ITP-111

| | |
|---|---|
| Mendoza, Mayen Sofia T. | Refrea, Mar John S. |
| Miranda, Kian Andrei L. | Sarino, Nhoel Ivan A. |
| Soriano France Hedrich O. | 4-ITB |

**Human-based Attacks: Phishing, Pretexting, Baiting, Impersonation, Insider Threats**

In the modern digital age, cybersecurity is not solely about protecting systems and networks through technological defenses—it is also about safeguarding the human element. As organizations adopt advanced encryption, firewalls, and intrusion detection systems, attackers have shifted focus toward exploiting human psychology rather than software vulnerabilities. Human-based attacks, also known as social engineering, manipulate trust, authority, curiosity, or fear to deceive individuals into revealing confidential information, granting access, or performing actions that compromise security. These attacks pose a serious threat because they bypass even the most sophisticated technical controls, making people the weakest link in the security chain.

The purpose of this research is to explore and analyze five major human-based attack vectors—phishing, pretexting, baiting, impersonation, and insider threats. Each of these attacks exploits human behavior in unique ways, yet all share a common goal: unauthorized access to information, systems, or financial assets. Understanding how these attacks operate, who they target, and what tools and techniques are used is vital to developing comprehensive defense strategies. By studying real-world incidents, such as corporate breaches and financial scams, this paper aims to highlight the tangible risks these attacks present to individuals, businesses, and governments.

In today's interconnected digital ecosystem, the significance of addressing human-based attacks cannot be overstated. Reports from organizations such as NIST, Verizon, and the FBI's Internet Crime Complaint Center (IC3) consistently show that the majority of successful cyber incidents involve some form of human manipulation. The global impact of these threats includes billions of dollars in financial losses, reputational damage, operational disruptions, and legal consequences. This research emphasizes that effective cybersecurity requires not just technical solutions but also education, policy, and cultural change to build a resilient human firewall against evolving social engineering threats.

## Methodology

This analysis employs a multi-source, evidence-based methodology to provide a comprehensive understanding of phishing, social engineering, and insider threat patterns. The approach integrates both technical frameworks and empirical breach data to ensure accuracy and real-world relevance. Primary technical guidance was drawn from the National Institute of Standards and Technology (NIST), specifically leveraging the *NIST Phish Scale* and related phishing resilience guidelines available through the NIST Computer Security Resource Center. These frameworks offer structured methods for evaluating phishing susceptibility and assessing the effectiveness of awareness and defense programs across organizations.

To ground the analysis in quantitative evidence, large-scale breach datasets such as the Verizon Data Breach Investigations Report (DBIR) 2024/2025 were incorporated. The DBIR provides statistical insights into threat vectors, attacker motivations, and the frequency of phishing and social engineering incidents across industries. This empirical data serves as a foundation for identifying emerging attack trends, measuring human factor vulnerabilities, and evaluating mitigation strategies.

Complementary sources from industry reports and investigative journalism further enhance the contextual depth of this study. Publications from WIRED, Krebs on Security, and SEC filings—particularly regarding high-profile incidents like Ubiquiti's breach—are used to illustrate real-world case studies of organizational compromise and response. Additionally, alerts and advisories from the Internet Crime Complaint Center (IC3) offer practical insights into current threat patterns and law enforcement observations.

For technical and procedural depth, guidance from the OWASP Foundation and established practitioner playbooks on social engineering and open-source intelligence (OSINT) techniques were analyzed. These materials inform the operational aspects of phishing prevention, incident response, and red team simulation exercises. Finally, academic literature from platforms like ResearchGate provided peer-reviewed insights into the psychological, organizational, and behavioral components of insider threats and social manipulation.

Throughout this study, only the most critical and "load-bearing" data points—such as verified statistics, documented financial losses, and authoritative framework recommendations—are emphasized and cited directly in the text. This approach ensures that every reference contributes meaningfully to understanding the evolving landscape of phishing and social engineering threats.

## 1. Phishing

### 1.1 Description & step-by-step

Phishing is the bulk mailing or targeted sending of deceptive messages (email, SMS, messaging apps) designed to trick recipients into revealing credentials, clicking malicious links, or opening attachments that install malware. Attackers adapt messages to look like trusted senders (banks, colleagues, services). The NIST Phish Scale classifies phish by difficulty to detect and recommends threat-aware training.

**Typical step flow:**

- Recon/OSINT: Gather targets' contact info and context (roles, vendors).
- Craft message: Create a believable pretext (invoices, password reset, HR notice).
- Delivery: Send email/SMS with malicious link or attachment (or credential harvesting page).
- Exploitation: Victim clicks link, enters credentials, or runs attachment.
- Post-compromise: Attacker uses credentials, moves laterally, exfiltrates data or initiates BEC/wire

  fraud.

Targets: Individuals, SMBs, enterprises, governments — virtually anyone with email or messaging access. DBIR shows social engineering and human error remain major contributors to breaches.

### 1.2 Tools & techniques

- Credential-harvesting pages (fake login portals), malicious Office macros, HTML smuggling.

- URL shorteners, domain spoofing, look-alike domains (typosquatting).

- Phishing kits sold on underground markets; automated mass mailing tools.

- Use of compromised legitimate services (Google Drive/OneDrive) to host payloads.

### 1.3 Case study: Large-scale phishing and BEC examples

Verizon DBIR / industry trend (summary): The DBIR repeatedly documents social engineering as a primary human element in many incidents — e.g., a significant fraction of breaches involves a non-malicious human element (falling for phishing, misconfiguration).

Specific example: Ubiquiti Networks (2015) — BEC following social engineering

When/where: Discovered June 5, 2015; U.S.-based company with Hong Kong subsidiary.

What happened: Attackers used targeted email impersonation to trick finance staff; transfers of ~$46.7M made to attacker accounts. SEC filing and reporting provide details.

Impact: ~$46.7M financial loss, investigations, tightened controls. This demonstrates how phishing/BEC techniques can cause catastrophic direct financial loss without any software exploit.

**1.4 Impact analysis**

**-** Financial: direct wire losses (millions), incident response, remediation costs; DBIR quantifies an increasing role of social engineering in costly attacks.

**-** Reputational: customer trust erosion once credentials or funds are lost.

- Operational: account lockouts, downtime, forensic investigations.

- Legal/Compliance: fines if regulated data exposed, contractual liability.

**1.5 Mitigations (per threat) — phishing**

**Technical:** Email filtering (DMARC, DKIM, SPF), anti-phishing gateways, URL rewriting, sandboxing attachments.

**Administrative:** Incident response playbooks for suspected phishing, least privilege for finance approvals (dual-approval for wire transfers).

**Human-centric:** Continuous security awareness training, phishing simulations informed by NIST Phish Scale, role-based training for high-risk roles (finance, HR).

**Framework mapping:** Controls align with NIST CSF (Protect / Detect / Respond), CIS Controls (email and web protections, identity management), and ISO 27001 ISMS controls.

**2. Pretexting**

**2.1 Description & step-by-step**

Pretexting involves creating a fabricated story or scenario—known as a *pretext*—to manipulate individuals into revealing confidential information or granting unauthorized access. The attacker carefully crafts a believable identity, such as posing as an IT support technician, a law enforcement officer, a bank representative, or even a trusted vendor or coworker, to establish

credibility and lower the target's defenses. Unlike typical phishing, which often relies on mass emails or messages, pretexting tends to be more personalized and interactive, requiring direct communication with the victim.

**Typical step flow:**

- Recon for context and authority.
- Build a believable story (e.g., "we need to reset your VPN now").
- Use social pressure/urgency and some legitimate detail to engender trust.
- Obtain information, credentials, or access.

Targets: High-value staff (IT, payroll, executives), customer support teams.

**2.2 Tools & techniques**

- Caller ID spoofing, VoIP to obfuscate origin, pre-recorded prompts, use of internal jargon to increase authenticity.

- OSINT to make pretext realistic (LinkedIn, company websites).

**2.3 Case study: Twitter (July 2020) — social engineering of employees**

- When/where: July 14–15, 2020; Twitter, U.S. and global impact.

- How: Attackers used phone-based spear-phishing (vishing) to trick a small number of Twitter employees into revealing credentials that allowed access to internal support tools; attackers posted tweets from high-profile accounts and attempted a Bitcoin scam. Twitter and regulatory investigations documented the attack and subsequent restrictions.

- Outcome/impact: Over 100 targeted accounts compromised; direct financial loss to victims of the Bitcoin scam (smaller sums than BEC but high reputational impact for Twitter). Arrests followed.

**2.4 Impact analysis**

**-** Financial: Can enable fraud or facilitate larger breaches; direct losses may be lower than BEC but facilitate more serious compromise.

**-** Reputational: Loss of control over high-profile accounts harms public trust (Twitter example).

- Operational: Forced rollback of privileges, temporary lockouts, process changes.

- Legal/Compliance: Potential regulatory scrutiny if data/processes abused.

### 2.5 Mitigations — pretexting

**Technical:** Strong multi-factor authentication (hardware tokens), contextual access controls (IP/geo policies), out-of-band verification for sensitive actions.

**Administrative:** Strict verification procedures for internal support requests (e.g., callback to known number), role separation for privileged tools, logging of all privileged actions.

**Human-centric:** Training on vishing and pretext recognition, regular drills, “no unique answer” policies (don’t reveal personal info).

**Framework:** NIST guidance on identity and access management, CIS Control 4 (access control) and 16 (incident response).

## 3. Baiting

### 3.1 Description & step-by-step

Baiting is a social engineering technique that exploits human curiosity, greed, or desire for incentives by offering a seemingly appealing reward to lure victims into compromising their own security. Unlike phishing or pretexting, which rely primarily on communication and deception, baiting introduces a physical or digital “bait”—such as a free USB flash drive, download link, or promotional offer—that appears harmless or beneficial. A classic example involves leaving an infected USB drive labeled *“Confidential – Payroll 2025”* in a public area, enticing someone to plug it into their computer out of curiosity or the expectation of reward. Once the device is connected, malware automatically installs, granting the attacker unauthorized access to sensitive files or network systems.

**Steps:**

- Attacker leaves physical media or posts enticing downloads.
- Victim picks up/trusts the bait and plugs USB or downloads file.
- Malware runs (autorun or social engineering to enable macros), establishing foothold.
- Targets: Office employees, visitors, events (conferences).

### 3.2 Tools & techniques

- Infected USB drives, malicious free downloads, poisoned software packages (supply-chain style baiting).

- Use of social media or job sites to bait jobseekers with malicious attachments.

### 3.3 Case study: Stuxnet-style supply chain vs. classic baiting

**-** While high-profile nation-state attacks use advanced supply-chain techniques, many incidents are low-tech: USB drop exercises in red teaming routinely succeed at compromising devices. Academic and practitioner red team reports repeatedly show physical baiting works in real environments (OSINT and red team guides document methodology).

### 3.4 Impact analysis

- Financial: Clean-up, reimaging devices, malware containment costs.

- Operational: Infection can cause production outages or unauthorized access enabling subsequent attacks.

- Reputational & legal: Sensitive data exfiltration triggers reporting obligations.

### 3.5 Mitigations — baiting

**Technical:** Disable autorun, endpoint protection that blocks unknown USB devices, application whitelisting.

**Administrative:** Acceptable use policies forbidding use of unknown media, visitor controls.

**Human-centric:** Awareness campaigns (don't plug unknown USBs), physical security for devices.

**Framework:** CIS Controls for endpoint security and removable media management.

## 4. Impersonation

### 4.1 Description & step-by-step

Impersonation is a targeted form of social engineering that involves an attacker posing as a trusted individual or legitimate entity to deceive victims into performing specific actions or disclosing sensitive information. While it shares similarities with pretexting, impersonation is more identity-focused, relying on the attacker's ability to convincingly mimic the identity, tone, and authority of a real person—such as a company executive, IT administrator, financial officer, or vendor representative. The goal is often to exploit the victim's trust and compliance with hierarchical or professional relationships.

**Flow:**

- Gather information (org chart, email formats).
- Forge sender identity (display name spoofing, look-alike domains, compromised accounts).
- Send targeted request with urgency.
- Victim acts (approves payments, shares data).

**4.2 Tools & techniques**

**-** Email spoofing, domain squatting, compromised inboxes, deepfake audio for phone impersonation.

**-** BEC attacks are a common impersonation use case; FBI/IC3 documents widespread losses.

**4.3 Case study: Business Email Compromise — scale & losses**

**-** IC3 / industry findings: BEC/EAC scams caused billions in losses globally (IC3 reported early estimates in billions; cumulative industry analyses put losses into the tens of billions over multiple years).

**-** Ubiquiti (2015) serves again as a strong example of impersonation that led to massive wire transfers.

**4.4 Impact analysis**

**-** Financial: Often large single losses (wire transfers).

**-** Reputational: Suppliers/customers may lose confidence.

**-** Operational & legal: Bank recovery efforts, litigation, regulation.

**4.5 Mitigations — impersonation**

**Technical:** DMARC/DKIM/SPF enforcement, secure email gateways that detect impersonation, transaction verification systems.

**Administrative:** Strong payment authorization policies (dual signoff, verification on independent channel), vendor onboarding checks.

## 5. Insider threats

### 5.1 Description & step-by-step

Insider threats occur when individuals within an organization—such as employees, contractors, or trusted third-party partners abuse their authorized access to systems, networks, or data, either intentionally (malicious) or unintentionally (negligent). Malicious insiders may act out of  while negligent insiders might expose information through. Unlike external attacks, insider threats are particularly dangerous because they originate from users who already possess legitimate access rights, making detection and prevention more difficult.

**Common patterns:** misuse of privileged credentials, unauthorized data exfiltration, accidental misconfiguration. Detection is challenging due to legitimate access overlap.

### 5.2 Tools & techniques

- Abuse of legitimate tools (SFTP, cloud storage), insider collusion with external actors, use of removable media.

- Attackers may bribe or coerce employees or leverage negligent behavior (weak passwords).

### 5.3 Case study: Snowden + government insider incidents (classic)

- Edward Snowden (2013): high-impact data exfiltration from NSA systems by a privileged insider. Snowden's case is widely documented and shows how insider access can bypass many perimeter controls; it drove investments in insider threat programs across governments.

### 5.4 Impact analysis

**- Financial**: remediation, litigation, loss of IP.

**- Reputational:** breaches of trust; regulatory fallout (especially where regulated personal data leaked).

**- Operational:** loss of sensitive capabilities, potential national security implications.

**- Legal/Compliance:** GDPR/HIPAA fines where personal data is involved.

### 5.5 Mitigations — insider threats

**Technical:** Least privilege, privileged access management (PAM), user and entity behavior analytics (UEBA), DLP, robust logging and SIEM.

**Administrative:** Clear policies for separation of duties, exit procedures that revoke access promptly, routine audits.

**Human-centric:** Employee assistance programs, monitoring for disgruntlement indicators (while respecting privacy), insider threat awareness training.

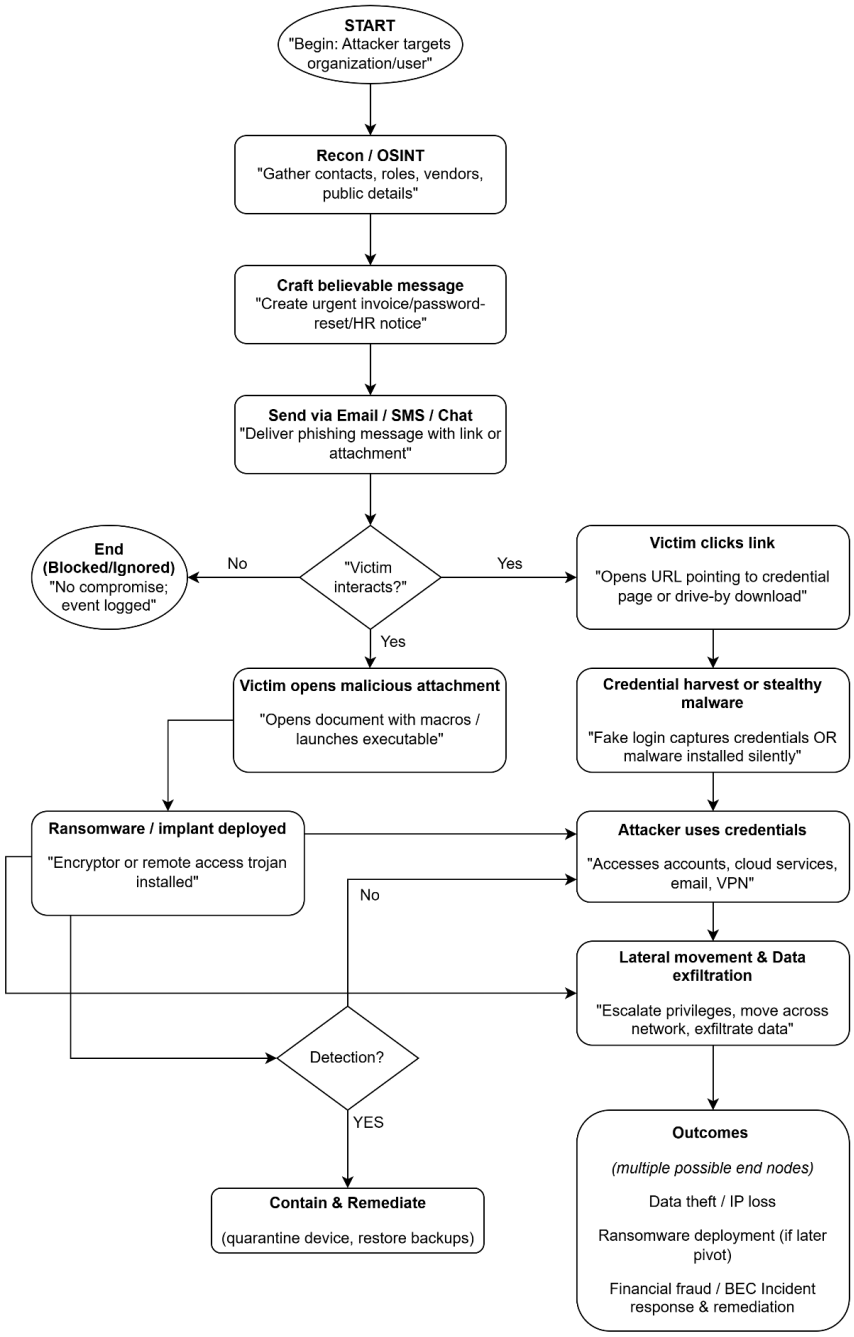
**Framework:** NIST and CCDCOE insider threat guidance provide frameworks for socio-technical defenses.

**Comparative overview table**

| Threat | Typical Targets | Example Case | Primary Impact | Top 3 Controls |
|---|---|---|---|---|
| Phishing | All (email users) | Ubiquiti/BEC, general DBIR trends | Financial loss, credential theft | Email auth (DMARC), phishing training, sandboxing |
| Pretexting (vishing) | IT, finance, support | Twitter 2020 | Account takeover, reputational harms | MFA, OOB verification, staff verification procedures |
| Baiting | Office staff, visitors | Red team USB drop incidents (industry reports) | Malware infection, lateral movement | Disable autorun, EPP, media policy |
| Impersonation (BEC) | Finance/vendors | Ubiquiti $46.7M loss | Large wire fraud | Dual approvals, vendor validation, secure email controls |

| | | | | |
|---|---|---|---|---|
| Insider threats | Privileged employees | Snowden (2013) | Data exfiltration, IP loss | PAM, UEBA, least privilege, exit controls |

## Diagrams (flowchart — phishing example)



## Best Practices & Framework Alignment

Mitigating human-based attacks requires aligning organizational defenses with established cybersecurity frameworks that emphasize a balance of technical, administrative, and human-centric controls. The **NIST Cybersecurity Framework (CSF)** provides a holistic approach structured around five core functions: **Identify, Protect, Detect, Respond, and Recover.** Within this model, organizations should identify key assets and user roles vulnerable to social engineering, protect identities through strong authentication and access controls, detect anomalies via continuous monitoring and behavioral analytics, respond promptly through documented incident response plans, and recover operations using tested backup and restoration strategies. (Source: NIST Computer Security Resource Center)

The **CIS Controls v8.1** offers a prioritized set of safeguards tailored to defend against prevalent attack vectors like phishing, pretexting, and insider threats. Key recommendations include implementing **email and web browser protections**, enforcing **multi-factor authentication (MFA)** for critical systems, and ensuring comprehensive **logging and endpoint defenses**. These controls directly support risk reduction by focusing on visibility, automation, and user awareness—areas commonly exploited in social engineering attacks. (Source: Center for Internet Security)

The **ISO/IEC 27001** standard complements these approaches by establishing an **Information Security Management System (ISMS)** that promotes continuous risk assessment, improvement, and adherence to organizational security policies. Through the ISMS framework, enterprises can ensure that preventive and detective measures are regularly audited, evaluated, and adapted to emerging social engineering tactics.

Operationally, organizations should **prioritize security training for high-risk roles**—particularly employees in **finance, HR, and IT**, who are frequent targets of phishing and pretexting. Regular **phishing simulations** and **awareness programs** should be integrated into performance metrics, using tools like the **NIST Phish Scale** to measure susceptibility and resilience. Additionally, **transaction verification controls**, **change management procedures**, and **executive approval protocols** should be enforced to limit the impact of impersonation and insider threats. (Source: NIST Computer Security Resource Center)

## Conclusion & key insights

**Human weakness is an attack surface.** Technical defenses alone are insufficient; attackers exploit trust and process weaknesses. DBIR and incident case studies confirm social engineering remains a dominant factor in breaches. Verizon

**High financial and reputational stakes.** Incidents like Ubiquiti's multi-million-dollar loss and the Twitter breach show social engineering can produce catastrophic and highly visible impacts. SEC+1

**Defense must be socio-technical.** Combine DMARC/SPF/DKIM, PAM, DLP, and UEBA with realistic training, verification procedures, and policy controls. Map controls to NIST CSF and CIS for prioritization. CIS+1

**Final recommendation:** Start with a risk-based program: identify top human-risk vectors (finance/BEC, privileged users), implement technical mitigations (email auth, MFA, PAM), apply administrative policies (dual approvals, incident playbooks), and scale human-centered training driven by phishing simulation results and the NIST Phish Scale.

References

**References (selected — APA)**

Note: below are the principal sources used in this report. I used them for factual data, case summaries, and best-practice guidance.


- Dawkins, S. (2023). *NIST Phish Scale User Guide* (NIST TN 2276). National Institute of Standards and Technology. https://csrc.nist.gov/pubs/tn/2276/final. NIST Computer Security Resource Center
- NIST. (2021). *Phishing* — Small Business Cybersecurity Corner. National Institute of Standards and Technology. https://www.nist.gov/itl/smallbusinesscyber/guidance-topic/phishing. NIST
- Verizon. (2024). *Data Breach Investigations Report (DBIR) 2024/2025 executive summaries and report.* Verizon Business. https://www.verizon.com/business/resources/reports/dbir. Verizon+1

- OWASP. (2020). *OSINT/Red Team: Phishing Techniques & Testing Guide.* OWASP chapter/guide. OWASP Foundation+1
- IC3 (FBI). (2019). *Business Email Compromise (BEC) — The $26 Billion Scam.* Internet Crime Complaint Center (IC3). https://www.ic3.gov/PSA/2019/PSA190910. Internet Crime Complaint Center
- Wired. (2020). *Inside the Twitter Hack—and What Happened Next.* Wired. https://www.wired.com/story/inside-twitter-hack-election-plan/. WIRED
- U.S. Securities and Exchange Commission (SEC). (2015). Ubiquiti Networks, Inc. 8-K filing re: fraud. https://www.sec.gov/Archives/edgar/data/1511737/000157104915006288/t1501817_8k.htm. SEC
- ResearchGate / academic surveys: (2023–2024). *Phishing and social engineering reviews and surveys.* Representative articles: “Phishing Attacks in Social Engineering: A Review” (2023) and IJAAS 2024 comprehensive survey. ResearchGate+1
- CCDCOE. (2018). *Insider Threat Study (NATO CCD COE).* https://ccdcoe.org/uploads/2018/10/Insider_Threat_Study_CCDCOE.pdf.