DEPI PROJECT

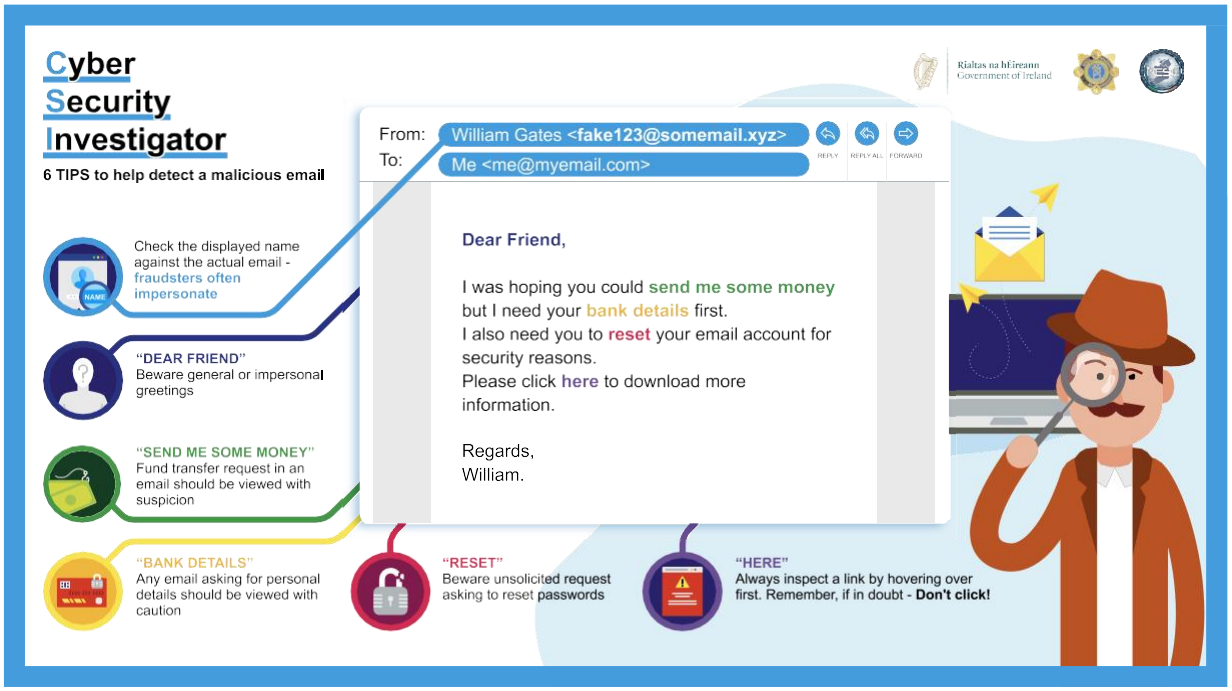
Email Phishing Analysis and Detection

Team Members :
Mahmoud Ahmed
Youssef Ahmed Fouad
Gamal Abd Elnaser Sayed
Supervisorer :
Eng/ Noureldin Essam

What is Phishing?

Phishing is a type of fraudulent activity carried out by cyber criminals whereby they send disguised e-mails to people or organisations purporting to be from reputable sources to influence the reader to click on links to dodgy websites or to give sensitive information away such as bank details, account passwords or credit card information.

Some phishing emails may contain viruses disguised as harmless attachments, which are activated when opened by the victim. The aim of this type of phishing attack could be something more specific, such as the theft of a business's sensitive data.



Criminals will use multiple mediums for delivering phishing lures such as:

Email Phishing: Malicious content delivered through email.

Smishing: SMS text messages to a mobile phone.

Vishing: Fraudulent voice phone calls.

Each of these lures will be designed to look genuine, and the sender will usually claim to be a person or organisation that you are familiar with to make it easier for them to gain your trust. It is becoming increasingly difficult to identify these social engineering attempts as attackers become more sophisticated. Attackers take advantage of people's social instincts, such as being helpful and efficient or their emotions such as fear or anger.

Avoid Being Phished

Criminals will check the internet for peoples publicly available information to make their phishing e-mails more convincing. By thinking about what personal information you and others have about you online there are some easy steps to take to make you a less likely target for a phishing e-mail attack.

All social media platforms provide in depth privacy settings for their users. You can review these settings within your social media accounts and make sure your information isn't publicly viewable.

It's not only you who can post information about yourself online. Be wary of what information your family, friends or work colleagues have posted about you. If necessary, ask them to remove any information about you.

Make a simple checklist you can remember easily by using our giveaway phishing signs (below) to help you to quickly scan e-mails that you aren't too sure about. If you are suspicious of an email that you have received, then report it to your IT administrator or e-mail provider and then delete the suspicious e-mail.

Giveaway Signs of Phishing

As cyber criminals make their phishing e-mails more convincing to try to quickly gain your trust, always pause to consider if an e-mail makes you suspicious. You can still stay one step ahead of them by remembering to scan for one or more of these giveaway phishing signs that could signify you are being targeted by a phishing e-mail.

Does the e-mail begin with a general or impersonal greeting such as 'Dear Friend' or 'valued customer'? If you aren't addressed by your name, then this could signal that the sender does not know you and should not have your e-mail address.

Check the senders email address by hovering your mouse over the 'from' address or clicking the down arrow beside the senders name to reveal more details about the sender. Does the name match the e-mail address and do they look legitimate? If not, the sender could be trying to impersonate somebody.

Is there a sense of urgency to the e-mail such as a request for your bank details or an action to take to avoid losing a service? Your bank or other familiar organisations will never make such requests from you in an e-mail.

Always check password reset or authentication requests sent to you by e-mail or SMS. You should only receive these requests if you have requested a password change or attempted to authenticate through your online account. Cyber criminals can send unsolicited requests to steal your passwords, if in doubt, don't click and report it to your account provider.

Are you being offered something for free or at a very well discounted rate? Ask yourself does this sound too good to be true? This tactic is used to panic you into thinking you might miss out on a good opportunity if you don't follow the e-mails instructions. If it sounds too good to be true, it probably is.

Some of these giveaway signs can also be present in text message scams (smishing) and fraudulent telephone voice calls (vishing). The general advice for these types of scams is as follows:

- Shortened and unrecognisable links are a sure giveaway, don't click the bait.
- Honest communications will never ask you to provide personal details.
- If it feels too good to be true or you aren't totally sure of something, then don't engage.
- Contact the organisation directly using their official phone number which should be on their official website to check if they have tried to contact you.

Have You Clicked the Bait?

If you think you've been the victim of a phishing e-mail and have already clicked a link, attachment or provided sensitive information then you can still take these actions to minimise the disruptive effects of the attack.

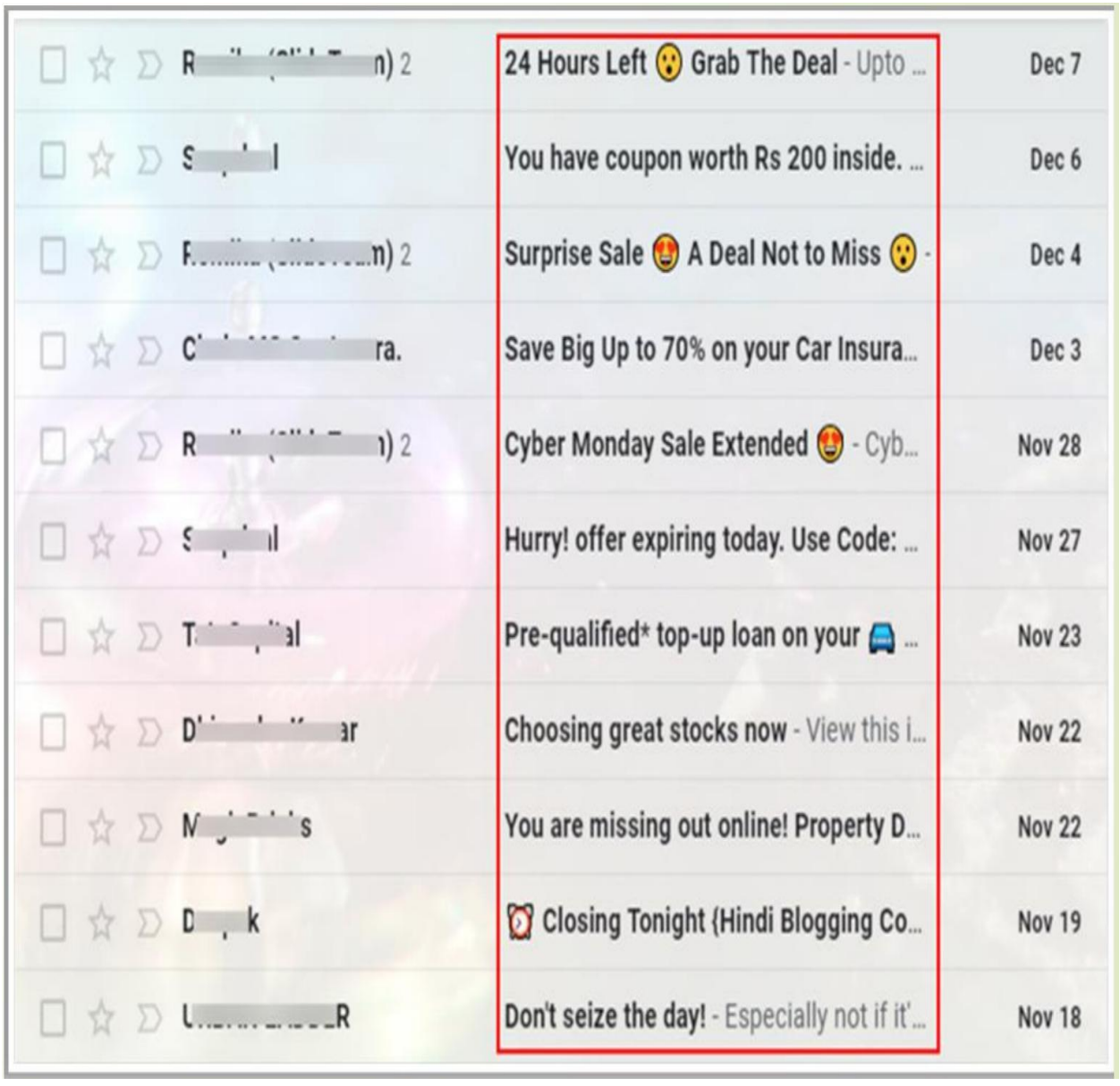
If you have provided sensitive information such as your password or bank details then change your passwords on all your accounts and contact your bank to get advice on what you should do next.

Use an antivirus software program to run a full scan of your device so it can attempt to uncover any possible viruses and try to remove them.

If you have been victim of a fraud, then you should report this to your local Garda station.

Spamming

- **Spam refers to undesired emails used to distribute malicious links and attachments, cause network congestion, perform phishing and financial frauds and so on.**
- **The spam may also consume bandwidth of the email servers causing DoS conditions.**
- **In the example the email address doesn't match the sender name or the content of message**



Glossary

Antivirus
Software that is designed to detect, stop and remove viruses and other kinds of malicious software.

Cyber-attack
Malicious attempts to damage, disrupt or gain unauthorised access to computer systems, networks or devices, via cyber means.

Cyber security
The protection of devices, services and networks — and the information on them — from theft or damage.

Multi-factor authentication
The use of two different components to verify a user's claimed identity.

Phishing
Untargeted, mass emails sent to many people asking for sensitive information (such as bank details) or encouraging them to visit a fake website.

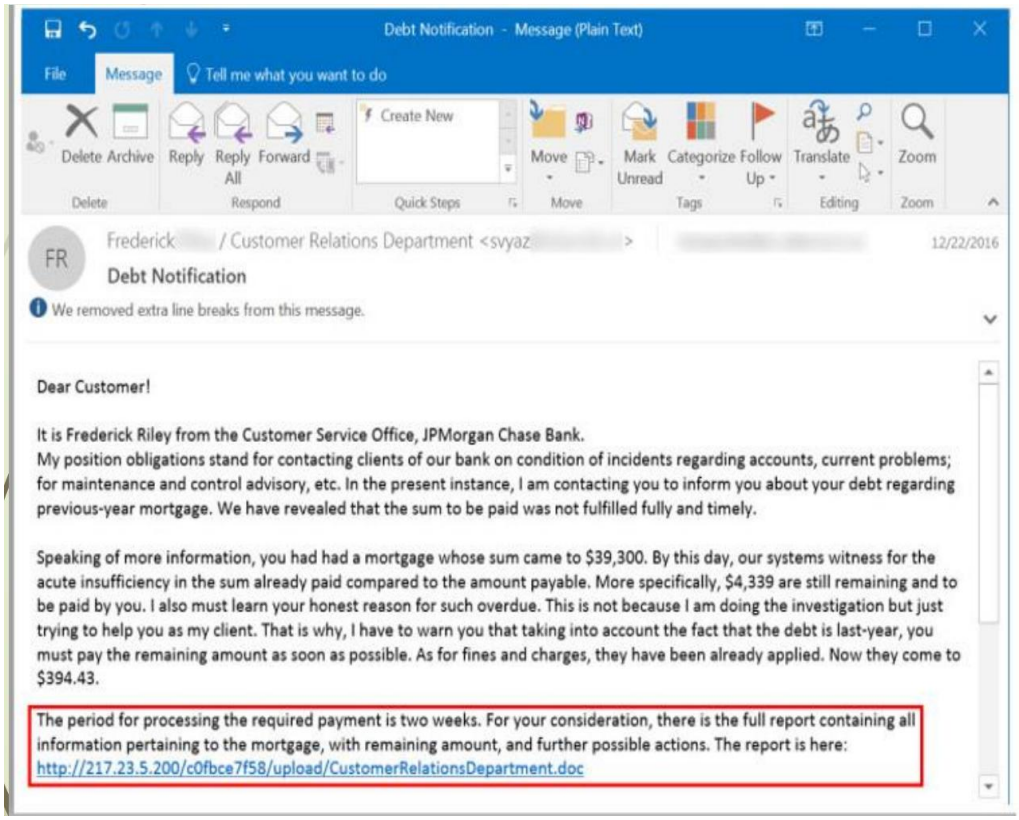
Smishing
Phishing via SMS: mass text messages sent to users asking for sensitive information (eg bank details) or encouraging them to visit a fake website.

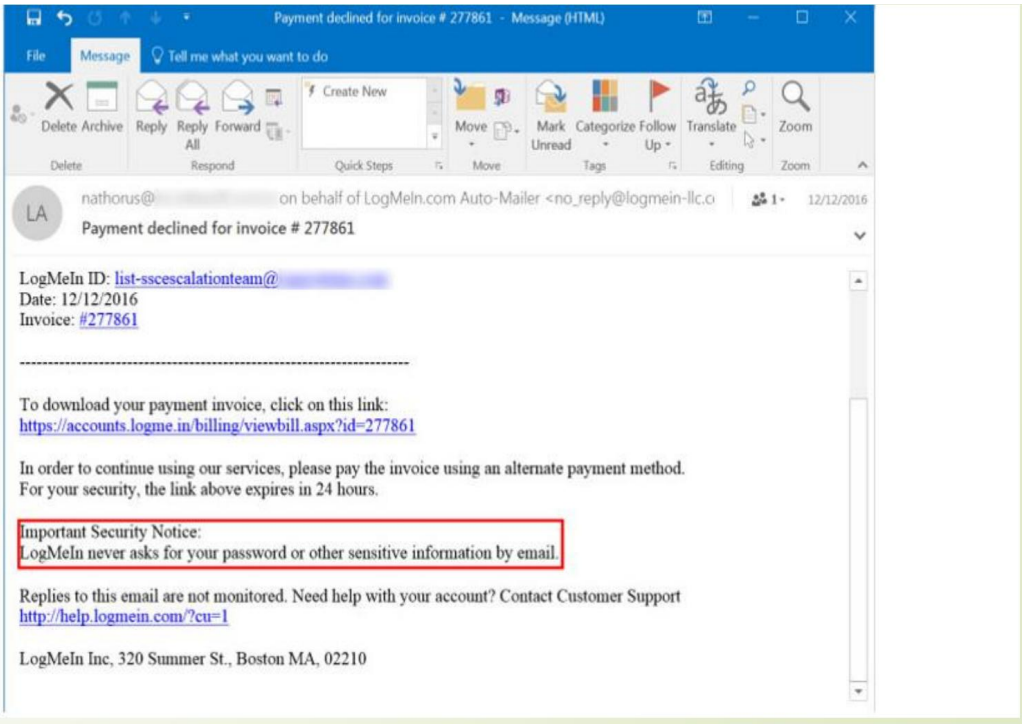
Vishing
The fraudulent practice of making phone calls or leaving voice messages purporting to be from reputable companies in order to induce individuals to reveal personal information, such as bank details and credit card numbers.

Virus
Programs which can self-replicate and are designed to infect legitimate software programs or systems. A form of malware.

Examples of phishing

Phishing involves fraudulently acquiring sensitive information (e.g., passwords, credit cards) by masquerading as a trusted entity.




Payment declined for invoice # 277861 - Message (HTML)
File
Message
Tell me what you want to do
Delete
Archive
Reply
Reply All
Forward
Create New
Move
Mark Unread
Categorize
Follow Up
Translate
Zoom
Delete
Respond
Quick Steps
Move
Tags
Editing
Zoom
LA
nathorus@ on behalf of LogMeIn.com Auto-Mailer <no_reply@logmein-llc.c
12/12/2016
Payment declined for invoice # 277861
LogMeIn ID: list-sscescalationteam@
Date: 12/12/2016
Invoice: #277861
To download your payment invoice, click on this link:
https://accounts.logme.in/billing/viewbill.aspx?id=277861
In order to continue using our services, please pay the invoice using an alternate payment method.
For your security, the link above expires in 24 hours.
Important Security Notice:
LogMeIn never asks for your password or other sensitive information by email.
Replies to this email are not monitored. Need help with your account? Contact Customer Support
http://help.logmein.com/?cu=1
LogMeIn Inc, 320 Summer St., Boston MA, 02210

Email Phishing Analysis and Detection

1. Introduction

Email phishing is one of the most widespread cybersecurity threats. Attackers frequently use fraudulent emails to deceive individuals into revealing sensitive information such as login credentials, financial details, or personal data. Phishing attacks often rely on psychological manipulation, spoofed sender identities, malicious links, fake landing pages, and urgent language.

This project, **Email Phishing Analysis and Detection**, provides a practical demonstration of how phishing emails can be analyzed, detected, and reported using automated scripts. The project integrates a combination of techniques including **keyword analysis**, **URL extraction**, **domain reputation checking**, **trusted-domain validation**, and **AI-based contextual analysis**.

The system also includes an **interactive CLI tool**, a **VirusTotal URL scanner**, and a **PDF report generator** to help users understand email risks quickly and effectively.

2. Project Objectives

The goal of this project is to develop a lightweight but robust phishing-detection demo that can:

1. Identify suspicious keywords commonly used in phishing attacks.
2. Extract and analyze URLs from email content.
3. Validate domains against a known set of trusted providers.
4. Query VirusTotal to check URL reputation.
5. Use AI (Groq-based LLM) to analyze and categorize email risk contextually.
6. Generate a structured report containing all findings.
7. Provide a clean CLI interface for users to operate the tool easily.

3. System Overview

The system architecture combines several modules working together:

3.1 Keyword Analysis Module

- Detects suspicious phrases typically used for manipulation or deception.
- Uses regular expressions for efficient text scanning.

3.2 URL Extraction Module

- Identifies embedded URLs within email bodies.
- Uses regex to extract HTTP/HTTPS links.

3.3 Trusted Domain Validation

- Compares extracted domains with a predefined whitelist.
- Alerts when URLs originate from suspicious or unknown domains.

3.4 VirusTotal Integration

- Queries VirusTotal API to:
    - Submit URLs for scanning
    - Retrieve analysis reports
    - Evaluate malicious or suspicious ratings
- Provides evidence-based threat evaluation.

3.5 AI-Based Contextual Analysis

- Uses Groq with an LLM (gpt-oss-120b) to classify email content as:
    - **SAFE**
    - **SUSPICIOUS**
    - **PHISHING**
- AI considers tone, intent, inconsistencies, linguistic features, and context.

3.6 PDF Reporting Module

- Generates a professional report summarizing:
    - Keywords
    - URLs
    - VirusTotal results
    - Overall risk analysis

4. Code Explanation – Module by Module

Below is a detailed description of the script you provided.

4.1 Importing Required Libraries

Your script imports multiple libraries for processing, scanning, UI formatting, and API communication.

Key categories:

- **Regex & parsing:** `re`, `tldextract`
- **Display enhancements:** `colorama`, `tabulate`, `tqdm`
- **File handling and storage:** `os`, `json`, `fpdf`
- **External APIs:** `requests`, `groq`
- **VirusTotal scanning:** Your custom module `vt_url_check`

Each library plays a specific role in modularizing the project.

Below is a detailed description of the script you provided.

5. Phishing Keyword Detection

5.1 Keyword Dataset

The `PHISHING_KEYWORDS` array contains more than 150 keywords categorized into:

- Urgency & threats
- Social engineering
- Fake financial claims
- Technical jargon
- Account compromise indicators
- Attachment/download lures

5.2 Detection Logic

```
for keyword in PHISHING_KEYWORDS:
    if re.search(rf"\b{keyword}\b", email_content, re.IGNORECASE):
        found_keywords.append(keyword)
```

- Case-insensitive matching
- Word-boundary search
- Optimized for large keyword lists

5.3 Purpose

Phishing commonly relies on emotional pressure and manipulated urgency. This module quickly highlights suspicious patterns.

6. URL Extraction and Domain Analysis

6.1 URL Extraction

```
url_pattern = r'(https?://[^\s]+)'
```

- Captures all URLs within the email content.

6.2 Domain Extraction

The library `tldextract` normalizes URLs to avoid inconsistencies.

Example:

```
https://mail.google.secure.com
→ google.secure.com
```

6.3 Trusted Domain Validation

A large trusted list is included covering:

- Social media
- e-commerce
- email providers
- cloud platforms
- banks
- government institutions

- educational entities

Unknown domains are flagged as suspicious.

7. VirusTotal URL Analysis Module

7.1 Workflow

1. Submit URL
2. Poll analysis status
3. Retrieve final reputation statistics
4. Extract `malicious` and `suspicious` votes from 90+ antivirus engines

7.2 Key Code

```
analysis_id = submit_url(url)
report = get_url_report(url_id)
stats = report["data"]["attributes"]["last_analysis_stats"]
```

7.3 Output Example

```
malicious=3
suspicious=5
stats={"malicious": 3, "suspicious": 5, "undetected": 38, ...}
```

7.4 Importance

This provides real-world, evidence-based threat scoring.

Example Workflow

Step 1 — User selects "Analyze Email File"

Step 2 — System reads file

Step 3 — Keyword scanning finds:

```
['urgent', 'account locked', 'verify your identity']
```

Step 4 — URL scanning finds:

http://login-security-paypal.com/session/reset

Step 5 — Domain check identifies:

```
login-security-paypal.com → NOT trusted
```

Step 6 — VirusTotal returns:

```
malicious=12, suspicious=4
```

Step 7 — AI concludes:

```
Verdict: PHISHING
```

Step 8 — Report is saved

Conclusion:

The Phishing Email Detection System combines keyword scanning, URL analysis, domain reputation checks, VirusTotal integration, and AI-driven evaluation to provide a robust defense against phishing attacks. Its multi-layered approach ensures accurate detection, actionable insights, and practical reporting. Continuous updates and future enhancements will further strengthen its effectiveness, making it a reliable and scalable cybersecurity solution.