BixeLab

30 April 2024

I.D. RISK ALERTS

Open Source Edition

Identity and Biometric Vulnerabilities | Threats and Risks | Mitigations

Welcome

The I.D. Risk Alerts newsletter (open source edition) is a quarterly newsletter with analysis on recent ID fraud, biometric threats and identity vulnerabilities that are already known in the public domain. A comprehensive bi-annual report will also be available to select BixeLab subscribers.

Australia: New Legislation and Payment Systems in Focus

Australia has new digital identity regulation and payment systems introduced in Nov 2023. As Australia spearheads major digital advancements, BixeLab CEO, Dr Ted Dunstone, offered crucial insights in a Senate hearing regarding the Digital ID Bill. The amended Bill has been passed by the Senate in March 24 and will now progress to the House of Representatives. "Digital ID makes it safer and easier for Australians to prove who they are online," The Minister for Finance, Senator Katy Gallagher said. Meanwhile, New South Wales advances towards a real-time digital payments system, with increased usage of the Customer Payment Platform (CPP) across key government sectors. It is set for completion within the next 12 to 18 months.

Open Source Face Image Quality (OFIQ): Enhancing Face Biometric System

The Open Source Face Image Quality (OFIQ) library, an implementation reference for ISO/IEC 29794-5 (a standard for face image data quality) is now accessible on GitHub. OFIQ is applicable to enrolment for biometric passports and EU's EES biometric kiosks. Simultaneously, the National Institute of Standards and Technology (NIST) has adapted its Face Analysis Technology Evaluation (FATE) Part 11: Face Image Quality Vector Assessment to align more closely with ISO 29794-5:2024. The British Standards Institution (BSI) has implemented OFIQ. If you are interested in trying OFIQ, we recommended to use Biometric Quality Assessment Tool (BQAT) - an open-source framework for generating quality statistics for biometric samples

Corporate Fraud via AI Deepfakes and Voice Cloning

Modality: **Face, Voice** Attack Instrument: **AI deepfake and voice cloning** Location: **HK, UK** Date of report: **February 2024**

A multinational corporation based in Hong Kong had a massive financial loss of approximately **$25 million USD**. The criminals involved used AI deepfakes to impersonate the enterprise's CFO and colleagues. A series of fifteen transactions occurred within the span of a week, after multiple calls between the employee in question and the deepfake-using criminals. Meanwhile, a similar weakness was exposed by The Times, highlighting the risks associated with AI voice cloning. A journalist posing as an attacker was able in 15 minutes to successfully clone a voice and then undertake transactions worth £250.

- **Potential Targets**: employees of corporations, especially those in finance or decision-making roles, are at significant risk. Encryption-based communication, online banking platforms, and digital workplaces utilising AI technology (especially voice and face prone to deepfakes) are also potential targets.
- **Mitigation Strategy**: expert training to increase scam awareness among stakeholders, implement verification procedures for major transactions, improve detection technologies through testing, enact regular audits, and strengthen backup and authentication measures.
- **Criticality: Extremely High** - the use of AI deepfakes and voice cloning in corporate fraud represents an extremely high security threat, due to their potential for substantial financial loss as well as and the increasing sophistication and accessibility of these technologies with the evolving cyber threat landscape.

Embracing Digital Transformation: Spain's ID Plan and EU's Digital Identity Wallets

Spain is modernising its identification system with the national police launching a comprehensive digital ID plan, featuring upgrades such as improved data verification, biometric enrolment, and introducing 'miDNI', a digital counterpart of the physical ID. Similarly, the European Union's digital identity wallet initiative is just one approval away from providing residents with greater data control and security. This program, backed by legislative support, offers free qualified electronic signatures and wallet interactions for EU wallet users.

Tech Giants and Adobe Pave the Way for Ensuring Authenticity of Digital Content

The Coalition for Content Provenance and Authenticity (C2PA) and major camera manufacturers have joined forces to counter the spread of AI-generated and manipulated content. Tech giants and camera makers like Nikon, Sony, Canon, Adobe, and Microsoft are implementing the Content Authenticity Initiative's (CAI) C2PA digital signature system or Content Credentials in 2024. This measure is developed to verify the origin and alteration of digital images, in response to the growth of manipulated and AI-generated content. While the CAI's system, combining metadata, watermarks, and content fingerprinting, offers a comprehensive solution for content provenance in images, videos, as well as audios, its wide-scale application requires significant educational efforts to empower users to distinguish between authentic and manipulated content.

AI and Biometrics Dominating Anti-Fraud Strategies: 2024 Report

Fraud prevention professionals are increasingly leveraging generative AI and machine learning (ML), according to the 2024 Anti-Fraud Technology Benchmarking Report by the Association of Certified Fraud Examiners (ACFE) and SAS. The report highlights that 83% of these professionals plan to integrate generative AI into their strategies over the next two years, despite current adoption standing at only 18%. The use of biometrics in fraud prevention has also risen by 14% since 2019, and adaptation of robotics has grown from 9% in 2019 to 20% recently.

National Election Commission Battles Deepfake Threats

Modality: **Face Video** Attack Instrument: **AI Deepfakes**
Location: **South Korea, U.S.** Date of report: **February 2024**

The National Election Commission (NEC) of South Korea detected 129 illegal deepfake posts related to the April general election. The deepfakes manipulated videos of candidates, distorting or altering their speeches. These deepfakes, circulated mainly on social media, misleadingly manipulated videos and speeches of opposing candidates. Despite efforts to remove these posts, a lack of advanced monitoring systems and limited manpower mean that many could remain undetected.

Similar fraudulent tactics have been observed internationally. In New Hampshire, United States, political campaigns were recently targeted by an extensive AI-powered robocall campaign. The deceptive scheme employed artificial intelligence to impersonate President Biden's voice, asking people to refrain from voting in what appears to be a clear attempt at voter suppression. Such illicit manoeuvres, deeply vested in the misuse of AI technology, underline the potential for malicious actors to fabricate high-profile voices or videos to drive political agendas and the influence of AI technologies on the integrity and fairness of democratic elections.

In response to these globally emerging problems, tech companies, including Google, have signed the '2024 AI Elections Accord'. Adobe is also taking significant steps to combat the spread of artificial intelligence (AI)-generated videos or deepfakes in the upcoming US presidential election. The aim of this accord is to develop technologies to counter such deepfakes and uphold the integrity of elections. Meanwhile, the US Federal Communications Commission (FCC) has enacted an immediate ban on the use of AI-generated voices in robocalls aimed at misleading voters.

- **Potential Targets:** key entities susceptible to manipulation through fraudulent impersonation tactics, involving artificial intelligence (AI) technologies, include voters, political campaigns, political candidates, electoral bodies and democratic processes.
- **Mitigation Strategy:** various strategies are recommended to safeguard the integrity of these targets. Enhance deepfake detection capabilities for social media platforms, increased cooperation with tech firms, and greater public awareness. Additionally, governments should require social media platforms to use independent AI monitoring tools for deepfake detection to safeguard election integrity. Implement legal prohibitions against AI-generated fraudulent voices in robocalls, enhance cyber infrastructure security for prompt detection and interception of attacks, and drive public awareness campaigns to educate individuals about verifying information from credible sources.
- **Criticality: Extremely High** - recent incidents in both South Korea and the United States demonstrated how advanced technologies like AI can be misused for disruptive political agendas, posing a significant threat to democratic processes and potentially leading to social unrest. It is critical to combat these threats with effective policy frameworks and regularly updated protective measures. Raising public awareness about the issue, along with escalating the continual monitoring and independent testing of defence strategies, is crucial. The integrity of democratic processes hangs in the balance, demanding swift and comprehensive action to mitigate this high-risk situation.

Big-Box Retail Synthetic ID Scam Shakes Security Landscape

Modality: **Identity Theft** Attack Instrument: **Synthetic ID generation** Location: **U.S.** Date of report: **February 2024**

A site called OnlyFake that produces synthetic IDs using AI, creating convincing images for just $15. OnlyFake can reportedly generate up to 20,000 fake IDs each day. The site successfully provided a realistic California driver's license image that bypassed the ID verification system of a cryptocurrency exchange. OnlyFake is not limited to U.S. IDs, as synthetic IDs from Switzerland, Canada, and Austria have also been found.

- **Potential Targets:** cryptocurrency exchanges, online platforms requiring ID verification, banks using eIDV, and individuals accessing these platforms.
- **Mitigation Strategy:** tackling the surge in synthetic ID calls for enhanced verification procedures and implementation of biometric authentication measures. Incorporating automated ID fraud detection, authentication methods, and ambiguous cases adjudication improves robustness. Strengthened encryption and data security for user IDs, alongside regular audits and penetration tests, is paramount. Continued monitoring and governance, following initial implementation of independently tested remote ID verification solutions are essential.
- **Criticality:** **Extremely High** - this case highlights the potential for easy and inexpensive large-scale synthetic ID generation. This pushes a potent threat onto systems that rely on traditional ID verification methodologies, inadequately equipped to confront such advanced fraud techniques.

Financial Fraud in India's Aadhaar-Enabled Payment Systems

Modality: **Identity Fraud** Attack Instrument: **Stolen or Manipulated Biometric Data** Location: **India** Date of report: **January 2024**

India's Aadhaar-Enabled Payment System (AEPS) has dealt with a spate of financial frauds, as noted by the Indian Cyber Crime Coordination Centre (I4C). Millions of bank account holders have been targeted by criminals, using the victims' biodata to gain unauthorised access to their accounts. This has put the security of the AEPS, a biometric-based system designed to promote secure transactions, under intense scrutiny.

- **Potential Targets:** Indian citizens needing government payments, banks, micro-finance institutions, and individual account holders linked to the AEPS.
- **Mitigation Strategy:** enhance data protection and use of biometric identification with robust requirements and testing through government procurement processes like the RFT. Additionally, implement real-time fraud detection, strengthen regular security audits, and boost user education on financial and data safety. Introduce minimum benchmark performance standards for these technologies and carry out continual risk reviews of sampled devices.
- **Criticality:** **High** - the significant number of individuals impacted underscores the urgency for enhanced security measures, and continual improvements, as the integrity of AEPS, crucial to India's financial stability, hinges on persistent safeguarding informed by regular system testing.

UK Identity Fraud Involving Fake Restaurants

Modality: **Identity Theft** Attack Instrument: **False IDs** Location: **UK** Date of report: **February 2024**

Identity fraud cases in the UK have sharply increased. As an example, the fraudulent misuse of stolen IDs linked to famous chefs to register cloned restaurants at the UK's Companies House. High-profile individuals including Heston Blumenthal, Yotam Ottolenghi, and the Ritz have been exploited by these fraudsters to open bank accounts and apply for loans, causing severe reputational and financial damage. More than 750 fake firms have been registered in the last six weeks, many under misspelled names.

- **Potential Targets:** high-profile chefs, restaurants, customers, and the broader business registration, verification systems, and banking institutions.
- **Mitigation Strategy:** increase robustness of identity validation during business and bank account registrations. Foster awareness among restaurateurs and customers about potential threats and scams. Ensure necessary regulatory intervention and oversight.
- **Criticality:** **High** - secure digital identities, stringent administrative checks, and a streamlined verification system. This incident emphasises the necessity of data cross-checking with different datasets and thorough ID validation processes.

Pervasive Cyber Attacks Against Tech Giants

Modality: **Trojan Attack on Biometric Systems, Spyware, Unauthorised Access** Attack Instrument: **Cyber Espionage** Location: **International** Date of report: **April 2024**

A strategic and comprehensive cyber campaign, orchestrated by a state-sponsored group has been unleashing a series of attacks on high-profile tech companies, with Apple, Microsoft, and leading biometric systems being their primary targets. Since Nov 23, the group known as Midnight Blizzard managed to lurk undetected in Microsoft's systems for three months, utilising a brute-force method to gain unauthorised access to sensitive source codes and high-ranking executives' emails. Meanwhile, Apple issued an urgent warning to its iPhone users across 92 countries about a novel spyware attack in April 2024. Further compounding the cyber onslaught was the deployment of the "GoldPickaxe" trojan that exploits biometric systems.

- **Potential Targets:** high-profile tech companies, corporate email accounts, source code repositories, biometric systems, and individual users are the primary targets.
- **Mitigation Strategy:** reinforce password and biometric security, encourage multi-factor authentication, ensure secure communication protocols, maintain robust system monitoring for swift detection and isolation of threats, governance should stipulate independent testing of presentation attack detection (PAD) and authentication methods to ensure their effectiveness.
- **Criticality:** **High** -the risk posed by pervasive cyber attacks, including Trojan attacks, is immense and can potentially cripple even the most sophisticated systems worldwide, affecting not just large organisations but also individual mobile users on a global scale.