Sniff-Phish: A Proactive Phishing Detection System Using Chrome Extension with URL, Domain, Content, and Machine Learning Analysis

Manish Chandra Sinha[1]; Dr. Amol Kumbhare[2]
[1]Research Scholar; [2]Associate Professor
College of Engineering, Dr. A. P. J. Abdul Kalam University, Indore (M.P.)
Email id: manish.sinha66@gmail.com

Abstract
Phishing is one of the most prevalent forms of social engineering attacks where hackers steal user credentials through deceptive URLs and email scams. With the rapid expansion of digitization, cybersecurity has become increasingly critical. This paper presents Sniff-Phish, a proactive phishing detection system implemented as a Google Chrome browser extension, developed in three progressive versions. Sniff-Phish 1.0 detects phishing URLs based on URL and domain-based features, achieving a detection rate of 97.2% against conventional phishing attacks, 76.52% against zero-day attacks, and 72.22% against spear-phishing attacks. Sniff-Phish 2.0 extends the framework to include lexical, page, and content-based feature analysis with 18 weighted parameters, improving zero-day detection to 89.12%. Sniff-Phish 3.0 integrates machine learning via a hybrid decision tree classifier, achieving a true positive rate of 99.6% with a false positive rate below 1%. All three versions outperform existing state-of-the-art anti-phishing tools, making Sniff-Phish a lightweight, client-side, and proactive solution against phishing attacks.

Keywords
Phishing Detection, Chrome Extension, Zero-Day Attack, Spear-Phishing, URL Analysis, Machine Learning, Decision Tree, Sniff-Phish.

1. Introduction

The Internet has transformed modern society by enabling new forms of communication, commerce, and education. However, this technological advancement has also created new opportunities for cybercriminals. Phishing, a form of social engineering attack, remains one of the most effective and widespread cybersecurity threats. Attackers craft deceptive URLs and fraudulent emails to trick unsuspecting users into revealing sensitive credentials, causing significant financial and data losses worldwide.

Phishing attacks are broadly categorized into three types: conventional phishing, which targets a large number of users indiscriminately; zero-day phishing, which exploits newly registered or unknown malicious domains that bypass blacklists; and spear-phishing, a targeted form of attack directed at specific individuals with privileged access or social authority. Existing anti-phishing tools primarily rely on blacklists, which suffer from inherent delays in identifying new phishing domains. Studies show that 50–80% of phishing domains are only added to blacklist databases after financial damage has already occurred.

This paper presents Sniff-Phish, a browser-integrated, client-side phishing detection system developed as a Google Chrome extension. The system employs a progressive three-version architecture: Sniff-Phish 1.0 (URL and domain feature analysis), Sniff-Phish 2.0 (enriched rule-based incremental construction), and Sniff-Phish 3.0 (hybrid machine learning classification). The system avoids dependence on blacklists and operates in real-time within the user's browser, providing an efficient, lightweight, and proactive solution.

2. Literature Review

A substantial body of research has addressed the phishing detection problem from multiple perspectives. The existing anti-phishing methodologies can be broadly classified into the following categories:

2.1 User Awareness-Based Approaches

Steve et al. (2007) proposed an interactive instructional game called 'Anti-Phishing Phil' to educate users. Kumaraguru et al. (2007) established integrated training proposals to impart awareness about phishing emails. While effective to a degree, these approaches are limited by the human factor, as studies show 53% of users remain vulnerable even after encountering phishing attacks.

2.2 Blacklist/Whitelist-Based Approaches

Prakash et al. (2010) proposed PhishNet, an approximate matching algorithm against known blacklists. Jain et al. (2016) designed a dynamic whitelist framework achieving a true positive rate of 86.02%. The major limitation of blacklist approaches is the delay between phishing domain creation and blacklist update, making them ineffective against zero-day attacks.

2.3 Visual Similarity-Based Approaches

Dunlop et al. (2010) proposed GoldPhish using image-based content analysis. Chen et al. (2010) applied page-content similarity for phishing detection. Chiew et al. (2015) utilized website logos as distinguishing features. These methods are computationally intensive and unsuitable for lightweight client-side deployment.

2.4 Machine Learning-Based Approaches

Mohammad et al. (2014) used self-structuring neural networks for phishing classification. Basnet and Doleck (2015) applied machine learning for URL-based detection. Feng et al. (2018) employed a novel neural network for phishing identification. Despite high accuracy, many ML-based systems require significant training overhead, making them less suitable for real-time browser environments.

2.5 Hybrid Approaches

Orunsolu et al. (2019) proposed a predictive model combining multiple techniques. Niakanlahiji et al. (2018) developed PhishMon using an ML framework for webpage detection. The research gap identified in the literature is the lack of a lightweight, real-time, client-side tool that combines heuristic feature analysis with machine learning to handle zero-day and spear-phishing attacks effectively.

3. Proposed Methodology

The proposed system Sniff-Phish is developed as a Google Chrome browser extension that operates at the client side. The extension analyses any URL or hyperlink right-clicked within the browser and returns a legitimacy-level report before the user accesses the link. The framework was developed progressively across three versions to address increasing levels of phishing attack sophistication.

3.1 Sniff-Phish 1.0 – URL and Domain Feature Analysis

Sniff-Phish 1.0 employs 11 URL and domain-based rules to determine whether a given link is phishing or legitimate. These features include:

- IP Address in URL – Presence of an IP address instead of a domain name.
- URL Length – Abnormally long URLs are suspicious.
- URL Shortening – Use of URL shortening services (e.g., bit.ly).
- @ Symbol in URL – Use of '@' to redirect to a malicious domain.

- Double Slash Redirect – Occurrence of '//' in URL path.
- Prefix/Suffix with Hyphen – Hyphens in domain names.
- Sub-domain Depth – Number of sub-domain levels in URL.
- HTTPS Status – Presence and validity of HTTPS certificate.
- Domain Registration Age – Newly registered domains flagged as suspicious.
- WHOIS Record – Missing or incomplete domain ownership information.
- Blacklist Status – Cross-check against known phishing blacklists.

Each parameter returns a weighted score and a collective legitimacy percentage is calculated to classify the URL as phishing, suspicious, or legitimate.

3.2 Sniff-Phish 2.0 – Feature-Enriched Rule-Based Framework

Sniff-Phish 2.0 expands the feature set to 20 parameters, adding page-based and content-based features. The WHOIS server and blacklist dependency from version 1.0 were removed to address GDPR compliance (Ferrante et al. 2018) and to improve detection against zero-day attacks. The four feature categories are:

1. URL-Based Features: Characteristics derived from the URL string itself.
2. Domain-Based Features: Information obtained from passive domain queries.
3. Page-Based Features: Website credibility and traffic-based attributes.
4. Content-Based Features: Analysis of web content requiring active tracking.

Weightage values for each parameter were empirically determined by analyzing 1,500 phishing URLs from PhishTank and 1,500 legitimate URLs from Alexa Top Sites. The weighted mean of all parameter scores determines the legitimacy level of the URL.

3.3 Sniff-Phish 3.0 – Hybrid Machine Learning Classification

Sniff-Phish 3.0 integrates a machine learning classifier with the heuristic rule-based engine. Six machine learning algorithms were evaluated: Decision Tree, Random Forest, Naive Bayes, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Logistic Regression. The Decision Tree with ensemble methods achieved the highest classification accuracy and was selected for the final hybrid model.

The hybrid architecture operates in four phases:

5. Phase I – Attribute Weightage Derivation: Phishing attribute weights are derived from empirical dataset analysis.
6. Phase II – Dataset Collection and Feature Extraction: Phishing URLs are collected from PhishTank, legitimate URLs from Alexa. Features are extracted and datasets labelled for supervised learning.
7. Phase III – Algorithm Selection and Training: Six ML algorithms are benchmarked; Decision Tree with ensemble is selected based on accuracy.
8. Phase IV – Chrome Extension Integration: The trained classifier is embedded into the Chrome extension for real-time prediction.

4. Experimental Results and Discussion

All experiments were conducted on an Intel Core i7 Q720 CPU @ 1.60 GHz laptop running 64-bit Windows 7 Home Premium with the latest version of Google Chrome, connected to the internet at an average download speed of 10 Mbps. Phishing datasets were sourced from PhishTank, IsitPhishing, Enron Email Dataset, Millersmiles Scam Report, and Tumblr Targeted Email Attacks. Legitimate datasets were obtained from Alexa Top Sites, Stuffgate Free Online Website Analyzer, and major search engines.

4.1 Performance Metrics

The system was evaluated using the following standard metrics: True Positive Rate (TPR), False Positive Rate (FPR), False Negative Rate (FNR), True Negative Rate (TNR), and overall Accuracy. TPR is defined as the proportion of phishing URLs correctly classified as phishing. FPR is the proportion of legitimate URLs misclassified as phishing.

4.2 Comparison of Sniff-Phish Versions

Table 1: Detection Rate Comparison across Sniff-Phish Versions

Attack Type	Sniff-Phish 1.0 TPR	Sniff-Phish 2.0 TPR	Sniff-Phish 3.0 TPR
Conventional Phishing	97.2%	97.13%	99.6%
Zero-Day Attack	76.52%	89.12%	96.8%
Spear-Phishing	72.22%	76.2%	94.4%
False Positive Rate	2.1%	1.5%	<1.0%
Overall Accuracy	97.2%	97.13%	99.2%

4.3 Comparison with Existing Anti-Phishing Tools

Sniff-Phish was compared against leading anti-phishing Chrome extensions including Avira Browser Safety, Netcraft Extension, BitDefender Traffic Light, Dr.Web Link Checker, and Web of Trust (WOT). The comparison was conducted on both conventional phishing and zero-day attack datasets.

Table 2: Comparison with Existing Anti-Phishing Chrome Extensions

Tool	Conventional TPR	Zero-Day TPR	FPR
Avira Browser Safety	94.1%	61.3%	3.2%
Netcraft Extension	92.5%	58.7%	4.1%
BitDefender Traffic Light	93.8%	63.0%	3.5%
Dr.Web Link Checker	90.2%	54.9%	5.3%
Web of Trust (WOT)	91.7%	57.4%	4.8%
Sniff-Phish 3.0 (Proposed)	99.6%	96.8%	<1.0%

4.4 Discussion

The results demonstrate that Sniff-Phish 3.0 significantly outperforms both its predecessor versions and existing commercial anti-phishing tools across all three attack categories. The incorporation of machine learning in the hybrid model particularly improves zero-day and spear-phishing detection rates, which are the most challenging attack types for blacklist-based systems. The incremental architecture ensures that each version builds upon the weaknesses identified in the previous one, resulting in a continuously improving detection system.

The false positive rate below 1% in Sniff-Phish 3.0 is a significant improvement over competing tools, reducing user inconvenience from false alarms. The lightweight, client-side deployment ensures that the system operates without requiring cloud-based or server-side infrastructure, making it suitable for real-world browser usage.

5. Conclusion and Future Work

This paper presented Sniff-Phish, a proactive phishing detection system implemented as a Google Chrome browser extension in three progressive versions. Sniff-Phish 1.0 achieved a detection rate of 97.2% using URL and domain features. Sniff-Phish 2.0 extended feature coverage to include page and content analysis, improving zero-day detection to 89.12%. Sniff-Phish 3.0 integrated a hybrid decision tree machine learning classifier, achieving an overall accuracy of 99.2% with a true positive rate of 99.6% and a false positive rate below 1%.

All three versions outperform conventional phishing detection methods and existing anti-phishing tools in terms of zero-day and spear-phishing detection, confirming the effectiveness of the proactive, feature-based, blacklist-independent approach. Being lightweight and client-side, the system is practical for real-world deployment without centralized infrastructure support.

5.1 Future Work

Several directions for future research are identified:

9. Investigating the impact of expanding the feature set and varying feature selection strategies on detection accuracy.
10. Evaluating the system on larger and more diverse datasets encompassing a wider variety of phishing attack vectors.
11. Incorporating deep learning algorithms such as LSTM or BERT-based models for improved contextual URL analysis, and adopting online dynamic learning to minimize false positive rates.
12. Extending the detection framework to mobile platforms and developing detection mechanisms for attacks targeted at mobile devices.
13. Exploring adversarial robustness to ensure the system remains effective against evolving evasion strategies employed by sophisticated phishing attackers.

References

[1] Abdelhamid, N, Ayesh, A & Hadi, W 2014, 'Multi-Label Rules Algorithm Based Associative Classification', Parallel Processing Letters, vol. 24, no. 1, pp. 1-21.

[2] Ahmed, A & Lina, Z 2017, 'Phishing environments, techniques, and countermeasures: A survey', Computers & Security, vol. 68, no. 1, pp. 160-196.

[3] Anjum, S, Antesar, M, Hossain, MA 2016, 'A literature review on phishing crime, prevention review and investigation of gaps', Proceedings of the tenth SKIMA Conference, pp. 9-15.

[4] Ankitkumar, J & Gupta, BB 2018, 'Towards detection of phishing websites on client-side using machine learning based approach', Telecommunications Systems, vol. 68, no. 4, pp. 687–700.

[5] Basnet, RB & Doleck, T 2015, 'Towards developing a tool to detect phishing urls: A machine learning approach', Proceeding of the 2015 IEEE CICT Conference, pp. 220–223.

[6] Chen, C, Dick, S & Miller, J 2010, 'Detecting visually similar web pages: Application to phishing detection', ACM Transactions on Internet Technology, vol. 10, no. 2, pp. 1-38.

[7] Chiew, KL, Chang, EH, Sze, SN & Tiong, WK 2015, 'Utilisation of website logo for phishing detection', Computers and Security, vol. 54, pp. 1-11.

[8] Dhamija, R, Tygar, JD & Marti, H 2006, 'Why phishing works', Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 581–590.

[9] Dunlop, M, Groat, S & Shelly, D 2010, 'GoldPhish: Using Images for Content-Based Phishing Analysis', Proceeding of the Fifth ICIMP Conference, pp. 123-128.

[10] Feng, F, Zhou, Q, Shen, Z, Yang, X, Han, L & Wang, J 2018, 'The application of a novel neural network in the detection of phishing websites', Journal of Ambient Intelligence and Humanized Computing, pp. 1-15.

[11] Ferrante & Anthony, J 2018, 'The impact of GDPR on WHOIS: Implications for businesses facing cybercrime', Cyber Security: A Peer-Reviewed Journal, vol. 2, no. 2, pp. 143-148.

[12] Fette, I, Sadeh, N & Tomasic, A 2007, 'Learning to detect phishing Emails', Proceedings of the International WWW Conference, pp. 649-656.

[13] Gupta, BB, Tewari, A & Jain, AK 2017, 'Fighting against phishing attacks: state of the art and future challenges', Neural Computing and Applications, vol. 28, no. 12, pp. 3629–3654.

[14] Jain, AK & Gupta, BB 2016, 'A novel approach to protect against phishing attacks at client side using auto-updated white-list', EURASIP Journal on Information Security, vol. 1, no. 9, pp. 1-11.

[15] KPMG 2019, What's next: Key cyber security considerations for 2019. Available from: <https://advisory.kpmg.us/>. [3 July 2019].

[16] Kumaraguru, P et al. 2007, 'Getting users to pay attention to anti-phishing education', Proceedings of the eCrime Researchers Summit, pp. 70–81.

[17] Mohammad, R, Thabtah, F & McCluskey, L 2014, 'Predicting phishing websites based on self-structuring neural network', Neural Computing and Applications, vol. 25, no. 1, pp. 443–458.

[18] Niakanlahiji, A, Chu, B & Al-Shaer, E 2018, 'PhishMon: A Machine Learning Framework for Detecting Phishing Webpages', Proceedings of the IEEE ISI Conference, pp. 220-225.

[19] Orunsolu, AA, Sodiya, AS & Akinwale, AT 2019, 'A predictive model for phishing detection', Journal of King Saud University - Computer and Information Sciences, vol. 5, pp. 1-16.

[20] Panos, L & Christof, E 2016, 'Machine Learning', IEEE Software, vol. 33, no. 5, pp. 110-115.

[21] Prakash, P, Kumar, MR, Kompella, R & Gupta, M 2010, 'PhishNet: Predictive Blacklisting to Detect Phishing Attacks', Proceedings of the IEEE INFOCOM, pp. 1-5.

[22] Steve, S et al. 2007, 'Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish', Proceedings :SOUPS Symposium, pp. 88–99.

[23] Zhang, Y, Jason, IH & Lorrie, FC 2007, 'Cantina: A content-based approach to detecting phishing web sites', Proceedings of the 16th WWW Conference, pp. 639–648.