Intelligent Cyber-Security
Case studies: Social engineering attacks

Dr. Alexandru ARCHIP

Gheorghe Asachi Technical University of Iasi

Artificial Intelligence (MSc., second year) – 2025 – 2026

alexandru.archip@academic.tuiasi.ro

In brief...

1. The attack pattern

2. Phishing
   - Weaknesses and vulnerabilities
   - Classic mitigation techniques
   - AI/ML defensive approaches

Outline

Definitions and examples

Social engineering [1]

Social engineering refers to the act of deceiving an individual into performing actions that compromise security, expose personal data, or damage the individual's own reputation.

Phishing is the most commonly employed technique for conducting *social engineering*-based attacks.

Effects

A successful social engineering/phishing attack could enable a malicious actor to:

- obtain sensitive information about a person or a company;
- gain unauthorised access to security credentials;
- deploy malware on the victim's IT equipment;
- conceal different types of fraud attempts.

Definitions and examples

- In 2023, the threat group Scattered Spider used social engineering and multi-factor authentication bypass to compromise internal systems at MGM Resorts International and Caesars Entertainment, obtaining employee credentials and customer data [3].
- In 2021, the Colonial Pipeline attack began with a phishing email, ultimately leading to ransomware deployment and the shutdown of major fuel pipeline operations in the United States [4].
- In 2023, Kroll reported emerging social engineering campaigns leveraging collaborative platforms such as Microsoft Teams, highlighting a shift towards more sophisticated phishing vectors [5].
- On 3 March 2025, the Microsoft Security Blog reported a tax-theme phishing campaign targeting U.S. accountants, which used PDF attachments that led to ZIP files containing a Windows shortcut that launched GuLoader and ultimately installed the Remcos backdoor [6].

* Slide generated with the assistance of ChatGPT.

Outline

Security considerations

- Phishing techniques are both an *attack* and an *attack vector*:
  - if the target of the malicious actor is the data themselves, *phishing* is an *attack*;
  - if *phishing* is used to deliver malware or other exploitation payloads, it is a component of the *attack vector*.
- In either case, it is observable during the *Reconnaissance*, *Delivery / Initial Access* and *Command&Control / Lateral Movement* stages of the CKC [7] and MITRE ATT&CK Matrix [8], respectively.
- it is considered one of the most dangerous attacks:

	October	**November**	**December**
Number of unique phishing Web sites (attacks) detected	345,881	313,288	329,954
Unique phishing email campaigns	28,327	27,668	33,899
Number of brands targeted by phishing campaigns	315	333	309

Table 1: Phishing statistics for Q4 2024 (taken from [9])

Security considerations

- Developing phishing pages has become remarkably simple.
  - Several GitHub repositories publicly provide *phishing kits* — core code required to develop deceptive web pages.
  - Brezeanu *et al.* [10] demonstrated that these *phishing kits* have been used in recent attacks with minimal modifications.
- A new threat emerges with the advancement of Generative AI solutions. Begou *et al.* [11] examined the resilience of LLMs against prompt injection attacks and showed that publicly available LLMs can be easily manipulated into generating sophisticated phishing campaigns.
  - Using only six distinct prompts, Begou *et al.* were able to successfully: i) clone a targeted website; ii) integrate code for stealing credentials; iii) obfuscate the code; iv) automate website deployment on a hosting provider; v) register a phishing domain name; and vi) integrate the website with a reverse proxy.

Weaknesses and vulnerabilities

Potential weaknesses:

- **Psychological** factors are the **root cause**:
  - users often focus only on the *look and feel* of a web page; if it resembles the expected page, they rarely check other indicators;
  - attackers exploit human emotions such as fear, panic, duty, happiness, or excitement to persuade individuals into clicking malicious links.
- **IT**-related **weaknesses** also contribute to the proliferation of such attacks:
  - XSS may be exploited as an attack vector to deliver deceptive pages;
  - authentication schemes may not enforce strict validation criteria (e.g., multi-factor authentication, trusted device/location checks);
  - protocol weaknesses and software vulnerabilities further expand the *phishing attack surface*.

Weaknesses and vulnerabilities

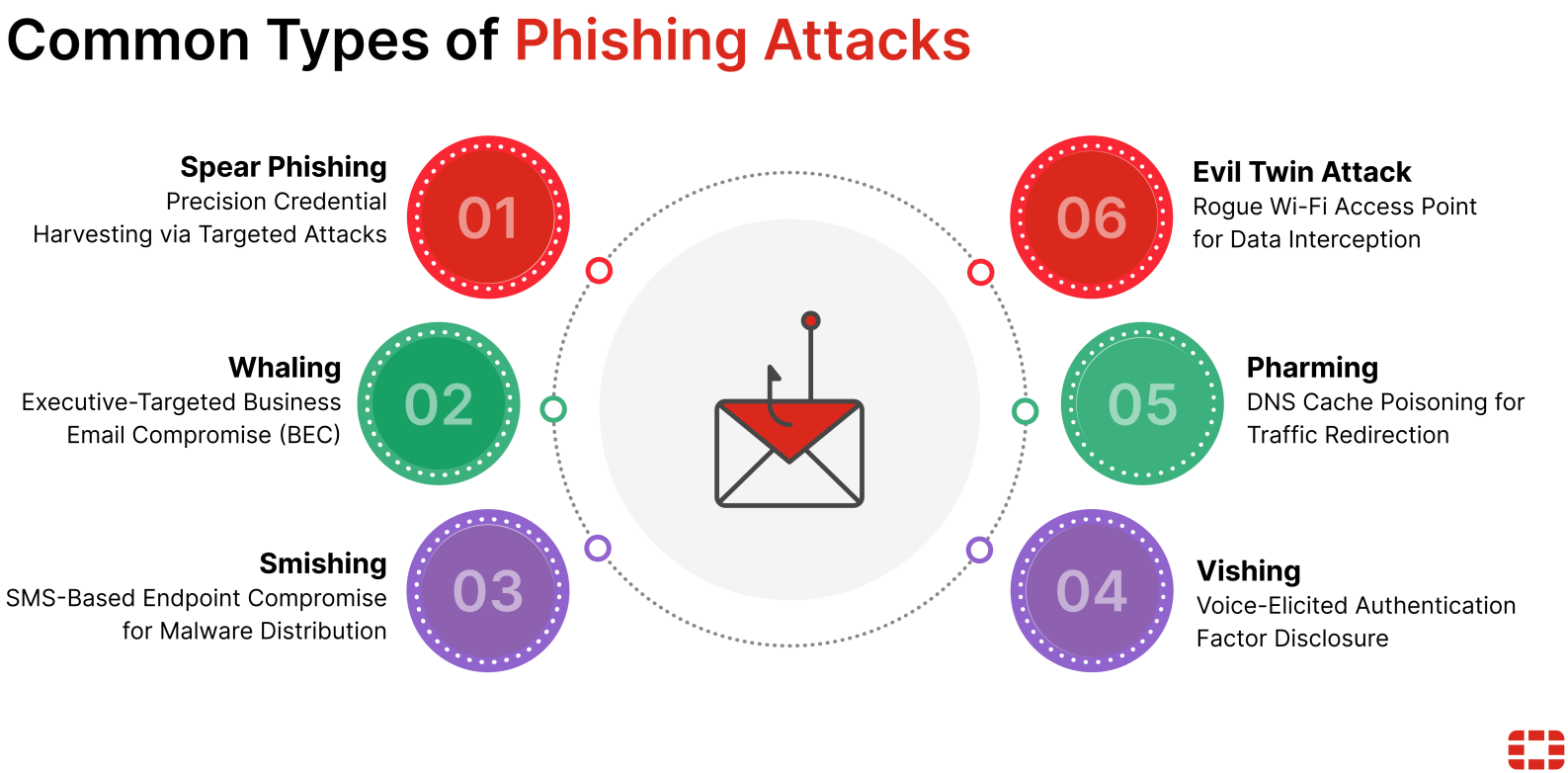

Figure 1: Phishing enabling vulnerabilities (image taken from [12])

Classic mitigation techniques

- Raising user awareness remains the most important mitigation measure.
  - Remember that *psychological factors* are the root cause!
- URL filtering approaches are the first and most common technical line of defence:
  - web service providers maintain allow/deny lists of known phishing websites; web browsers include components that check requested URLs against these deny lists before performing the request;
  - some approaches employ regular-expression-based analysis to determine whether the requested URL resembles a known legitimate one for a given web authority.
- Web browsers with embedded antivirus or antimalware modules perform basic code analysis steps:
  - various on- and off-page factors are analysed and checked against known threat signatures;
  - identified patterns are then compared with known threats and blocked if necessary.
- Mitigating known attacks that could serve as attack vectors is also highly important (e.g. detecting and preventing XSS reduces the likelihood of phishing delivery through this channel).

AI/ML defensive approaches

Apruzzese *et al.* [13, Section 3.3 *Machine Learning in Phishing Detection*] provide a comprehensive review of ML/AI applications in phishing detection. Key findings include:

- there are two distinct goals in phishing detection — identifying fraudulent websites and detecting delivery mechanisms (e.g. email, SMS);
- detecting malicious websites involves analysing data extracted from URLs, HTML/CSS/JavaScript code, the visual representation of webpages, and various external factors (e.g. DNS or WHOIS queries);
- email analysis relies primarily on textual content inspection, often complemented by attachment scanning;
- in most scenarios, supervised ML techniques are preferred, as labelling legitimate webpages is considered a relatively straightforward task.

Phishing website detection

Sahingoz *et al.* [14] focus on URL analysis:

Features NLP-based features computed on the words within the URL;

words include components such as domain names, special expressions (e.g. `www`, `com`), and random strings;

Word Vector and hybrid NLP-Word Vector embeddings were also tested.

Classification algorithms seven algorithms, including Naive Bayes, Random Forest, and kNN.

Training and validation 10-fold cross-validation; algorithms were used with default parameter values.

Results Random Forest achieved a 97.98% accuracy and 99% recall when trained on NLP features.

Phishing website detection

Niakanlahiji *et al.* [15] describe *PhishMon*:

Features

HTTP: header field names, number of header fields, and number of non-standard headers;

Code complexity: minified or obfuscated code, number of external script blocks, number of inline script blocks, number of DOM event handlers, and whether the URL is redirected;

Digital certificate: whether the certificate passes browser validation, number of included subdomains, longevity, and issuer.

Classification algorithms CART, kNN, AdaBoost, and Random Forest.

Training and validation 10-fold cross-validation.

Results Random Forest achieved an accuracy of 95.4% with a false positive rate of only 1.3%.

Phishing website detection

Corona *et al.* [16] propose *DeltaPhish*:

Features URL components, HTML code, and links between the homepage and the analysed page;

visual rendering of the analysed page and the website's homepage, using *histogram of oriented gradients* and color histograms.

Classification algorithms SVM for both text data and visual snapshots; results are combined using an SVM with a Radial Basis Function (RBF) kernel.

Training and validation 5-fold cross-validation.

Results 95% detection rate with a false positive rate of 1%.

Phishing email detection

Fang *et al.* [17] implement *Themis*:

Features character-level and word-level representations of email headers and body. Word2Vec is used to derive token embeddings.

Classification algorithms Recurrent Convolutional Neural Networks with an Attention Mechanism.

Training and validation traditional training-validation-test split, combined with 10-fold cross-validation on the training-validation set for parameter fine-tuning.

Results 99% recall, 99.664% precision, and 99.331% F1-score.

Phishing website detection – unsupervised

Brezeanu *et al.* introduce *Phish Fighter*:

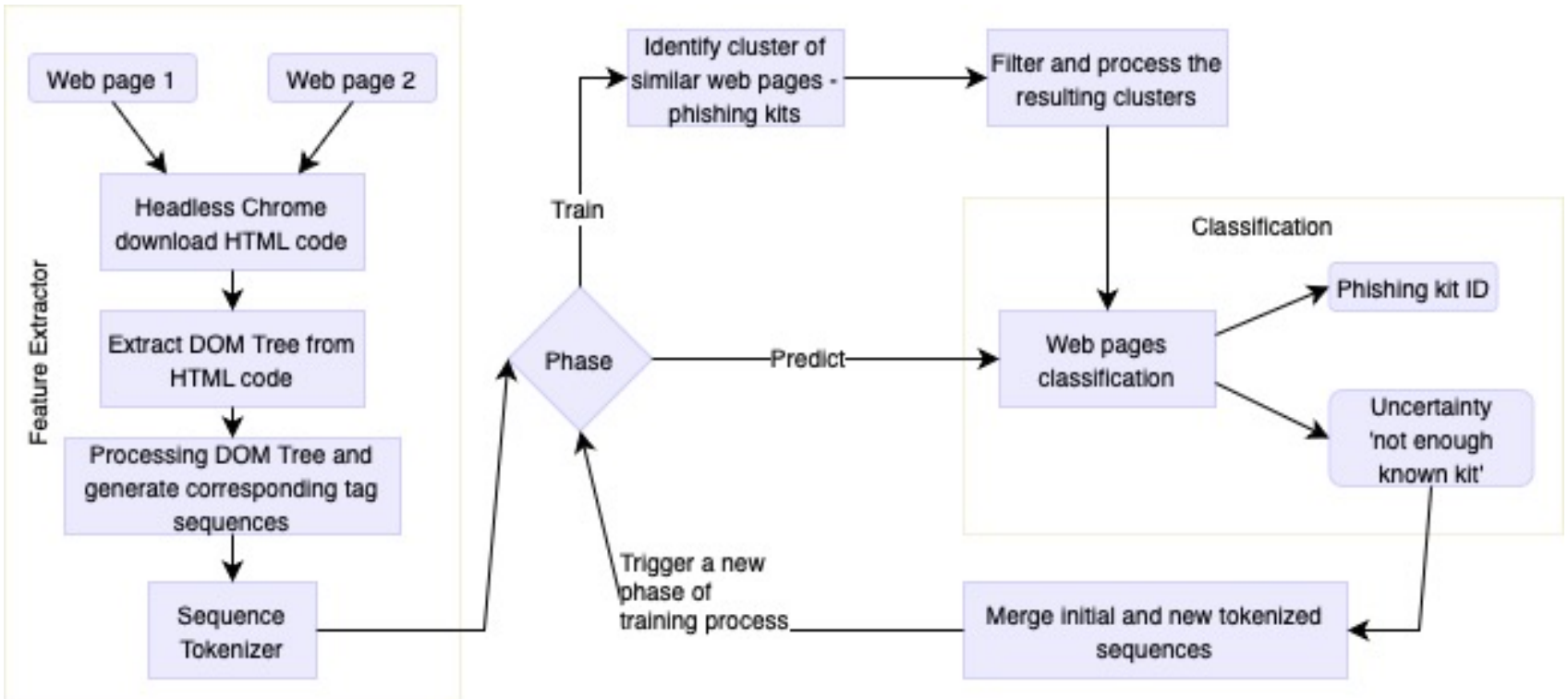

Figure 2: Phish Fighter's architecture (taken from [10])

Phishing website detection – unsupervised

Features	DOM tag sequence derived from the <body> of the HTML code.
Machine learning algorithms	clustering: AGNES, with centroid linkage and silhouette score dynamic threshold;
	classification: Random Forest with 128 decision trees.
Training and validation	traditional training-validation split;
	tests have also been conducted on pages collected from different sources *after* training had been finalised.
Results	Precission: micro 95.13%, weighted 91.55%
	Recall: micro 95.13%, weighted 95.13%
	F1-score: micro 95.13%, weighted 93.30%

Bibliography

1. National Institute of Standards and Technology. (n.d.). *Social engineering*. In *NIST Computer Security Resource Center Glossary*. National Institute of Standards and Technology. https://csrc.nist.gov/glossary/term/social_engineering
2. National Cyber Security Centre. (n.d.). *Phishing and scams*. In *NCSC Collection*. National Cyber Security Centre. https://www.ncsc.gov.uk/collection/phishing-scams
3. Wikipedia. (2024). *Scattered Spider*. In *Wikipedia, the free encyclopedia*. Wikimedia Foundation. https://en.wikipedia.org/wiki/Scattered_Spider
4. BlueVoyant. (2023). *8 Devastating Phishing Attack Examples and Prevention Tips*. https://www.bluevoyant.com/knowledge-center/8-devastating-phishing-attack-examples-and-prevention-tips
5. Kroll. (2023). *Q3 2023 Threat Landscape Report: Social Engineering*. https://www.kroll.com/en/reports/cyber/threat-intelligence-reports/q3-2023-threat-landscape-report-social-engineering

Bibliography

6. Microsoft. (2025). *Threat actors leverage tax-season to deploy tax-themed phishing campaigns*. Microsoft Security Blog. https://www.microsoft.com/en-us/security/blog/2025/04/03/threat-actors-leverage-tax-season-to-deploy-tax-themed-phishing-campaigns/
7. Lockheed Martin. (n.d.), *Cyber Kill Chain*, https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html
8. MITRE Corporation. (n.d.). *MITRE ATT&CK®*. In *MITRE ATT&CK Knowledge Base*. MITRE Corporation. https://attack.mitre.org/
9. Anti-Phishing Working Group. (n.d.). *Phishing Activity Trends Reports*. In *APWG Research*. Anti-Phishing Working Group. https://apwg.org/trendsreports
10. Gabriela Brezeanu, Alexandru Archip, and Codruț-Georgian Artene. (2025). *Phish Fighter: Self Updating Machine Learning Shield Against Phishing Kits Based on HTML Code Analysis*. *IEEE Access*, 13, 4460-4486. https://doi.org/10.1109/ACCESS.2025.3525998

Bibliography

11. Nils Begou, Jérémy Vinoy, Andrzej Duda, and Maciej Korczyński. (2023). *Exploring the Dark Side of AI: Advanced Phishing Attack Design and Deployment Using ChatGPT*. In *Proceedings of the 2023 IEEE Conference on Communications and Network Security (CNS)*, 1-6. https://doi.org/10.1109/CNS59707.2023.10288940

12. Fortinet. (n.d.). *19 Types of Phishing Attacks*. In *Fortinet Cyber Glossary*. Fortinet. https://www.fortinet.com/resources/cyberglossary/types-of-phishing-attacks

13. Apruzzese, G., Laskov, P., Montes de Oca, E., Mallouli, W., Brdalo Rapa, L., Grammatopoulos, A. V. and Di Franco, F. (2023). *The Role of Machine Learning in Cybersecurity*. *Digital Threats*, 4(1), Article 8. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3545574

14. Sahingoz, Ö. K., Buber, E., Demir, Ö., and Diri, B. (2019). *Machine learning based phishing detection from URLs*. *Expert Systems with Applications*, 117, 345357. https://doi.org/10.1016/j.eswa.2018.09.029

Bibliography

15. A. Niakanlahiji, B.-T. Chu, and E. Al-Shaer. (2018). *PhishMon: A Machine Learning Framework for Detecting Phishing Webpages*. In *Proceedings of the 2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 220–225. IEEE. https://doi.org/10.1109/ISI.2018.8587410

16. I. Corona, B. Biggio, M. Contini, L. Piras, R. Corda, M. Mereu, G. Mureddu, D. Ariu, and F. Roli. (2017). *DeltaPhish: Detecting Phishing Webpages in Compromised Websites*. arXiv preprint arXiv:1707.00317.

17. Y. Fang, C. Zhang, C. Huang, L. Liu, and Y. Yang. (2019). *Phishing Email Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism*. *IEEE Access*, 7, 56329–56340. https://doi.org/10.1109/ACCESS.2019.2913705.