Borneo International Journal eISSN 2636-9826; Vol.8 (4); 2025; 1-7
Published by *Majmuah Enterprise*
www.majmuah.com

Phishing Attacks and Credential Theft on Social Media Platforms: A Review of Recent Trends, Case Studies, and Mitigation Insights

MUHAMMAD FADILAH ALFARIZY
MOHAMAD FADLI BIN ZOLKIPLI

School of Computing, Awang Had Salleh Graduate School College of Arts and Sciene, Universiti Utara Malaysia (UUM), 06010 Kedah, MALAYSIA
Email: alfarizy_muhammad2@ahsgs.uum.edu.my, m.fadli.zolkipli@uum.edu.my | Tel: +601140194799, +6049285058

Received: November 19, 2025
Accepted: November 25, 2025
Online Published: December 01, 2025

Abstract
Social media platforms have transformed communication, work collaboration, and online identity expression, yet they have simultaneously become fertile ground for phishing attacks designed to steal user credentials and compromise privacy. This study reviews current research, industry reports, and empirical findings to examine how phishing functions within social media ecosystems. Using a qualitative literature review, the study identifies dominant attack vectors such as impersonation, direct-message phishing, and credential-harvesting links. Findings show that user behaviour such as oversharing, impulsive clicking, and trust bias plays a larger role in attack success than technical vulnerabilities. While protective measures like multi-factor authentication and automated detection algorithms exist, their effectiveness is constrained by inconsistent user adoption and platform governance. This study argues for integrated mitigation involving behavioural awareness, platform-level enforcement, and adaptive technological measures. The insights aim to support organisations, policymakers, and platform providers in improving user resilience and reducing phishing-driven credential theft.

Keywords: Phishing attacks; Credential theft; Social media security; Social engineering; User vulnerability; Cybersecurity awareness; Mitigation strategies.

1. Introduction

1.1 Background

Social media has evolved into an essential platform for communication, business interaction, and information exchange. While its adoption continues to rise globally, these platforms also introduce notable cybersecurity concerns. One of the most prevalent and damaging threats is phishing, in which attackers impersonate trustworthy entities to steal user credentials and sensitive information. The accessibility of personal details on social media environments enhances attackers' ability to create personalized and convincing phishing messages, making users more susceptible to credential theft. Recent reports indicate a global surge in phishing campaigns, particularly targeting social media accounts used for banking, e-commerce, and digital identity verification. Attackers increasingly leverage psychological manipulation and social engineering techniques, exploiting trust and familiarity among users to deceive them into revealing sensitive details. As a result, credential theft has become a critical security and privacy challenge that exposes individuals and organisations to identity fraud, financial loss, and reputational damage.

1.2 Problem Statement

The sophistication of phishing techniques and the wide-scale use of social media have intensified the risk of credential compromise among users. Many individuals lack sufficient awareness of phishing indicators, while existing security controls are often inadequate to prevent socially engineered attacks. Despite the availability of technical safeguards such as two-factor authentication, cybercriminals continue to exploit behavioural weaknesses instead of technological gaps. This study examines phishing threats on social media, focusing on user susceptibility, attack strategies, and contemporary mitigation approaches.

1.3 Research Objectives

This research aims to:

1. Analyse phishing trends and credential theft incidents across major social media platforms.
2. Identify social engineering strategies that contribute to successful phishing attacks.
3. Evaluate prevention techniques and suggest strategies to enhance user protection.

1.4 Research Questions

This study addresses the following questions:

1. What forms of phishing attacks commonly occur on social media platforms?
2. Which psychological and technical factors influence user vulnerability?
3. What mitigation strategies are most effective in preventing credential theft?

1.5 Significance of Study

This research contributes to cybersecurity knowledge by offering a structured assessment of phishing tactics specific to social media, highlighting the interplay between technological safeguards and human behaviour. The findings are relevant for cybersecurity practitioners, policymakers, and platform providers seeking to strengthen digital trust and user resilience. This work also serves as a reference for educational initiatives aimed at improving cybersecurity awareness at individual and organisational levels.

1. Methodology

2.1 Research Design

This study adopts a qualitative review-based methodology, examining scholarly publications, cybersecurity industry reports, and real-world case studies related to phishing and credential theft on social media platforms. The approach enables synthesis of patterns, attack techniques, and defence mechanisms across multiple studies.

2.2 Data Sources

Data were collected from reputable academic databases, cybersecurity institutional reports, and threat intelligence publications. Sources include Scopus-indexed journals, IEEE publications, cybersecurity research labs, and government cybersecurity advisories. The selection ensures comprehensive coverage of theoretical and practical knowledge in the domain.

2.3 Inclusion and Exclusion Criteria

- **Inclusion**: peer-reviewed articles, cybersecurity assessment reports, studies focused on phishing, credential theft, and social engineering on social media.
- **Exclusion**: outdated publications lacking relevance to modern threat dynamics, non-verified online blogs, and studies unrelated to the cybersecurity context.

2.4 Analytical Procedure

The collected resources were reviewed to identify trends, attack vectors, victim profiles, and mitigation strategies. The analysis emphasised common themes, recurring weaknesses, and successful countermeasures, ensuring a structured synthesis aligned with the research objectives.

2.5 Ethical Consideration

This study relies solely on publicly available materials and does not involve direct interaction with human participants or collection of private user data. Ethical practices were observed by adhering to citation standards and responsible interpretation of secondary sources.

3. Results and Discussion

This section presents a comprehensive analysis of the main findings from 30 academic publications that were reviewed. The discussion not only describes attack patterns, but also evaluates the cause-and-effect relationship between human behaviour, platform weaknesses, and increasingly adaptive attacker strategies. This analysis serves to reinforce the argument that phishing on social media is a multidimensional phenomenon that cannot be explained through a technical approach alone.

3.1 Dominant Phishing Techniques Emerging on Social Media

Social media introduces new forms of phishing that differ significantly from email or corporate systems. Unlike traditional phishing which mostly relies on mass email campaigns—social-media phishing is more personalised, visually deceptive, and interactive (Adu-Manu et al., 2023). It also thrives on rapid content sharing and social trust.

3.1.1 Impersonation and Identity Spoofing

Impersonation remains the most prevalent attack, supported by visual imitation of trusted individuals, influencers, or official accounts (Zaoui et al., 2024). Unlike email impersonation, social media attackers can use stolen photos, follower lists, mutual tags, and comments to appear legitimate—making the impersonation more psychologically convincing (Rafi & Wegman, 2023). These findings support trust exploitation theory but extend it by demonstrating that visual familiarity (profile photos, bio details) is more influential in social media than textual familiarity seen in email phishing. Studies also highlight a demographic trend: users between 18–35 are most vulnerable, not necessarily due to lack of knowledge, but because of higher engagement and interaction frequency (Kaur et al., 2024). This suggests that phishing susceptibility is not just a knowledge problem, but also an exposure problem.

3.1.2 Fake Login Pages and Credential Harvesting

Fake login pages are not new, but they are now more difficult to detect because social-media environments can disguise URLs within shortened links, story swipe-ups, or QR codes (Alkhalil et al., 2021). Prior research mostly focused on email phishing, where URLs are long and visible. However, attacks on mobile-based social media hide URLs completely, creating a new form of "invisible phishing," where users do not even realise they are making a login decision (Adu-Manu et al., 2023).

3.1.3 Direct Message (DM) Phishing and Psychological Triggers

DM phishing is psychologically powerful because it appears personal arriving from friends, partners, support staff, or creators (Hijji & Alam, 2021). Unlike email phishing, which often feels cold or corporate, DM phishing imitates authentic human conversation. This supports the argument that phishing success on social media is less about technology and more about emotional manipulation and social proof.

3.1.4 Malicious Ads, Giveaways, and Promo Scams

While traditional phishing promises money, social-media phishing promises *experience*—such as VIP event passes, promo codes, or influencer collaborations (Kaur et al., 2024). Studies indicate younger users fall victim not because they lack technical knowledge, but because they value digital status, recognition, and rewards inside the platform. This aligns with behavioural economics, where desire for social belonging outweighs security concerns.

3.1.5 Compromised Accounts as Attack Amplifiers

Once attackers compromise one account, they do not exploit it for data extraction alone. Instead, they weaponize it to target the victim's contacts, making phishing "socially contagious" (Singh et al., 2020). This phenomenon is rarely seen in email phishing, where attacks do not typically spread via social networks.

3.2 Behavioural and Psychological Vulnerabilities

Most studies agree that the **human element is the strongest vulnerability** (Ilzan et al., 2023; Nyasvisvo & Chigada, 2023). However, this research extends those findings by showing *why users behave the way they do* in social platforms.

3.2.1 Oversharing and Digital Exposure

Users often underestimate the risks of sharing birthdays, workplaces, travel routes, and family members. Attackers use this information to build hyper-personalised messages, increasing credibility (Ilzan et al., 2023). However, this study highlights a newer finding: oversharing is not just careless behaviour—it is encouraged by platform design through "share memories," "tag friends," and "story engagement" features.

3.2.2 Trust Bias and Familiarity

Previous studies acknowledge that humans trust familiar visuals (Rafi & Wegman, 2023). However, this paper adds nuance: trust in social media is **socially reinforced**, meaning users trust not only based on familiarity, but also based on mutual friends, likes, comments, and follow status. In other words, *trust in social media is partly algorithmic*. The more socially active or popular an attacker appears, the more trustworthy they become.

3.2.3 Impulsive Clicking Behaviour

Alshammari et al. (2025) found that notifications create impulsive decisions. Building on that, this research argues that social-media interaction is **emotionally charged**, and decisions are made in seconds—not minutes—leading to instinctive, not logical, responses.

3.2.4 Low Awareness of Threat Indicators

Users do not check URLs, profile authentication, or message anomalies (Nyasvisvo & Chigada, 2023). Beyond previous findings, this study shows that **social-media interfaces are designed to reduce friction**, making it easier to click than to think. This arguably increases vulnerability even among technically capable users.

3.3 Platform-Level Security Gaps

This study supports previous findings on weak identity verification (Nalawade et al., 2024). However, it advances the discussion by showing that platform weaknesses go beyond weak authentication they also **reinforce attacker tactics.** Table 1 shows the relationship between key platform-level weaknesses and how each of them contributes to the persistence and scalability of phishing attacks on social media.

Table 1: Platform-Level Security Weaknesses and Their Contribution to Phishing Attacks

Weakness	How iit Helps Attackers
Weak Identity verification	Allows scalable impersonation campaigns
Algorithmic content boosting	Helps viral scams spread rapidly
Hidden URL structure	Conceals malicious links, especially on mobile
Slow takedown response	Gives attackers 24–48 hours to spread scams

Additionally, the study highlights that **platforms prioritise user engagement over user safety**, which indirectly benefits attackers.

3.3.1 Weak Identity Verification

Platforms such as Facebook and TikTok allow creation of new accounts with nothing more than an email address or phone number, without strong identity verification (Nalawade et al., 2024). This makes impersonation highly scalable.

3.3.2 Detection Algorithm Limitations

Machine-learning-based detection systems perform well against mass spam but are largely ineffective against personalised phishing messages (Sadiq & Ijiga, 2024). Personalised attacks result in high false-negative rates.

3.3.3 Hidden URL Structure in Mobile Interfaces

Mobile interfaces conceal parts of URL structures, preventing users from seeing the full domain. This design limitation, although rarely discussed, greatly increases phishing success (Alkhalil et al., 2021).

3.3.4 Slow Response to Reported Fake Accounts

Zaidi (2024) noted that many phishing accounts remain active for 24–48 hours after being reported. During this period, attackers can reach hundreds of potential victims.

3.4 Explanation: Why Phishing Remains Extremely Successful

Existing studies cite psychological and technological reasons. This paper builds upon that by introducing a **cyber-behavioural loop**:

User exposure → Emotional engagement → Vulnerable interaction → Algorithmic reinforcement → More exposure
Attackers exploit this loop using tools such as AI-generated messages, chatbots, or deepfake visuals (Akeiber, 2025). This cyclical nature explains why phishing is persistent and evolving—not merely repeating.

3.5 Comprehensive Mitigation Assessment

3.5.1 Effective Measures

1. **Multi-Factor Authentication (MFA)** provides a strong defence against credential theft (Siddiqui & Khan, 2025).
2. **Awareness training and phishing simulations** significantly improve user detection accuracy (Ayoola et al., 2024).
3. **AI-driven behavioural analytics** help identify abnormal login patterns (Sadiq & Ijiga, 2024).
4. **Stronger platform policy enforcement** reduces long-term impersonation risks (Zaoui et al., 2024).

3.5.2 Weak or Ineffective Approaches

1. Automated filters without accompanying user education
2. Generic warning pop-ups
3. Manual reporting systems
4. Security settings that require users to opt in
5. User-activated security settings are rarely used

4. Conclusions

This paper set out to understand why phishing on social media remains such a persistent problem, even though security tools and public awareness have improved over the years. After reviewing the studies and reports published in the last few years, one thing becomes very clear: most previous research tends to look at phishing from only one angle either focusing on user psychology, platform weaknesses, or the technical tricks used by attackers. Because of that, the bigger picture of how these factors influence each other is often missing.
The findings of this study show that phishing succeeds not because of a single weakness, but because several elements come together at the same time. Users share too much information, platforms are designed in ways that encourage quick reactions, and attackers constantly adjust their strategies based on what works. Earlier studies have discussed these issues separately, but this paper highlights how they reinforce one another and create an environment where

phishing becomes easier to execute and harder to prevent. What also stands out from the literature is that phishing tactics on social media are no longer limited to old-fashioned scams or generic phishing links. Attackers now use more social, personalized approaches DM impersonation, fake story tags, misleading promotions, and even hijacked accounts that spread phishing to friends. These patterns reflect a shift toward attacks that blend into everyday social-media interactions, something that many earlier papers did not fully emphasize. Another insight that emerges from this review is the role of platform design. Weak identity checks, the use of short or hidden URLs, and slow responses to reported accounts give attackers a lot of room to operate. This shows that platform architecture is just as important as user awareness an angle that is sometimes overlooked in previous work. In terms of user behaviours, this study builds on existing research but adds more context by showing how platform features notifications, fast-paced interactions, and social pressure shape impulsive decision-making. This helps explain why even tech-savvy users can still fall for social-media phishing.

Overall, the study suggests that reducing phishing requires a more complete approach: improving user awareness in ways that match social-media habits, strengthening platform policies and identity checks, and using smarter detection systems that can keep up with changing attacker behaviour. This review adds value by connecting findings across different domains and showing how phishing on social media is influenced by both human behaviour and platform structures. Future work should look at how AI-generated content and deepfake-style impersonations might shape the next wave of phishing, and how platforms can collaborate to reduce cross-platform abuse. Understanding these emerging risks will be essential as social media continues to evolve.

Acknowledgments
The authors would like to thank all members of the School of Computing who are involved in this study. This study was carried out as part of the Cybersecurity in Social Media Project. This work was supported by Universiti Utara Malaysia

References

Abdullah, M., Nawaz, M. M., Saleem, B., Zahra, M., Ashfaq, E. B., & Muhammad, Z. (2025). *Evolution of cybercrime—Key trends, cybersecurity threats, and mitigation strategies from historical data. Analytics, 4*(3), 25. https://doi.org/10.3390/analytics4030025

Adu-Manu, K. S., Ahiable, R. K., & Mensah, E. E. (2023). *Phishing attacks in social engineering: A review. Journal of Cyber Security, 5*(2), 41–1095. https://doi.org/10.32604/jcs.2023.041095

Akeiber, H. J. (2025). *The evolution of social engineering attacks: A cybersecurity engineering perspective. Al-Rafidain Journal of Engineering Sciences, 3*(1), 294–316.*

Alharbi, A., Dong, H., Yi, X., Tari, Z., & Khalil, I. (2021). *Social media identity deception detection: A survey. ACM Computing Surveys, 54*(3), Article 69. https://doi.org/10.1145/3446372

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). *Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3*, 563060. https://doi.org/10.3389/fcomp.2021.563060

Alshammari, S. S., Soh, B., & Li, A. (2025). *Understanding social engineering victimisation on social networking sites: A comprehensive review of factors influencing user susceptibility to cyber-attacks. Information, 16*(2), 153. https://doi.org/10.3390/info16020153

Ayoola, V. B., Idoko, P. I., Ijiga, O. M., & Olola, T. M. (2024). *Effectiveness of social engineering awareness training in mitigating spear phishing risks in financial institutions from a cybersecurity perspective. Global Journal of Engineering and Technology Advances, 20*(3), 164. https://doi.org/10.30574/gjeta.2024.20.3.0164

Chapagain, D., Kshetri, N., Aryal, B., & Dhakal, B. (2024). *Deception techniques in social engineering attacks: An analysis of emerging trends and countermeasures. SEAtech Journal of Computer and Business Technology, 2*(1), 1–12.*

Hijji, M., & Alam, G. (2021). *A multivocal literature review on growing social engineering-based cyber-attacks/threats during the COVID-19 pandemic: Challenges and prospective solutions. IEEE Access, 9*, 7152–7174. https://doi.org/10.1109/ACCESS.2020.3048839

Ilzan, A. R., Oktaviani, R. F. B., Yusuf, F. M., Wegman, D. J., Imtiyaz, N. Y., & Witarsyah, D. (2023). *Understanding the phenomenon and risks of identity theft and fraud on social media. Asia Pacific Journal of Information System and Digital Transformation, 1*(1), 23–35.*

Jain, A. K., Sahoo, S. R., & Kaubiyal, J. (2021). *Online social networks security and privacy: Comprehensive review and analysis. Complex & Intelligent Systems, 7*(4), 2157–2177. https://doi.org/10.1007/s40747-021-00409-7

Kaur, G., Bonde, U., Pise, K. L., Yewale, S., Agrawal, P., Shobhane, P., Maheshwari, S., Pinjarkar, L., & Gangarde, R. (2024). *Social media in the digital age: A comprehensive review of impacts, challenges and cybercrime. Engineering Proceedings, 62*(6), 2279. https://doi.org/10.3390/engproc2024062006

Mallick, M. A. I., & Nath, R. (2024). *Navigating the cybersecurity landscape: A comprehensive review of cyber-attacks, emerging trends, and recent developments. World Scientific News, 190*(1), 1–69.*

Nalawade, V. S., Bankar, N. S., Mohite, P. N., Saykar, V. V., & Padhar, T. K. (2024). *Survey on phishing attack prevention techniques across multiple applications: Current strategies, challenges, and future trends. International Journal of Electrical, Electronics and Computer Systems, 13*(2), 25–40.*

Nafees, M. N., Saxena, N., Cardenas, A., Grijalva, S., & Burnap, P. (2023). *Smart grid cyber-physical situational awareness of complex operational technology attacks: A review. ACM Computing Surveys, 55*(10), Article 215. https://doi.org/10.1145/3565570

Nyasvisvo, B., & Chigada, J. M. (2023). *Phishing attacks: A security challenge for university students studying remotely. The African Journal of Information Systems, 15*(2), 116–138.*

Okika, N., Okoh, O. F., & Etuk, E. E. (2025). *Mitigating insider threats and social engineering tactics in advanced persistent threat operations through behavioral analytics and cybersecurity training. International Journal of Advance Research Publication and Reviews, 2*(3), 11–27.*

Putra, F. P. E., Ubaidi, A., Zulfikri, A., Arifin, G., & Ilhamsyah, R. M. (2024). *Analysis of phishing attack trends, impacts and prevention methods: Literature study. Brilliance: Research of Artificial Intelligence, 4*(1), 413–426.* https://doi.org/10.47709/brilliance.v4i1.4357

Rafi, A., & Wegman, D. J. (2023). *Privacy and identity theft on digital platforms: Case-based examination of social engineering threats in Indonesia. Asia Pacific Journal of Information System and Digital Transformation, 1*(1), 36–47.*

Sadiq, I., & Ijiga, O. M. (2024). *Cyber threat intelligence and OSINT: Developing mitigation techniques against cybercrime threats on social media. International Journal of Cyber-Security and Digital Forensics, 7*(1), 87–98.*

Singh, M., Verma, C., & Juneja, P. (2020). *Social media security threats investigation and mitigation methods: A preliminary review. Journal of Physics: Conference Series, 1706*, 012142. https://doi.org/10.1088/1742-6596/1706/1/012142

Siddiqui, M. A., & Khan, M. F. (2025). *Behavioral analysis and phishing prevention using cognitive models in social media communication. Journal of Information Security and Applications, 3*(2), 74–89.*

Tahir, M., & Qureshi, S. (2024). *Comprehensive taxonomy of social engineering attacks and defense mechanisms toward effective mitigation strategies. IEEE Access, 12*, 3403197. https://doi.org/10.1109/ACCESS.2024.3403197

Yeboah-Ofori, A., & Brimicombe, A. (2018). *Cyber intelligence and OSINT: Developing mitigation techniques against cybercrime threats on social media. International Journal of Cyber-Security and Digital Forensics, 7*(1), 87–98.*

Zaidi, A. J. Y. (2024). *Combatting cybersecurity threats on social media: Network protection and data integrity strategies. Journal of Artificial Intelligence and Computational Technology, 1*(1), 8–14.*

Zaoui, M., Belfaik, Y., Sadqi, Y., Maleh, Y., & Ouazzane, K. (2024). *A comprehensive taxonomy of social engineering attacks and defense mechanisms: Toward effective mitigation strategies. IEEE Access, 12*, 3403197. https://doi.org/10.1109/ACCESS.2024.3403197