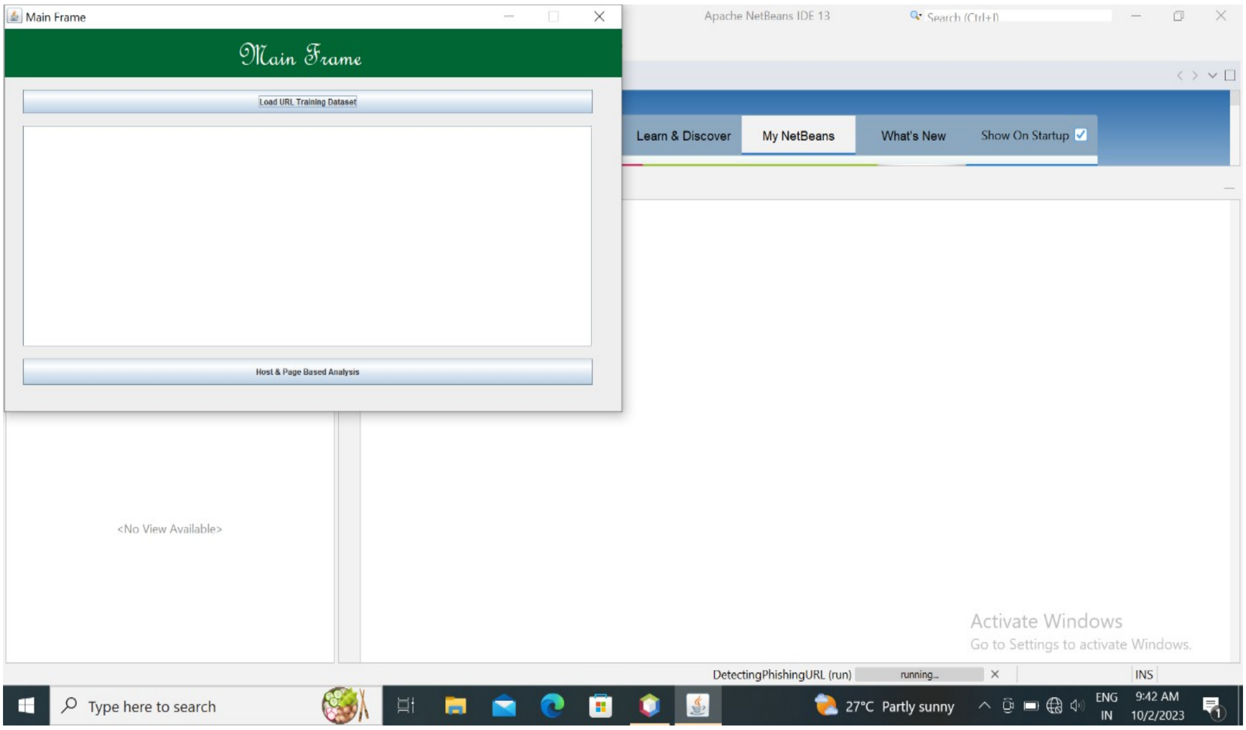
Response time is the total amount of time it takes to respond to a request for service. That service can be anything from a memory fetch, to a disk IO, to a complex database query, or loading a full web page. Ignoring transmission time for a moment, the response time is the sum of the service time and wait time. Response time may refer to:

$$\text{Response Time Calculation} = TP \div TP + FN$$

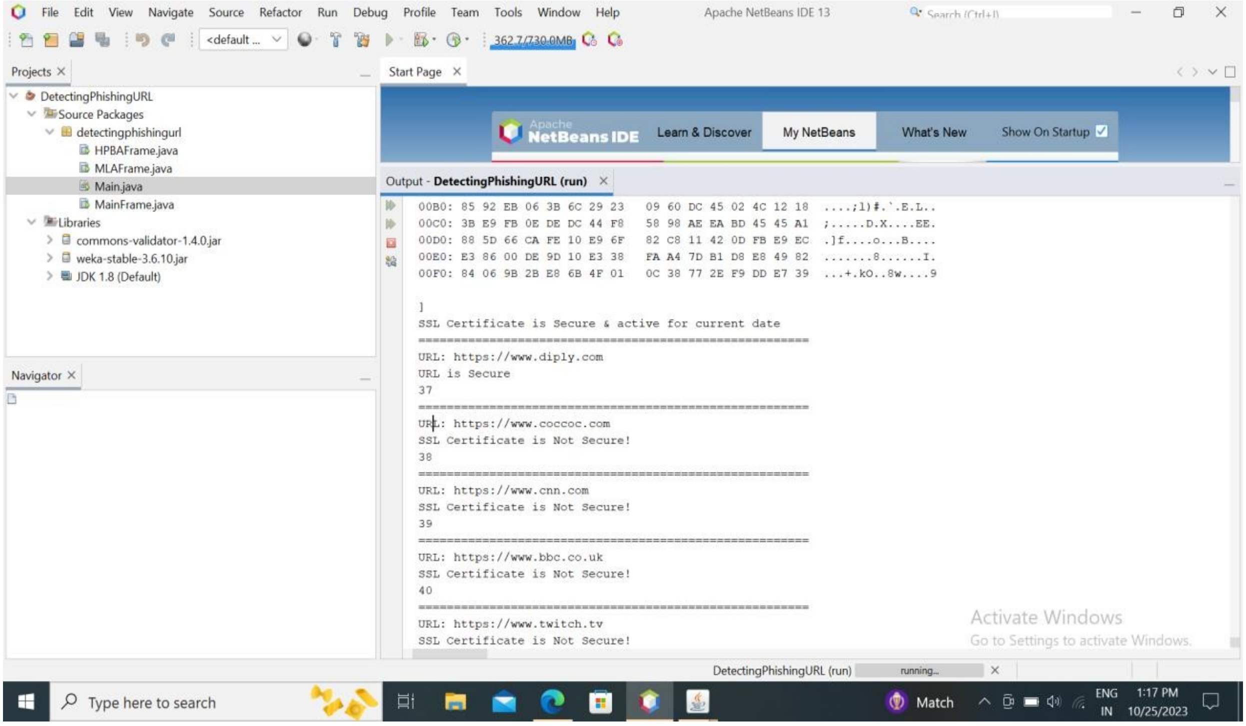
User Feedback

This module is used to add user feedback about Risky Web Services. Feedback is essential to the working and survival of all regulatory mechanisms found throughout living and nonliving nature, and in man-made systems such as education and economy.

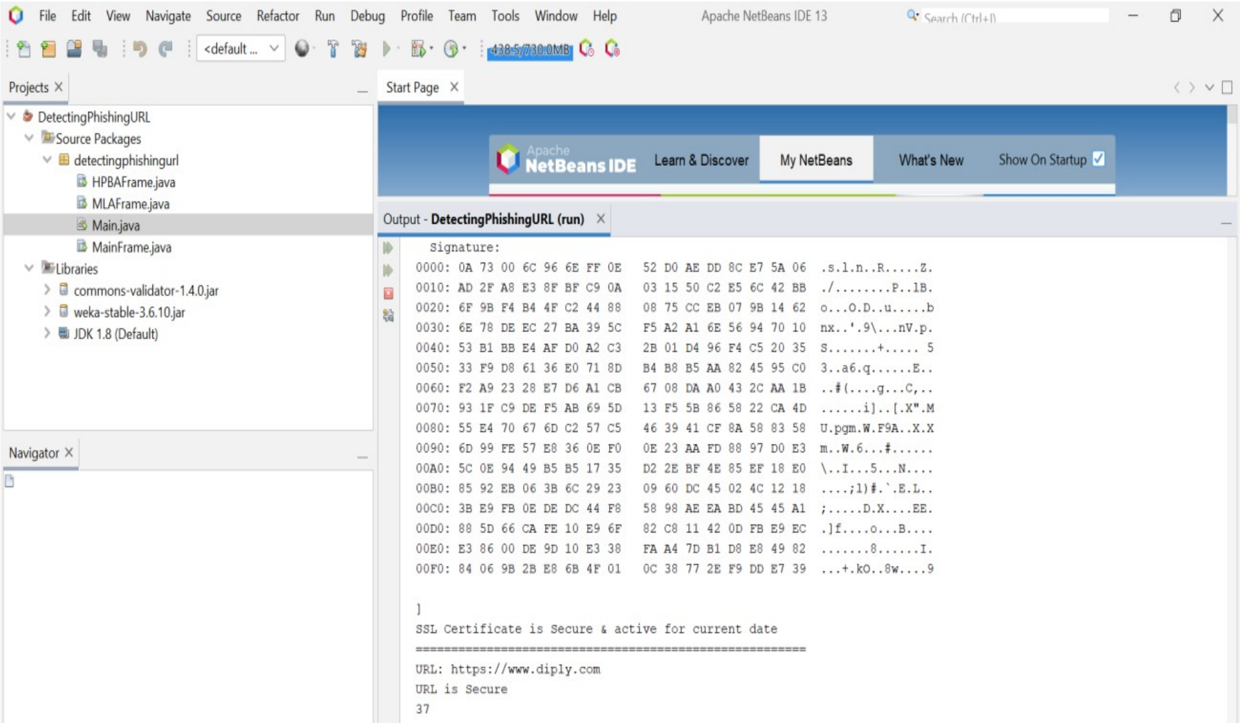
Result and Analysis



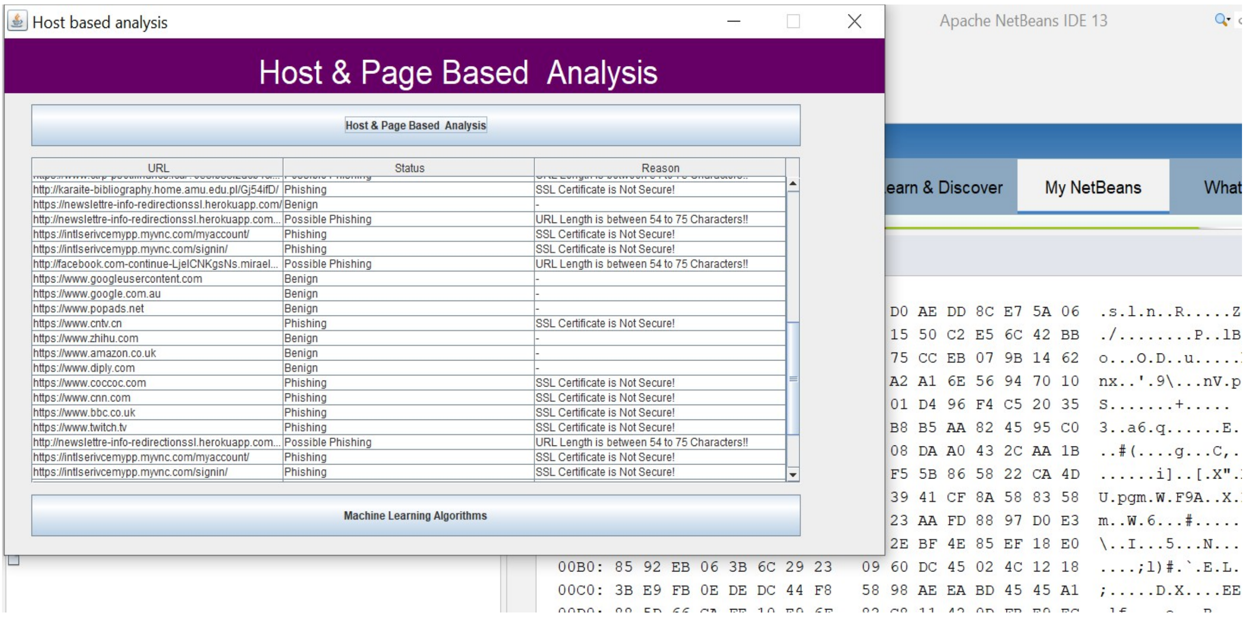
By clicking the load URL training Dataset, the dataset is chosen to train. After the dataset is loaded click Host and Page based analysis.



```
00B0: 85 92 EB 06 3B 6C 29 23   09 60 DC 45 02 4C 12 18  ....;l)#.`.E.L..
00C0: 3B E9 FB 0E DE DC 44 F8   58 98 AE EA BD 45 45 A1  ;.....D.X....EE.
00D0: 88 5D 66 CA FE 10 E9 6F   82 C8 11 42 0D FB E9 EC  .]f....o...B....
00E0: E3 86 00 DE 9D 10 E3 38   FA A4 7D B1 D8 E8 49 82  .......8......I.
00F0: 84 06 9B 2B E8 6B 4F 01   0C 38 77 2E F9 DD E7 39  ...+.kO..8w....9

]
SSL Certificate is Secure & active for current date
=======================================================
URL: https://www.diply.com
URL is Secure
37
=======================================================
URL: https://www.coccoc.com
SSL Certificate is Not Secure!
38
=======================================================
URL: https://www.cnn.com
SSL Certificate is Not Secure!
39
=======================================================
URL: https://www.bbc.co.uk
SSL Certificate is Not Secure!
40
=======================================================
URL: https://www.twitch.tv
SSL Certificate is Not Secure!
```



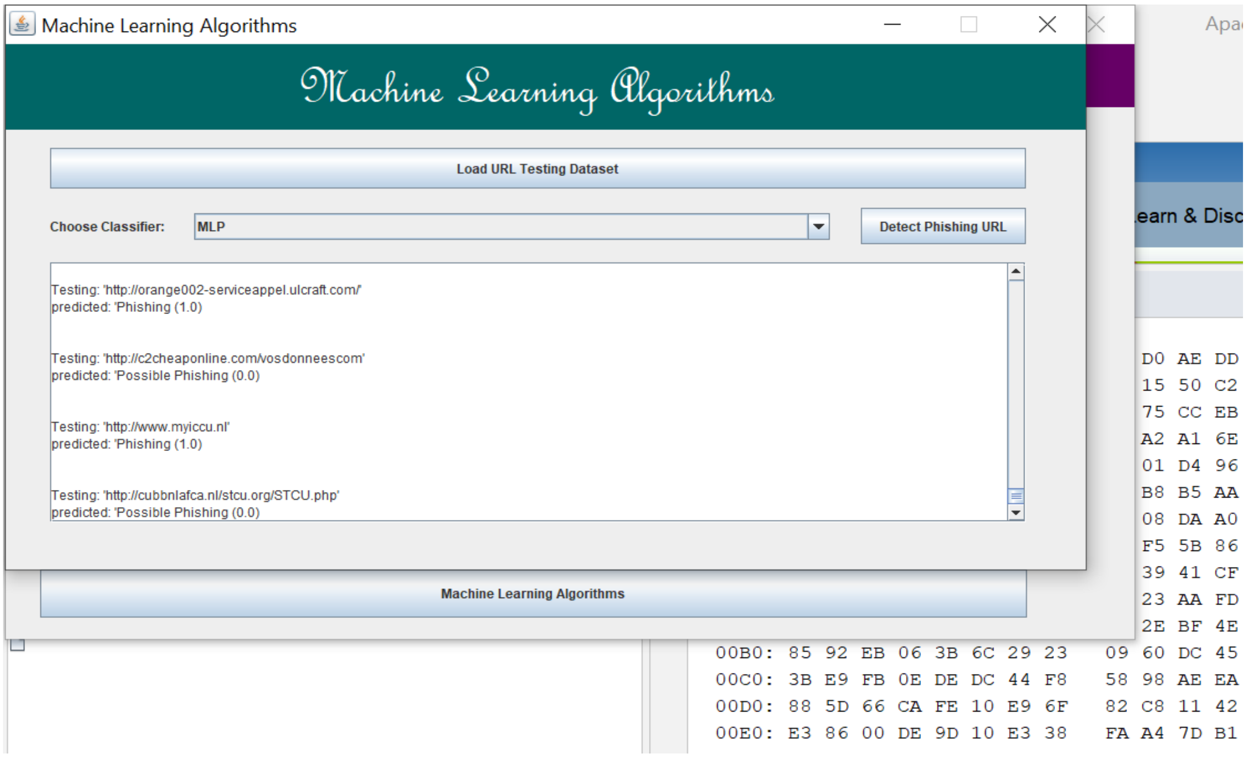
```
  Signature:
0000: 0A 73 00 6C 96 6E FF 0E   52 D0 AE DD 8C E7 5A 06  .s.l.n..R.....Z.
0010: AD 2F A8 E3 8F BF C9 0A   03 15 50 C2 E5 6C 42 BB  ./........P..lB.
0020: 6F 9B F4 B4 4F C2 44 88   08 75 CC EB 07 9B 14 62  o...O.D..u.....b
0030: 6E 78 DE EC 27 BA 39 5C   F5 A2 A1 6E 56 94 70 10  nx..'.9\...nV.p.
0040: 53 B1 BB E4 AF D0 A2 C3   2B 01 D4 96 F4 C5 20 35  S.......+..... 5
0050: 33 F9 D8 61 36 E0 71 8D   B4 B8 B5 AA 82 45 95 C0  3..a6.q......E..
0060: F2 A9 23 28 E7 D6 A1 CB   67 08 DA A0 43 2C AA 1B  ..#(....g...C,..
0070: 93 1F C9 DE F5 AB 69 5D   13 F5 5B 86 58 22 CA 4D  ......i]..[.X".M
0080: 55 E4 70 67 6D C2 57 C5   46 39 41 CF 8A 58 83 58  U.pgm.W.F9A..X.X
0090: 6D 99 FE 57 E8 36 0E F0   0E 23 AA FD 88 97 D0 E3  m..W.6...#......
00A0: 5C 0E 94 49 B5 B5 17 35   D2 2E BF 4E 85 EF 18 E0  \..I...5...N....
00B0: 85 92 EB 06 3B 6C 29 23   09 60 DC 45 02 4C 12 18  ....;l)#.`.E.L..
00C0: 3B E9 FB 0E DE DC 44 F8   58 98 AE EA BD 45 45 A1  ;.....D.X....EE.
00D0: 88 5D 66 CA FE 10 E9 6F   82 C8 11 42 0D FB E9 EC  .]f....o...B....
00E0: E3 86 00 DE 9D 10 E3 38   FA A4 7D B1 D8 E8 49 82  .......8......I.
00F0: 84 06 9B 2B E8 6B 4F 01   0C 38 77 2E F9 DD E7 39  ...+.kO..8w....9

]
SSL Certificate is Secure & active for current date
=======================================================
URL: https://www.diply.com
URL is Secure
37
```



Host & Page Based Analysis

URL	Status	Reason
http://karaite-bibliography.home.amu.edu.pl/Gj54ifD/	Phishing	SSL Certificate is Not Secure!
https://newslettre-info-redirectionssl.herokuapp.com/	Benign	-
http://newslettre-info-redirectionssl.herokuapp.com...	Possible Phishing	URL Length is between 54 to 75 Characters!!
https://intlserivcemypp.myvnc.com/myaccount/	Phishing	SSL Certificate is Not Secure!
https://intlserivcemypp.myvnc.com/signin/	Phishing	SSL Certificate is Not Secure!
http://facebook.com-continue-LjelCNKgsNs.mirael...	Possible Phishing	URL Length is between 54 to 75 Characters!!
https://www.googleusercontent.com	Benign	-
https://www.google.com.au	Benign	-
https://www.popads.net	Benign	-
https://www.cntv.cn	Phishing	SSL Certificate is Not Secure!
https://www.zhihu.com	Benign	-
https://www.amazon.co.uk	Benign	-
https://www.diply.com	Benign	-
https://www.coccoc.com	Phishing	SSL Certificate is Not Secure!
https://www.cnn.com	Phishing	SSL Certificate is Not Secure!
https://www.bbc.co.uk	Phishing	SSL Certificate is Not Secure!
https://www.twitch.tv	Phishing	SSL Certificate is Not Secure!
http://newslettre-info-redirectionssl.herokuapp.com...	Possible Phishing	URL Length is between 54 to 75 Characters!!
https://intlserivcemypp.myvnc.com/myaccount/	Phishing	SSL Certificate is Not Secure!
https://intlserivcemypp.myvnc.com/signin/	Phishing	SSL Certificate is Not Secure!

After the URL is trained, it displays the status of the trained URL whether there is phishing or Possible phishing along with the reason. Next click machine learning algorithms. It displays another page.


Machine Learning Algorithms
Machine Learning Algorithms
Load URL Testing Dataset
Detect Phishing URL
Machine Learning Algorithms
Apa
earn & Disc

Choose Classifier: MLP

```
Testing: 'http://orange002-serviceappel.ulcraft.com/'
predicted: 'Phishing (1.0)


Testing: 'http://c2cheaponline.com/vosdonneescom'
predicted: 'Possible Phishing (0.0)


Testing: 'http://www.myiccu.nl'
predicted: 'Phishing (1.0)


Testing: 'http://cubbnlafca.nl/stcu.org/STCU.php'
predicted: 'Possible Phishing (0.0)
```

```
                                               D0 AE DD
                                               15 50 C2
                                               75 CC EB
                                               A2 A1 6E
                                               01 D4 96
                                               B8 B5 AA
                                               08 DA A0
                                               F5 5B 86
                                               39 41 CF
                                               23 AA FD
                                               2E BF 4E
00B0: 85 92 EB 06 3B 6C 29 23   09 60 DC 45
00C0: 3B E9 FB 0E DE DC 44 F8   58 98 AE EA
00D0: 88 5D 66 CA FE 10 E9 6F   82 C8 11 42
00E0: E3 86 00 DE 9D 10 E3 38   FA A4 7D B1
```

Here, is the output for the phishing URL using MLP algorithm. It displays the output as predicted phishing or predicted possible phishing.

Conclusion

In conclusion, the proposed system to quantify and predict the behavior of Malicious Web Services (MWSs) using Multilayer Perceptron (MLP) has the potential to significantly improve the accuracy, performance, and cost-effectiveness of security applications that rely on MWSs. The system works by first extracting a set of features from the behavior of MWSs. These features are then fed into an MLP model, which is trained to predict the behavior of MWSs in terms of response time. Once the MLP model is trained, it can be used to rank MWSs in a quantitative manner. This ranking can be used to identify the most reliable MWSs, which are those that are most likely to exhibit predictable and consistent behavior. The proposed system has a number of advantages over traditional methods of malicious URL detection, such as blacklists and heuristics. First, it is more effective at detecting new and emerging threats. Second, it can predict the behavior of MWSs in terms of response time. Third, it can help to reduce the cost and complexity of managing MWSs.

References

[1] Stefano Calzavara, Alvise Rabitti, Alessio Ragazzo, and Michele Bugliesi. Testing for integrity flaws in web sessions. In Computer Security - 24th European Symposium on Research in Computer Security, ESORICS 2019, Luxembourg, 23-27 September 2019, pages 606–624.

[2] Stefano Calzavara, Mauro Conti, Riccardo Focardi, Alvise Rabitti, and Gabriele Tolomei. Mitch: A machine learning approach to the blackbox detection of CSRF vulnerabilities. In IEEE European Symposium on Security and Privacy, EuroS&P 2019, Stockholm, Sweden, 17-19 June 2019, pp. 528–543.

[3] M. Mohammadi, S. Yazdani, M.H. Khanmohammadi, K. Maham. Financial Reporting Fraud Detection: An Analysis of Data Mining Algorithms. Int. J. Financ. Manag. Account, 2020, vol. 4, p. 12.

[4] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, L. Cilar. Interpretability of machine learning based prediction models in healthcare. arXiv 2020, arXiv:2002.08596.

[5] D. Collaris, J.J. van Wijk. ExplainExplore: Visual Exploration of Machine Learning Explanations. In Proceedings of the 2020 IEEE Pacific Visualization Symposium (PacificVis), Tianjin, China, 3–5 June 2020, pp. 26–35.