

Phishing Email Awareness Systems

Posted on January 27, 2020

View this essay online:

<https://academic-master.com/phishing-email-awareness-systems/>

Introduction

Phishing email awareness systems are most effective when they are designed as socio-technical defenses rather than as tests of whether individual employees can spot suspicious spelling or formatting. Modern attacks may use compromised accounts, legitimate cloud services, QR codes, convincing payment requests, telephone follow-up, stolen session tokens, or artificial intelligence to create messages that appear credible. The 2026 Verizon Data Breach Investigations Report continues to identify social engineering, phishing, stolen credentials, and exploitation of technical weaknesses as recurring contributors to breaches, showing why organizations need layered controls rather than one universal classifier (Verizon, 2026). A resilient program combines sender authentication, secure email gateways, phishing-resistant identity controls, employee reporting, realistic training, analyst review, and rapid incident response. It should also measure message difficulty and business impact instead of judging awareness only through click rate. The research problem is therefore not simply how to identify a deceptive message, but how to reduce the likelihood that one successful deception becomes an organizational breach while maintaining usable communication and a culture in which employees report mistakes quickly.

Threat Model and Research Questions

Phishing is a broad category of deceptive communication intended to induce an unsafe action such as disclosing credentials, authorizing a payment, installing malware, opening a malicious document, or transferring sensitive information. Spear phishing targets a selected person or group using contextual details; business email compromise often impersonates executives, vendors, or employees to redirect money or data; credential phishing uses fraudulent login pages; and malware campaigns rely on attachments or links. These categories overlap, which means one defensive control cannot address every scenario. Attachment sandboxing may help detect malicious files but do little against a legitimate supplier account that has been compromised and is now requesting a bank-account change. The research should therefore ask which technical and contextual features distinguish high-risk messages in the target organization, how automated scoring should interact with human judgment, whether employees can report suspicious messages quickly, how realistic simulations affect behavior, and which response controls limit harm after interaction. These questions produce a testable system rather than promising that a single “specially crafted algorithm” will solve phishing universally.

An attacker's lifecycle also clarifies where defensive layers belong. Reconnaissance may reveal executive names, suppliers, current projects, organizational language, and technology through public websites, social platforms, prior breaches, or compromised accounts. The attacker then crafts a pretext and delivers a message through email or a related channel. The recipient interprets cues and chooses whether to click, scan, reply, pay, authenticate, or report. If the deception succeeds, the attacker may steal credentials, establish persistence, alter mailbox rules, exploit internal trust, or pursue financial or data theft. Defenders can intervene before, during, and after each step. Domain authentication can reduce direct spoofing; gateways can analyze reputation, URLs, and attachments; passkeys or hardware-backed authentication can reduce the value of stolen passwords; transaction verification can block fraudulent transfers; and incident-response procedures can revoke sessions and remove related messages. The system should be judged by how well these layers work together rather than by the accuracy of one detection model.

Layered Technical Architecture

A practical anti-phishing architecture begins with email authentication using SPF, DKIM, and DMARC so receiving systems can evaluate whether messages using an organization's domain are authorized and whether protected content has been altered. These controls make direct domain spoofing more difficult, although they do not stop look-alike domains or malicious messages sent from a genuinely compromised account. A secure email gateway can add reputation checks, attachment inspection, URL analysis, language signals, threat intelligence, and time-of-click protection. Automated triage may then group similar messages, calculate risk, and prioritize analyst review. Machine-learning models can support classification, but they require version control, representative training data, drift monitoring, privacy safeguards, and comparison with baseline filters. Performance reporting should include precision, recall, false-positive rate, false-negative rate, and time to decision rather than only overall accuracy. High-impact actions such as deleting an executive message or blocking a critical vendor should not rely blindly on an opaque score without defined review and rollback procedures.

Identity controls reduce the damage of messages that evade filtering. Phishing-resistant authentication methods such as properly implemented passkeys or security keys provide stronger protection than one-time codes that users can be persuaded to enter into an attacker-controlled site. Multifactor authentication should also be supported by conditional access, device trust, session monitoring, least privilege, and rapid session revocation because attackers may steal authenticated sessions or exploit repeated prompts. Employees need an easy reporting interface inside the mail client or mobile application. The report should preserve headers and relevant content, provide confirmation to the employee, and route the message into automated and human triage. Reporting must remain available even when the user is uncertain; a control labeled only for confirmed phishing may discourage useful submissions. The same interface should support urgent disclosure after a click or credential entry, because rapid containment is more valuable than hiding an error out of fear of punishment.

Human-Centered Awareness and the NIST Phish Scale

Awareness training should prepare employees for realistic decisions rather than teach a checklist of obvious grammatical errors. Useful warning signs include unexpected urgency, unusual authority, requests to bypass procedure, changed payment details, credential requests, mismatched domains, unfamiliar attachments, and communication that does not fit the sender's normal behavior. Yet sophisticated attacks may contain correct grammar and genuine branding, so employees should learn to verify high-risk requests through a trusted channel obtained independently of the message. Training should also be role-specific. Finance staff need scenarios involving invoices and bank changes, administrators may receive document-sharing lures, and executives may face legal or acquisition pretexts. The National Institute of Standards and Technology's Phish Scale helps program managers rate the human detection difficulty of simulated messages by considering observable cues and the message's alignment with the recipient's work context. A low click rate on an obvious simulation is not equivalent to strong performance on a highly convincing message, so difficulty-adjusted interpretation is essential (Dawkins & Jacobs, 2023).

Simulations should be used as practice and process testing rather than public humiliation. Campaigns that exploit bereavement, medical fear, layoffs, or other personal vulnerabilities merely to increase clicks can damage trust and discourage reporting. Immediate coaching after interaction is more useful than punitive ranking, and exercises should test the whole system: whether employees report the message, whether analysts identify similar messages, whether compromised credentials can be revoked, and whether business procedures prevent fraudulent payments. Metrics should therefore extend beyond click rate to reporting rate, time to first report, credential submission, repeat susceptibility, message difficulty, time to containment, false positives, and the consequence of any successful interaction. Qualitative interviews can explain why a message appeared credible and whether warnings were understood. An awareness program should demonstrate measurable improvement in organizational resilience, not simply generate a dashboard showing that employees clicked fewer easy simulations.

Incident Response, Business Controls, and Research Method

When an employee reports that they clicked a link, entered credentials, opened a suspicious file, or approved a transaction, the response should be immediate and nonjudgmental. Depending on the event, security staff may reset credentials, revoke sessions, inspect mailbox rules, isolate a device, search for related messages, preserve evidence, block malicious infrastructure, monitor financial activity, and notify affected parties. Prompt disclosure is a protective behavior and should be reinforced. Business-process controls remain essential because technical email security cannot

prevent every fraudulent request. Bank-account changes can require verification through a known telephone number, large payments can require separation of duties, and requests for sensitive records can follow an approval workflow that cannot be bypassed through an urgent executive email. Attackers frequently exploit organizational hierarchy and reluctance to challenge authority, so a secure culture explicitly authorizes employees to verify unusual requests regardless of the apparent sender's seniority.

A mixed-method research design can evaluate the proposed system more credibly than a literature review or one simulation campaign. The study can begin with anonymized incident data, architecture review, employee interviews, and analysis of existing simulations. A prototype triage model can then be tested on labeled messages and compared with current filters using technical performance metrics. A phased pilot can measure reporting time, analyst workload, false positives, containment speed, and employee experience, while interviews after exercises can explain behavioral decisions that quantitative logs cannot. Privacy and ethics require clear notice about the purpose of monitoring, limited retention, restricted access, and safeguards preventing simulation data from becoming general productivity surveillance. Employees with disabilities or different language backgrounds should be able to participate fairly. The objective is to measure whether layered controls reduce the likelihood and consequence of phishing while preserving trust, communication, and organizational usability.

Conclusion

Phishing email awareness systems should be designed as integrated defenses rather than as attempts to identify one weak employee or build one perfect algorithm. Modern phishing blends social engineering with compromised accounts, cloud infrastructure, business processes, identity systems, and increasingly convincing language, which makes layered protection necessary. Domain authentication, secure gateways, URL and attachment inspection, phishing-resistant authentication, simple reporting, analyst review, transaction verification, and rapid response each reduce a different part of the risk. Human training remains essential when it is realistic, ethical, role-specific, and interpreted in light of message difficulty through tools such as the NIST Phish Scale. Research should compare technical detection, employee reporting, business-process controls, incident consequences, and privacy outcomes rather than relying on click rate alone. No system can prevent every deceptive message, but a well-designed architecture can make successful compromise less likely, detect it faster, and prevent one mistake from becoming an uncontrolled breach.

References

Dawkins, S., & Jacobs, J. (2023). *NIST Phish Scale user guide* (NIST Technical Note 2276). National Institute of Standards and Technology.

National Institute of Standards and Technology. (2022). *Federal cybersecurity awareness programs: A*

mixed methods research study (NISTIR 8420).

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2015). The design of phishing studies. *Human-Centric Computing and Information Sciences*, 5, 1-17.

Verizon. (2026). *Data breach investigations report*.