Risk control

Social Engineering and Deepfakes

In partnership with

GALLAGHER BASSETT
GUIDE. GUARD. GO BEYOND.

Risk Management Partners
rmpartners.co.uk

Social Engineering and Deepfakes

Introduction

Society is changing rapidly, especially with the increased usage of information technology. With that comes opportunity for bad actors to take advantage. One of the key risks in this context is 'social engineering.'

Social engineering is a strategy used by individuals or groups to manipulate and deceive people into revealing sensitive information or performing actions that compromise their security. It relies on psychology and human behaviour, rather than technical know-how[1].

There are two different methods to consider, either:

- Using psychological manipulation to get further access to an IT system such as impersonating an important client via a phone call to lure the target into browsing a malicious website to infect the organisation's workstation
- Using IT technologies to obtain banking credentials via a phishing attack to steal the organisation's money

The rapid increase in the use of IT technologies has accelerated the use of such techniques in cyber-attacks.

Common Techniques

A variety of tactics[2] are used, these include:

- **Pretexting** which involves using a false justification for asking the victim to do something by impersonating IT Support and asking for a password.
- **Baiting** encourages the victim by using a lure such as a USB flash drive infected with a key logger (a form of malware that keeps track of and records keystrokes as a person types) left on a desk.
- **Quid Pro Quo** involves asking the victim to give a password in return for financial gain.
- **Tailgating** involves a person following someone into a sensitive area, using a device to copy the identity of a Radio Frequency ID pass.
- **Water-Holing** is where the hacker takes advantage of trusted websites people regularly visit.
- **Phishing** involves trying to acquire usernames, passwords, and credit card information by masquerading as a trustworthy organisation via bulk email which tries to avoid an IT system's spam filters.
- **Spear Phishing** is a focussed attack, via email on a particular person with the goal to penetrate the organisation's defences.
- **Honey Traps** use a trick to encourage men to interact with a fictional female online.
- **Scareware / Rogue Security Software** which is a form of malware that encourages the user to pay for the fake or simulated removal of malware.
- **Whaling** where a type of phishing attack exploits the influence of senior executives over lower-level roles, such as CEOs over financial executives or assistants.
- **Pharming** where individuals are redirected to a malicious site that impersonates it by exploiting system vulnerabilities that match domain names with IP addresses.
- **Pretexting** in which the con artist gains a victim's trust, typically by creating a backstory that makes them sound trustworthy. It is often used at an early stage of more complex social engineering attack.
- **Vishing / Voice Phishing** is an attack that uses the phone. Often the person receives a recorded message telling them their bank account have been compromised. The victim is then prompted to enter their details via their phone's keypad, giving them access to their accounts.

Targeting

The target of an attack is often chosen through targeting email addresses or through information available on social media, often due to a significant internet presence.

Psychological Attack Scenarios[3]

- **Fear** is used where a person is threatened with legal action, tricking the person into complying with the instruction, logging into a malicious site to rectify a mistake.
- **Greed** is used as a bait where the email suggests that a person can get something for nothing, via a carefully worded email asking for bank information to transfer funds.
- **Curiosity** can be used where an email includes attachments related to a high-profile media event encouraging the document to be opened and trigger the malware.
- **Helpfulness** can be a tactic where an email asks a person to send someone the password to a financial system so that a 'manager' can make sure everyone is paid.
- **Urgency** is often used with time pressure to encourage action without crosschecking identities.

One recently used tactic involved an invoice being sent from a contractor to a public sector organisation via an email that was intercepted. The invoice amount and bank details were changed. The change in bank account was checked and the payment made, however, afterwards the bank raised it as an exceptional transaction which brought to light the fraud.

Defending the Organisation

A key defence in cyber security[4] is to train staff to learn the psychological triggers and other identifiers.

Encourage staff to be suspicious of unsolicited communications and unknown people. Everyone should also check whether emails come from a verified source by double-checking the senders' name and look out for identifiers such as spelling or grammar errors. Opening suspicious email attachments should be discouraged.

When encouraged as a sense of urgency to send sensitive information, take the time pressure away before providing sensitive information only after the appropriate checks have been made.

Check website security before submitting information, even if it seems legitimate; and pay particular attention to URLs and sites that look genuine, but web addresses are subtly different from the legitimate site they are seeking to imitate.

An organisation should already maintain an established cyber threat strategy with specific measures in place. As humans are the target, make sure to engage with employees to:

- Build awareness and a positive security culture.
- Test the effectiveness of guidance and training.
- Reinforce technological cyber security measures.

Deepfakes

Deepfakes are an emerging trend within a wider aspect of synthetic media which is generated utilising a form of artificial intelligence / machine learning to create believable, realistic videos, pictures, audio, and text of events which never happened. A real problem exists that despite being a fake, the video quality may be good enough that a casual viewer might be convinced it is authentic.

Deepfakes use neural networks to generate realistic visual and auditory elements to create deceptive and misleading content. It can then be used to produce false information, spread misinformation, and deceive targeted groups of unsuspecting people.

According to some research[5] the volume of deepfake online content surged by 300% between 2022 and 2023. This worrying increase may suggest that the trend will continue in the years to come and present even greater difficulties to organisations and societies as the technology becomes cheaper and more easily available.

Terminology

There are various terms currently in use that need some explanation, namely 'deepfake', 'cheapfake' and 'shallowfake'.

'Cheapfakes' also known as "shallowfakes" are audio-visual manipulations created using cheaper, more accessible software. These techniques are less expensive, require less technical skill, and are available on a larger scale. One example is using video editing software to slow down footage making the speech slurred so that the viewer thinks the high-profile person on camera is drunk.

The term "deepfakes" is derived from the fact that the technology involved in creating this style of manipulated content involves the use of deep learning techniques. Deep learning is a subset of machine learning techniques, which are themselves a subset of artificial intelligence (AI).

AI generated text is another type of deepfake that is a growing challenge. The primary threat of deepfakes is that they are used to simulate or alter a specific individual's digital representation. However, this threat is not restricted to deepfakes alone but incorporates the entire field of synthetic media and their use in spreading disinformation.

Deepfakes videos or images often feature people who have been digitally altered, whether it is their voice, face, or body, so that they appear to be "saying" something else or are someone else entirely.

Recent developments, however, have now made deepfake technology far less resource intensive, and so more accessible to the general population.

The rise of deepfakes in part is because people are more likely to believe what they see. Synthetic media such as manipulated photos and audio and video deepfakes, can be especially convincing and dangerously effective.

The Deepfake Threat

Deepfakes continue to pose a threat for individuals and industries, including potential large-scale impacts to nations, governments, businesses, and society, such as social media disinformation campaigns operated at scale by well-funded nation state actors. Experts from different disciplines[6] whose research interests involve deepfakes tend to agree that the technology is rapidly advancing, and the cost of producing top-quality deepfake content is reducing. As a result, an emerging threat is developing where the attacks will become easier and more successful, and the efforts to counter and mitigate these threats will need governments, industries, and society to co-ordinate their efforts to combat this.

Deepfake Usage

Deepfake technology is being used for a wide variety of purposes, including:

- Scams and hoaxes.
- Election manipulation.
- Social engineering.
- Automated disinformation attacks.
- Identify theft and fraud.
- Celebrity pornography.

The scam or hoax is typically a false video of a senior official admitting to criminal activity, such as financial crimes, or making false claims about an organisation's activity. The time and cost to disprove such accusations could have a major impact on the organisation's brand or public reputation.

Deepfakes are increasingly being used in major attempts at extortion and fraud. A recent social engineering incident tricked a Chief Executive Officer into believing he was speaking to the Company Group Chief Executive[7]. The deepfake voice impersonated the Chief Executive and convinced the individual to transfer a substantial sum of money to a third-party bank account.

Automated disinformation attacks can be used to spread conspiracy theories and incorrect theories about political and social issues which can be difficult to refute.

Threat Assessment

A simple framework for an organisation to prioritise deepfake risk is to consider three aspects: severity (the level of harm caused by the deepfake), scale (how widespread the harm is) and resilience (the ability of the target to withstand the impact).

- **Individual:** The impact of a deepfake on an individual is potentially severe and long-lasting, and many individuals may not have the resilience or resources to 'bounce back' from an attack, particularly given the difficulty in having content removed from the internet.
- **Organisational:** The organisational impact of deepfakes can vary widely. Certain organisations could be badly damaged by a successful deepfake attack, but many will already have resources and processes in place that could be adapted to respond to threats involving deepfakes.
- **Societal:** The societal impact of deepfakes might cause a gradual erosion of trust without having a sufficiently direct effect on enough people or organisations to trigger a response. However, a growing number of countries have seen increasing political polarisation and volatility caused by digital misinformation.

Disinformation and Misinformation

Disinformation and misinformation are emerging as significant cyber threats to businesses worldwide[8]. Fake news stories, hoaxes and propaganda have been used to encourage public cynicism and distrust.

- Disinformation is verifiably false information created and disseminated with intent to deceive, whereas misinformation is false information shared without malicious motive.
- Risks to businesses targeted by disinformation or misinformation include financial losses; erosion of market confidence and customer, client, and employee trust; and slowdowns in rolling out new technology or products.
- The best defence is to develop a strategy for strengthening company reputation and addressing falsehoods as part of the overall response plan.

Identifying Deepfakes

Deepfakes can be spotted by recognising unusual activity or unnatural movement, including:

- Unnatural or lack of eye movement which reacts to the person they are speaking with.
- A lack of regular blinking or no blinking at all.
- Unnatural facial expressions and facial morphing.
- Unnatural body shape as the focus is more on the face.
- Unnatural hair.
- Abnormal skin colours.
- Awkward head and body positioning.
- Inconsistent facial positions.
- Odd lighting, shifts in lighting between frames, misplaced shadows, or discoloration of the image.
- Bad lip-syncing which fails to match the spoken words.

Taking Positive Action

Good basic security procedures are effective at dealing with deepfake threats.

For instance, having automatic checks built into any process for disbursing funds would have stopped many deepfake and similar frauds.

Organisations can also:

- Ensure employees know about how deepfaking works and the challenges it can pose.
- Educate employees on how to spot a deepfake.
- Make sure the organisation is media literate and uses reliable news sources.
- Maintain good protocols - "trust but verify". A sceptical attitude to voicemail and videos can help avoid many traps.

Basic cyber-security best practice will play a vital role when it comes to minimizing the risk:

- Regular backups protect data against ransomware and gives the ability to restore damaged data.
- Using different, strong passwords for different accounts means just because one network or service has been broken into does not mean any others have been compromised.
- Use a good security package to protect networks, laptops, and smartphones against cyber threats. A package that provides anti-virus software, a Virtual Private Network to stop Wi-Fi connections being hacked, and protection for webcams is advisable.

References

1. 'Cybersecurity: social engineering'. European Council. Available here: https://www.consilium.europa.eu/en/policies/cybersecurity-social-engineering/
2. 'What is social Engineering – Common Methods' Available here: https://www.knowbe4.com/what-is-social-engineering/
3. 'Types of social engineering attacks' Available here: https://terranovasecurity.com/blog/examples-of-social-engineering-attacks/
4. 'Examples & Prevention Tips.' Available here: https://www.itgovernance.co.uk/social-engineering-attacks
5. 'Sumsub Research: UK Deepfake Incidents Surge 300% from 2022 to 2023', Available at: https://www.prnewswire.co.uk/news-releases/sumsub-research-uk-deepfake-incidents-surge-300-from-2022-to-2023-301999013.html
6. 'UK energy boss conned out of £200,000 in 'deep fake' fraud.' Available here: https://www.cityam.com/uk-energy-boss-conned-out-of-200000-in-deep-fake-fraud/
7. 'The increasing threat of deepfake identities.' The Department for Homeland Security. Available here: https://www.dhs.gov
8. 'The threat mis and disinformation pose to business.' Available here: Bank of America

Further information

For access to further RMP Resources you may find helpful in reducing your organisation's cost of risk, please access the RMP Resources or RMP Articles pages on our website. To join the debate follow us on our LinkedIn page.

Get in touch

For more information, please contact your broker, RMP risk control consultant or account director.

contact@rmpartners.co.uk

Risk Management Partners

The Walbrook Building
25 Walbrook
London EC4N 8AW

020 7204 1800
rmpartners.co.uk

FP255-2025