

Email Security: A Systematic Survey

Vaibhavi Talawnekar, Prof. Sheetal Gawande, Diksha Patil, Arya Patil, Vinay Saini

Student Department of Information Technology, Pillai College of Engineering, New Panvel

Professor Department of Computer Engineering, Pillai College of Engineering, New Panvel

Student Department of Information Technology, Pillai College of Engineering, New Panvel

Student Department of Information Technology, Pillai College of Engineering, New Panvel

Student Department of Information Technology, Pillai College of Engineering, New Panvel

Abstract

E-mail is becoming a tool we use in our daily life for communication like people, organizations, businesses and government establishments. Email is a very beneficial tool, but is the most vulnerable service for cyber threats such as spoofing, phishing, business email compromise, spreading malwares, viruses, trojans, worms etc. Old email security methods, which are primarily focused on rule-based filtering methodologies are often not effective when it comes to increase of advancements and developing attacks on email. To conquer these drawbacks, researchers are exploring different methodologies like ML and hybrid techniques that can identify mails that are intended to cause harm

to systems more productively and precisely. This systematic survey confers the leading threats that are caused to security of email communication where it highlights the sub domains- phishing and spoofing of email. It reviews the recent developments in intelligent email threat detection, compares the existing techniques based on strengths and limitations and identifies the existing research gaps. The survey also discusses future directions for research into more reliable, scalable and adaptive email security solutions to address the emerging cyber threats.

Key Words: Email Security, Email Spoofing, Email Phishing, Machine Learning

1. INTRODUCTION

Email remains a necessary means of exchanging personal, academic and professional information. But it is fundamentally vulnerable, despite its critical role. The issue is with the Simple Mail Transfer Protocol (SMTP), designed in the early days of the internet, and

1.1 Email Security

lacking any built-in process to verify the real identity of a sender. This fundamental weakness lets attackers impersonate trusted entities, spread malware, run phishing campaigns and easily commit financial fraud.

Broadly defined, email security is the collection of technologies, protocols, and operational practices used to protect email communications from unauthorized access, spoofing, spam, malware, and phishing. At its heart, information security is built on four pillars: confidentiality, integrity, authentication and availability. Modern shielding methods combine with established authentication quality protocols (SPF, DKIM and DMARC) with intelligent behavioral examination activated by machine learning to remarkably increase detection accuracy and reinstate trust in digital communication.

1.2 Statistics of Email Security Threats

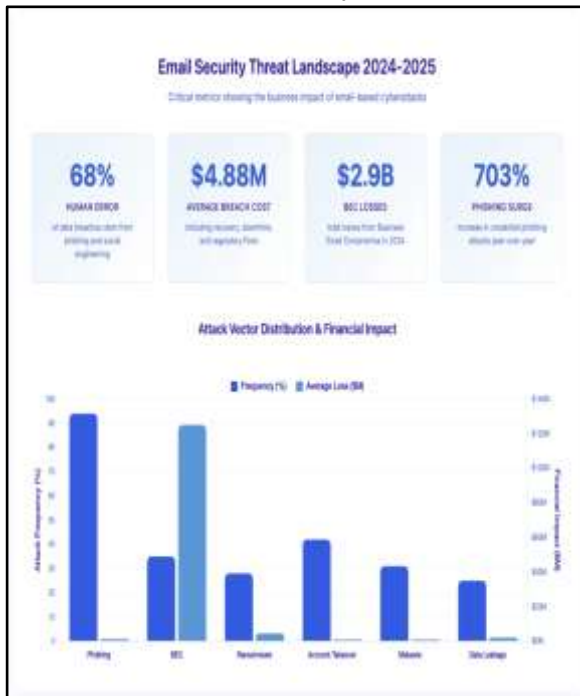


Fig 1.2.1 Email security threats

The visualization that presented illustrates the recent obstacles email security is facing. It conveys that phishing remains the most common ambush used to breach email security, having the largest financial impact on Business Email Compromise (BEC). The core technique used to implement these attacks remains spoofing, a well known method where the attacker disguises as harmless to trick victims so they will reveal confidential information or information that can lead to fraudulent payments. Such threats emphasize the importance of authentication protocols that lead to security by verifying the legitimacy of the person who dispatched the mail. The protocols SPF, DKIM and DMARC help organizations to take preventive measures to block spoofed mails before the incident is done.

1.3 Email Phishing

Email phishing is a type of cyber-attack which uses email contents for malevolent purposes to replicate verified organizations, institutions and identity of individuals. They bait victims by offering large amounts of monetary prizes, and link suspicious links in the contents of email messages. They sometimes disguise themselves as banks or any other large organization, in order to take advantage of human psychology. The different types involved in phishing include:

- **Bulk Phishing:** Casts a wide, untargeted net.
- **Spear Phishing & Whaling:** Extremely attacked, very costly threats

transmitted through important figures in organizations like executives, directors etc.

- **Clone Phishing & Business Email Compromise (BEC):** Highly advanced identification trickery focused at seizing workflows of trusted and legitimate business organizations. Present-day strategy utilizes combination of URLs with new attack techniques like by linking zero-day malwares which are capable of avoiding authentication protocols and infiltrate the system. To overcome this problem the recent research decided to substantially use machine learning and deep learning models. These techniques proved to be helpful as it can analyze, identify patterns of user and text as well as structure of URLs in emails. The results are satisfactory as they show accuracy rate remarkably, putting a halt to attempts of phishing before they can reach the victims inbox.

1.4 Email Spoofing

Email spoofing is a technique used by attackers to change the structural information of an email so that it appears to have been sent from a trusted domain. The idea is to relax the guard of the recipient so that they interact with malicious payloads or reveal confidential data. By default, SMTP has no sender validation, allowing attackers to spoof the email meta-data and change the “from” address to impersonate virtually any domain.

To address this vulnerability the industry has a triad of defensive protocols:

1. **Sender Policy Framework (SPF):**
Checks if the mail server sending the message is explicitly authorized by the owner of the domain.
2. **DomainKeys Identified Mail (DKIM):**
Adds cryptographic signature to ensure the message content was not modified in transit.
3. **Domain-based Message Authentication, Reporting and Conformance (DMARC):**
Ties SPF and DKIM together, providing domain owners with strict policies for handling authentication failures and visibility via automated reporting.

Although valuable, these protocols are not silver bullets. Research has shown that clever SMTP manipulation, protocol edge cases, and human errors in domain configuration can be used by attackers to regularly bypass them.

This means that in next-generation email security frameworks these authentication checks are increasingly being combined with machine learning classifiers to catch fake messages that look technically valid.

1.5 Email Header Analysis

Users interact with the body of an email, but every email has a header. The header is a rich source of technical metadata that captures how the email was created and delivered across the web. The body is the message, but the header is the digital paper trail. This creates an accurate repository for detecting irregularity and identification of spoofed mails. The responsible header fields are:

- **From, Return-Path, and Reply-To** (Sender identities)
- **Received** (The path from where the email was directed.)
- **Message-ID** (The special identity generated by the directed server.)
- **Authentication-Results** (The outcome that is achieved after the authentication protocols (SPF, DKIM, DMARC) checks are performed)
- **Subject and Date**

The inspection on these fields which are performed by the security systems are for revealing the signs like sender does not match domains, suspicious servers, abnormal patterns used for routing or imitated identifications. This method is considered best because it uses meta-information for analyzing the header and is a speedy process.

They do not need to go through the contents of email which speeds up the process as well as provides privacy. The main concern presented here is between privacy and speed which made analyzing mail header a foundation for ML-based spoofing detection models.

2. PRISMA MODEL

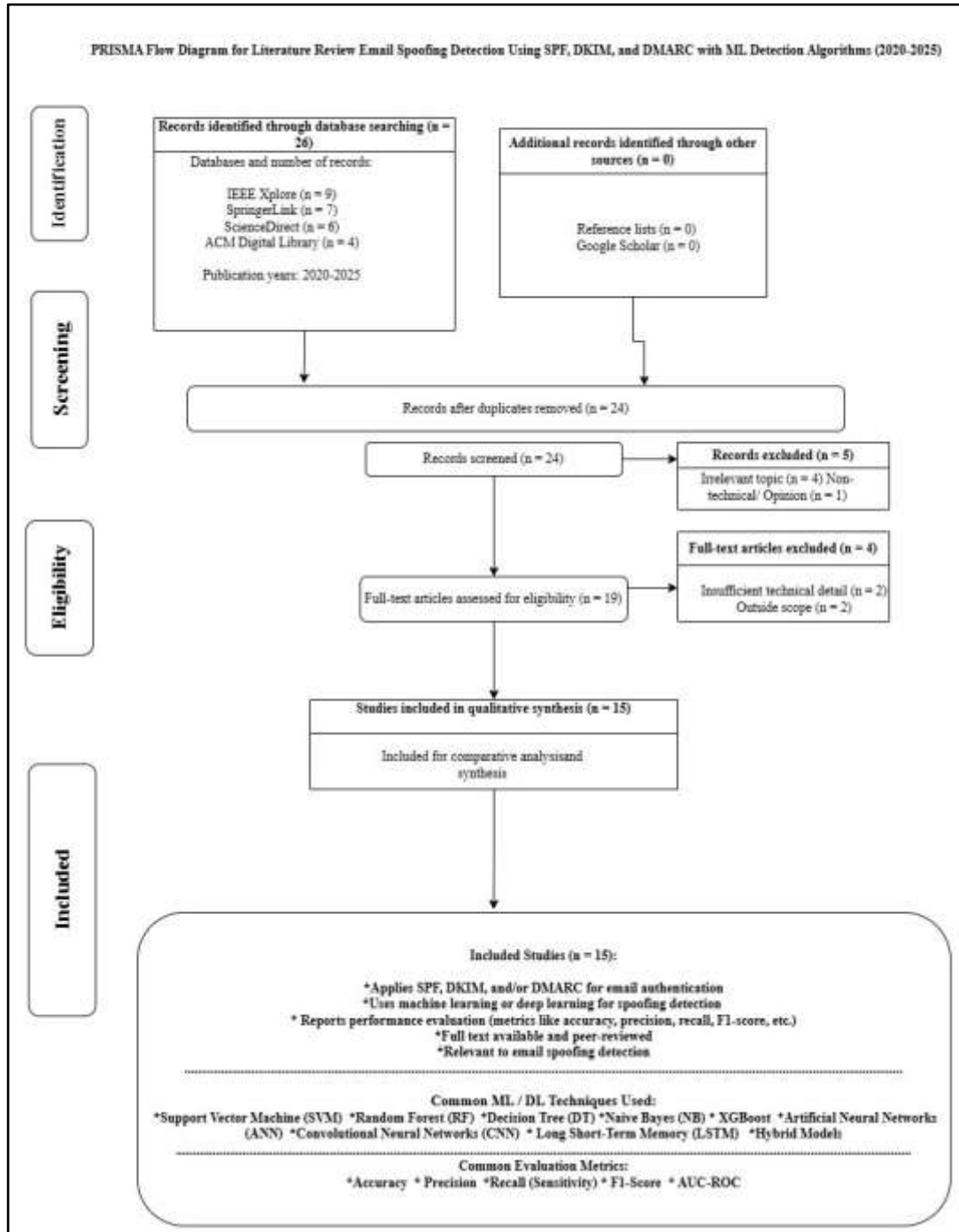


Fig 2.1 Paper selection for conducting survey

This visualization presented as a PRISMA flow diagram is to inform the process of paper selection for conducting this survey. As presented, 15 papers are

relevant and used for this survey after excluding them from 26 spotted papers. The selection criteria is clearly stated in the figure presented.

3. RESEARCH TRENDS

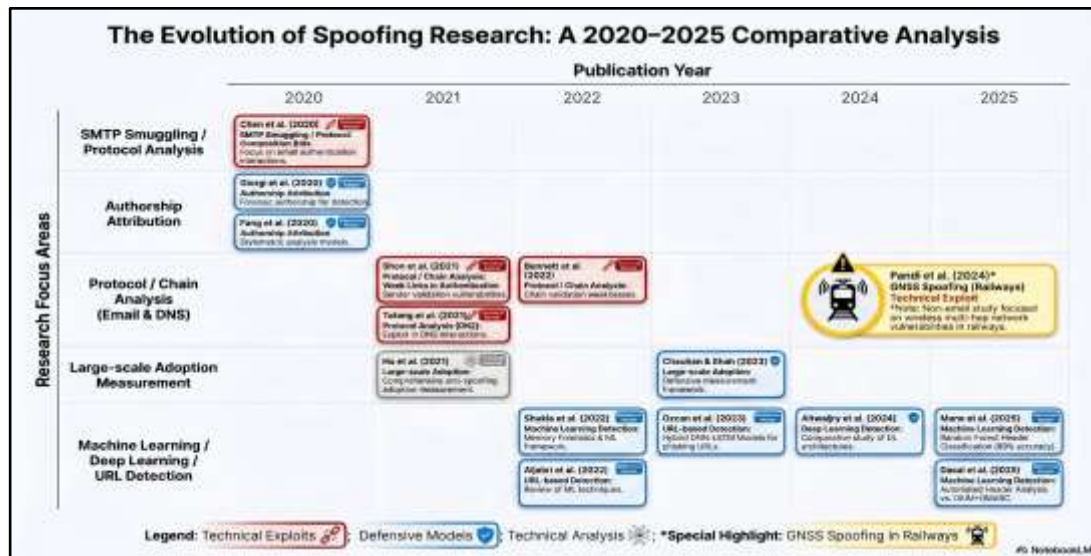


Figure 3.1 Research Trends of Email Spoofing

The visualization presented in this part evaluates the trends that evolve related to research for email spoofing (2020 to 2025). As discussed in earlier sections the prior research focused solely on authentication protocols to

detect spoofing but as time machine learning and deep learning methods were used for detection of spoofing sometimes combined with prior and new methods together for better results.

I.Email Spoofing

4.LITERATURE REVIEW

Email spoofing is a technique to impersonate the identity of trusted individuals in the world of technology. Invader (sender) tries to form a communication by sending mail from an unknown

source impersonating as a trusted source to the victim (receiver) but in actuality it is fraudulent. From numerous different angles researches have analyzed this threat to security:

- **Authentication-based Approaches:**

The foundation of this approach is policy frameworks and cryptographic checks like protocols for header analysis. The analysis is performed, by paper Email Spoofing Defence Techniques, designates the challenges of present-day assumption [13]. The advanced attack of paper SMTP Smuggling exhibits how minor oddity in managing the protocol sanctions attackers to create a diversion for spoofed mails [1].

- **Machine Learning (ML)-based Approaches:**

For detecting the abnormalities researchers apply ML, working apart from inflexible rule-based systems. The paper GNSS-based Train positioning Recognizing spoofing attack formed on Gradient boosting

(BO-LightGBM) for complex signatures recognition [10]. At the data-level, the paper Power of Machine Learning to Prevent Spoofing uses a combination of NLP with identifiers (classifiers) for detecting duplicitous vocabulary [9]. To resolve this problem with ‘black box’, the paper XAI for Improving ML-Based Spoofing Detectors presents explainable AI, which leads to transparency of detection outputs [2].

- **Hybrid Approaches:** Since there is no agreed method we use Hybrid systems where a combination of intelligent algorithms and protocol checks are applied. The paper Reliable Email Spoofing Detection uses a combination of header analysis with SVMs and neural networks [3], and this paper Comparative Study of Spoofing Detection conveys that with

the combination of automated header scrutiny with authentication protocols provides a rigid shielding than utilizing one of them solely [4].

II. Email Phishing

The trick of manipulating human psychology while blending it with technical deception is known as phishing. Spoofing is a technique where as phishing is an attack which manipulates the victims by exploiting their trust, greed so they reveal their confidential information.

- **Machine Learning (ML)-based Approaches:** Phishing content transforms quickly so solutions presented to this problem are deep learning models. The paper Deep Learning based Phishing URLs and Email Detection utilizes methods which implement RNNs, CNNs and Transformers for analyzing implanted URLs and content [7]. Also this

paper Cloud-based Email Phishing Attack differentiates LSTM networks, Naïve Bayes and SVM from the perspective of infrastructure for classifying uncertain campaigns of phishing [11].

- **Hybrid Approaches:** They use a chain of multiple classifiers to enhance accuracy. Phishing Detection via Hybrid Machine Learning introduces the LSD ensemble model, which combines predictions from Logistic Regression, SVM, and Decision Trees [8]. The paper Real-Time Phishing Detection on URL Features advances by pacing forward for scrutinizing linguistic features countering domain status metrics using a multi-classifying pipeline permitting real-time protection by countering suspicious and malicious links [6].

Table 4.1 Literature Review

Category	Paper (Year)	Detection Approach	Work Done	Methodology	Technology Used	Success	Limitations
Spoofing	Composition Kills (2020) [15]	Authentication-Based	Studied weaknesses in email sender authentication.	Protocol analysis	SPF, DKIM, DMARC	Identified 18 authentication bypass techniques.	Requires protocol-level changes.
	Weak Links in Authentication Chains (2021) [14]	Authentication-Based	Analyzed spoofing attacks on major email providers.	Attack analysis	SPF, DKIM, DMARC	Discovered 14 spoofing attack types.	No practical detection framework.
	Email Spoofing Defence Techniques (2022) [13]	Authentication-Based	Reviewed deployment of authentication protocols.	Literature review & protocol evaluation	SPF, DMARC, DNS Analysis	Measured real-world adoption of security protocols.	Focuses on protocol adoption rather than detection.

	Email Spoofing: In Today's Era (2022) [12]	Authentication-Based	Demonstrated spoofing despite authentication.	Experimental analysis	SPF, DKIM, DMARC	Showed spoofed emails can bypass authentication.	Results vary across providers.
	Cross-Domain Spoofing Detection (2024) [10]	ML-Based	Proposed optimized spoofing detection model.	ML classification	BO-LightGBM	Achieved 97.2% accuracy.	Requires extensive tuning.
	ML, Data Science & NLP to Prevent Spoofing (2024) [9]	Hybrid-Based	Combined ML, NLP, and domain verification.	ML + NLP + Domain Verification	NLP, ML, Domain Verification	Improved malicious content detection.	Increased system complexity.
	SMTP Smuggling (2025) [1]	Authentication-Based	Demonstrated protocol-level bypass attacks.	Experimental attack demonstration	SMTP Manipulation, SPF, DMARC	Successfully bypassed authentication mechanisms.	Limited to SMTP smuggling attacks.
	Comparative Analysis of Email Spoofing Detection (2025) [4]	Authentication-Based	Comparing authentication and header analysis.	Comparative analysis	SPF, DKIM, DMARC, Header Analysis	Showed authentication alone is insufficient.	Provider-dependent performance.
	Reliable Email Spoofing Detection (2025) [3]	ML-Based	Detected spoofed emails using header analysis.	Header analysis + ML classification	Random Forest	Achieved 94–96% accuracy.	Depends on correct SPF/DMARC configuration.

	XAI for ML-Based Spoofing Detectors (2025) [2]	ML-Based	Improved detection using Explainable AI.	Explainable Machine Learning	SHAP, LIME, XAI	Over 90% detection performance.	Vulnerable to adversarial attacks.
Phishing	Cloud-based Email Phishing Attack (2023) [11]	ML-Based	Compared ML and DL models for phishing detection.	ML/DL comparison	SVM, Naïve Bayes, LSTM	Achieved 99.62% accuracy.	High computational cost.
	Hybrid ML Based Phishing Detection (2024) [8]	Hybrid-Based	Combined multiple ML models for phishing detection.	Ensemble Machine Learning	Logistic Regression, SVM, Decision Tree	Improved accuracy and reduced false positives.	Cannot detect all phishing attacks.
	Real-Time Phishing Detection (2025) [6]	ML-Based	Detected phishing using URL-based ML.	URL-based ML classification	Logistic Regression, SVM, Decision Tree	Achieved 97.3% accuracy.	Less effective on new URLs.
	Deep Learning for Phishing Detection (2025) [7]	ML-Based	Applied deep learning for phishing detection.	Deep Learning	CNN, RNN, Transformers	Achieved 95–98% accuracy.	Requires large datasets and high computational resources.
	Staying Ahead of Phishers (2025) [5]	Hybrid-Based	Reviewed hybrid and ensemble phishing detection methods.	Survey/Literature Review	ML, Deep Learning, GANs, Ensemble Models	Hybrid models showed better robustness and accuracy.	Survey paper without implementation.

5. CONCLUSION

Email Security is considered a major concern as the attacks related to phishing and spoofing are getting advanced which makes the security system difficult to identify them. This survey conducted on email security analyzes the latest research conducted on this topic and learns that by creating a combination of different techniques like machine learning, deep learning and header analysis can provide a better solution for securing the email security. This evaluation also noticed the growth of Explainable AI (XAI), which provides a better clarity and simplifies AI-driven decisions. Despite promising outcomes, there are still challenges which include new attacking strategies, limitation of

datasets to conduct research, increase in computational requests and danger from hostile attacks. The future research should concentrate on points like availability of dataset primarily based on headers which will help in detecting spoofing as well as expand adaptive, adjustable and smart security remedies that can defend themselves against advanced attacks in real-time. As this survey highlights the combination of precise identification while enhancing clarity and functional effectiveness which can help to create secure and trusted communication for email service used by organizations, institutions, businesses and individuals.

ACKNOWLEDGEMENT

We hereby thank to our Guide Prof. Sheetal Gawande for giving us knowledge & spending their valuable time in making us understanding the concept for this survey.

REFERENCES

- [1] C. Wang et al., "Email Spoofing with SMTP Smuggling: How the Shared Email Infrastructures Magnify this Vulnerability," Proc. USENIX Security Symposium, 2025. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity25/presentation/wang-chuhan>
- [2] N. H. Imam, "XAI for Improving the Security of ML-Based Spoofing Email Detectors," Journal of Information Security and Cybercrimes Research, vol. 8, no. 2, pp. 111–128, 2025. [Online]. Available: <https://journals.seu.edu.sa/jiscr>
- [3] D. Mane et al., "Reliable Email Spoofing Detection using Enhanced Cybersecurity Approaches," Advances in Nonlinear Variational Inequalities, vol. 28, no. 3s, 2025. [Online]. Available: <https://www.anvi.org>
- [4] C. C. Dasal et al., "Comparative Analysis of Email Spoofing Detection Using Automated Header Analysis and Authentication Schemes," Cognizance Journal of Multidisciplinary Studies, vol. 5, no. 10, 2025. [Online]. Available: <https://cognizancejournal.com>
- [5] S. Kavya and D. Sumathi, "Staying ahead of phishers: a review of recent advances and emerging methodologies in phishing detection," Artificial Intelligence Review, vol. 58, 2025. [Online]. Available: <https://doi.org/10.1007/s10462-024-11055-z>
- [6] A. Shukla et al., "Real-Time Phishing Detection Using Machine Learning on URL Features," in Proc. IEEE ICONAT, 2025, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/ICONAT66879.2025.11362893>
- [7] V. S. Pandi et al., "Detecting Phishing URLs and Email using Deep Learning: Developing Cybersecurity," in Proc. IEEE INCSST, 2025, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/INCSST64791.2025.11210288>
- [8] N. Siva et al., "Phishing Detection System through Hybrid Machine Learning Based on URL," in Proc. IEEE INCET, 2024, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/INCET61516.2024.10593373>
- [9] S. R. Chanthati, "How the power of machine learning, data science and NLP can be used to prevent spoofing and reduce financial risks," Global Journal of Engineering and Technology Advances, vol. 20, no. 2, 2024. [Online]. Available: <https://doi.org/10.30574/gjeta>

- [10] S. Sowmya and A. S. V., “Cross-Domain Spoofing Detection Using BO-LightGBM,” ITDECC Conference, 2024. [Online]. Available: <https://ieeexplore.ieee.org>
- [11] U. A. Butt et al., “Cloud-based email phishing attack using machine and deep learning algorithm,” Complex & Intelligent Systems, vol. 9, pp. 3043–3070, 2023. [Online]. Available: <https://doi.org/10.1007/s40747-022-00760-3>
- [12] P. D. Chauhan and A. M. Shah, “Email Spoofing: In Today’s Era,” Research Square Preprint, 2022. [Online]. Available: <https://www.researchsquare.com>
- [13] M. D’Angelone, “Email spoofing defence techniques: a comprehensive review and development of a novel measurement tool,” MSc thesis, National College of Ireland, 2022. [Online]. Available: <https://norma.ncirl.ie>
- [14] K. Shen et al., “Weak Links in Authentication Chains: A Large-scale Analysis of Email Sender Spoofing Attacks,” in Proc. USENIX Security Symposium, 2021. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity21/presentation/shen-kaiwen>
- [15] J. Chen, V. Paxson, and J. Jiang, “Composition Kills: A Case Study of Email Sender Authentication,” in Proc. USENIX Security Symposium, 2020. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity20/presentation/chen-jianjun>