##### A Project report on

###### **DETECTING MALICIOUS SITES**

###### A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

**Bachelor of Technology**

**in**

**Computer Science and Engineering**

Submitted by

G. NAVEEN

(20H51A0564)

I.SREE ANVITA

(20H51A0566)

FARHEEN

(20H51A0591)

Under the esteemed guidance of

Ms.B.Gayathri

(Assistant Professor)



**Department of Computer Science and Engineering**

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**

(UGC Autonomous)

\*Approved by AICTE \*Affiliated to JNTUH \*NAAC Accredited with A+ Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

#### 2020- 2024

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



#### CERTIFICATE

This is to certify that the Major Project Phase I report entitled **"Detecting Malicious sites"** being submitted by G.NAVEEN(20H51A0564),I.SREE ANVITA(20H51A0566), FARHEEN(20H51A0591) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

###### The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

**Ms.B.GAYATHRI Dr. Siva Skandha Sanagala**

**Assistant Professor Associate Professor and HOD**

**Dept. of CSE Dept. of CSE**

#### ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Ms.B.GAYATHRI, Assistant Professor** , Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

G.NAVEEN 20H51A0564

I.SREE ANVITA 20H51A0566

FARHEEN 20H51A0591

Detecting Malicious sites

**TABLE OF CONTENTS**

**CHAPTER NO. TITLE PAGE NO.**

ABSTRACT ii

**1** **INTRODUCTION** 1 1.1 Problem Statement 2

1.2 Research Objective 2

1.3 Project Scope and Limitations 3

**2** **BACKGROUND WORK** 4

2.1 Web Filters and Blacklists 5

2.1.1.Introduction 5

2.1.2.Merits,Demerits and Challenges 5

2.1.3.Implementation of Web Filters and Blacklists 6

2.2 Verify the URL 7

2.2.1.Introduction 7

2.2.2.Merits,Demerits and Challenges 7

2.2.3.Implementation of Verify the URL 8

2.3 Email Filtering 9

2.3.1.Introduction 9

2.3.2.Merits,Demerits and Challenges 9,10

2.3.3.Implementation of Email Filtering 10

**3 RESULTS AND DISCUSSION** 11 3.1 Result and discussion 12

**4** **CONCLUSION** 13

4.1 Conclusion 14

**5** **REFERENCES** 15

5.1 References 16

CMRCET B. Tech (CSE) Page No i

Detecting Malicious sites

# **ABSTRACT**

There are number of users who purchase products online and make payment through various websites.There are multiple websites who ask user to provide sensitive data such as username, password or credit card details etc. often for malicious reasons. This type of websites is known as phishing website. n order to detect and predict phishing website, we proposed an intelligent, flexible and effective system that is based on using classification Data mining algorithm. We implemented classification algorithm and techniques to extract the phishing data sets criteria to classify their legitimacy. 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. Once user makes transaction through online when he makes payment through the website our system will use data mining algorithm to detect whether the website is phishing website or not. 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. With the help of this system user can also purchase products online without any hesitation.

CMRCET B. Tech (CSE) Page No ii

Voice Accent Detection Model

# **CHAPTER 1**

**INTRODUCTION**

CMRCET B. Tech (CSE) Page No 1

Detecting Malicious sites

**CHAPTER 1**

**INTRODUCTION**

**1.1.Problem Statement**

There are number of users who purchase products online and make payment through various websites .There are multiple websites who ask user to provide sensitive data such as username, password or credit card details etc. often for malicious reasons. This type of websites is known as phishing website. Phishing is an internet scam in which an attacker sends out fake messages that look to come from a trusted source. A URL or file will be included in the mail, which when clicked will steal personal information or infect a computer with a virus. Traditionally, phishing attempts were carried out through wide-scale spam campaigns that targeted broad groups of people indiscriminately. It has become one of the most powerful and harmful of all the cyber-attacks. Because of lack of adequate identification techniques and protective methods, phishing methods are becoming common day by day.

**1.2.Research Objective**

The objective of this project is to train machine learning models and deep neural nets on the dataset created to predict phishing websites. Both phishing and benign URLs of websites are gathered to form a dataset and from them required URL and website content-based features are extracted. The performance level of each model is measures and compared.

CMRCET B. Tech (CSE) Page No 2

Detecting Malicious sites

**1.3. Project Scope and Limitattions**

**Project Scope:**

1. Objective: The primary objective of this project is to design and implement a comprehensive system to detect and prevent phishing websites to safeguard users' sensitive data and personal information.
2. Inclusions:

a. Phishing Website Detection: Develop algorithms and mechanisms to identify potential phishing websites.

b. Real-time Monitoring: Implement real-time monitoring of web traffic for phishing indicators.

c. Alerting and Notification: Notify users and system administrators when a phishing attempt is detected.

d. Data Protection: Develop methods to safeguard users' sensitive data and personal information.

e. Reporting and Analysis: Provide reporting and analysis tools to track and understand phishing trends and patterns

1. Exclusions:

a. Physical security measures are not within the scope of this project.

b. This project does not include the development of end-user web browsers or email clients.

c. Legal and law enforcement actions against phishing perpetrators are not part of the scope.

**Limitations:**

1. User Education: While the project may include user education and awareness materials, it cannot guarantee that all users will follow recommended security practices. Human error remains a significant factor in phishing incidents.
2. User feedback: Relying solely on automated detection may miss emerging threats. User feedback may be essential to enhance the system's accuracy.

CMRCET B. Tech (CSE) Page No 3

Detecting Malicious sites

**CHAPTER 2**

**BACKGROUND WORK**

CMRCET B. Tech (CSE) Page No 4

Detecting Malicious sites

**CHAPTER 2**

**BACKGROUND WORK**

**2.1 Web Filters and Blacklists**

**2.1.1 Introduction**

Web filters and blacklists are a fundamental component of cybersecurity efforts designed to thwart the activities of malicious actors on the internet, particularly in the realm of phishing websites. These databases comprise an extensive repository of known malevolent web addresses and domains. Security software, as well as web browsers, rely on these databases to automatically prevent users from accessing or interacting with potentially harmful websites.

**2.1.2 Merits, Demerits and Challenges**

**Merits:**

1. Well-maintained blacklists can be effective at blocking known phishing sites.
2. Low Resource Overhead: These are relatively low-resource solutions that are easy to implement.
3. Well-maintained blacklists offer robust protection by blocking access to websites with a known history of phishing activities.
4. User-Friendly: These solutions are user-friendly, as they operate in the background without requiring active engagement from the end-users. Users do not need to take any specific actions to benefit from this protection.
5. Low Resource Overhead: Implementing and maintaining web filters and blacklists is relatively resource-efficient. They do not require extensive computational power or frequent updates compared to some other solutions, making them cost-effective.

**Demerits:**

1. Varied Response: Not all users may be receptive to or capable of understanding security education.
2. Human Error: Even educated users can make mistakes.

CMRCET B. Tech (CSE) Page No 5

Detecting malicious sites

**Challenges:**

1. Changing Threat Landscape: Phishing tactics constantly evolve, requiring ongoing user Education.
2. False Positives: One of the most significant challenges is the potential for false positives, where legitimate websites are mistakenly identified as phishing sites. This can lead to user frustration and can impact productivity.
3. Limited Coverage: These solutions are only as effective as the databases they rely on. If a phishing website is not yet known or has not been added to the blacklist, users will not be protected.

**2.1.3 Implementation of Web Filters and Blacklists:**

1. Planning and Requirements Analysis: Before implementation, it's essential to define the project's scope and objectives. Determine the specific needs of the organization or user base, such as the type of web filtering required and the scale of deployment.
2. Selection of Web Filtering Solutions: Choose appropriate web filtering and blacklist management solutions. This could involve selecting commercial products, subscribing to threat intelligence services, or developing custom filters and blacklists in-house.
3. Database Maintenance and Updates: Regularly update the blacklist and web filtering databases to keep up with emerging threats and ensure that the most current information is used for blocking access to malicious websites.
4. User Education and Awareness: Implement an education and awareness program for end-users. It's essential to communicate the purpose and benefits of the web filtering and blacklist system, as well as inform users about potential false positives and the importance of not attempting to bypass these security measures.
5. Testing and Tuning: Prior to full deployment, perform rigorous testing to ensure that the web filtering and blacklist system is functioning correctly without causing significant false positives or negatives. Fine-tune the system to balance security and usability.

CMRCET B. Tech (CSE) Page No 6

Detecting malicious sites

**2.2** **Verify the URL**

**2.2.1 Introduction**

The URL Verification project centers on the critical task of ensuring the legitimacy and security of web addresses that users interact with. This initiative combines various technological tools and strategies to assess the authenticity of URLs, protecting users from malicious or fraudulent websites. By scrutinizing the URL's components, SSL certificates, and database records, the project aims to offer a layered defense against phishing attempts and other cyber threats, providing users with peace of mind and a safer online experience.

**2.2.2 Merits, Demerits and Challenges**

**Merits:**

1. Real-time Protection: URL verification offers real-time assessment, preventing users from accessing known malicious sites.
2. User-Friendly: Users benefit without actively engaging in complex security protocols.
3. Wide Adoption: Many browsers and security software adopt URL verification, providing broad coverage.
4. Proactive Defense: It offers proactive protection, preventing users from even encountering potential threats.
5. Resource Efficiency: Implementation and maintenance of URL verification systems are resource-efficient.

**Demerits:**

1. False Positives: A major challenge is the potential for legitimate websites to be incorrectly flagged as malicious, leading to user frustration.
2. Lag Time: URL verification systems may suffer from delays in updating their databases, leaving users vulnerable to emerging threats.

CMRCET B. Tech (CSE) Page No 7

Detecting malicious sites

1. Limited Coverage: These systems are effective only against known threats, unable to detect zero-day attacks.
2. False Sense of Security: Relying solely on URL verification might make users overconfident, potentially reducing their vigilance against phishing attempts.
3. Maintenance Effort: Keeping databases up-to-date and accurate is an ongoing effort that demands resources.

**Challenges:**

1. Zero-Day Attacks: URL verification may not catch new, previously unknown phishing tactics.
2. Evasion Techniques: Phishing websites can use evasion techniques to avoid detection by URL verification systems.
3. Database Accuracy: The effectiveness of URL verification depends on the accuracy and timeliness of the underlying databases.
4. User Awareness: Users need to be educated about the importance of URL verification.

**2.2.3 Implementation of Verify the URL**

The implementation of URL verification involves several key steps, including planning and requirements analysis, selection of appropriate verification systems, integration with existing infrastructure, maintenance of up-to-date databases, user education, continuous monitoring, and incident response procedures. Fine-tuning, regular auditing, and compliance with relevant regulations are also vital aspects of the implementation process. This comprehensive approach reduces the risk of users encountering phishing attempts and contributes to a more secure online environment.

CMRCET B. Tech (CSE) Page No 8

Detecting Malicious sites

**2.3** **Email Filtering**

**2.3.1 Introduction**

Email filtering is a critical cybersecurity project designed to safeguard users and organizations from the ever-present threat of malicious emails, including phishing attempts, malware distribution, and spam. This project utilizes advanced technologies and strategies to automatically assess incoming and outgoing emails, identifying and isolating potentially harmful content. By analyzing the content, sender reputation, attachments, and embedded links, email filtering aims to ensure that only safe and legitimate emails reach users' inboxes while blocking or quarantining suspicious or malicious messages. This not only enhances security but also improves overall email productivity and efficiency by reducing the time and resources required to deal with spam and threats.

**2.3.2** **Merits, Demerits and Challenges**

**Merits:**

1. Reduced Threat Exposure: Email filtering significantly reduces the risk of users falling victim to phishing attacks and malware distribution, enhancing overall cybersecurity.
2. Improved Productivity: Filtering out spam and malicious emails reduces the clutter in users' inboxes, enabling them to focus on legitimate correspondence.
3. Real-time Protection: Advanced email filtering systems offer real-time detection and response to emerging threats.
4. Automatic Operation: Users benefit from protection without the need for manual intervention.
5. Cost Efficiency: Effective email filtering can reduce the costs associated with handling spam and email-borne threats.

CMRCET B. Tech (CSE) Page No 9

Detecting Malicious sities

**Demerits:**

1. False Positives: Email filtering may occasionally flag legitimate emails as suspicious, leading to potential communication issues.
2. Missed Threats: Advanced phishing or malware attacks can sometimes bypass email filters.
3. Overly Permissive Filters: Overly lenient filtering settings may allow some malicious content to slip through.

**Challenges:**

1. Zero-Day Threats: Emerging threats not yet identified by email filtering systems can be challenging to detect.
2. Evasion Techniques: Attackers continually develop new tactics to bypass email filters, requiring system updates.
3. Configuring and Tuning: Proper configuration and tuning of filters are essential to minimize false positives and false negatives.

**2.2.3 Implentation of Email Filtering**

1. Selection of Filtering Solutions: Choose and deploy email filtering software or services, which may include on-premises solutions or cloud-based services.
2. Configuration and Tuning: Fine-tune filtering rules to balance security and usability, reducing the chances of false positives while maintaining protection.
3. User Training: Conduct user education and awareness programs to ensure users understand the role of email filtering and how to handle filtered messages.
4. Monitoring and Incident Response: Implement monitoring and response mechanisms to address and investigate suspicious emails and potential threats.
5. Compliance with Regulations: Ensure that email filtering practices align with relevant data privacy and security regulations.

CMRCET B. Tech (CSE) Page No 10

Detecting Malicious sites

**CHAPTER 3**

**RESULTS AND DISCUSSION**

CMRCET B. Tech (CSE) Page No 11

Detecting Malicious sites

**CHAPTER 3**

**RESULTS AND DISCUSSION**

The primary objective of Detecting Malicious sites project was to develop an intelligent system capable of effectively detecting and predicting phishing websites, thereby enhancing the overall security of online transactions. By leveraging classification data mining algorithms, we aimed to classify websites as either legitimate or malicious based on a set of criteria. These criteria primarily revolved around URL and domain identity, as well as security and encryption factors.

One of the key outcomes of our project was the performance evaluation of the classification algorithm employed. We compared the effectiveness of our data mining approach with traditional classification algorithms. The results demonstrated that our system outperformed conventional methods in detecting phishing websites. This indicates that data mining algorithms, when properly implemented, can significantly contribute to the security of online transactions.

Our system exhibited a higher detection rate for phishing websites by considering multiple characteristics, such as the URL and domain identity. The utilization of this extensive set of criteria allows for a more comprehensive assessment of a website's legitimacy. This multi-dimensional approach is vital in the ever-evolving landscape of online threats. It enables our system to identify sophisticated phishing attempts that may mimic legitimate websites.

One of the major challenges in any phishing detection system is the avoidance of false positives, where legitimate websites are incorrectly classified as malicious. Our project focused on fine-tuning the data mining algorithm and criteria to minimize such false positives. The results showed a considerable reduction in these cases, indicating that our system effectively balances security and usability.

A significant advantage of our system is that it empowers users to make online purchases without hesitation.

CMRCET B. Tech (CSE) Page No 12

Detecting Malicious sites

CHAPTER 4

**CONCLUSION**

CMRCET B. Tech (CSE) Page No 13

Detecting malicious sites

**CHAPTER 4**

**CONCLUSION**

In conclusion, Our project aimed to fortify online transaction security by employing data mining to detect phishing websites. By employing classification algorithms that consider URL, domain identity, security, and encryption, our system outperformed traditional methods, yielding higher detection rates with fewer false positives. This approach enhances user trust, empowering them to transact online with confidence. The practical applications are wide-reaching, benefiting industries, businesses, and individuals. E-commerce enterprises, in particular, can create a more secure transaction environment that fosters customer trust. Our commitment remains steadfast: continuous refinement and vigilance to safeguard users in an evolving digital landscape where the assurance of security is paramount, By harnessing the power of classification algorithms, we have significantly improved the detection of phishing websites. Our comprehensive approach, which considers URL, domain identity, security, and encryption, outshines traditional methods and reduces false positives, instilling user confidence. The practical applications span various industries, offering increased security for businesses and individuals alike. In the e-commerce sector, our solution paves the way for a more secure transaction environment, nurturing customer trust. Our unwavering commitment to continuous refinement ensures that we remain at the forefront of safeguarding users in today's dynamic digital landscape. In an age where online security is paramount, our project stands as a beacon of assurance and trust..

CMRCET B. Tech (CSE) Page No 14

Detecting malicious sites

**REFERENCES**

CMRCET B. Tech (CSE) Page No 15

Detecting malicious sites

## REFERENCES

1. https://scholarworks.rit.edu/cgi/viewcontent.cgi?article=11869&context=theses
2. <https://is.muni.cz/th/cfw06/Master_s_thesis.pdf>
3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9133026/>
4. <https://github.com/topics/email-spam-filter>
5. https://www.codeproject.com/Questions/409516/How-to-check-if-given-url-is-existed

## 

CMRCET B. Tech (CSE) Page No 16