International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Phishing Website Detection

Nivyashree R[1], Bhoomika S R[2], Chiranth H M[3], Bhavan N Gowda[4], Chakram Janya H U[5]
[1]Asst. Professor,Department of Computer Science & Engineering, Malnad College of Engineering
[2]Department of Computer Science & Engineering, Malnad College of Engineering
[3] Department of Computer Science & Engineering, Malnad College of Engineering
[4] Department of Computer Science & Engineering, Malnad College of Engineering
[5] Department of Computer Science & Engineering, Malnad College of Engineering

---

Abstract - This literature survey examines software-based phishing detection techniques, a critical area of cybersecurity. With phishing attacks growing rapidly each year, this study explores the phishing ecosystem, current statistics, automatic detection schemes, feature analysis, datasets, algorithms, and evaluation metrics. Emphasis is given to the challenges in feature robustness, handling adaptive attacks, and limitations in large-scale data processing. The survey also identifies research gaps in addressing new attack vectors, offering insights for future directions.

***Key Words*:** Phishing, Phishing Detection, Literature Survey, Cybersecurity, Machine Learning.

1.INTRODUCTION

Phishing has emerged as one of the most prevalent forms of cyber-attacks, exploiting social engineering to deceive users and steal sensitive information. The constant evolution of phishing tactics poses significant threats to online security. This paper presents a comprehensive review of literature on web phishing detection, emphasizing software-based approaches. The study outlines key techniques, highlights research gaps, and proposes future research directions to strengthen defenses against phishing threats.

2. Body of Paper

2.1 Phishing Techniques and Impact

Phishing involves deceptive practices, often mimicking trusted websites. Dou et al. (2018) report a 36.29% average annual growth in phishing sites over six years, peaking at 97.36% growth in recent years. The economic implications are severe, with users losing trust in online systems (Sunil & Sardana, 2017). Mao et al. (2017) emphasize the visual mimicry of phishing pages as a critical technique for deceiving users.

2.2 Detection Techniques

Several methods have been developed to detect phishing attacks. A PageRank-based approach by Sunil and Sardana (2017) achieved 98% accuracy using heuristic-based detection. Mao et al. (2017) introduced "Phishing-Alarm," leveraging Cascading Style Sheet (CSS) features to analyze page similarity. Dou et al. (2018) conducted a systematic review on phishing detection, covering algorithms, datasets, features, and evaluation metrics. A hybrid model using ensemble predictions and clustering for phishing categorization (2019) also shows promise.

2.3 Challenges and Gaps

Despite progress, existing systems face limitations. Dou et al. (2018) noted issues with evaluating feature robustness and adapting to changing datasets and attack methods. Sunil and Sardana (2017) highlighted shortcomings of blacklist-based systems that fail to identify new phishing sites. Mao et al. (2017) discuss evasive tactics such as using images or invisible content. Neupane et al. (2016) conducted an fMRI study revealing lower cognitive activity in impulsive users when exposed to phishing.

Literature Review

A. Systematization of Knowledge (SoK) [1]
Zuochao Dou et al. provide a systematic review of existing software-based phishing detection techniques. The authors present a taxonomy, review datasets and features used, and evaluate machine learning algorithms in phishing detection.

- **Problem Identified**: Difficulty in evaluating feature robustness and handling large-scale datasets.
- **Gap**: Existing methods degrade over time due to evolving attack patterns.

B. PageRank-Based Detection [2]
Sunil and Sardana propose a lightweight detection approach using Google's PageRank and GTR heuristics to classify phishing sites.

- **Problem Identified**: Blacklisting fails to detect new or evolving phishing sites.
- **Gap**: Limited scalability; lacks adaptability to unseen attacks.

C. Network Function Virtualization (NFV) [3]
Li and Chen present a survey on NFV and its integration with software-defined networking (SDN), enabling dynamic and scalable security services.

- **Problem Identified**: Lack of efficient API design for NFV communication.
- **Gap**: Mismatch between network demand and provisioning capabilities.

D. Phishing-Alarm Using Page Component Similarity [4]
Mao et al. introduce a CSS-based phishing detection system implemented as a browser extension to assess visual similarity.

- **Problem Identified**: Vulnerability to image-based content replacement.
- **Gap**: High-performance cost for image-based comparison; limited resilience to advanced evasion techniques.

E. Intelligent Anti-phishing Strategy Model [5]
This paper presents an ensemble learning model combined with hierarchical clustering for phishing detection using large-scale real-world data.

- **Problem Identified**: High complexity and false positives in traditional models.

- **Gap**: Difficulties in real-time processing and deployment on large datasets.

F. Neural Markers of Cybersecurity [6]

Neupane et al. use fMRI studies to analyze user brain responses during phishing and malware detection tasks.

- **Problem Identified**: Impulsive behavior reduces detection accuracy.
- **Gap**: Experimental results may not fully translate to real-world scenarios.

Feature / Paper	[1]	[2]	[3]	[4]	[5]	[6]
Focus	ML Review	Heuristic	Network Infra	Visual Similarity	Ensemble Learning	Neuro-Cognitive
Dataset Scale	Comprehensive	Small	N/A	Real-world	Large-scale	Controlled
Accuracy	Varies	98%	N/A	High	High	Behavior-focused
Scalability	Limited	Medium	High	Medium	Medium	Low
Gap	Robustness	Blacklist fails	API design	Image evasion	Real-time detection	Generalization

3. CONCLUSIONS

Phishing remains a persistent threat to online security. While methods like PageRank analysis and CSS-based detection offer high accuracy, challenges such as adaptive attacks and large-scale dataset handling remain. This literature review underscores the need for robust, scalable, and intelligent phishing detection systems to address evolving threats effectively.

ACKNOWLEDGEMENT

The authors express their gratitude to the researchers and developers whose contributions have significantly advanced the field of phishing detection.

REFERENCES

1. Dou, Z., Khalil, I., Khreishah, A., Al-Fuqaha, A.: Systematization of Knowledge (SoK): A Systematic Review of Software Based Web Phishing Detection.

2. Sunil, A.N.V., Sardana, A.: A PageRank Based Detection Technique for Phishing Web Sites.

3. Mao, J., Tian, W., Li, P., Wei, T., Liang, Z.: Phishing-alarm: Robust and Efficient Phishing Detection via Page Component Similarity.

4. Li, Y., Chen, M.: Software-defined Network Function Virtualization: A Survey.

5. An Intelligent Anti-phishing Strategy Model for Phishing Website Detection.

6. Neupane, A., Saxena, N., Maximo, J.O., Kana, R.: Neural Markers of Cybersecurity: An fMRI Study of Phishing and Malware Warnings.

BIOGRAPHIES

Mrs. Nivyashree R, currently working as Asst. Professor in the Dept. of CSE at Malnad College of Engineering, Hassan.

Bhoomika S R, pursuing engineering from the branch of CSE at Malnad College of Engineering.

bhoomikasr17@gmail.com

Chiranth H M, pursuing engineering from the branch of CSE at Malnad College of Engineering.

chiruchiranthhm@gmail.com

Bhavan N Gowda , pursuing engineering from the branch of CSE at Malnad College of Engineering.

bhavanng@gmail.com

Chakram Janya H U, pursuing engineering from the branch of CSE at Malnad College of Engineering.

chakramjanya69@gmail.com