Title: Beyond the Firewall: Human Centric Threat Modeling for the Age of Social Engineering

Abstract: For decades, enterprise threat modeling has fixated on technical flaws: software vulnerabilities, protocol weaknesses, and architectural gaps. This paper argues that this view is now dangerously obsolete. We contend the primary attack vector has decisively shifted from machine exploitation to human manipulation, rendering a human-centric approach to security not as a supplement, but as an essential pillar of modern enterprise defense. Through a qualitative, comparative analysis of the landmark security breaches at Twitter (2020), Uber (2022), and MGM Resorts (2023), this research deconstructs the attackers' recurring playbook. Our analysis exposes a consistent pattern of social engineering that methodically dismantles robust technical controls by exploiting cognitive biases, inherent human trust, and procedural gaps. In response, this paper introduces the Human-E.S.C.A.P.E. Threat Model, a new, structured framework for modeling these human-centric threats. We conclude by outlining actionable mitigation strategies derived from this model, designed to fortify an organization's most critical asset: its people.

Keywords: Cyber security, Threat Modeling, Social Engineering, Human Factors, Enterprise Security, Insider Threats, Vishing, Phishing.

I. INTRODUCTION

The modern landscape of cyber warfare presents a stark paradox. Organizations invest fortunes in firewalls and advanced endpoint protection, yet their defenses are routinely dismantled by an attacker with a convincing voice and a clever lie [1]. A growing body of evidence shows that both sophisticated Advanced Persistent Threat (APT) groups and agile cyber criminals have reached the same conclusion: compromising a human is often far more efficient and cost-effective than developing a novel exploit for a hardened system [2].

This reality has fueled the ascent of social engineering as a primary attack strategy. It is the art of psychological manipulation, turning an individual's innate helpfulness or natural deference to authority into a critical security vulnerability. While the tactics are timeless, their seamless integration into complex, multi-stage cyber attacks now defines modern corporate breaches. The most catastrophic incidents of data loss or ransomware deployment now frequently begin not with a brute-force technical assault, but with a disarmingly simple, deceptive human interaction [3].

Despite this clear and present danger, our defensive strategies and threat modeling frameworks have been slow to adapt. Paradigms such as STRIDE and PASTA, while invaluable, were engineered to diagnose technical systems, not human psychology [4]. They excel at identifying flaws in code but are ill equipped to model the messy, unpredictable, and exploitable nature of human behavior. This creates a fundamental disconnect in enterprise security: digital fortresses are hardened while the human manned gates are left comparatively unprotected.

This paper confronts this gap by championing a human-centric threat model. To build a truly resilient defense, organizations must learn to see their security posture through the eyes of an attacker who targets people, not ports. To build this case, we will explore the blind spots in traditional threat modeling, deconstruct the attack chains of recent high profile breaches, and synthesize these events to reveal a common pattern of human exploitation. By examining these attacks through a human-focused lens, we will deliver an evidence based road map for security leaders to address this persistent and growing threat.

II. BACKGROUND AND RELATED WORK

To grasp the nature of human centric threats, one must understand both the craft of the social engineer and the doctrines of threat modeling that so often fail to account for it.

A. The Social Engineer's Toolkit

At its core, social engineering is a discipline of applied psychology, not technology. It preys on fundamental human drivers such as helpfulness, trust, and fear of authority [5]. As the legendary hacker Kevin Mitnick observed, "It's easier to fool a person than a computer" [6]. The primary methods include:

- Phishing: Email-based attacks disguised as legitimate communication, designed to harvest credentials or deliver malware. Spear phishing is a highly targeted variant aimed at specific individuals or roles [7].
- Vishing (Voice Phishing): The use of telephony to impersonate a trusted entity such as an IT support agent or a new executive to coax sensitive information from a target [8].
- SMiShing (SMS Phishing): Phishing attacks conducted via text message, often leveraging a sense of urgency to compel users to click malicious links.
- Pretexting: The creation of a believable scenario, or pretext, to justify the attacker's requests. This narrative is often meticulously built from details gathered via open-source intelligence (OSINT), for instance, from professional networking sites [9].
- MFA Fatigue: A modern technique used after an attacker has obtained a user's password. The attacker triggers repeated MFA push notifications, bombarding the target until they relent out of annoyance or confusion and approve the request [10].

These methods are rarely used in isolation. A sophisticated campaign often begins with OSINT to construct a credible pretext, which is then deployed in a vishing call designed to socially engineer an MFA bypass.

B. Where Traditional Threat Models Fall Short

Threat modeling provides a structured methodology for identifying security risks. A framework like Microsoft's STRIDE is invaluable for developers, prompting them to consider Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege. Its focus, however, remains on technical systems. It struggles profoundly when the "spoofed" entity is not a server but a person's trust over a phone line.

Consider this: how does STRIDE model the risk of an IT help desk agent who, driven by a desire to be helpful, overrides a cumbersome verification protocol? All too often, this systemic vulnerability is dismissed as a simple "user error" [11]. This categorization is a critical failure. It mischaracterizes a systematic, repeatable attack vector as a random accident, thus preventing organizations from designing the robust, process-level defenses that are desperately needed.

C. The Human Factor

Insights from psychology and human-computer interaction (HCI) are vital. Cognitive biases, such as our deference to perceived superiors (authority bias) or our tendency to see what we expect to see (conformation bias), render individuals vulnerable [12]. This is compounded by "security fatigue," a phenomenon where users, bombarded with constant security warnings and procedures, grow apathetic. In this state, they are far more likely to engage in risky behavior, such as approving an unexpected MFA prompt simply to make an alert go away and get on with their work [13]. An effective threat model cannot treat these human factors as edge cases; they are core, exploitable vulnerabilities.

D. Foundational Psychological Principles in Social Engineering

A robust analysis of human-centric attacks requires grounding in established psychological principles. Social engineers do not invent new human weaknesses; they exploit pre-existing, predictable cognitive pathways. Foundational work in the psychology of influence by researchers like Robert Cialdini identifies several key principles frequently weaponized by attackers. These include:

Authority: Individuals are highly likely to comply with requests from perceived authority figures. An attacker impersonating an IT administrator or a corporate executive leverages this deep-seated deference [16].

Liking: People are more easily persuaded by individuals they like or feel a connection with. Attackers build rapport by being friendly, feigning common interests, or adopting a helpful persona, thereby lowering the target's defenses [16].

Reciprocity: The principle that people feel obligated to give back after they have received something. An attacker offering "help" with a manufactured IT problem can create a sense of obligation in the target to comply with a subsequent request [16].
These principles, combined with the cognitive shortcuts (heuristics) and biases described by Tversky and Kahneman [12], form the theoretical bedrock of modern social engineering.

III. METHODOLOGY

To ground this analysis in empirical reality, this research employs a qualitative, comparative case study methodology. This approach is exceptionally well-suited for exploring the complex, multi-faceted nature of real world cyber attacks, allowing for a deep examination of the "how" and "why" behind these events without requiring access to proprietary corporate data [14].

A. Case Selection

We selected three landmark security breaches: Twitter (2020), Uber (2022), and MGM Resorts (2023). These cases were chosen for their recency, significant impact, and the well-documented nature of their attack chains. Crucially, they were selected because they all pivot on the social engineering of IT support or adjacent help desk functions, providing a clear, comparable through line for analysis. While other breaches involving social engineering have occurred (e.g., at Okta), these three cases offer the most publicly detailed accounts of how human-facing support processes were systematically compromised to bypass modern technical defenses like MFA, making them ideal subjects for building a human-centric threat model.

B. Data Analysis Process

Data was collated from a range of public sources, including in-depth reporting from outlets like The Wall Street Journal and WIRED, official corporate statements and SEC filings, and analyses from government agencies such as CISA. A thematic analysis was performed on these source materials. Each case was systematically coded to deconstruct the attack chain using a consistent analytical framework focused on the human element, examining: 1) the Initial Vector (the specific social engineering tactic used), 2) the Exploited Vulnerability (the psychological trigger or procedural gap leveraged), 3) the Bypassed Defense (the technical control rendered ineffective), and 4) the Resulting Impact. This structured coding enabled a systematic comparison across cases, as summarized in Table I.

IV. RESULTS: CASE STUDY ANALYSIS

Our analysis reveals that in each case, attackers strategically chose to target a human vulnerability as the path of least resistance.

A. The Twitter Hack (2020): Weaponizing Trust in the Help Desk

- Initial Vector: The breach began with a coordinated series of vishing calls. Posing as internal IT staff, the attackers guided employees to a fake internal website to harvest their corporate credentials [8].
- Exploited Vulnerability: The attackers preyed on trust by invoking the Authority principle [16]. Posing as legitimate IT staff, they created a scenario where cooperation was the path of least resistance for employees accustomed to following instructions from technical support.
- Bypassed Defense: Even with MFA in place, the attackers simply requested the MFA codes during the vishing call. The employees, believing the interaction was legitimate, complied. The human operator, not a software flaw, was the point of failure for the MFA control.
- Resulting Impact: Armed with legitimate, MFA verified credentials, the attackers accessed Twitter's powerful internal tools, hijacked high profile accounts, and executed a cryptocurrency scam, causing immense reputational damage.

B. The Uber Breach (2022): An Attack on Patience and Trust

- Initial Vector: This was a two-pronged assault on human psychology. An attacker, having purchased an employee's credentials on the dark web, initiated an MFA fatigue attack, spamming the employee with push notifications for over an hour [10]. Concurrently, the attacker reached out on WhatsApp, pretexting as an IT staff member and instructing the employee to accept the prompt to resolve the "issue."

- Exploited Vulnerability: This attack brilliantly leveraged both security fatigue and trust. The target, annoyed by the incessant notifications, was offered a simple solution by a person they believed was there to help.
- Bypassed Defense: The attackers did not break MFA; they socially engineered it. The system performed exactly as designed, but the human at the end of the chain was manipulated into becoming an unwitting accomplice.
- Resulting Impact: Once inside, the attacker discovered a network share containing PowerShell scripts with hardcoded administrative credentials. This catastrophic failure of access control enabled sweeping privilege escalation and the compromise of Uber's G-Suite, Slack, and sensitive security reports.

C. The MGM Resorts Attack (2023): Hacking the Help Desk

- Initial Vector: The attack began with simple reconnaissance. The "Scattered Spider" hacking group reportedly used LinkedIn to identify an MGM employee. They then placed a single, 10-minute vishing call to the IT help desk, convincingly impersonating that employee to request a password reset [15].
- Exploited Vulnerability: The critical vulnerability was not a single employee's mistake but a systemic, procedural weakness in the IT help desk's identity verification process. The attacker's convincing pretext exploited the agent's desire to provide efficient customer service, leading them to neglect strict security checks.
- Bypassed Defense: The attack circumvented MFA and Identity and Access Management (IAM) entirely. Rather than attempting to break the controls, the attackers co-opted the organization's own legitimate support processes to have MFA reset and re-enrolled on a device they controlled.
- Resulting Impact: With a legitimate foothold, the attackers deployed ransomware, triggering a corporate meltdown. Hotel room keys, slot machines, and payment systems failed, paralyzing operations and costing the company an estimated $100 million.

TABLE I. SUMMARY OF CASE STUDY FINDINGS

Case	Primary Social Engineering Tactic(s)	Exploited Human Vulnerability	Bypassed Technical Control
Twitter (2020)	Vishing, Pretexting	Trust in authority, Helpfulness	Multi-Factor Authentication (MFA)
Uber (2022)	MFA Fatigue, Vishing, Pretexting	Security fatigue, Trust, Annoyance	Multi-Factor Authentication (MFA)
MGM (2023)	Vishing, Pretexting (OSINT-based)	Helpfulness, Procedural weakness at help desk	Identity and Access Management (IAM)

V. DISCUSSION

The analysis across these three incidents reveals a disturbing and predictable pattern. These were not isolated failures but the foreseeable outcomes of a security posture that neglects the human dimension.

A. Synthesis of Findings: The Human Attack Surface

1. **The IT Service Desk as a High-Value Target:** Our analysis reveals the IT service desk as the modern enterprise's soft underbelly. In two of the three cases (MGM and Twitter), this was the fulcrum of the attack. Attackers understand that service desks are staffed by humans evaluated on their ability to resolve issues quickly. This makes them a prime target for social engineering and a gateway for credential resets and MFA bypasses.
2. **Trust as a Weapon:** Every one of these attacks pivoted on the exploitation of trust. The attackers' primary tool was not malware but credibility. This demonstrates that in a human-centric model, trust itself must be treated as a managed vulnerability, not an assumed state.

3. **The Illusion of Technical Controls:** All three companies had deployed MFA, yet it failed to prevent the breach in every case. This reveals the dangerous illusion of purely technical security. A technical control is only as strong as the human processes that govern it. Threat models must scrutinize not just the lock, but also the process for creating a new key.

B. A Proposed Framework: The Human-E.S.C.A.P.E. Threat Mode

he preceding analysis reveals the inadequacy of technical-only threat models. To address this, we propose a new, structured framework for modeling human-centric threats: The Human-E.S.C.A.P.E. Threat Model. This model forces security architects to analyze threats not just as code exploits, but as orchestrated campaigns against people and processes.
E.S.C.A.P.E. is an acronym for the six stages of a human-centric attack:

1. **E**ntry Vector: How is the initial contact made? (e.g., Vishing call, SMiShing text, Phishing email).
2. **S**ocial Tactic: What is the specific social engineering technique used? (e.g., Pretexting, MFA Fatigue).
3. **C**ognitive Bias: Which psychological principle is being leveraged? (e.g., Authority, Liking, Urgency).
4. **A**uthority Abuse: What legitimate authority or process is being impersonated or co-opted? (e.g., IT Help Desk, New Executive Onboarding).
5. **P**rocedural Gap: Which specific policy or procedural weakness allows the attack to succeed? (e.g., Weak identity verification, lack of mandatory callback).
6. **E**vasion of Controls: Which technical security control is ultimately bypassed or rendered irrelevant? (e.g., MFA, IAM)

APPLICATION OF THE E.S.C.A.P.E. MODEL TO THE MGM RESORTS BREACH

Stage	Analysis of the MGM Breach
Entry Vector	Vishing (Phone call to IT help desk).
Social Tactic	OSINT based Pretexting (Impersonating a known employee found on LinkedIn).
Cognitive Bias	Authority (Help desk agent's deference to an employee) and Helpfulness.
Authority Abuse	Co-opting the legitimate IT Help Desk password reset process.
Procedural Gap	Inadequate identity verification protocol for remote password resets.
Evasion of Controls	IAM and MFA were bypassed entirely by resetting credentials

VI. CONCLUSION

This analysis leads to one compelling conclusion: the primary battlefield for enterprise security has shifted from silicon to psychology. This paper has demonstrated, through the stark lessons of Twitter, Uber, and MGM, that social engineering is a central pillar of advanced cyber attacks.

To combat these threats, we introduced the Human-E.S.C.A.P.E. Threat Model, a novel framework designed to shift defensive thinking from technical configurations to human processes. By forcing an analysis of the psychological principles, procedural gaps, and co-opted authorities that attackers leverage, this model enables organizations to identify and mitigate their true weakest links. The solutions are not just technical, but are deeply rooted in behavior, culture, and process.

Future work should focus on developing a quantifiable risk model for help desk identity verification processes, factoring in variables such as request urgency, employee role, and the sensitivity of requested resources. However, the path forward requires an immediate change in mindset. We must stop viewing our people as the weakest link and start treating them as the most critical defense layer to be trained, empowered, and defended with the same rigor we apply to our most sensitive data.

REFERENCES

[1] ENISA Threat Landscape 2023, European Union Agency for Cybersecurity, Oct. 2023. [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023

[2] K. Poulsen, "The Human Factor: The Weakest Link in Cyber security," IEEE Security & Privacy, vol. 18, no. 1, pp. 5-7, Jan.-Feb. 2020.

[3] Verizon, "2023 Data Breach Investigations Report," Verizon Enterprise Solutions, 2023. [Online]. Available: https://www.verizon.com/business/resources/reports/dbir/

[4] A. Shostack, Threat Modeling: Designing for Security. Hoboken, NJ: John Wiley & Sons, 2014.

[5] C. Hadnagy, Social Engineering: The Art of Human Hacking. Indianapolis, IN: Wiley Publishing, Inc., 2010.

[6] K. D. Mitnick and W. L. Simon, The Art of Deception: Controlling the Human Element of Security. Indianapolis, IN: Wiley Publishing, Inc., 2002.

[7] P. Krombholz, H. Hobel, M. Huber, and E. Weippl, "Advanced social engineering attacks," Journal of Information Security and Applications, vol. 22, pp. 113-122, Jun. 2015.

[8] U.S. Department of Justice, "Three Individuals Charged For Alleged Roles In Twitter Hack," Ofce of Public Affairs, Jul. 31, 2020. [Online]. Available: https://www.justice.gov/opa/pr/three-individuals-charged-alleged-roles-twitter-hack

[9] S. E. F. van der Walt and R. G. von Solms, "A pretexting-based social engineering model," in Proc. 10th Int. Conf. Information Warfare and Security, 2015, pp. 316-323.

[10] Uber, "Update on our security incident," Uber Newsroom, Sep. 19, 2022. [Online]. Available: https://www.uber.com/newsroom/update-on-our-security-incident/

[11] M. A. Sasse, "Users are not the enemy," Communications of the ACM, vol. 49, no. 10, pp. 40-46, Oct. 2006.

[12] A. Tversky and D. Kahneman, "Judgment under Uncertainty: Heuristics and Biases," Science, vol. 185, no. 4157, pp. 1124-1131, 1974.

[13] S. R. R. Kirlappos, I. Kirlappos, and M. A. Sasse, "'Security fatigue': The effects of usability on security," in Proc. IEEE e-Crime Researchers Summit, 2013, pp. 1-6.

[14] R. K. Yin, Case Study Research and Applications: Design and Methods, 6th ed. Thousand Oaks, CA: SAGE Publications, 2018.

[15] K. Poulsen and R. McMillan, "The Anatomy of a Devastating Cyberattack," The Wall Street Journal, Sep. 28, 2023. [Online]. Available: https://www.wsj.com/story/the-anatomy-of-the-devastating-cyberattack-on-mgm-and-caesars-57a151f1

[16] R. B. Cialdini, Influence: The Psychology of Persuasion. New York: Collins, Rev. ed., 2007.