ANALYSIS OF DEEPFAKES PROBLEM FOR BANKS AND FINANCIAL INSTITUTIONS

Abstract. This research explores gross losses from payment card fraud worldwide and considers the main modern fraud methods. The issue of payment fraud will be exacerbated by the digitalization of economic relations, in particular the introduction by banks of the concept of "Bank-as-a-Service", which will increase the burden on payment services.

The goal of this research is to highlight and classify main vectors of the deepfakes fraudulent attacks for payment systems in contemporary economy.

The methodological basis of the research is general and specific scientific techniques and methods. In particular, a systematic approach, a method of analysis and synthesis, taxonomy, financial technologies, anti-fraud methods.

The results of our study confirm that new ways of payment fraud are constantly emerging. The development of anti-fraud methods takes a lot of resources, but it allows us to maintain the credibility of banks and payment organizations and reducing losses from fraud.

Keywords: Fraud, deepfakes, payment, cards, anti-fraud, synthetic identities, biometric.

Problem statement. As a result of the global pandemic in 2020, many companies had to remotely set up processes to comply with government regulations and continue to operate.

An attacker can easily intercept the video stream from the phone's camera and use deepfake technology for malicious purposes. However, there is no way to bypass
the Face ID facial recognition system in this way. The Apple ID system uses depth sensors and verifies identity not only based on appearance, but also on the physical shape of the face.

The deepfakes problem firstly was described in 2018 as a fakes technology that allows to make a hyper-realistic video, audio and photo fakes without (or with a little) trace of manipulations.

With the help of a deepfake, a fraudster can impersonate another person in a bank. Deepfake technology allows anyone to bypass the facial recognition system. Therefore deepfakes are actively used by criminals in the financial environment to create synthetic identities and forge documents. Such attacks will become more and more realistic as the technology improves in the coming years. Deepfake technology can also be used to attack a business, creating compromising video or audio deepfakes, allegedly recorded by a company executive.

While it is becoming increasingly difficult for fraudsters to forge real passports due to modern security measures, digital facial images represent a new approach to document fraud. For example, technology can combine or transform the face of a person who owns a passport and the face of a person who wants to obtain a passport illegally, increasing the chances of successfully passing an identity check.

According to Europol's forecast, criminals will begin to use deepfakes even more actively in the future and are already using them right now. Financial institutions and law enforcement will need to significantly improve skills and technology to keep up with criminals' use of deepfakes [1].

Thus, the deepfakes problem is extremely dangerous for the development of financial technologies. However, it is rather poorly studied, since it appeared just recently. This determines the actuality of this study, which is aimed at researching and classifying the main vectors of deepfakes attacks.

Analysis of the resent research and publications. The problem of deepfakes was highlighted less than 5 years ago. The term itself spread around 2018, although it is difficult to ascertain who used it first. In the original interpretation, “deepfakes” called hyper-realistic videos using face swaps that leave little trace of manipulation [2]. The main direction of using these technologies was the creation of fictitious news style content that is fabricated to deceive the public [3]. Early examples of deepfakes focused on political leaders, actresses, comedians, and entertainers having their faces weaved into adults videos [4].

Deepfake technology can generate video of a person saying anything, without the consent of the person whose image and voice is involved [5]. One of the main factors in the emergence of technology is the increase in the performance of computers, which began to allow the creation of realistic fakes at home [6]. The early
emergence of deepfake technology was completely reviewed in [7]. In conclusion this author says that deepfake technology can create a fraudulent identity. Later, studies appear that consider the problem of deepfakes in the context of fraud [8], and methods to detecting deepfakes [9].

At the same time, user identification systems, which are used by the real banks and financial institutions, do not provide sufficient counteraction to deepfakes, as indicated in [10]. Researchers highlighted that new generation of AI that could pose a serious threat to companies through deepfakes.

The purpose of the study is to highlight and classify main vectors of the deepfakes fraudulent attacks for payment systems in contemporary economy.

Main results. Payment card fraud is limited by card expiry dates, limits, and security notifications. The methods of fraud detection can help find and stop fraudulent transactions made by perpetrators, but almost useless, when customers unwittingly transfer money to criminals by themselves. Fig. 1 shows the actual fraudulent payment schemes active in the Single Euro Payments Area (SEPA) countries.

![Fig. 1. Value of fraud types as a share of total card fraud using cards issued within SEPA](source: Developed by author based on the data of ECB [11])

As we can see from Fig. 1, the share of Card-not-Present (CNP) fraud accounting for 80% of the total value of fraud. It increased by 5 p.p. over the past 5 years. In contrast, the share of fraud at ATMs and POS terminals decreased to 5% and 15% of the total value of fraud respectively.
Considering that the total losses from card fraud are estimated at €1.87 billion, the losses from CNP are €1.5 billion.

Despite the fact that up to 80% of the total volume of fraudulent transactions is carried out through e-commerce and other traditional channels, new types of fraud are a big problem. Fig. 2 shows the main deepfake problems.

**Fig. 2. Classification of the main deepfake problems**

*Source: Authors own development.*

Next, consider the main vectors of attacks on card payment systems.

**Vector 1. Toxic credit cards and loans.**

The Covid-19 epidemic has forced many financial and credit institutions to reconsider their approach to doing business. Previously, many organizations had their own physical points of service and customer identification widely distributed in all major cities, in banks, offices of credit organizations or simply in crowded places - for example, at shopping centers.

Under the conditions of strict quarantine and subsequent restrictions, companies met their customers and began to offer remote methods of service and identification. This innovation reduced the level of checks on customer profiles and documents, which gave rise to a new wave of fraudulent attacks on financial and credit institutions.

Organizations that issued instant credit cards or loans in the form of payment cards with a credit limit were attacked. Conventionally, attacks can be divided into several categories.
First-party fraud
In this case, the scammer uses fake income documents to get a higher credit limit or a cash loan. Documents confirming the place of residence can also be forged in order to confuse and throw off the trace of the exchange system of creditors who exchange data on unscrupulous borrowers, as well as complicate the search for a borrower by collection organizations.

Typically, such a scam can be carried out once, taking loans in a short time from many available organizations, and then “disappear” by turning off the phone.

After the name and surname of such a client enter the bases of unscrupulous borrowers exchanged by credit organizations, repeated loans will no longer be issued.

Third party fraud
This method implies that the loan is obtained on behalf of a person whose documents were stolen in digital or physical form.

An attack from a third party is most often obtained in the form of a transfer to a card number of a criminal, because a credit institution cannot obtain the name of the real owner by the card number, and banks do not disclose such information. Or, if the attacker has access to the victim's online banking, the received loan in the form of wire transfer funds is withdrawn to a front organization or an individual who is engaged in cashing out.

Contacting the police and searching for an individual or the owner of a “one-day” company who receives the victim's funds can take several years and not produce results. At the same time, the victim’s credit reputation also suffers, because most likely the victim will find out that a loan was opened for the victim when the debt has already been sold to a collection agency and the client’s profile is included in the database of unscrupulous borrowers from which it is almost impossible to remove yourself without going to the courts and other bureaucracy.

Such an attack can be multiplied by an attacker as many times as much attacker can to steal the victim's documents or access to his account.

Vector 2. Synthetic Identities
This type of attack involves taking loans and credits for a person who never existed. This is the most technologically advanced method in the attackers' arsenal.

With the development of artificial intelligence systems and machine learning systems, it became possible to fully synthesize the profile of a living person, including photo, video and voice.

This attack vector also uses fake IDs that attackers acquire on the black market, as well as utility bills that confirm the fact of residence, and also synthesizes “account statements” from well-known banks confirming the solvency of the borrower.

Actuality of the Synthetic Identities problem shows the data of Federal Trade Commission (USA) (table 1).
### Dynamics of the most common types of identity theft

<table>
<thead>
<tr>
<th>Type of identity theft</th>
<th>Number of reports in 2020</th>
<th>% of total</th>
<th>Change from 2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government documents or benefits fraud</td>
<td>406,375</td>
<td>24.5%</td>
<td>1663.0%</td>
</tr>
<tr>
<td>Credit card fraud</td>
<td>393,207</td>
<td>23.7%</td>
<td>44.7%</td>
</tr>
<tr>
<td>Other identity theft</td>
<td>353,152</td>
<td>21.3%</td>
<td>63.7%</td>
</tr>
<tr>
<td>Loan or lease fraud</td>
<td>204,967</td>
<td>12.3%</td>
<td>95.8%</td>
</tr>
<tr>
<td>Employment or tax-related fraud</td>
<td>113,529</td>
<td>6.8%</td>
<td>149.2%</td>
</tr>
<tr>
<td>Phone or utilities fraud</td>
<td>99,539</td>
<td>6.0%</td>
<td>19.2%</td>
</tr>
<tr>
<td>Bank fraud</td>
<td>89,476</td>
<td>5.4%</td>
<td>52.4%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,660,245</strong></td>
<td><strong>100%</strong></td>
<td><strong>106.8%</strong></td>
</tr>
</tbody>
</table>

*Source: Developed by author based on the data of FTC [12]*

The attack is carried out in two stages. First, an account is opened for a synthetic person in a financial institution, in the second stage, loans are received from another company, and sometimes from the same one that opened the account, but such cases are rare. Companies rarely issue loans immediately to clients with new profiles without a credit history, but they are always ready to receive a transfer from another financial institution in the form of a loan.

Thus, the attacker deceives 2 financial institutions at once - the one that opened the account and the one that issued the loan.

After receiving a loan, the withdrawal scheme is similar to the “Third Party Fraud” scheme - the funds are transferred to a company or a person who is an accomplice of the attacker.

The problem with this attack is that the actual borrower does not exist at all and the number of attempts by the attacker is limited only by how many synthetic profiles the systems or personnel of both companies that are being attacked will miss.

This method of attack is the most dangerous at the moment because as techniques are constantly being improved, it becomes more and more difficult to identify synthetic personalities.

In 2019, an executive of a UK-based energy firm thought he was speaking on the phone with his boss, the CEO of the firm’s German parent company, who asked him to send €220,000 to a Hungarian supplier [13]. The caller said the request was urgent, directing the executive to pay within an hour, which he did. Instead of his boss, the executive spoke to a voice recording generated by artificial intelligence-based software that successfully impersonated the CEO.

Live facial biometric data many digital-only banks rely on to authenticate their customers is not fraud-proof either. Cybercriminals found a way to recreate 3D
models of faces using recorded videos that can be used to log in by generating head tilts and turns on demand.

Social engineering plays a significant role in modern fraud cases. A man behind an Instagram account with 2.5 million followers flaunting his opulent lifestyle told people they could earn as much as him by sending him money [14]. He was arrested after stealing over 400 million from individuals and businesses worldwide [15].

There are many examples of money flippers on social media that promise to turn $100 into $1000, $500 into $5000, and so on [16]. Suffice to say that people don't get their investments back. If the recipient is not blacklisted, has a business, and receives money regularly, training a system to detect such type of fraud is challenging, if not impossible, for now.

Payment card and identity fraud are closely tied to criminal activities that aim to launder money and conceal identities. Modern compliance and anti-money laundering investigations check social media accounts for suspicious posts and activities. To get around these checks, criminals buy inexpensive accounts created and maintained for a few years to develop a plausible online identity.

People who want to take on a different identity can buy a passport and a new identity with social media accounts, diplomas, and other documents for relatively little money (Table 2). On one side, it's easier to obtain a new identity than ever before. On the other, regulators and service providers are tightening security and making it more difficult to evade their checks.

### Table 1. The minimum cost of a brand new identity

<table>
<thead>
<tr>
<th></th>
<th>USA</th>
<th>Canada</th>
<th>Australia</th>
<th>UK</th>
<th>Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>New identity: Passport, ID Card and Birth Certificate</td>
<td>$1152</td>
<td>$1175</td>
<td>$1355</td>
<td>$1255</td>
<td>$1125</td>
</tr>
<tr>
<td>Education: High School Diploma and Bachelor’s Degree</td>
<td>0.1699 BTC</td>
<td>0.1699 BTC</td>
<td>0.1699 BTC</td>
<td>0.1699 BTC</td>
<td>0.1699 BTC</td>
</tr>
<tr>
<td>Finance: Bank account, Credit Card and 5000 of Counterfeit Currency</td>
<td>$115+ 0.0984 BTC</td>
<td>$115+ 0.0984 BTC</td>
<td>$115+ 0.2593 BTC</td>
<td>$115+ 0.1162 BTC</td>
<td>$115+ 0.0894 BTC</td>
</tr>
<tr>
<td>Total</td>
<td>$1267+ 0.2683 BTC</td>
<td>$1290+ 0.2683 BTC</td>
<td>$1470+ 0.4292 BTC</td>
<td>$1370+ 0.2861 BTC</td>
<td>$1240+ 0.2593 BTC</td>
</tr>
</tbody>
</table>

*Source: compiled by the author on the basis of [17]*

### How to fight synthetic personality theft

Business verticals most vulnerable to synthetic personalities is a Financial Organisations, Electronic Money Institutions, Banks, Landing, Marketplace, Government, Gambling, and Casinos.
Know Your Client (KYC) approach. This is the first layer of defense for any financial organization, required by regulators and risk management to avoid sanctions or penalties, especially with online fraud operations. Regulators strictly prohibit serving a certain group of persons who is under sanctions or can’t prove a source of income or their personality. Also, for Anti-Money Laundering (AML) purposes, an individual’s identity must be in compliance with laws and regulations.

The fraudster techniques continuously evolving and for now, it’s really not enough just to get the scans of clients' documents and believe it’s a real person and documents not faked.

In the current world, from point of view of a financial organization, any new clients must be in a category with a zero-trust strategy. To verify clients' personalities, financial organizations have several new technologies based on biometry and machine learning technologies [18], but they are also vulnerable.

Vector 3. Technical State of Face Recognition and Liveliness Detection Technology

Anti-fraud technologies often come late compared to the techniques used by scammers.

This is understandable because before you can come up with a way to protect yourself, you need to discover and recognize a new attack method.

A study of the current technical literature has not found any academic analysis of the vulnerability of liveliness detection systems to Deepfakes attacks.

Both technologies, face recognition, and liveliness detection is under active development and are currently far from ideal.

Popular open-source face recognition systems are vulnerable to deepfake attacks with 95% FAR (False Acceptance Rate). The detail can be found in [19].

This problem also affects services deployed by enterprise cloud providers, with FARs ranging from 78% to 99.9%. The details can be found in Tariq, Shahroz, Sowon Jeon, and Simon S. Woo. [20].

The top-rated model measured in academia cannot detect deepfakes with a FAR of 0.86, which is a drastic drop from the certified quality of 0.173 FAR and the accuracy of 0.988.

*Optical character recognition (OCR) ID verification*

During the registration of a new client or so-called onboarding, the client is asked to provide some form of identification, such as a passport, national ID, or driver's license for recognition and verification. The next step also asks for service invoices so that the client can confirm their actual address of residence, but this is not always required.

During the verification process, the provided data is sent to some technical solution that performs authentication by checking important features of the document.
such as signatures, holograms, stamps, the shape of the corners of the document, and others. Information from the document is extracted using text recognition tools (OCR) and checked against the data that the client filled in the registration form.

**Biometric KYC**

1. Face detection.

During the process of registering a new client, the onboarding system requires the user to take a selfie. The registration system, after receiving the client's selfie, matches the face from the selfie with the image on the ID card uploaded earlier in the identification ID stage. Next, the onboarding system makes a decision based on the degree of face matching. If the level of accuracy is higher or meets the minimum required parameters, the face from the ID and the selfie are considered to match and the client can proceed to the next step, for example, Liveliness detection.

2. Liveliness detection

Unfortunately, the method of matching a face from a selfie with a face on a document can no longer provide a sufficient level of confidence, since photo-based face synthesis systems already successfully pass such checks. A believable synthetic selfie can be generated based on even a single photo that an attacker can easily get from a social profile potential victim or generate a fundamentally new face for both documents and selfies.

The new approach checks whether the person being tested is alive. The person who is in front of the camera is checked for reality. For this, the presence of breathing, blinking, or even the presence of a heartbeat is determined. The onboarding system may ask the user to perform the actions listed above, such as blinking, smiling, or performing head movements in different directions in a random order so that attackers cannot synthesize a previously prepared video that will perform the above actions.

Academic research is important, however, it is most often carried out in controlled environments using synthetic datasets, which is actually far from reality.

Commercial solutions and real-life datasets collected by financial institutions are most often closed and not published, so we can only guess about the real effectiveness of closed commercial solutions compared to open data obtained from open source solutions and datasets.

As an actual example of the vector 3 attack we can refer the deepfake attack that Chinese fraudsters performed against China’s taxation system in April 2021.

They used facial images purchased on the black market to create synthetic identities and set up a shell company that issued fake tax invoices for about $76.2M USD.

They managed to spoof face recognition and liveness checks by using a special phone (available for $250) to hijack the mobile camera, injecting pre-made deepfake models good enough to accomplish their goals [21].
Conclusions. The results of the study show that deepfakes fraud technologies have now emerged, which are a serious threat to the security systems of banks and financial institutions:

- most passive solutions cannot guarantee protection against deepfakes;
- all face matching services can be fooled by face swapping;
- no solution can detect real-time face swap in ID photos.

In an attempt to balance convenience and security, security is losing. Customers don’t like long passwords and additional verification methods. They frankly don’t care if their information leaks because “they have nothing to hide and don’t have that much money anyway.” To slow down the advancement of financial fraud, anti-fraud systems based on artificial intelligence methods are currently being developed and implemented. In fact, the competition between methods of fraud and anti-fraud continues all the time. The development of anti-fraud methods takes a lot of resources, but it allows us to maintain the credibility of financial institutions.

References:
10. Wiggers K. (2021) Study warns deepfakes can fool facial recognition. [Online]. Available at: https://venturebeat.com/2021/03/05/study-warns-deepfakes-can-fool-facial-recognition/


### Література


