Lecture 4: Social Engineering and the Human Factor

Theme: Malware and Social Engineering **Technical University of Moldova**
Lecturer: Maxim Masiutin, Associate Professor

Introduction

Welcome back. In our previous lecture, we explored the technical world of malware. Today, we shift focus to what security professionals often call the weakest link in any security system: the human element.

Social engineering is the art of manipulating people into performing actions or divulging confidential information. While technical attacks exploit software vulnerabilities, social engineering exploits human psychology. And as it turns out, human psychology has many vulnerabilities that cannot be patched with software updates.

The statistics are sobering. According to the 2025 Verizon Data Breach Investigations Report, the human element is involved in around 60 percent of all breaches. The FBI's Internet Crime Complaint Center reported 859,532 complaints in 2024 with total losses of 16.6 billion dollars, up 33 percent year over year.

Perhaps most alarming: according to research by KnowBe4, a security awareness training provider, AI-generated phishing emails now have a 54 percent click-through rate, vastly outperforming the 12 percent click rate of human-crafted phishing attempts. The combination of artificial intelligence and social engineering is creating threats of unprecedented sophistication.

Let us explore how these attacks work and how we can defend against them.

Part 1: The Psychology of Social Engineering

Social engineering succeeds because it exploits fundamental aspects of human psychology. Understanding these principles helps us recognize and resist manipulation.

Principles of Influence

Robert Cialdini, a social psychology researcher, identified six principles of influence that social engineers routinely exploit:

1. Reciprocity

Humans feel obligated to return favors. If someone does something for us, we feel compelled to do something in return.

Attack example: An attacker posing as IT support "helps" an employee with a fake problem, then asks for their password as part of the "solution." The employee feels obligated to comply after receiving "help."

2. Commitment and Consistency

Once we commit to something, we tend to follow through to appear consistent.

Attack example: An attacker gets a small commitment first ("Can you just confirm you work in the accounting department?"), then escalates ("Since you're in accounting, you can approve this urgent wire transfer, right?").

3. Social Proof

We look to others' actions to determine correct behavior, especially in uncertain situations.

Attack example: "Everyone else in your department has already updated their passwords using this link. You're the last one."

4. Authority

We tend to comply with requests from authority figures or those who appear to have authority.

Attack example: An email appearing to come from the CEO demands immediate action. Employees hesitate to question authority, especially under time pressure.

5. Liking

We are more likely to comply with requests from people we like or find attractive.

Attack example: Attackers build rapport through friendly conversation before making requests. They may research targets to find common interests.

6. Scarcity

We value things more when they are scarce or available for limited time.

Attack example: "This security update is only available for the next 30 minutes before your account is suspended." Time pressure prevents careful consideration.

Additional Psychological Factors

Fear and Urgency

Fear triggers our fight-or-flight response, which reduces analytical thinking. Urgent requests exploit this by demanding immediate action.

"Your account has been compromised! Click here immediately to secure it or lose all your data!"

Curiosity

Humans are naturally curious. We want to see what is behind the curtain.

"You won't believe what your coworker said about you in this video..."

Greed

The promise of something for nothing appeals to our desire for gain.

"You've won a free iPhone! Just enter your details to claim it."

Helpfulness

Most people want to be helpful. Attackers exploit this by asking for "just a small favor."

"I'm locked out of my account and really need to get this report to the boss. Can you just log in for me quickly?"

Part 2: Phishing Attacks

Phishing is the most common form of social engineering, using deceptive messages to trick recipients into revealing information or taking harmful actions.

Email Phishing

Traditional phishing uses mass emails impersonating legitimate organizations:

- Banks requesting account verification
- Service providers warning of account suspension
- Delivery companies with shipping notifications
- IT departments requiring password resets

The APWG (Anti-Phishing Working Group), an international coalition of industry and law enforcement, tracked over one million phishing attacks in Q1 2025, the highest quarterly total since late 2023.

Characteristics of phishing emails:

- Sense of urgency
- Generic greetings ("Dear Customer")
- Suspicious sender addresses

- Poor grammar or spelling (though AI is eliminating this indicator)
- Mismatched or suspicious URLs
- Requests for sensitive information

Spear Phishing

Spear phishing targets specific individuals using personalized information. Unlike mass phishing, spear phishing researches targets carefully.

Attackers gather information from:

- Social media profiles
- Company websites
- LinkedIn connections
- Public records
- Previous data breaches

A spear phishing email might reference:

- Your actual job title and department
- Recent company announcements
- Names of your colleagues or managers
- Projects you're working on
- Conferences you attended

This personalization dramatically increases success rates.

Whaling

Whaling targets high-value individuals, typically executives or those with financial authority. These attacks are highly researched and carefully crafted.

Common whaling scenarios:

- Fake urgent requests from the CEO
- Fraudulent legal notices
- Fake merger and acquisition documents
- Board meeting materials

Executives are targeted because they:

- Have authority to approve transactions
- Access sensitive information
- May have less technical security awareness

- Are often too busy for careful scrutiny

Business Email Compromise (BEC)

BEC attacks impersonate business associates to manipulate employees into transferring funds or revealing information.

Common BEC scenarios:

- **CEO fraud**: Fake executive requests wire transfer
- **Vendor impersonation**: Fake invoice with changed payment details
- **Attorney impersonation**: Urgent legal matter requiring payment
- **Account compromise**: Attacker uses actual compromised email account

According to the FBI Internet Crime Complaint Center (IC3), BEC has caused over 55 billion dollars in losses globally since 2013. A single successful attack can transfer millions before detection.

Part 3: Voice-Based Attacks (Vishing)

Vishing (voice phishing) uses phone calls to manipulate victims. With the rise of AI voice cloning, vishing has become dramatically more dangerous.

Traditional Vishing

Classic vishing attacks include:

- Fake tech support calls ("We detected a virus on your computer")
- IRS or tax authority impersonation
- Bank fraud department impersonation
- Prize or lottery scams

Vishing is effective because:

- Real-time conversation creates pressure
- Caller ID can be spoofed
- Voice conveys authority and urgency
- Harder to verify than email

AI-Powered Vishing

The landscape changed dramatically with AI voice cloning. According to CrowdStrike, a cybersecurity company, vishing attacks surged 442 percent in late 2024, driven by AI deepfakes.

How AI vishing works:

1. Attackers collect voice samples from public sources (videos, podcasts, earnings calls)
2. AI models clone the voice with high fidelity
3. Attackers call employees impersonating executives
4. Real-time voice generation makes conversations convincing

Real-world case: In February 2024, a finance worker at Arup, a multinational engineering consultancy, transferred 25 million dollars to fraudsters after attending what appeared to be a legitimate video conference call. Every face on the screen was real, every voice matched perfectly, but all were AI-generated deepfakes created by attackers who cloned executive voices and faces using publicly available footage.

According to the 2025 State of the Phish report by Proofpoint, an email security company, vishing now affects 30 percent of organizations, with deepfake-powered executive impersonation up 15 percent.

Part 4: Text-Based Attacks (Smishing)

Smishing (SMS phishing) uses text messages to deliver attacks. As people become more suspicious of email, attackers shift to SMS.

Why Smishing Works

- Higher open rates than email (90%+ open rate)
- Smaller screens make URL inspection difficult
- Less spam filtering on mobile
- People trust text messages more than email
- Sense of immediacy

Common Smishing Attacks

- Bank alerts about suspicious transactions
- Delivery notifications with tracking links

- Prize or reward notifications
- Account verification requests
- Two-factor authentication code theft

Smishing Techniques

Package delivery scams: "Your package could not be delivered. Reschedule here: [malicious link]"

Bank alerts: "Unusual activity detected on your account. Verify immediately: [link]"

MFA (Multi-Factor Authentication) bypass: "Your verification code is 123456. If you didn't request this, call [attacker's number]"

Part 5: QR Code Phishing (Quishing)

Quishing uses QR codes to deliver phishing attacks. According to Hoxhunt, a human risk management platform, QR code phishing rose over 500 percent as corporate "QR normalcy" spread.

Why Quishing Works

- QR codes hide the destination URL
- Users cannot preview the link before scanning
- Smartphones may not display full URLs
- Corporate adoption of QR codes has normalized scanning
- Bypasses email security filters

Quishing Scenarios

Physical quishing:

- Fake parking tickets with QR codes for "payment"
- Malicious QR codes placed over legitimate ones
- QR codes in fake official notices

Digital quishing:

- QR codes in phishing emails (bypass text-based filters)
- Fake two-factor authentication expiration notices
- Corporate document QR codes

Real examples:

- Fake parking tickets in San Francisco directing to credential theft sites
- Washington University QR code scam stealing credentials
- Fake Microsoft 2FA expiring emails with QR codes

Defending Against Quishing

- Use QR scanners that preview URLs
- Be suspicious of unexpected QR codes
- Verify physical QR codes haven't been tampered with
- Type URLs manually when possible

Part 6: Multi-Channel Attacks

Modern attackers no longer rely on single-channel attacks. Instead, they create elaborate schemes using multiple touchpoints to build credibility.

How Multi-Channel Attacks Work

1. **Initial contact**: Email confirms an "urgent security update"
2. **Reinforcement**: SMS reminder about the same issue
3. **Social validation**: LinkedIn message from fake IT colleague
4. **Escalation**: Teams/Slack message expressing concern
5. **Final push**: Deepfake voice call from "manager" demanding action

After multiple touchpoints, victims are far more likely to trust the request because it has been "confirmed" through various channels.

Characteristics

- Consistent narrative across channels
- Each touchpoint reinforces legitimacy
- Time pressure builds throughout
- Multiple apparent sources agree on urgency
- Exploits our tendency to trust corroborated information

Part 7: AI in Social Engineering

Artificial intelligence has transformed social engineering from a labor-intensive craft to an industrialized threat.

AI-Generated Phishing

The 2025 Egress Phishing Threat Trends Report, published by Egress, an email security vendor, indicates that 82.6 percent of phishing emails analyzed between September 2024 and February 2025 contained AI-generated content.

According to Abnormal Security, an AI-based email security provider, by October 2025 AI-generated phishing became the top enterprise email threat, surpassing ransomware, insider risk, and traditional social engineering combined.

Why AI phishing is effective:

- Perfect grammar and spelling
- Contextually appropriate content
- Personalization at scale
- Mimics writing style of impersonated individuals
- Generates unique variations to evade filters

The numbers: According to KnowBe4 research, AI-generated phishing emails have a 54 percent click-through rate compared to 12 percent for human-crafted attempts.

Deepfake Technology

Audio deepfakes clone voices from samples:

- Earnings calls, interviews, and podcasts provide source material
- Real-time generation enables live conversations
- Quality continues improving rapidly

Video deepfakes create convincing fake video:

- Virtual meeting impersonation
- Fake video messages from executives
- Combined with voice cloning for complete impersonation

According to the ENISA (European Union Agency for Cybersecurity) Threat Landscape 2024 report, the dark web trade in deepfake tools rose 223 percent between Q1 2023 and Q1 2024.

GenAI in Attacker Workflows

Throughout 2025, Generative AI (GenAI) became central to attacker operations:

- More convincing AI-generated phishing
- Faster automated reconnaissance
- Fraud schemes enhanced by synthetic media
- Lower barrier to entry for attackers

According to the 2025 SoSafe Human Risk Review, published by SoSafe, a security awareness platform provider, more than 86 percent of organizations have already encountered at least one AI-related phishing or social engineering incident.

Part 8: Physical Social Engineering

Social engineering extends beyond digital channels to physical attacks.

Pretexting

Pretexting creates a fabricated scenario to engage victims. The attacker assumes a false identity with a plausible backstory.

Examples:

- Posing as IT support to gain building access
- Impersonating delivery personnel
- Pretending to be a new employee
- Acting as an auditor or inspector

Effective pretexts include:

- Detailed background story
- Appropriate appearance and props
- Confidence and authority
- Knowledge of organizational details

Tailgating and Piggybacking

Tailgating is following an authorized person through a secure entrance.

Techniques:

- Carrying boxes ("hands full" excuse)

- Wearing similar attire to employees
- Arriving during busy entry times
- Appearing to search for access card

Piggybacking is similar but with the authorized person's awareness (they hold the door out of politeness).

Baiting

Baiting leaves infected media for victims to find and use.

Common baits:

- USB drives labeled "Confidential" or "Salary Information"
- CDs or DVDs left in common areas
- Devices labeled with company logos

Curiosity drives victims to plug in found devices, triggering malware installation.

Dumpster Diving

Searching through trash for valuable information:

- Discarded documents
- Old hardware with data
- Sticky notes with passwords
- Organization charts and directories

Proper document destruction policies counter this threat.

Part 9: Defense Strategies

Defending against social engineering requires a combination of technical controls, policies, and human awareness.

Security Awareness Training

The most effective defense is educated users who recognize and resist attacks.

Training should cover:

- Recognition of phishing indicators
- Verification procedures for unusual requests
- Reporting suspicious activities

- Safe handling of unexpected contacts
- Real-world examples and scenarios

Training approaches:

- Regular awareness sessions
- Simulated phishing campaigns
- Gamification and engagement
- Immediate feedback on mistakes
- Metrics tracking improvement

Verification Procedures

Establish procedures for verifying unusual requests:

Out-of-band verification:

- Call back on known numbers (not numbers provided in suspicious messages)
- Use alternate communication channels
- Verify through established contacts

Multi-person authorization:

- Large transactions require multiple approvals
- Sensitive actions need supervisor verification
- Wire transfers have mandatory confirmation procedures

Technical Controls

While social engineering targets humans, technical controls help:

Email security:

- Spam and phishing filters
- DMARC (Domain-based Message Authentication, Reporting and Conformance), DKIM (DomainKeys Identified Mail), SPF (Sender Policy Framework) authentication
- External email warnings
- Link protection and sandboxing

Phone security:

- Caller verification procedures
- Training on spoofed caller ID
- Recording of sensitive calls

Access controls:

- Multi-factor authentication
- Visitor management systems
- Badge requirements for entry

Creating a Security Culture

Long-term defense requires cultural change:

- Make security everyone's responsibility
- Remove stigma from reporting mistakes
- Celebrate security-conscious behavior
- Leadership demonstrates commitment
- Regular communication about threats

Part 10: Practical Exercises

Let me walk through some scenarios to apply what we have learned.

Scenario 1: Suspicious Email Analysis

You receive an email:

```
From: IT-Security@yourcompany.com.suspicious-domain.com
Subject: URGENT: Password Expiration in 24 Hours

Dear Employee,

Your corporate password will expire in 24 hours. To prevent account
lockout, please update your credentials immediately using the secure
link below:

[Update Password Now]

Failure to act will result in loss of access to all company systems.

IT Security Team
```

Red flags:

- Domain mismatch (suspicious-domain.com)
- Generic greeting ("Dear Employee")

- Urgency pressure
- Threat of consequence
- Suspicious link

Correct response: Do not click the link. Report to IT security. Verify through official IT channels if concerned about password expiration.

Scenario 2: Phone Call Verification

Your phone rings. The caller claims to be from your bank's fraud department: "We've detected suspicious activity on your account. To verify your identity, please confirm your account number and the last four digits of your Social Security Number."

Red flags:

- Unsolicited call
- Requesting sensitive information
- Creating urgency about "suspicious activity"

Correct response: Hang up. Call your bank using the number on your card or statement. Never provide sensitive information to incoming callers.

Scenario 3: Executive Wire Transfer Request

You receive an email from the CEO asking you to urgently wire funds to a new vendor: "I need you to process a wire transfer of $50,000 to this new vendor immediately. I'm in a meeting and can't be disturbed. This is confidential - don't discuss with anyone else."

Red flags:

- Urgency
- Request for secrecy
- Unusual request from executive
- Cannot be verified through normal channels

Correct response: Follow established verification procedures regardless of apparent sender. Contact the CEO through known channels. Never bypass verification for "urgent" requests.

Conclusion

Today we explored social engineering and the human factor:

1. **Psychology of manipulation**: Cialdini's principles and emotional triggers
2. **Phishing varieties**: Email phishing, spear phishing, whaling, BEC
3. **Vishing**: Voice-based attacks amplified by AI voice cloning
4. **Smishing**: SMS-based attacks with high open rates
5. **Quishing**: QR code phishing exploiting normalized QR usage
6. **Multi-channel attacks**: Coordinated attacks across multiple touchpoints
7. **AI transformation**: GenAI making attacks more convincing and scalable
8. **Physical attacks**: Pretexting, tailgating, baiting
9. **Defense strategies**: Training, verification procedures, technical controls, culture

The human element will always be part of security. Our goal is not to eliminate human involvement but to empower people to recognize and resist manipulation.

In our next lecture, we will examine access control models and identity management.

Discussion Questions

1. How can organizations balance security verification procedures with operational efficiency?
2. What ethical considerations arise when conducting simulated phishing campaigns?
3. As AI makes deepfakes increasingly convincing, what new verification methods will be needed?

Thank you for your attention. See you in our next session.

Review Questions

1. Explain Cialdini's six principles of influence and how each can be exploited in social engineering attacks.
2. Compare and contrast phishing, spear phishing, and whaling. What makes each effective?
3. What is Business Email Compromise (BEC) and why has it become one of the most financially damaging attack types?
4. How are AI and deepfake technologies changing the social engineering threat landscape?
5. Describe QR code phishing (quishing). Why is it particularly effective in today's environment?

6. What makes multi-channel social engineering attacks more dangerous than single-channel attacks?
7. What elements should an effective security awareness training program include?
8. Describe three physical social engineering techniques and the countermeasures for each.

Key Terms

- **Adversarial AI**: Attacks that target machine learning-based security systems to evade detection or cause misclassification
- **Baiting**: Leaving infected media or offering enticing items to lure victims
- **BEC**: Business Email Compromise, an attack impersonating executives or partners to authorize fraudulent transactions
- **CEO Fraud**: A form of business email compromise where attackers impersonate senior executives to authorize fraudulent transactions
- **Deepfake**: AI-generated synthetic media that convincingly mimics real people
- **Dumpster Diving**: Searching through discarded materials for useful information
- **Phishing**: Fraudulent communication designed to trick recipients into revealing information or taking harmful actions
- **Phishing Simulation**: Controlled exercises that test employee susceptibility to phishing attacks as part of security training
- **Pretexting**: Creating a fabricated scenario to manipulate a target into providing information or access
- **Quishing**: Phishing attacks using QR codes to direct victims to malicious sites
- **Security Awareness Training**: Educational programs designed to teach employees to recognize and respond to security threats
- **Smishing**: SMS-based phishing attacks
- **Social Engineering**: The psychological manipulation of people into performing actions or divulging confidential information
- **Spear Phishing**: Targeted phishing directed at specific individuals or organizations
- **Tailgating**: Following an authorized person through a secure entrance without proper credentials
- **Vishing**: Voice-based phishing attacks, often using phone calls
- **Whaling**: Phishing attacks specifically targeting senior executives

References and Further Reading

- NIST SP 800-50: Building an Information Technology Security Awareness and Training Program
- Anti-Phishing Working Group (APWG) Reports
- FBI Internet Crime Complaint Center (IC3) Annual Reports
- SANS Security Awareness Resources