IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303014 |

Survey on Machine Learning Defence against Url Web Phishing

Jayachandra T, Dr. A. Rengarajan

2nd year MCA (ISMS), School of Computer Science and IT, Jain (Deemed-to-be University), Bangalore, India

Associate Professor, School of Computer Science and IT, Jain (Deemed-to-be University), Bangalore, India

ABSTRACT: The prevalence of web phishing poses a significant threat to online security, as attackers continuously devise sophisticated methods to deceive users into divulging sensitive information or executing malicious actions. A primary tactic employed in phishing attacks is the use of deceptive URLs, which mimic legitimate websites to trick unsuspecting victims. In response to this evolving threat landscape, machine learning (ML) techniques have emerged as promising tools for automatically identifying and mitigating URL-based phishing attacks. This survey paper provides an in-depth examination of the state-of-the-art ML approaches utilized in defending against URL web phishing.

The survey begins with an overview of phishing attacks, emphasizing the importance of defending against URL-based phishing due to its widespread prevalence and detrimental impact on individuals, organizations, and society. The role of ML in enhancing web security by enabling automated detection and response mechanisms is highlighted, underscoring the need for robust defenses against phishing threats.

In the background and related work section, the paper discusses existing techniques for phishing detection and the utilization of ML in web security applications. Previous surveys on phishing defense are reviewed to contextualize the current state of research and identify gaps in knowledge. Subsequently, the survey delves into methodologies employed in ML-based URL phishing defense, categorizing approaches based on feature extraction techniques, classification algorithms, dataset characteristics, and evaluation metrics. Various feature extraction methods, including lexical, content-based, and host-based features, are explored alongside popular classification algorithms such as decision trees, support vector machines, neural networks, ensemble methods, and deep learning architectures. Additionally, the characteristics of datasets used for training and evaluation purposes, as well as commonly employed evaluation metrics such as accuracy, precision, recall, and F1-score, are analyzed. Challenges and limitations inherent in ML-based URL phishing defense are discussed, including issues related to imbalanced datasets, generalization to new phishing techniques, adversarial attacks, privacy concerns, and scalability. Finally, the paper outlines future directions for research, including the integration of explainability in ML models, ensemble learning for improved robustness, transfer learning and domain adaptation, integration of human feedback mechanisms, and addressing ethical and societal implications.

This survey paper provides a comprehensive overview of the current landscape of ML-based URL phishing defense, offering insights into methodologies, challenges, and future research directions. By synthesizing existing knowledge and identifying areas for advancement, this paper aims to contribute to the ongoing efforts to enhance web security and protect users against phishing threats.

I. INTRODUCTION

In the digital age, where the Internet serves as an indispensable tool for communication, commerce, and entertainment, the prevalence of cyber threats looms large, casting shadows over the seamless fabric of cyberspace. Among these threats, phishing attacks stand out as a pervasive menace, continually evolving in sophistication and scale. Phishing, a form of cybercrime wherein attackers masquerade as trustworthy entities to deceive users into divulging sensitive information such as usernames, passwords, and financial details, poses substantial risks to individuals, businesses, and society at large. At the heart of many phishing attacks lies the deceptive use of Uniform Resource Locators (URLs), the web addresses that users click on to access websites. While legitimate URLs serve as gateways to desired online destinations, malicious URLs serve as conduits for nefarious activities, leading unsuspecting victims down treacherous paths fraught with fraud and exploitation. Recognizing the pivotal role URLs play in facilitating

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303014 |

phishing attacks, cybersecurity researchers and practitioners have turned to machine learning (ML) techniques as a means of fortifying defenses against this insidious threat.

The symbiotic relationship between phishing attacks and URLs underscores the urgency of devising robust defenses capable of discerning legitimate URLs from their malicious counterparts. Traditional approaches to URL phishing detection, reliant on static blacklists and heuristic rules, have proven inadequate in combating the dynamic and polymorphic nature of modern phishing campaigns. Consequently, the application of ML algorithms, endowed with the ability to learn patterns and characteristics from data, has emerged as a promising frontier in the ongoing battle against web phishing. Against this backdrop, this survey paper embarks on a comprehensive exploration of the landscape of ML-based defenses against URL web phishing. By synthesizing existing research, analyzing methodologies, and identifying challenges and opportunities, this paper endeavors to shed light on the state-of-the-art in URL phishing detection and pave the way for future advancements in the field of web security.

1.1 Phishing Attacks Overview

Phishing attacks represent a form of social engineering wherein attackers exploit human psychology and trust to deceive victims into performing actions that compromise their security or privacy. These actions may include clicking on malicious links, entering credentials into counterfeit websites, or downloading malware-infected attachments. Phishing attacks often leverage familiar communication channels such as email, text messages, social media platforms, and instant messaging services, making them difficult to discern from legitimate communications.

The success of phishing attacks hinges on the ability of attackers to craft convincing messages and counterfeit websites that closely mimic legitimate entities, such as banks, e-commerce platforms, or government agencies. By leveraging persuasive language, compelling visual elements, and spoofed domain names, attackers aim to create a false sense of urgency or authority that compels victims to act impulsively without questioning the authenticity of the request.

1.2 Importance of Defending Against URL Web Phishing

The proliferation of URL-based phishing attacks underscores the critical importance of deploying effective defenses to safeguard users and organizations from falling victim to such schemes. The consequences of phishing attacks can be severe, ranging from financial losses and identity theft to reputational damage and regulatory penalties. Moreover, phishing attacks often serve as entry points for more sophisticated cyber threats, such as ransomware, data breaches, and business email compromise (BEC) scams, further amplifying the need for robust defense mechanisms.

In the context of URL web phishing, where attackers manipulate web addresses to deceive users into visiting counterfeit websites or downloading malicious content, the stakes are particularly high. Unlike other forms of phishing that rely solely on social engineering tactics, URL-based phishing attacks exploit vulnerabilities in the web infrastructure itself, making them inherently more challenging to detect and mitigate. As such, the development of MLbased defenses tailored specifically to identify and neutralize malicious URLs represents a crucial line of defense in the ongoing battle against web phishing.

1.3 Role of Machine Learning in Phishing Defense

Machine learning, a subfield of artificial intelligence (AI) that focuses on enabling computers to learn from data and make predictions or decisions without explicit programming, offers a powerful arsenal of tools for combating phishing attacks. Unlike traditional rule-based approaches that rely on static signatures or heuristics, ML algorithms possess the ability to adapt and evolve in response to changing threats, making them well-suited for detecting novel and previously unseen phishing patterns.

In the context of URL web phishing, ML techniques can be leveraged to analyze various features and characteristics of URLs, such as lexical components, structural properties, and behavioral patterns, to distinguish between legitimate and malicious URLs. By training on labeled datasets comprising known examples of phishing and legitimate URLs, ML models can learn to generalize from past observations and make accurate predictions about the trustworthiness of unseen URLs encountered in the wild.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303014 |

Moreover, the interdisciplinary nature of ML, drawing upon concepts from statistics, mathematics, and computer science, allows researchers to leverage diverse techniques and algorithms, ranging from traditional classifiers such as decision trees and support vector machines to more advanced deep learning architectures such as convolutional neural networks and recurrent neural networks. This versatility enables ML-based phishing detection systems to adapt to the evolving tactics and techniques employed by attackers, thereby enhancing their efficacy and resilience in the face of sophisticated adversaries.

II. BACKGROUND

In the contemporary digital landscape, where the internet serves as an integral component of daily life for billions worldwide, the threat of cyberattacks looms large. Among the myriad of cyber threats, phishing stands out as one of the most pervasive and insidious forms of attack. Phishing attacks are designed to deceive individuals into divulging sensitive information such as passwords, financial data, or personal identification details by masquerading as legitimate entities. The attackers often employ various psychological tactics and sophisticated techniques to trick unsuspecting users into taking actions that compromise their security and privacy. The fundamental principle underlying phishing attacks is the exploitation of trust. By impersonating trusted entities like banks, social media platforms, or government agencies, cybercriminals aim to convince users to interact with malicious content, typically delivered through email, social media messages, or websites. One common technique utilized in phishing campaigns is the use of deceptive URLs (Uniform Resource Locators), which mimic legitimate web addresses to trick users into visiting malicious websites. These deceptive URLs often redirect users to counterfeit webpages designed to steal sensitive information or deliver malware.

As the prevalence and sophistication of phishing attacks continue to evolve, traditional defense mechanisms such as blacklisting known malicious URLs or using static rule-based heuristics have become inadequate. Consequently, there has been a growing interest in leveraging advanced technologies, particularly machine learning (ML), to enhance the detection and mitigation of phishing threats. Machine learning offers the promise of automated, scalable, and adaptive solutions capable of analyzing large volumes of data to discern patterns and anomalies indicative of phishing activities. Machine learning algorithms are well-suited for phishing detection tasks due to their ability to learn from historical data and extract meaningful features that discriminate between legitimate and malicious URLs. By training on labeled datasets containing examples of both phishing and legitimate URLs, ML models can learn to generalize patterns and characteristics associated with each class, thereby enabling the automated identification of potential phishing attempts in real-time.

Furthermore, the rapid proliferation of online services and the increasing complexity of phishing techniques necessitate continuous adaptation and improvement of defense mechanisms. Machine learning techniques provide a flexible framework for iteratively refining detection capabilities, as they can be trained on updated datasets to incorporate emerging trends and novel attack vectors.

Previous research efforts have explored various aspects of machine learning in the context of phishing defense, including feature extraction techniques, classification algorithms, dataset characteristics, and evaluation metrics. However, despite significant advancements in the field, several challenges and limitations persist, hindering the widespread adoption and efficacy of ML-based phishing detection systems. Addressing these challenges requires a holistic understanding of the interplay between machine learning algorithms, cybersecurity principles, and the evolving threat landscape. By delving into the background of machine learning in web security and phishing detection, this survey aims to provide a comprehensive foundation for further exploration of the topic, elucidating the opportunities, challenges, and future directions in leveraging ML for defense against URL web phishing.

III. METHODOLOGIES IN ML-BASED URL PHISHING DEFENSE

The defense against URL web phishing through machine learning (ML) entails a multifaceted approach involving various methodologies aimed at accurately identifying and classifying phishing URLs. These methodologies encompass feature extraction techniques, classification algorithms, dataset characteristics.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303014 |

1.2 Feature Extraction Techniques

Feature extraction plays a crucial role in representing URLs in a format conducive to machine learning models' analysis. Several feature extraction techniques have been explored in the context of URL phishing defense:

1. Lexical Features: These features capture the structural elements of URLs, such as the length of the URL, the presence of special characters, and the frequency of certain characters or character combinations. Lexical features provide valuable insights into the inherent characteristics of phishing URLs, enabling classifiers to discern patterns indicative of malicious intent.

2. Content-based Features: Content-based features involve analyzing the textual content of URLs, including keywords, domain names, and path segments. Techniques such as term frequency-inverse document frequency (TF-IDF) and n-gram analysis are commonly employed to extract meaningful features from URL text. By identifying suspicious keywords or domain anomalies, content-based features contribute to distinguishing phishing URLs from legitimate ones.

3. Host-based Features: Host-based features focus on attributes associated with the domain hosting the URL, such as domain age, registrar information, and geographical location. These features leverage domain reputation services, WHOIS databases, and DNS records to assess the trustworthiness of URLs based on their hosting environments. Host-based features provide valuable context for determining the likelihood of a URL being involved in phishing activities.

3.2 Classification Algorithms

Classification algorithms form the core of ML-based URL phishing defense systems, tasked with accurately categorizing URLs as either phishing or legitimate. A variety of algorithms have been explored for this purpose:

1. Decision Trees: Decision tree algorithms recursively partition the feature space based on threshold values, forming a tree-like structure that facilitates intuitive rule-based classification. Decision trees offer transparency and interpretability, making them suitable for understanding the decision-making process in URL phishing detection systems.

2. Support Vector Machines (SVM): SVMs are supervised learning models that separate data points using hyperplanes in high-dimensional feature spaces. By maximizing the margin between different classes, SVMs excel at capturing complex decision boundaries, thereby effectively distinguishing between phishing and legitimate URLs.

3. Neural Networks: Neural networks, particularly deep learning architectures, have gained prominence in URL phishing detection due to their ability to automatically learn hierarchical representations of input data. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are commonly employed to extract features from URL text and detect subtle patterns indicative of phishing attempts.

4. Ensemble Methods: Ensemble methods combine multiple base classifiers to improve predictive performance through aggregation or boosting techniques. Random forests and gradient boosting machines (GBMs) are popular ensemble methods in URL phishing defense, leveraging the diversity of constituent classifiers to enhance robustness and generalization capabilities.

5. Deep Learning Architectures: Deep learning architectures, such as deep neural networks (DNNs) and recurrent neural networks (RNNs), leverage multiple layers of abstraction to learn intricate representations of URL features. These architectures excel at capturing complex patterns and correlations in large-scale URL datasets, enabling state-of-the-art phishing detection performance.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303014 |

3.3 Dataset Characteristics

The characteristics of the dataset used for training and evaluating ML models significantly influence the effectiveness and generalization capabilities of URL phishing defense systems:

1. Publicly Available Datasets: Publicly available datasets, such as the PhishTank and OpenPhish repositories, provide labeled examples of phishing URLs collected from diverse sources. These datasets facilitate benchmarking and comparative analysis of different phishing detection approaches, fostering transparency and reproducibility in research.

2. Synthetic Datasets: Synthetic datasets generated using data augmentation techniques or adversarial perturbations enable researchers to explore the robustness of ML models against various evasion strategies. By synthesizing realistic phishing scenarios, synthetic datasets contribute to assessing the resilience of URL phishing defense systems under different threat landscapes.

3. Real-world Datasets: Real-world datasets sourced from live phishing campaigns or security incident reports offer insights into the evolving tactics and techniques employed by adversaries. These datasets capture the dynamic nature of URL phishing threats, enabling the evaluation of ML models' efficacy in detecting emerging phishing patterns and trends.

Challenges and Limitations

1. Imbalanced Datasets:

One of the significant challenges in training effective machine learning models for phishing URL detection is dealing with imbalanced datasets. Phishing datasets often contain a disproportionate number of legitimate URLs compared to phishing URLs. This class imbalance can lead to biased models that favor the majority class, resulting in poor performance in identifying phishing URLs. Techniques such as oversampling, undersampling, or using advanced algorithms like cost-sensitive learning are employed to address this issue.

2. Generalization to New Phishing Techniques:

Phishers continually evolve their tactics to bypass traditional detection mechanisms. As a

Result, machine learning models trained on historical phishing data may struggle to generalize to new and emerging phishing techniques. Adversarial learning techniques, where models are trained to recognize adversarial examples crafted to deceive them, can enhance robustness against novel phishing attacks. Nonetheless, ensuring the adaptability of ML models to evolving phishing tactics remains a persistent challenge.

3. Adversarial Attacks:

Adversarial attacks pose a significant threat to the effectiveness of machine learning-based phishing detection systems. Attackers can manipulate or craft malicious URLs specifically designed to evade detection by trained models. Adversarial examples often exploit vulnerabilities in feature representations or model architectures to deceive classifiers. Robust machine learning techniques such as adversarial training, feature obfuscation, or input sanitization are essential for mitigating the impact of adversarial attacks. However, designing models resilient to sophisticated adversarial manipulations without sacrificing performance on legitimate data remains an ongoing research challenge.

4. Privacy Concerns:

Machine learning-based URL phishing detection systems often rely on access to sensitive user data, such as browsing history or email communication. Privacy-preserving mechanisms are crucial to protect user confidentiality while still enabling effective phishing detection. Federated learning, differential privacy, and homomorphic encryption are among the techniques used to address privacy concerns in machine learning models. However, striking a balance

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303014 |

between preserving user privacy and ensuring model efficacy poses inherent trade-offs that require careful consideration.

IV. CONCLUSION

In this survey paper, we have explored the landscape of machine learning (ML) techniques for defending against URL web phishing, a pervasive threat in the realm of cybersecurity. Our analysis revealed a rich diversity of methodologies, ranging from feature extraction techniques to sophisticated classification algorithms, all aimed at identifying and thwarting phishing attacks in real-time. One of the key insights gleaned from our investigation is the pivotal role that ML plays in enhancing web security by automating the detection of phishing URLs. Traditional approaches often rely on static rule-based systems or manually curated blacklists, which struggle to keep pace with the rapidly evolving tactics employed by cybercriminals. In contrast, ML offers a dynamic and adaptive framework capable of learning patterns and anomalies inherent in phishing URLs, thereby enabling proactive defense mechanisms. Feature extraction emerges as a critical component in ML-based phishing detection systems, with lexical, content-based, and host-based features being widely utilized to capture the distinctive characteristics of malicious URLs. These features serve as input to various classification algorithms, ranging from classical methods like decision trees and support vector machines to more advanced techniques such as neural networks and ensemble learning. The effectiveness of these algorithms is further bolstered by the availability of diverse datasets, including publicly available repositories, synthetic datasets, and real-world collections, enabling robust model training and evaluation.

Despite the progress made in ML-based URL phishing defense, several challenges and limitations persist. Imbalanced datasets, wherein the number of phishing URLs is significantly lower than legitimate ones, pose a substantial hurdle to model generalization and performance. Moreover, the dynamic nature of phishing attacks necessitates continuous adaptation of ML models to emerging threats, requiring robust mechanisms for scalability and real-time updates. Additionally, concerns related to privacy, adversarial attacks, and ethical implications underscore the need for holistic approaches that prioritize not only accuracy but also transparency, fairness, and user trust. Looking ahead, several avenues for future research and innovation emerge. Incorporating explainability techniques into ML models can enhance their interpretability, enabling stakeholders to understand the rationale behind phishing predictions and fostering trust in automated defense systems. Ensemble learning methodologies hold promise in improving the robustness and resilience of detection algorithms by leveraging diverse models and input features. Transfer learning and domain adaptation techniques offer avenues for leveraging knowledge from related domains to enhance the generalization capability of phishing detection models. Furthermore, integrating human feedback mechanisms can leverage the expertise of security professionals and end-users, augmenting the capabilities of ML systems and facilitating rapid response to emerging threats. The intersection of machine learning and web security presents a fertile ground for innovation in combating URL web phishing. By addressing the challenges outlined in this survey and embracing emerging opportunities, researchers, practitioners, and policymakers can collectively advance the stateof-the-art in phishing defense, safeguarding the integrity and trustworthiness of online ecosystems for generations to come.

REFERENCES

[1] Amiri, I. S., Akanbi, O. A., & Fazeldehkordi, E. (2014). A machine-learning approach to phishing detection and defense. Syngress.
[2] Tang, L., & Mahmoud, Q. H. (2021). A survey of machine learning-based solutions for phishing website detection. Machine Learning and Knowledge Extraction, 3(3), 672694.
[3] Kulkarni, A. D., & Brown III, L. L. (2019). Phishing websites detection using machine learning.
[4] James, J., Sandhya, L., & Thomas, C. (2013, December). Detection of phishing URLs using machine learning techniques. In 2013 international conference on control communication and computing (ICCC) (pp. 304-309). IEEE.
[5] Korkmaz, M., Sahingoz, O. K., & Diri, B. (2020, July). Detection of phishing websites by using machine learning-based URL analysis. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.
[6] Deshpande, A., Pedamkar, O., Chaudhary, N., & Borde, S. (2021). Detection of phishing websites using Machine Learning. International Journal of Engineering Research & Technology (IJERT), 10(05).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details