ETHICAL

HUMAN HACKING

DINA BOARD

TABLE OF CONTENTS

ABOUT

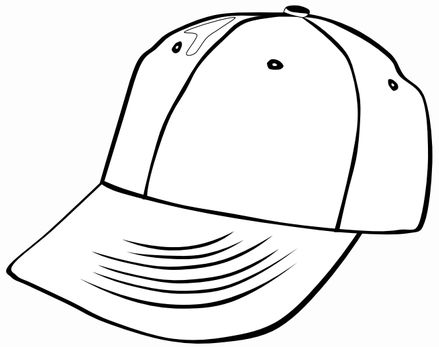
Hacking humans, like hacking computers, involves identifying and exploiting vulnerabilities—but rather than software flaws, it targets human behaviour, psychology, and trust through social engineering and manipulation.

Hacking humans is a process which exploits vulnerabilities in human psychology and behaviour to gain unauthorized access to sensitive information, manipulate opinions, or influence actions [1].

Humans are the final weakest link when it comes to information security - how do we ethically test them?

Education and training is one of the biggest and strongest ways to mitigate human-targeted attacks by improving awareness, reducing susceptibility to social engineering tactics, and fostering critical thinking about security risks.

ETHICAL CONSIDERATIONS & CONCERNS



Social engineering attacks may have unintended after-effects on the victim, which may be so severe that they can lead to suicide or other forms of trauma. Thus, ethical concerns related to such attacks, as well as their consequences, could be well minimized if the right actions are taken post the attack. [3]

Social engineering attacks performed in a penetration testing or social engineering research environment are not intended to cause harm to the victim or to make malicious use of the information gathered in the attack. [3]

However, research into ethical penetration testing of all categories is currently heavily lacking, particularly in the area of psychological and emotional impact on individuals targeted in social engineering exercises. While technical penetration testing has well-established frameworks and best practices, ethical guidelines for human-focused testing remain underdeveloped.

This gap underscores the need for further interdisciplinary research that integrates cybersecurity, psychology, and ethics to establish clear standards for minimizing harm while effectively assessing security vulnerabilities. Developing such frameworks will help ensure that social engineering assessments strengthen security awareness without causing undue distress or violating ethical boundaries.

WHY ETHICAL TESTING IS IMPORTANT

To mitigate concerns (see right), to remain compliant with security regulations, not cause lasting damage, and overall improve the process of testing humans, ethics becomes crucial.

Ethics create a system of appropriate and inappropriate testing practices and help testers use a systematic approach that prioritizes security improvements while minimizing harm. Ethical guidelines ensure that social engineering tests are conducted responsibly, respecting individuals' privacy, emotional well-being, and legal boundaries. They also help build trust between security teams/ organizations and employees - helping change the culture from fear to learning.

CONCERNS

- Informed consent and authorization
- Psychological and emotional impact
- Legal and regulatory compliance
- Avoiding lasting damage to reputation
- Controlled use of red team findings
- Transparency in post-assessment reporting

COMMON VULNERABILITIES

PSYCHOLOGICAL AND EMOTIONAL MANIPULATION & BEHAVIOURAL AND HABIT BASED WEAKNESSES

HUMAN EMOTIONS

Attackers often impersonate trusted entities to exploit human emotions like fear or trust, creating a sense of urgency to persuade victims into providing access to sensitive information. This social engineering tactic preys on human emotions and exploiting general trust. Fear and urgency are powerful motivators, as they can cause an individual to behave impulsively without considering consequences [1]. Being aware of how your own emotions can be used against you is crucial in recognizing and resisting manipulation tactics, allowing you to pause, verify requests, and make informed security decisions rather than reacting impulsively.

BEHAVIOURAL AND HABIT BASED WEAKNESSES & OTHER

- Trust in authority - people usually easily comply with authority figures
- Social validation and peer pressure - individuals tend to follow group behaviour
- Curiosity and desire for free offers - people are tempted by opportunities
- Desire to be helpful - individuals tend to naturally assist others
- Habit and routine compliance - people tend to follow familiar processes without question
- Cognitive biases (familiarity and halo effect) - trust is built on familiarity, not verification

COMMON VULNERABILITIES

OVER RELIANCE ON SYSTEMS AND LACK OF AWARENESS

SOCIAL MEDIA

Studies have found that users with a higher level of involvement in a given social media are more susceptible to social engineering attacks and the user's level of involvement positively influences the user's trust in the social media. More involvement results in more trust [2]. Through reconnaissance and OSINT, an adversary can use an individual's social media easily against them in order to attack. To mitigate this, one should take steps to minimize their digital footprint and be mindful of the information they share on social media platforms - even ones they trust.

LACK OF AWARENESS

- Failure to recognize social engineering tactics - individuals may be unfamiliar with common attack techniques
- Assuming security is someone else's responsibility - similar to bystander apathy or the assumption that cybersecurity is solely handled by IT
- Not knowing the value of your own information - many people underestimate how small pieces of information can be used for targeted attacks
- Lack of verification culture - blindly trusting requests without verifying authenticity

CASE STUDIES

The next two slides contain two real-life red team engagements.

Please read each scenario and consider how you would approach the scenario in an ethical manner before going to the next slide. Remember, for all scenarios you have a letter of permission from executives that is your “get out of jail” card.

After each case slide, you will find a slide describing what was actually done and whether the team was successful.

CASE I

You have been tasked with gaining access to an office building and installing a rogue device on the company's internal network as part of a physical red team assessment.

Security Measures in Place:

- Building Security: Checks for employee badges upon entry.
- Secondary Security Personnel: Have access to the company's internal HR system and can validate badge authenticity if necessary.

Additional Information:

- You possess an authorization letter from the company's executives to present if your activities are discovered.

Your Task:

Given the limited information available, how would you approach gaining access to the building while remaining undetected? What tactics would you consider to blend in, bypass security, and achieve your objective?

CASE I - WHAT HAPPENED

- **Recon & Preparation:** Noticed all employees wore visible badges. Produced near-identical badges with our photos to blend in and bypass initial scanning.
- **Infiltration:** Accessed a senior floor, made three discreet attempts to plug in a rogue device, and blended in by working at an empty desk while posing as an employee from another branch.
- **Adaptive Tactic:** When notified the device was not connecting, I returned to the main floor, claimed an urgent meeting and a malfunctioning scanning card, and was allowed into a secure area despite some suspicion. I then installed the rogue device in the secure area on an open network jack and exited quickly, pretending to be on an intense call.
- **Outcome:** The rogue device remained undetected on the network for some months—even without activity—demonstrating the risk of prolonged, covert access and exposing significant gaps in physical and network security. Gaining undetected access to the building and work desk demonstrates severe vulnerabilities in physical security, potentially allowing unauthorized access to sensitive areas and data.

CASE II

As part of a red team engagement, you have been tasked with attempting to reset the password of a senior employee at a large organization. This organization outsources its account management to a third-party provider, and you have access to their Account Support phone number. Additionally, your colleagues have compiled a list of employee names and email addresses during their engagement with the same client.

Your Task:

How would you prepare for the call to Account Support in order to successfully reset “your” password? What pretext would you use, and how would you handle potential verification questions? What strategies would you employ to increase your chances of success while remaining inconspicuous?

CASE II - WHAT HAPPENED

- **Recon & OSINT:** Selected three target names and gathered extensive information using social media (LinkedIn, Instagram, Facebook). Brainstormed and documented typical identity-confirmation questions for password resets.
- **Prepared Exits:** Prepared two background noises (baby crying, dog barking) as pretext for an exit if needed.
- **Pretext:** While there is a mobile application available for password reset, I have lost my phone and need help with the reset.
- **Call Execution:** Called Account Support, impersonating the employee with the most available information. When faced with an unanswerable question, played the sounds and requested a callback to "calm my crying baby." Hung up, gathered the missing details using social media, and then called back to answer all questions correctly.
- **Outcome:** Successfully reset the employee's password, demonstrating how thorough OSINT combined with creative social engineering can bypass verification procedures during account support interactions.

COMMON TACTICS

Pretexting

Creating a fabricated scenario or persona (a pretext) to gain the victim's trust and elicit confidential information. Thus, an attacker might impersonate a co-worker or vendor, citing a plausible reason (e.g., updating a system) to request sensitive data.

Baiting

Offering something enticing (such as a free download, prize, or physical item) to lure victims into compromising security. Thus, an attacker might leave a USB labelled "Confidential" in public areas, enticing individuals to plug them into their computers, or they might send out an offer for a free downloadable report that requires users to enter their credentials or personal details before accessing it.

Distraction and Diversion

Creating a diversion that draws attention away from the attacker's true objective. An attacker might engage a security guard or employee in a distracting conversation about an unrelated issue, allowing a colleague to slip through security unnoticed or tamper with physical assets.

Tailgating (Piggybacking)

Following an authorized individual into a restricted area by taking advantage of the human tendency to hold doors open for others. An attacker might wait for an employee to enter a secure facility and then closely follow them without presenting their own credentials or follow an employee in with full hands.

Impersonation

Assuming the identity of a trusted person or authority figure to bypass security protocols. For example, pretending to be a high-level executive to request immediate action from employees, such as transferring funds or disclosing confidential information.

Elicitation

Engaging in casual conversation to subtly extract confidential information without the target realizing it. While waiting in a lobby or elevator, an attacker might strike up a general conversation about the organization's security procedures or recent internal changes, collecting tidbits of information that can be pieced together later.

BEST PRACTICES

Clear Authorization and Scope

Every exercise is fully authorized by the organization, with documented scope and clear boundaries. This prevents unauthorized actions and ensures that all parties are aware of the testing, reducing legal and reputational risks.

Prioritize Minimal Disruption and Do No Harm

Design your tactics to avoid disrupting normal business operations and causing undue stress or damage. The goal is to test defences and identify vulnerabilities—not to interrupt or compromise the daily functioning of the organization or harm its employees.

Provide Constructive Feedback and Foster Learning

Document all findings and share them with the organization in a clear, constructive manner that highlights both strengths and areas for improvement. Follow up with recommendations for training and policy enhancements. This approach supports a culture of continuous improvement and education, ensuring that the red team exercise contributes to the overall security posture rather than simply exposing weaknesses.

Safeguard Confidentiality and Data Privacy

Handle any sensitive information gathered during the exercise with strict confidentiality, using it only for reporting and improvement purposes. Respect for privacy is paramount. This practice protects individual and organizational data from being exposed or misused.

Use Realistic, Yet Non-Exploitive Tactics

Craft scenarios (like social engineering or phishing tests) that mimic real threats without crossing into areas that could lead to actual compromise, such as irreversible actions. Ensures that red teaming remains a safe simulation, providing valuable insights while avoiding the risk of true harm or a breach of trust.

Leave the person and organization better off after your red team engagement, not worse.

YOUR TURN

The next three slides contain three hypothetical red team engagements.

Please read each scenario and consider how you would approach the scenario in an ethical manner before going to the next slide. Remember, for all scenarios you have a letter of permission from executives that is your “get out of jail” card.

After each case slide, you will find a slide describing what could have been done.

RED TEAM THIS I

Your target organization recently conducted security awareness training, emphasizing the importance of verifying identities before sharing sensitive information. Your goal is to convince an employee to provide you with internal Wi-Fi credentials over the phone.

Your Task:

1. What pretext would you use to make the employee trust you?
2. How would you handle objections or skepticism?
3. What indicators might alert a well-trained employee that they are being socially engineered?
4. What security measures should be in place to prevent employees from falling for this attack?

ANSWERS

Pretext to gain trust:

- Pretend to be from IT support, claiming there's an urgent network issue affecting employees.
- Reference a real company initiative, such as a recent security upgrade or policy change, to add legitimacy.
- Use publicly available information (e.g., company directory, LinkedIn profiles) to personalize your approach.

Indicators that this is a social engineering attack:

- Unexpected requests for sensitive information over the phone.
- A sense of urgency or pressure to act quickly.
- Caller reluctance to verify their identity through official channels.

Handling objections or skepticism:

- Apply psychological tactics like urgency ("We need this resolved before the executives arrive for a critical meeting.").
- Offer to "verify" your identity by providing publicly available company information.
- Redirect suspicion by feigning frustration with technical issues to elicit sympathy.

Security measures to prevent this attack:

- Implement verification protocols for IT support calls (e.g., requiring employees to call IT directly).
- Enforce multi-factor authentication (MFA) to limit the impact of credential theft.
- Conduct ongoing phishing simulations to train employees in spotting social engineering tactics.

RED TEAM THIS II

You are conducting a red team assessment at a corporate office. The company frequently works with outside vendors for maintenance and deliveries. Your objective is to gain access to a restricted floor by posing as a vendor.

Your Task:

1. What type of vendor role would be most believable for this pretext?
2. What verbal and physical cues would help you appear legitimate?
3. How might a well-trained receptionist or security guard identify you as a potential threat?
4. What policies should the company enforce to prevent unauthorized vendor access?

ANSWERS

Most believable vendor role:

- A role that is commonplace but not highly scrutinized, such as an IT technician, office supplies delivery person, or building maintenance staff.
- A role that has time-sensitive tasks, such as fire extinguisher inspections, HVAC maintenance, or emergency repairs.

Verbal and physical cues to appear legitimate:

- Dress appropriately for the vendor role (e.g., uniform, ID badge, clipboard).
- Use industry jargon to sound credible (e.g., “I’m here to check the network drop in Conference Room B.”).
- Display confidence and urgency—people are less likely to question someone who acts like they belong.

How trained employees can detect a potential threat:

- The vendor has no prior appointment in the visitor system.
- They avoid verification procedures or become defensive when asked for details.
- They appear too eager to bypass security controls (e.g., avoiding the check-in desk, waiting for doors to be held open).

Security policies to prevent unauthorized vendor access:

- Require vendor pre-registration and check all IDs upon arrival.
- Train employees to verify vendor credentials with a known point of contact inside the company.
- Implement escort policies for all third-party vendors in sensitive areas.

IMPORTANT ETHICAL OVERVIEW FOR RED TEAM THIS II

- Red teamers must ***never*** impersonate emergency services (e.g., police, firefighters) or real government agencies due to legal risks.
- Any access gained ***should be documented but not exploited*** —for example, never tamper with actual systems.
- The objective is to ***identify policy weaknesses, not deceive employees*** into making serious security breaches.

RED TEAM THIS III

You have been tasked with launching a phishing attack against employees of a multinational corporation. Your goal is to craft an email that appears to come from the CEO, requesting urgent action from the recipient.

Your Task:

1. What elements would you include to make the email appear legitimate?
2. What psychological triggers (e.g., urgency, authority) would increase the likelihood of success?
3. What red flags should employees look for to identify this as a phishing attempt?
4. What technical and policy-based defenses should the organization have in place to prevent such attacks?

ANSWERS

Elements of a convincing phishing email:

- Spoofed sender address that looks similar to the CEO's real email (e.g., ceo@company.com vs. ceo@c0mpany.com).
- Urgency and authority, such as: "I need this report in the next 30 minutes—please reply ASAP."
- Minimal details that could trigger suspicion, keeping the request simple (e.g., "Confirm your credentials to access the executive dashboard.").

Psychological triggers used to increase success:

- Authority bias: Employees feel obligated to comply with executive requests.
- Urgency: Creates pressure to act without verifying legitimacy.
- Fear of consequences: Employees may fear missing a critical deadline.

Red flags that indicate a phishing attempt:

- Slight email domain mismatches or spelling errors.
- Unusual attachment types or links leading to unrecognized sites.
- Requests for sensitive information that wouldn't normally be asked via email.

Technical and policy-based defenses against phishing:

- Email filtering solutions to block spoofed or suspicious emails.
- Security awareness training to help employees recognize phishing tactics.
- Out-of-band verification (e.g., calling the sender before taking action).

CONCLUSION

Overall, further research into ethical human hacking is required to better navigate the current information security climate and what is to come - as humans remain the weakest link.

However, what we do know is that following strictly agreed upon guidelines within a team, a clearly defined scope, getting proper authorization, and leaving the individuals and organizations better off (not worse!) after the engagement is a great start to ethical human hacking engagements.

After all, it is human manipulation.

DISCUSSION QUESTIONS

What are the ethical boundaries of physical red teaming, and how can organizations ensure that assessments remain ethical and legal?

Based on the case studies presented, what were the key security failures that allowed breaches to occur, and how could they have been prevented?

How can an organization balance security awareness training with real-world red team exercises to improve overall security posture?

REFERENCES

1. Albladi, S. M., & Weir, G. R. S. (2020). Predicting individuals' vulnerability to social engineering in social networks. Cybersecurity (Singapore), 3(1), 1–19. https://doi.org/10.1186/s42400-020-00047-5
2. Daniel C, Sipper JR (2023) Hacking Humans: The Art of Exploiting Psychology in the Digital Age. J Res Dev. 10: 224. https://www.longdom.org/open-access/hacking-humans-the-art-of-exploiting-psychology-in-the-digital-age-101458.html
3. Hatfield, J. M. (2019). Virtuous human hacking: The ethics of social engineering in penetration-testing. Computers & Security, 83, 354–366. https://doi.org/10.1016/j.cose.2019.02.012
4. Matherly, M. C. (2020, October). The Group Psychology of Red Teaming. Army University Press. https://www.armyupress.army.mil/Journals/Journal-of-Military-Learning/Journal- of-Military-Learning-Archives/October-2020/Matherly-Red-Teaming/
5. Montañez, R., Golob, E., & Xu, S. (2020). Human Cognition Through the Lens of Social Engineering Cyberattacks. Frontiers in psychology, 11, 1755. https://doi.org/10.3389/fpsyg.2020.01755
6. Mouton, F., Malan, M. M., Kimppa, K. K., & Venter, H. S. (2015). Necessity for ethics in social engineering research. Computers & Security, 55, 114–127. https://doi.org/10.1016/j.cose.2015.09.001
7. Okenyi, P. O., & Owens, T. J. (2007). On the Anatomy of Human Hacking. Information Systems Security, 16(6), 302–314. https://doi.org/10.1080/10658980701747237
8. Vimal, Mani. (2022). Strengthening cybersecurity with red team engagements. ISACA Journal, 1. https://www.isaca.org/resources/isaca-journal/issues/2022/volume-1/strengthening-cybersecurity-with-red-team-engagements
9. Wilcox, H., & Bhattacharya, M. (2020). A human dimension of hacking: Social engineering through Social Media. IOP Conference Series: Materials Science and Engineering, 790(1), 012040. https: doi.org/10.1088/1757-899x/790/1/012040