Journal of Science Research International (JSRI)
ISSN: 2456-6365 (Online)

OPEN ACCESS

A TREND ANALYSIS OF PHISHING ATTACKS USING PUBLIC INCIDENT REPORTS (2019–2025): CYBERSECURITY THREAT DETECTION ON SOCIAL MEDIA USING NLP AND URL INTELLIGENCE

Simran Shinde, Milind Santhosh Gawali, Sreelakshmi Roshan Kattikulam, Nikita Shankarrao Bhokare and Sarthak Uttam Dhumal

Department of Computer Science,

Pillai College of Arts, Commerce and Science (Empowered Autonomous), New Panvel

**Corresponding author E-mail: simranshinde@mes.ac.in*

Received: 14 January 2026	Revised: 22 February 2026	Accepted: 01 April 2026	Published: 16 April 2026

DOI: **https://doi.org/10.5281/zenodo.19612983**

Abstract:

Phishing attacks remain among the most persistent and disruptive cybersecurity threats in the digital era, continually evolving despite advances in security technologies and increasing user awareness. Unlike conventional cyberattacks that exploit technical vulnerabilities, phishing uniquely targets human behavior and cognitive biases, making it particularly difficult to counter through technical controls alone. This research presents a detailed trend analysis of phishing attacks based on publicly available incident reports from 2019 to 2025, drawing on sources including APWG reports, CERT-In advisories, and PhishTank. The findings reveal a consistent increase in phishing incidents throughout the study period, with a pronounced surge during the COVID-19 pandemic driven by widespread digital dependency and increased online activity. Email-based phishing remains the dominant attack vector, while smishing, fraudulent websites, and social media-based phishing show steady growth. Financial institutions, e-commerce platforms, and individual users remain the most targeted entities. These results emphasize the limitations of purely technical defenses and highlight the need for continuous user awareness training, multi-factor authentication, and proactive cybersecurity policies.

Keywords: *Phishing Attacks, Cybersecurity, Trend Analysis, Social Engineering, Smishing, Multi-Factor Authentication, Incident Reports,*

1. Introduction

Phishing, a form of cybercrime, is not a recent phenomenon; however, its scale and sophistication have increased significantly over the past decade. The term "phishing" originated in the mid-1990s and was initially associated with attackers attempting to steal login credentials for early online services [1,2]. Over time, phishing techniques

have evolved alongside technological advancements, shifting from simple, deceptive emails to complex, multi-stage attacks that closely mimic legitimate communication channels. Today, phishing is recognized as a critical cybersecurity concern due to its widespread impact on individuals, organizations, and governments.

Phishing attacks can be broadly categorized by delivery method and target audience. Email phishing remains the most common form, where attackers send bulk emails impersonating trusted organizations. Spear phishing is a more targeted variant in which attackers customize messages for specific individuals or organizations, often using personal or organizational information to increase credibility. Whaling attacks focus on high-level executives or senior officials, aiming to gain access to sensitive corporate or financial information. Smishing and vishing represent phishing attacks conducted through SMS messages and voice calls, respectively, exploiting the trust users place in mobile communication [9].

One of the key reasons phishing continues to succeed is its exploitation of human psychology. Unlike technical attacks that require advanced technical skills, phishing relies on manipulating emotions such as fear, urgency, curiosity, and trust. For example, messages warning users of account suspension or of suspicious activity often pressure recipients to act quickly without verifying authenticity. This psychological manipulation allows phishing attacks to bypass even sophisticated security infrastructures. Another factor contributing to the persistence of phishing attacks is the increasing complexity of digital ecosystems. Users today interact with multiple online platforms, manage numerous accounts, and receive a high volume of digital communication daily. This information overload reduces users' ability to evaluate every message they receive critically. Attackers take advantage of this environment by crafting messages that blend seamlessly into routine digital interactions.

The rise of remote work and cloud-based services has further amplified phishing risks. Employees working remotely often rely heavily on email, collaboration tools, and cloud platforms, creating new opportunities for attackers to impersonate internal communications or service providers [10]. Understanding phishing as a socio-technical problem rather than a purely technical one is essential for effective mitigation. While technological defenses play a crucial role, long-term solutions require addressing human behavior, organizational culture, and policy frameworks. This study contributes to this understanding by analyzing phishing trends over an extended period, offering insights into how phishing attacks have evolved and why they remain effective despite increasing awareness.

2. Literature review

A comparative analysis of industry reports and academic research reveals notable differences in focus and methodology. Industry reports such as those published by APWG and Verizon primarily emphasize statistical trends, real-world incident data, and sector-specific impacts [1,2,3,4,5]. These reports are valuable for understanding the scale of phishing attacks and identifying emerging patterns. However, they often lack in-depth theoretical analysis or methodological transparency, limiting their applicability in academic contexts.

In contrast, academic studies tend to focus on controlled experiments, detection algorithms, and behavioral models. Machine learning-based detection approaches have been widely proposed, utilizing techniques such as decision trees, neural networks, and deep learning models [9]. While these studies contribute to technical innovation, their reliance on limited datasets and laboratory environments raises concerns regarding scalability and real-world effectiveness. Attackers frequently modify phishing techniques, rendering static detection models less effective over time.

Behavioral research provides critical insights into why phishing attacks continue to succeed. Studies examining user behaviour highlight that awareness alone does not guarantee secure behaviour [10]. Even trained users may fall victim to phishing when under stress, facing time constraints, or experiencing cognitive overload. This finding challenges the assumption that education alone can eliminate phishing risks and underscores the need for layered defense strategies.

Research conducted during global crises, particularly the COVID-19 pandemic, demonstrated how attackers rapidly adapt to contextual changes. Pandemic-themed phishing campaigns exploited public anxiety and uncertainty, resulting in significantly higher engagement rates [6,7]. These studies emphasize the importance of situational awareness in phishing defense strategies, as attackers are highly responsive to social, economic, and political events.

Despite extensive research, existing literature exhibits several limitations. Many studies analyze phishing activity over short timeframes, often focusing on a single year or specific event. This approach limits the ability to identify long-term trends and evolutionary patterns. Additionally, there is limited integration of data from multiple public sources, resulting in fragmented insights.

The lack of longitudinal studies that combine industry data with academic analysis represents a significant research gap. This study addresses this gap by synthesizing data from multiple public sources over seven years, enabling a more comprehensive understanding of phishing evolution. By focusing on trends rather than isolated incidents, the research provides a broader perspective on the persistence and adaptability of phishing attacks.

3. Method

The use of secondary data analysis in this study offers several advantages. Publicly available phishing incident reports provide access to large-scale datasets collected over extended periods, allowing for trend identification and comparative analysis. Unlike primary data collection, which may be limited by sample size and ethical constraints, secondary data enables the examination of real-world phishing activity across diverse sectors and geographic regions.

Reliability and validity were key considerations in selecting data sources. Reports published by organizations such as APWG and CERT-In are widely recognized for their credibility and methodological rigor [1,2,3,4,5,6,7]. These organizations collect data from multiple stakeholders, including security vendors, financial institutions, and incident response teams, enhancing data reliability. Cross-referencing data from multiple sources further strengthens validity by reducing source-specific bias.

Ethical considerations were also addressed in this research. The study relies exclusively on anonymized, publicly available data, ensuring that no personal or sensitive information is exposed.

No direct interaction with phishing victims or attackers was involved, eliminating ethical concerns related to consent or privacy. The research adheres to academic integrity standards by properly citing all data sources.

The analytical approach focused on descriptive statistics rather than inferential modeling. This decision aligns with the study's objective of identifying patterns and trends rather than predicting future incidents. By categorizing incidents based on year, attack vector, and target sector, the analysis provides a structured overview of phishing evolution. While secondary data analysis offers valuable insights, it also presents limitations. Reporting inconsistencies, underreporting, and variations in classification methods across sources may affect accuracy. These limitations were mitigated through data cleaning and careful interpretation of results.

3.1 Research design

This study adopts a descriptive and analytical research design based on secondary data analysis. The primary objective is to identify long-term trends and patterns in phishing attacks rather than to establish causal relationships.

Secondary data analysis was chosen because phishing incident data is widely available through reputable public sources and because direct data collection involving real phishing victims would raise ethical and privacy concerns.

3.2 Data sources

Data for this research were collected from multiple authoritative and publicly accessible sources to ensure reliability and completeness. The primary data sources include:

- Anti-Phishing Working Group (APWG) Reports (2019–2025): Comprehensive global statistics on phishing activity, including attack volumes, delivery mechanisms, and targeted industries (1–5).
- CERT-In Cybersecurity Advisories: Issued by the Indian Computer Emergency Response Team, documenting phishing incidents, emerging threats, and recommended mitigation strategies (6,7).
- PhishTank Open Dataset: A community-driven platform maintaining verified phishing URLs and incident reports (8).
- Public Cybersecurity Summary Reports: Additional insights from publicly available cybersecurity reports published by trusted organizations (9,10).

Data for the year 2025 includes information available up to the first quarter, as complete annual data was not available at the time of analysis.

3.3 Data analysis procedure

The collected data was subjected to a systematic cleaning process to remove duplicate records, false positives, and irrelevant entries. Phishing incidents were categorized based on year of occurrence, attack vector (email, SMS, websites, and social media), and targeted sector (financial services, individuals, e-commerce, government, healthcare, and others). Descriptive statistical techniques such as frequency distribution, percentage analysis, and trend comparison were applied to identify patterns over time. The results were summarized using tables and graphical descriptions to enhance clarity and interpretability.

4. Results

A closer examination of year-wise data reveals distinct phases in phishing activity. In 2019, phishing attacks were already prevalent, primarily targeting financial services and individual users through email-based campaigns [1]. During this period, attackers relied heavily on generic phishing emails with limited personalization.

The onset of the COVID-19 pandemic in 2020 marked a significant escalation in phishing activity. Attackers rapidly shifted their focus to pandemic-related themes, exploiting public demand for information and services.

Reports from 2020 and 2021 indicate a sharp rise in phishing campaigns impersonating healthcare organizations, government agencies, and remote work platforms [2,3,6,7].

Figure 1 illustrates year-wise trends in phishing attack frequency. A sharp rise is observed during the COVID-19 pandemic period (2020–2021), with continued growth at a more stable rate in subsequent years.

In the post-pandemic phase (2022–2025), phishing activity continued to grow, albeit at a more stable rate. Attackers increasingly adopted sophisticated techniques, including brand impersonation, shortened URLs, and

encrypted phishing websites [4,5]. The use of HTTPS-enabled phishing sites increased, reducing users' ability to distinguish between legitimate and malicious websites.

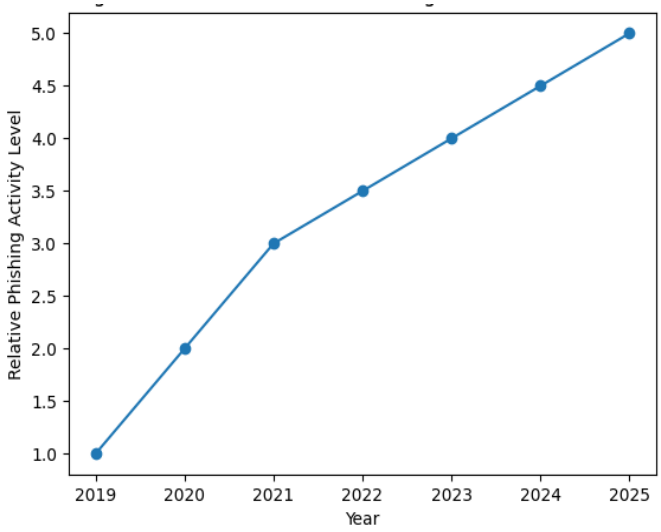

Figure 1: Year-wise trends in phishing attack frequency

The analysis of attack vectors indicates gradual diversification. While email remains dominant, mobile-based phishing gained traction due to increased smartphone usage. Social media platforms emerged as significant channels for phishing distribution, particularly targeting younger users.

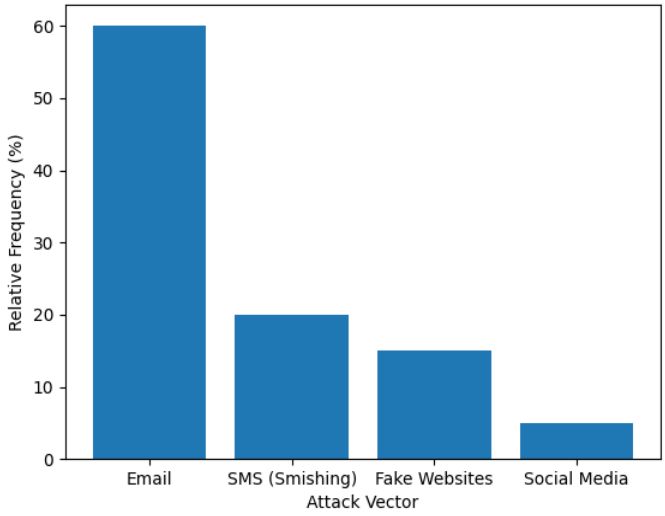

Figure 2: Sector-wise Distribution of Phishing Targets

Figure 2 illustrates the sector-wise distribution of phishing targets. Financial institutions remain the most targeted, while individuals and small businesses represent a broadening attack surface.

SSector-wise analysis shows sustained targeting of financial institutions, reflecting attackers' continued focus on monetary gain. However, the growing targeting of individuals and small businesses indicates a broadening attack surface and reduced barriers to entry for attackers.

5. Discussion

The expanded results reinforce the conclusion that phishing attacks are adaptive and resilient. The shift from generic phishing emails to highly contextualized campaigns demonstrates attackers' increasing understanding of user behaviour and environmental factors. This adaptability explains why phishing remains effective despite technological advancements.

Comparing these findings with prior studies reveals consistency with industry observations regarding the dominance of email phishing and the rise of mobile-based attacks [9,10]. However, the long-term perspective offered by this study provides deeper insight into how these trends develop gradually rather than abruptly.

From a practical standpoint, the findings highlight the importance of continuous cybersecurity awareness rather than one-time training programs. Users must be educated to recognize evolving phishing techniques across multiple platforms. Organizations should complement awareness initiatives with technical controls such as multi-factor authentication and real-time monitoring.

Overall, the discussion emphasizes that phishing prevention requires an integrated approach that combines technology, policy, and human-centered strategies.

To further contextualize the findings of this study, it is necessary to examine phishing attacks not only as isolated technical incidents but as part of a broader socio-technical phenomenon.

Phishing thrives at the intersection of technology, human behaviour, and organizational processes. While security technologies continue to advance, attackers strategically focus on exploiting predictable human responses, which remain relatively constant over time.

One important observation from the longitudinal data is the gradual professionalization of phishing campaigns. Early phishing attacks were often poorly written, generic, and easy to identify. Over time, attackers have adopted more sophisticated language, improved visual design, and legitimate branding elements to enhance credibility. Modern phishing emails frequently use corporate logos, professional formatting, and accurate organizational terminology, making them difficult for users to distinguish from legitimate communications.

Another notable trend is the increasing personalization of phishing messages. With the widespread availability of personal information on social media platforms and data breaches, attackers can tailor messages to specific individuals or organizations. Personalized phishing, particularly spear phishing, significantly increases success rates because victims are more likely to trust messages that reference familiar names, roles, or recent activities.

The role of automation in phishing operations has also expanded. Cybercriminals now use automated tools to generate large volumes of phishing emails, manage phishing websites, and analyse victim responses. Automation reduces the cost and effort required to conduct phishing campaigns, allowing attackers to target a broader audience while maintaining effectiveness.

From an organizational perspective, phishing represents a major operational and financial risk. Successful phishing attacks can result in direct financial losses, reputational damage, regulatory penalties, and long-term erosion of customer trust.

Mobile phishing, or smishing, deserves particular attention due to its rapid growth. Smartphones have become primary devices for communication, banking, and online transactions. Users often interact with mobile messages quickly and with less scrutiny than emails, increasing susceptibility to phishing.

Multi-factor authentication (MFA) has emerged as one of the most effective technical controls against phishing. Even when credentials are compromised, MFA can prevent unauthorized access. However, attackers have begun developing techniques to bypass MFA, such as real-time phishing proxies [9]. These dynamics further illustrate the adaptive nature of phishing and the need for layered defence strategies.

The long-term trend analysis conducted in this study demonstrates that phishing attacks do not evolve randomly but follow broader technological and societal shifts. Major events such as the COVID-19 pandemic act as catalysts that accelerate existing trends rather than creating entirely new ones. Understanding these patterns enables more proactive and adaptive defense strategies.

Overall, the expanded analysis confirms that phishing is not merely a technical problem but a systemic challenge requiring coordinated responses from individuals, organizations, and policymakers. Technical defences must be complemented by behavioural interventions, organizational culture changes, and supportive policy environments. Only through a holistic approach can the long-term impact of phishing be effectively mitigated.

Conclusion

This study provided a comprehensive longitudinal analysis of phishing attack trends from 2019 to 2025 based on publicly available incident reports. The findings confirm that phishing continues to be a persistent and evolving cybersecurity threat, primarily driven by the exploitation of human behaviour and the continuous adaptation of attack techniques. Despite significant advancements in security technologies, cybercriminals remain effective by leveraging psychological manipulation and emerging digital communication platforms.

The study highlights that technical solutions alone are insufficient to mitigate phishing risks. Sustained user awareness programs, the adoption of strong authentication mechanisms such as multi-factor authentication, and the implementation of proactive cybersecurity policies are essential for reducing susceptibility to phishing attacks. By examining long-term trends rather than isolated incidents, this research enhances understanding of the evolving nature of phishing and provides valuable insights for strengthening cybersecurity resilience among organizations and individual users.

References

1. Anti-Phishing Working Group. (2019-21). *Phishing activity trends report: Fourth quarter 2019*-21.
2. Anti-Phishing Working Group. (2023). *Phishing activity trends report: Fourth quarter 2023*.
3. Anti-Phishing Working Group. (2025). *Phishing activity trends report: First quarter 2025*.
4. Indian Computer Emergency Response Team. (2020). *Advisory on phishing attacks*. Ministry of Electronics and Information Technology, Government of India.
5. Indian Computer Emergency Response Team. (2021). *Cybersecurity advisory on COVID-19-related phishing campaigns*. Ministry of Electronics and Information Technology, Government of India.
6. PhishTank. (2025). *PhishTank phishing verification dataset*. Open Web Application Security Project.
7. Gupta, B. B., Arachchilage, N. A. G., & Psannis, K. E. (2018). Defending against phishing attacks: Taxonomy of methods, current issues, and future directions. *Telecommunication Systems, 67*(2), 247–266.
8. Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 373–382.