

RIPHAH INTERNATIONAL UNIVERSITY



Faculty of Computing FINAL YEAR PROJECT PROPOSAL & PLAN

Phishing attack detection and prevention tool

Project Team

Full Name of Student	SAP Id	Program	Contact Number	Email Address
Shawar Ahmed	27973	BSCYB	03456970824	27973@students.riphah.edu.pk
Syed Hussain Abid	33356	BSCYB	03348610002	33356@students.riphah.edu.pk

Sir Hummayun Raza
Lecturer

Phishing attack detection and prevention tool

Change Record

Phishing attack detection and prevention tool

Change Record

Author(s)	Version	Date	Notes	Supervisor's Signature
Shawar, Hussain	1.0	3/9/2024	Original Draft	
—	2.0	7/9/2024	Changes Based on Feedback From Supervisor	
—	3.0	25/9/2024	Changes Based on Feedback From Faculty	
			Added Project Plan	
			Changes Based on Feedback From Supervisor	

Project Proposal

Project Title: Phishing attack detection and prevention tool

Introduction and Background: Phishing is a cybercrime in which a criminal sends a fake e-mail that looks to come from a well-known and reputable company or organization, requesting personal information such as bank passwords, usernames, etc.

Phishing has been a major security threat in which there is a huge loss for companies as well as customers. These phishing attacks are increasing day by day due to lack of efficient detection techniques and effective preventive measures.

Educational institutions, which rely heavily on digital communication tools like email, have increasingly become prime targets for such attacks. Students, faculty, and administrative staff are often exposed to phishing emails that impersonate school authorities, financial aid services, or educational platforms, making the education sector vulnerable to these attacks.

Phishing attacks typically involve fraudulent emails that attempt to trick recipients into revealing sensitive information or clicking on malicious links. Attackers often impersonate trusted entities, such as school administrator or service providers, to deceive individuals into sharing confidential data. In the context of educational institutions, phishing scams have evolved to exploit various aspects of academic life, from fake scholarship offers to fraudulent online portals that mimic official school websites.

Existing Systems/ Survey/ Literature Review: There are many phishing detection and prevention tools available today but not any tool specific for education domain.

- **Phishing Trends in Education:** Investigate how phishing attacks are evolving in the education sector. For instance, the use of:
 - Fake educational portals.
 - Impersonation of university officials.
 - Spear-phishing targeting professors or administrators.
- **Existing Tools:** Currently, there are no widely recognized phishing detection tools designed exclusively for the education sector. However, several existing general-purpose tools are commonly used by educational institutions to protect against phishing attacks, and they can be customized or tailored to better suit the needs of the education domain.

Problem Statement: There is no tool that specifically built for phishing in the education domain, our final-year project could focus on developing a **phishing detection and prevention tool** that addresses these specific challenges. We could analyze educational phishing datasets, create custom detection algorithms, and design an interface that integrates seamlessly into educational institutions. This would not only fill a gap in the current tool landscape but also provide a more targeted solution for phishing protection in educational environments.

Objectives: The primary objective of this tool is to **enhance cyber security within educational institutions** by:

1. **Detecting Phishing Attacks:**
 - Implementing real-time detection mechanisms that can automatically identify phishing attempts through various techniques such as rule-based filters, machine learning, and behavioral analysis.
 - Analyzing incoming emails, links, and communication channels used within the educational environment (e.g., learning management systems, faculty-student communications).
2. **Preventing Phishing Attacks:**
 - Providing preventive measures, such as email filtering, suspicious URL blocking, and the integration of multi-factor authentication (MFA) to safeguard users from falling victim to phishing attempts.
3. **Reducing Human Error:**
 - Minimizing the chances of users being deceived by phishing emails or messages by automatically flagging or blocking malicious content before it reaches their inbox.
4. **Improving Overall Security Posture:**
 - Strengthening the cyber security framework of educational institutions, reducing vulnerabilities, and ensuring the protection of sensitive student, faculty, and administrative data.
5. **Providing User-Friendly Solutions:**
 - Ensuring the tool is easy to use and integrate into existing platforms, making it accessible for non-technical users such as students, educators, and administrative staff.

Proposed Solution:

1. **Custom Phishing Detection Algorithms**

- **Domain Analysis:** Implement a module to verify the sender's domain by comparing it against a known list of trusted educational domains. This will flag suspicious or newly created domains that might be used in phishing attempts.
- **Keyword-Based Detection:** Develop an algorithm to scan for commonly used phishing keywords in emails or messages, such as "urgent", "account locked", or "click here", particularly in educational contexts. These keywords can be expanded through continuous research.
- **URL Blacklist Integration:** Incorporate a regularly updated blacklist of known malicious URLs and websites. The tool can verify embedded links within emails and web pages against this list, alerting users if any malicious links are detected.

2. Interface Design and Integration

- **User-Friendly Interface for Educational Institutions:** The tool will feature an intuitive dashboard that integrates seamlessly with existing platforms like Learning Management Systems (LMS), email clients, and student portals.
- **Role-Based Access Control:** Allow different user levels (e.g., students, staff, IT administrators) with access to specific features, ensuring the right individuals have the appropriate tools to mitigate risks.

3. Educational Content for Prevention

- **Integrated Phishing Awareness Module:** Implement a built-in training feature that educates users on how to recognize phishing attempts. This could be part of an onboarding process for students and staff and include regular quizzes and updates on phishing trends specific to the educational sector.

4. Cross-Institution Collaboration

- **Data Sharing Between Educational Institutions:** Develop a system for educational institutions to share anonymized phishing data, such as new phishing campaigns targeting multiple schools, allowing the tool to evolve and improve detection.

Methodology:

Multi-Layer Approach: Algorithm and Design

In this section, a detailed description of the proposed solution is discussed. The multi-layer approach considers the strengths and weaknesses of existing phish-detecting algorithms (Blocklist, Blacklist, Phishtank and Whitelist) from different perspectives. These mentioned selected strategies have the capability of collecting data on phishing activities on the internet. These data are stored in their respective databases and accessible to users on the internet. The disadvantage of these selected approaches is that they do not have a unified database. This makes it difficult for users to get a valid and best solution to determining whether a website or an email is phishing or not phishing.

For example, if a suspected phishing website is registered in the domain blacklist for being a phishing website and a user who knows nothing about the Blocklist went authenticates the link in the domain Blacklist where the link is not yet reported as a phishing website for the platform to register. This action will then mislead the user into believing that the suspected website is authentic. This problem applies to Domain Blacklist, PhishTank, and Domain Whitelist platforms as well. It is for this reason that in this work, a combination of the four approaches in conjunction has been designed in a multi-layer framework to curb phishing activities. This will make URL authentication easy and more effective for the users. Also, users will not have to move from approach to approach; there will be efficient utilization of time surfing the internet and authenticating every domain before a page load or reload.

Implementation Plan:

Phase	Duration	Completion Time
Project Planning	2 weeks	Week 2
Literature Review & Research	2 weeks	Week 4
Data Collection & Preprocessing	2 weeks	Week 6
Feature Extraction & Selection	2 weeks	Week 8
Model Development	2 weeks	Week 12
Phishing Prevention Development	2 weeks	Week 14
Testing & Evaluation	2 weeks	Week 18
User Acceptance Testing	2 weeks	Week 22
Deployment & Integration	2 weeks	Week 26
Continuous Monitoring & Updates	2 weeks	Week 28
Final Report & Documentation	2 weeks	Week 34

- **Tools/Software:** Python (Scikit-learn, TensorFlow), IDEs (PyCharm, VSCode), email server for testing, cloud computing if needed.
- **Dataset:** Phishing datasets (e.g., PhishTank, SpamAssassin).

Evaluation Plan:

1. Unit Testing

- **Objective:** Verify that individual components of the tool (e.g., rule-based filters, URL blacklist, email parsing module) work as expected.
- **Method:**
 - Each module (e.g., the email content analyzer, URL checker) will be tested separately.

2. Controlled Simulations

- **Objective:** Test the tool's ability to detect and prevent phishing attacks in a simulated environment.
- **Method:**
 - A set of simulated phishing emails and benign emails will be used to test the system.
- **Criteria:**
 - The tool should detect and block a high percentage of phishing emails while allowing legitimate emails to pass through.

3. Performance Benchmarking

- **Objective:** Measure the tool's performance in terms of accuracy, speed, and resource usage.
- **Method:**
 - Conduct load testing to see how the tool performs with a large volume of emails in a short period.
- **Criteria:**
 - The tool should handle a significant volume of emails efficiently with minimal delay.

Project Scope/ Expected Outcomes: The tool will provide protection for different user groups within educational institutions, such as students, faculty, and administrative staff, while offering a user-friendly dashboard for IT administrators to monitor phishing threats and manage incidents. With a focus on real-time detection, user training, and behavioral analysis, this project aims to significantly reduce phishing incidents in educational institutions and improve overall cyber security awareness.

The primary goal of this project is to design, develop, and deploy a phishing detection and prevention tool aimed at protecting educational institutions from phishing attacks. The tool will focus on:

- **Detecting and preventing phishing emails** that specifically target educational users (students, faculty and administrators).

Conclusion and Future Work: The proposed **Phishing Attack Detection and Prevention Tool** is designed to protect educational institutions from phishing attacks without relying on artificial intelligence (AI). The system will utilize a combination of **rule-based detection** and **content analysis techniques** to identify and prevent phishing attempts. This tool will focus on detecting phishing emails, URLs, and suspicious patterns in email content to provide robust protection.

The primary objectives include:

1. **Detecting phishing attacks** using predefined rules, such as checking email headers, sender authenticity, and keyword-based content analysis.
2. **Preventing phishing attacks** by filtering suspicious emails, blocking URLs that are blacklisted, and alerting users to potential phishing risks.
3. **Enhancing user awareness** through notifications and warnings when phishing attempts are detected, encouraging better cybersecurity practices.
4. **Securing educational environments** by safeguarding personal data and institutional resources from phishing-related breaches.

The project methodology involves manual rule-based filtering, URL validation techniques, and integration with existing email platforms and security protocols. A clear timeline has been developed to ensure smooth implementation and testing.

References:

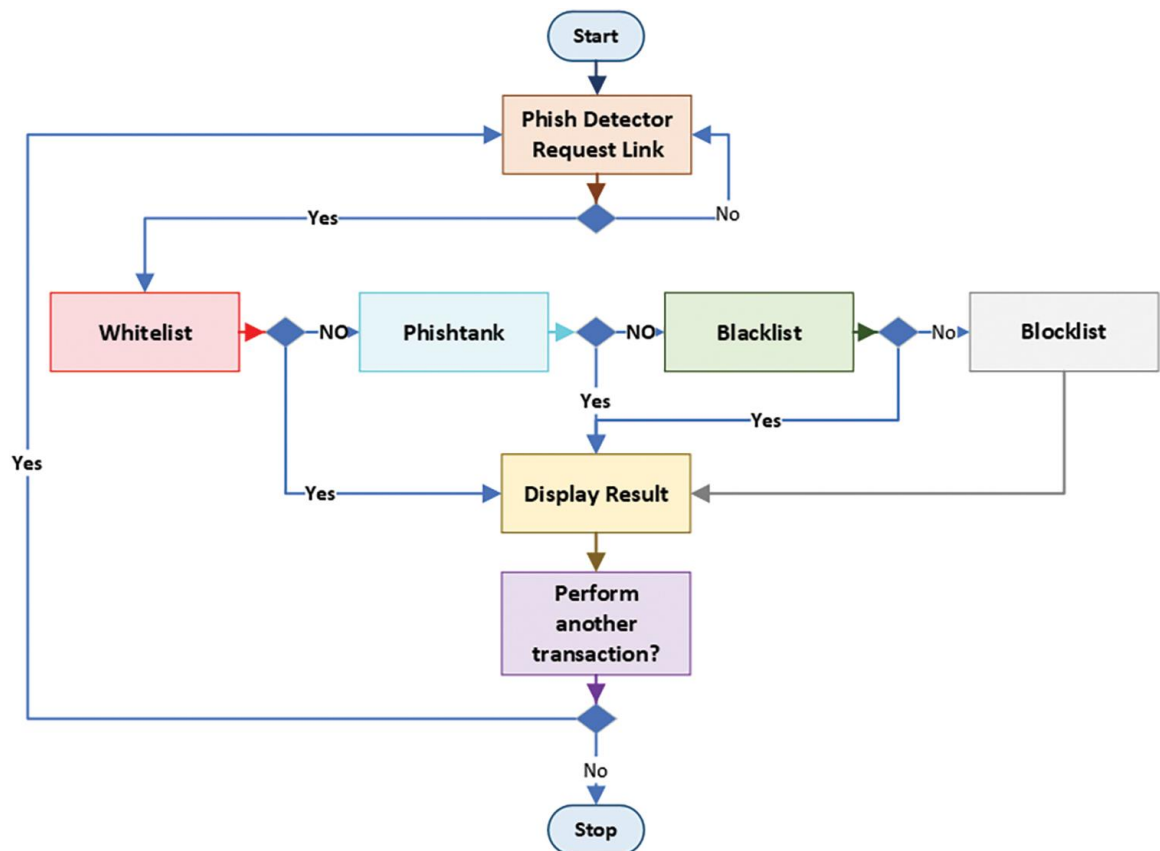
1. S. Venkatesha, K. R. Reddy and B. R. Chandavarkar, "Social engineering attacks during the COVID-19 pandemic," *SN Computer Science*, vol. 2, no. 2, pp. 1–9, 2021. <https://doi.org/10.1007/s42979-020-00443-1> [Google Scholar] [PubMed] [CrossRef]
2. K.S. Adu-Manu and R.K. Ahiable, "Detecting Phishing Using a Multi-Layered Social Engineering Framework," *J. Cyber Secur.*, vol. 5, no. 1, pp. 13-32. 2023.
3. A. Jamil, K. Asif, Z. Ghulam, M. K. Nazir, S. Mudassar Alam et al., "MPMPA: A mitigation and prevention model for social engineering based phishing attacks on Facebook," in *IEEE Int. Conf. on Big Data (Big Data)*, Seattle, WA, USA, pp. 5040–5048, 2018. <https://doi.org/10.1109/BigData.2018.8622505> [Google Scholar] [CrossRef]
4. M. Hijji and G. Alam, "A multivocal literature review on growing social engineering based cyber-attacks/threats during the COVID-19 pandemic: Challenges and prospective solutions," *IEEE Access*, vol. 9, pp. 7152–7169, 2021. <https://doi.org/10.1109/ACCESS.2020.3048839> [Google Scholar] [PubMed] [CrossRef]

Appendices: System Architecture Diagram

This diagram provides a visual representation of the tool's architecture, showcasing how the components interact with each other within an email system environment.

System Design:

- **Email Server Integration:** The tool is integrated with the email server (e.g., Postfix, Sendmail) to intercept and scan emails.
- **Rule-Based Detection Engine:** Uses predefined rules to analyze email content, headers, and URLs.
- **Blacklist Module:** Checks URLs and domains against a constantly updated blacklist (e.g., from PhishTank).
- **Alert System:** Sends warnings to users and administrators when phishing emails are detected.

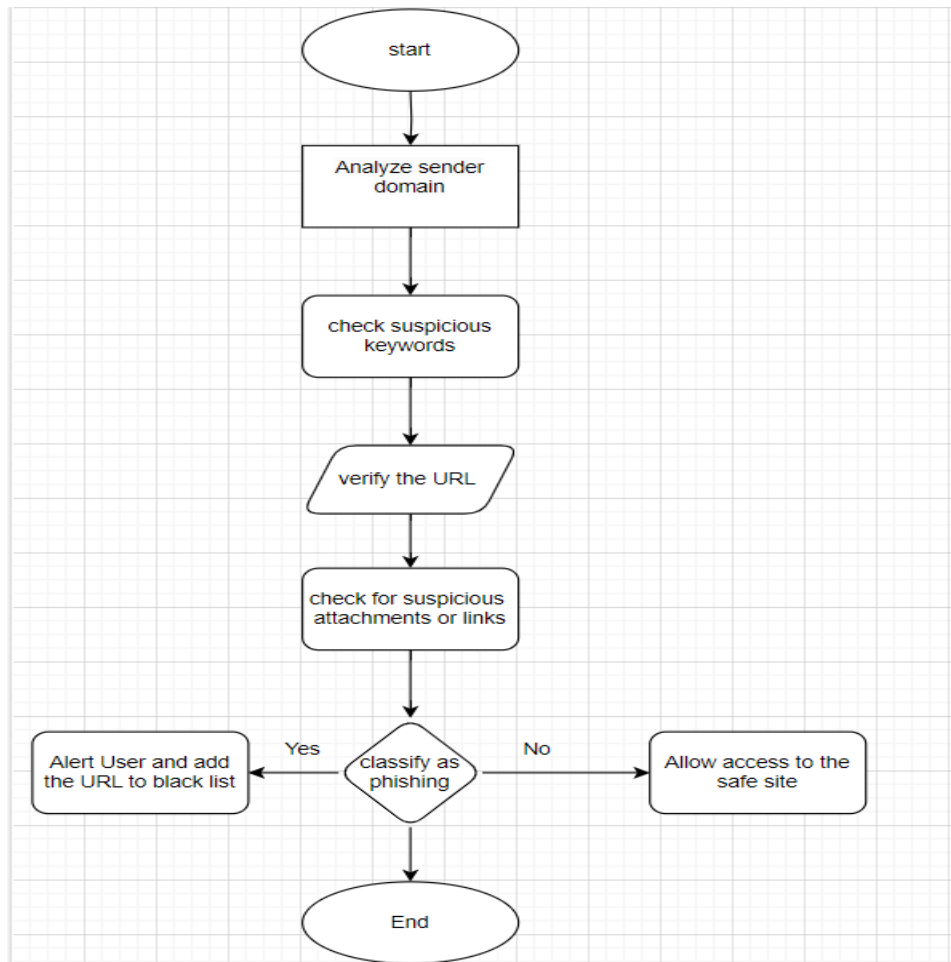


Phishing Detection Flowchart

This flowchart illustrates the decision-making process that the detection engine follows when analyzing emails and determining whether they are phishing or legitimate.

Phishing Detection Steps:

- **Step 1:** Analyze sender domain.
- **Step 2:** Check for suspicious keywords in the subject and body.
- **Step 3:** Verify URLs in the email body against the blacklist.
- **Step 4:** Check for suspicious attachments or links.
- **Step 5:** Classify as phishing or legitimate.

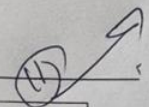


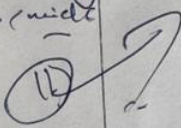
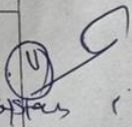
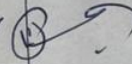
List of Faculty Proposed Changes

Project Title

List of Faculty Proposed Changes

Phishing attack detection and prevention tool

Supervisor's Signature: 

Proposed Change	Proposed By	Supervisor's Decision
Approved/Disapproved and/or Comments	Name of Faculty Member(s) who proposed this change	Approved/Disapproved and/or Comments
What type of phishing attacks will the detector focus on? Will it detect specific types such as spear-phishing, whaling, or general phishing campaigns? What detection methodologies will be employed? Will it use signature-based detection, machine learning, heuristic analysis, or a combination? How will the system identify phishing indicators? Will it analyze URL patterns, email metadata, sender behavior, or the content of the message (e.g., suspicious links, attachments)? How will the system distinguish between legitimate and phishing emails or websites? What criteria or algorithms will be used to avoid false positives?	Dr. Javed Iqbal	<ul style="list-style-type: none"> - specific attack related to .edu.pk. - no machine learning. - spam emails will 
What source did you concern to address the problem. Justify how your system is different and what value is added to improve the system. Research on Technical aspect. Conduct the survey of stakeholders.	Mr. Mueed Mirza	<ul style="list-style-type: none"> specific to .edu.pk. by survey of existing systems 
Lms integration, authentic source reference required	Mr. Yawar Abbas	<ul style="list-style-type: none"> open source only LMS. 
Need to understand the dynamic of the dataset	Mr. Awais Nawaz	<ul style="list-style-type: none"> NO-DATA SET.
accepted.	Mr. Tabbasum Javed	<ul style="list-style-type: none"> AI

Roles & Responsibility Matrix:

The purpose of roles & responsibility matrix is to identify who will do what.

WBS #	WBS Deliverable	Activity #	Activity to Complete the Deliverable	Duration (# of Days)	Responsible Team Member(s) & Role(s)
1	Phase 1: Planning and Requirements Gathering	1.1	Project Initiation		Shawar Ahmed and Syed Hussain Abid
		1.2	Requirements Gathering		
		1.3	Feasibility Study		
		1.4	Project Planning		
2	Phase 2: Design and Development	2.1	Architectural Design		Shawar Ahmed and Syed Hussain Abid
		2.2	User Interface Design		
		2.3	Database Design		
		2.4	Development and Coding		
		2.5	Integration		
3	Phase 3: Testing and Quality Assurance	3.1	Unit Testing		Shawar Ahmed and Syed Hussain Abid
		3.2	Integration Testing		
		3.3	System Testing		

4	Phase 4: Deployment and Launch	4.1	Deployment Planning		Shawar Ahmed and Syed Hussain Abid
		4.2	Deployment		
		4.3	Launch		

Approval

Approval


Project Supervisor

Comments Satisfactory

Name: H. M. A. H.

Date: _____

Signature: _____



Project Coordinator

Comments _____

Name: _____

Date: _____

Signature: _____