Borneo International Journal eISSN 2636-9826; Vol.8 (4); 2025; 42-46
Published by *Majmuah Enterprise*
www.majmuah.com

Security Breaches of Celebrity and Corporate Social Media Accounts: Risk Dynamics, Impact, and Preventive Frameworks

NORSYAZWANI BINTI MOHD PUAD and MOHAMAD FADLI BIN ZOLKIPLI,
School of Computing, College of Arts and Sciences, Universiti Utara Malaysia (UUM), 06010 Sintok, Kedah, MALAYSIA
Email: wani.mp13@gmail.com, m.fadli.zolkipli@uum.edu.my | Tel: +60134292055 | +60177247779

Received: November 19, 2025
Accepted: November 28, 2025
Online Published: December 01, 2025

Abstract

Social media platforms are now essential channels for organizations and public figures to connect, build brands, and create economic value. Yet, the immense visibility of celebrity and corporate accounts makes them irresistible targets for cybercriminals. Attackers seek quick financial gain, the ability to spread misinformation, and cause significant reputational damage. This paper investigates the common routes to compromise, including deceptive phishing schemes, password reuse, platform weaknesses, and psychological manipulation (social engineering). By reviewing prominent real-world cases and scholarly findings, we uncover the complex blend of psychological, technical, and governance issues driving these breaches. Finally, we propose a layered socio-technical mitigation framework that integrates technology, organizational policies, individual behavior, and platform accountability. The findings underscore a critical need to better protect these influential digital identities from ongoing exploitation.

Keywords: social media security, celebrity breaches, corporate cybersecurity, identity theft, social engineering

1. Introduction

1.1 Background of the Study

Platforms like Instagram, TikTok, Facebook, and X (formerly Twitter) have evolved into digital worlds where public figures and major organizations engage directly with massive global audiences. A verified checkmark signals trust, making these accounts incredibly powerful communication tools (Han, 2024). Unfortunately, this high level of trust and reach also makes these profiles attractive targets for cyber-attacks, which can spread rapidly and inflict deep, lasting reputational harm (Sears & Cunningham, 2024).

1.2 Problem Statement

Despite knowing better and having access to numerous security recommendations, highly visible celebrities and large corporations are still routinely targeted and successfully breached. The core weakness often lies in human error: things like using the same password everywhere, being slow to respond when a breach occurs, neglecting Multi-Factor Authentication (MFA), and poorly managing access given to internal staff or third-party teams (Zhang et al., 2022). These gaps consistently allow attackers to seize control.

1.3 Research Objectives

The goals of this research are straightforward:

- To analyse the most significant real-world account breaches involving influential celebrities and corporations.
- To synthesize academic knowledge surrounding social media cybersecurity risks.
- To develop and propose effective, holistic, socio-technical security solutions.

1.4 Significance of the Study

This study bridges the gap between cybersecurity theory and the practical, high-stakes risks faced by influential accounts. By offering a socio-technical perspective, this research provides valuable insights for cybersecurity teams, media management agencies, and platform developers seeking to secure powerful digital identities.

2.0 Literature Review

2.1 Social Media–Cybersecurity Nexus

Threats on social media are successful largely because they exploit human trust and our reliance on instant engagement (Koohang et al., 2021). Research by Herath et al. (2022) indicates that users often underestimate the danger posed by seemingly simple threats like phishing, impersonation, and sharing malicious links.

2.2 Breach Mechanisms and Attack Surface

Attackers commonly exploit several weak points:

- Easy-to-guess or reused passwords
- Social engineering (tricking the people managing the account)
- SIM-swaps (hijacking a phone number to bypass MFA)
- Credential stuffing (using credentials stolen from other data breaches) (Cremer et al., 2022; Zhang et al., 2022)

The notorious 2020 Twitter attack was a stark reminder of how exploiting internal administrative tools can lead to a massive, platform-wide compromise.

2.3 Human Factors and Social Engineering

Social engineering relies on manipulating our inherent trust and sense of urgency. Azam and Yusoff (2020) highlight how psychological cues—such as arousing curiosity or feigning authority—can significantly increase the chance that a malicious link will be clicked. The constant pressure on influencers to be immediately responsive only makes them more vulnerable to impulsive mistakes (Peterson & Roth, 2021).

2.4 Reputational Dynamics and Media Effects

When a breach hits, intense media coverage drastically magnifies the brand damage and erodes consumer confidence (Syed, 2019). For organizations associated with a celebrity, the negative fallout can intensify, sometimes leading to notable fluctuations in stock value (Han, 2024).

2.5 Theoretical Lenses

The socio-technical security model offers a useful lens, stressing that effective protection must harmoniously balance people, technology, and organizational policies (Cremer et al., 2022). Furthermore, social cybersecurity frameworks specifically address threats related to the spread of misinformation and identity manipulation (Mulahuwaish, 2025).

3.0 Methodology

To ensure the analysis and proposed framework are robust, this paper utilizes a qualitative, systematic literature review (SLR) and case study approach.

3.1 Data Collection and Selection

The research gathered data from three primary sources:

i. **Academic Databases**: Scopus, Web of Science, and IEEE Xplore were searched using key terms ("social media breach," "celebrity account compromise," "corporate cybersecurity," "social engineering") to synthesize scholarly findings on risk dynamics.
ii. **Public Incident Reports**: Official post-mortem reports, regulatory filings, and reputable cybersecurity news sources (e.g., Krebs on Security, TechCrunch) were used to document the technical mechanisms and impact of high-profile breaches (e.g., the 2020 Twitter incident).
iii. **Security Framework Documentation**: Existing, publicly available security guidelines (e.g., NIST, ISO 27001) were reviewed to inform the structure and components of the proposed preventive framework.

3.2 Analytical Approach

The analysis employed a thematic content analysis technique, specifically focusing on identifying recurring failure points across all documented breaches. These failure points were categorized into the three primary components of the socio-technical model: Technology, People (Human Factors), and Policy/Governance. This triangulation allowed for the formulation of targeted and evidence-based preventive recommendations.

4.0 Discussion and Analysis

4.1 Incident Patterns and Case Illustrations

High-profile compromises often follow a pattern: they involve attackers pretending to be the platform itself or exploiting the credentials of authorized team members. Once inside, attackers use administrative access gained via phishing or insider tools to launch fraudulent financial schemes (Sharevski & Kessell, 2023).

Clear examples include:

- Verified accounts on X used to promote elaborate cryptocurrency scams.
- Stolen celebrity Instagram accounts repurposed for bogus investment fraud.
- Corporate customer service accounts posting unauthorized, malicious policy updates.

4.2 Key Risk Drivers in High-Profile Contexts

The primary factors driving these security failures are organizational and human:

Table 1: Key Risk Drivers in High-Profile Contexts

Risk Driver	Description (Why it happens)
Social Engineering	Criminals exploit fame, trust, and the need for urgency.
Weak Authentication	Failure to use MFA; recycling old, compromised passwords.
Insider/Third-Party Access	Poor security standards maintained by external agencies or internal employees.
Platform Design Gaps	Technical flaws in the platform allowing attackers to gain elevated privileges.
(Yoon et al., 2022; Farazmanesh et al., 2022)	

4.3 The Wide-Ranging Impact

The consequences extend far beyond technical downtime:

- Individuals suffer significant anxiety and threats to self-identity (Sears & Cunningham, 2024).
- Businesses face a sharp decline in reputation and stakeholder trust (Syed, 2019).
- False statements can trigger dangerous misinformation cascades with real-world public consequences (Sharevski & Kessell, 2023).

4.4 Socio-Technical Preventive Framework

This layered framework creates robust defenses by addressing people, processes, and technology, working in harmony.

4.4.1 Identity & Access Security (Technology Focus)

i. MFA Mandatory for all Users: This is the baseline defence. Every person with administrative or publishing access must use Multi-Factor Authentication.
ii. Security Key Authentication: For the most sensitive, high-profile accounts, mandatory use of a physical security key (FIDO2/U2F) provides the strongest defence against sophisticated phishing and SIM-swapping.
iii. Password Manager Enforcement: Organizations must enforce the use of enterprise-grade password managers to generate and store unique, complex credentials for every platform.
iv. Least-Privilege Role Access: User permissions should be strictly defined and limited to only what is necessary for their job, minimizing the potential damage of a single compromised account.

4.4.2 Workflow Hygiene (Policy & Process Focus)

i. Split Publishing Rights (Separation of Duties): Critical actions should require authorization from two different roles to prevent single-person error or malice.
ii. Pre-Launch Verification for Important Posts: A formal, documented verification process must precede the publishing of sensitive, market-moving, or promotional content.
iii. Audit Logs for all Posting Activity: Implement detailed logging on all administrative actions and posts to enable rapid forensic tracing during an incident.

4.4.3 Cyber Awareness & Culture (People Focus)

i. Mandatory, Targeted Training: Training must simulate platform impersonation attempts, SIM-swap social engineering calls, and spear-phishing designed to exploit urgency and fame.
ii. Fake Login Simulation Drills: Conduct internal simulated phishing attacks that mimic platform login pages to test and reward employee vigilance.
iii. Zero-Trust Engagement Culture: All internal and external requests must be treated as potentially malicious and verified via an out-of-band channel.

4.4.4 Platform Governance (Accountability Focus)

i. Secure Defaults (Auto-MFA): Platforms should default high-risk, verified accounts to the most secure settings automatically.
ii. Robust Insider Access Monitoring: Platforms must track and audit internal employee access to high-profile accounts and tools.
iii. Anomaly Detection at Scale: Machine learning must be used to immediately flag and suspend activity that deviates from a celebrity's typical behavior, such as sudden login attempts from unusual locations or the posting of suspicious financial content.

4.5 Incident Response and Communication

Upon compromise, a rapid and transparent response is essential (Ruohonen et al., 2024):

- Immediate session invalidation to boot the attacker.
- A transparent public warning to followers.
- Forensic investigation.
- Rapid restoration via platform support.

5.0 Ethical Considerations

This research primarily utilizes publicly available, aggregated data on breaches (e.g., media reports, public regulatory filings). However, studying breaches involving public figures and corporations requires careful ethical consideration:

- **Privacy of Victims**: While the accounts discussed are public-facing, care was taken to focus only on the technical and procedural failure points, not on the personal or non-public data of the individuals or employees involved in the incidents.
- **Non-Disclosure of Vulnerabilities**: The paper focuses on known, high-level vulnerabilities (e.g., phishing, lack of MFA) and provides generalized mitigation strategies. It strictly avoids disclosing or exploiting any specific, unpatched platform vulnerabilities that could aid further malicious activity.
- **Reputational Harm**: The analysis of past incidents is framed as academic case studies for learning, minimizing any further unnecessary reputational harm to the involved parties.

6.0 Conclusion and Future Work

6.1 Conclusion

Highly influential social media accounts are vital communication assets, and attacks against them result in massive digital and real-world harm. This research confirms that security vulnerabilities are not just technical problems; they stem from a failure to address the combined weaknesses in human behavior, organizational structure, and platform design. Adopting a cohesive socio-technical strategy is essential to improve overall security posture and ensure responsible digital influence management.

6.2 Limitations and Future Work

Limitations:

- **Data Transparency**: The analysis relies primarily on publicly reported data. The true technical depth of many corporate breaches is often obscured by non-disclosure agreements or legal privilege, potentially leading to an incomplete understanding of certain technical failures.

- **Focus on Detection:** This study focuses heavily on prevention. Future work should more thoroughly investigate the economic efficacy and speed of various real-time anomaly detection systems deployed by the platforms themselves.

Next Research Phase: Future Direction

Building on the findings of the Socio-Technical Preventive Framework, the most impactful next phase of research is to empirically test a key vulnerability: third-party access. Next Research Topic: The Efficacy of Zero-Trust Models in Managing Third-Party Social Media Agencies: A Comparative Study of Celebrity and Corporate Risks

Research Rationale: The "Insider/Third-Party Access" risk driver (Section 4.2) is a persistent and common attack vector. High-profile accounts are rarely managed by a single person, relying instead on numerous external PR firms, marketing agencies, and media teams.

Research Objectives:

i. **Evaluate Current Access Protocols**: Analyze how current celebrity and corporate teams grant, monitor, and revoke access for external agencies.

ii. **Compare Zero-Trust Implementation**: Empirically compare the breach resilience of organizations using a strict Zero-Trust (ZT) model (temporary, restricted, monitored access) versus those using traditional, persistent access credentials.

iii. **Develop Best Practice Metrics**: Establish quantifiable metrics for monitoring third-party agency security performance, focusing on speed of breach detection and credential rotation frequency.

Acknowledgments

The author is grateful for the continuous guidance and expert support received from educators and cybersecurity professionals during the execution of this research.

References

Alharthi, A., & Al-Ghamdi, S. (2023). Digital identity theft and user behavior on social networks. Information & Computer Security, 31(4), 589–605.

Azam, F., & Yusoff, R. (2020). Human factors influencing social engineering vulnerability in online platforms. Asian Journal of Information Technology, 19(4), 345–354.

Cremer, F., et al. (2022). Cyber risk and cybersecurity: A systematic review of data availability. Information Systems Frontiers, 24(6), 1387–1402.

Egele, M., Stringhini, G., Kruegel, C., & Vigna, G. (2015). Detecting compromised accounts on social networks. NDSS Proceedings.

Elmas, T., Overdorf, R., & Aberer, K. (2020). Misleading repurposing on Twitter. ICWSM Proceedings.

Farazmanesh, F., Foroutan, F., & Bidgoly, A. J. (2022). Authorship verification for compromised accounts. IEEE ICMLA Proceedings.

Han, J. H. (2024). Reputation and breach coverage in celebrity firms. Strategic Management Journal, 45(2), 311–330.

Hammouchi, H., et al. (2024). STRisk socio-technical breach model. ICIS Proceedings.

Herath, T., Khanna, P., & Ahmed, M. (2022). Cybersecurity practices in social media. Journal of Cybersecurity and Privacy, 2(1), 1–18.

Koohang, A., Floyd, K., & Paliszkiewicz, J. (2021). Privacy, trust, and awareness in social media. Issues in Information Systems, 22(2), 133–145.

Mouncey, E. (2025). Instagram phishing awareness. Int. J. Cyber Behavior, 15(1), 12–27.

Mulahuwaish, A. (2025). Social cybersecurity threat landscape. Computers & Security, 139, 103155.

Peterson, J., & Roth, E. (2021). Influencer account hijacking analysis. Journal of Marketing Communications, 27(7), 654–672.

Rehman, S., & Farooq, U. (2023). User defense behaviors in social media. Telematics and Informatics, 83, 101966.

Ruohonen, J., et al. (2024). Breach communication and media narratives. Digital Threats, 5(1).

Sears, C. R., & Cunningham, D. R. (2024). Breach stress outcomes. Journal of Cybersecurity and Privacy, 4(3), 594–614.

Sharevski, F., & Kessell, B. (2023). Hacktivist information operations. SOUPS Proceedings.

Syed, R. (2019). Social media misuse and reputation risk. Computers in Human Behavior, 93, 366–377.

Yoon, S., Choi, M., & Park, J. (2022). Crisis communication after cyber incidents. Public Relations Review, 48(5), 102234.

Zhang, X., et al. (2022). Corporate data breach patterns. International Journal of Information and Computer Security, 19(3/4), 402–442.