Yevhenii SHTEFANIUK[1], Ivan OPIRSKYY[2], Ihor IVANCHENKO[3], Kateryna PINDEL[4]

Opiekun naukowy: Ivan OPIRSKYY[2]

DEEPFAKE – NEW TECHNOLOGY FOR IMPERSONATION IN CYBERATTACKS

Summary: The following work contains the analysis of the main principles of DeepFake technology, the possible usages of this technology in cyberattacks and recommendations for protecting against them.

Keywords: DeepFake, cybersecurity, impersonation, face recognition, identification, voice imitation

DEEPFAKE - NOWA TECHNOLOGIA IMITACJI W CYBERATAKACH

Streszczenie: Poniższa praca zawiera analizę głównych zasad technologii DeepFake, możliwe zastosowania tej technologii w cyberatakach oraz zalecenia dotyczące ochrony przed nimi.

Słowa kluczowe: DeepFake, cyberbezpieczeństwo, podszywanie się, rozpoznawanie twarzy, identyfikacja, imitacja głosu

1. Introduction

We can now see the widespread adoption of computer information systems. At the same time, the requirements for the security of such systems, in particular those used in the banking and economic spheres, are getting higher.

[1] Lviv Polytechnic National University, Graduate Student of Information Protection Department, yevhen.sht@gmail.com
[2] DSc, Lviv Polytechnic National University, Professor of Information Protection Department, iopirsky@gmail.com
[3] PhD, National Aviation University, Associated Professor of IT-security Department, igor-p-l@ukr.net
[4] Student of IT-security Department, National Aviation University

For successful user identification and authentication modern security systems rely on several factors, which are sufficiently robust in most cases. Despite this, human factor exploitation is one of the most effective tools for hacking. This is based on the social engineering attacks. Their main purpose of these attacks is to persuade the user to take the action necessary for the attacker. Different methods are used to do this, from fake phone numbers to social networking accounts. Recently, however, the development of machine learning technology has ushered in a completely new method for cyberattack - using of DeepFake technology.

In 2018, machine learning was firstly used to generate fake adult videos [1]. Since then, DeepFake technology has evolved rapidly and achieved significant results. It can be used not only for video generation but also for realistic voice imitation. The basis of DeepFake's work is the use of artificial neural networks such as Generative Adversarial Networks (GANs). Its principle of work is the application of two neural networks that compete with each other within the framework of a zero-sum game. One network (generator) generates potential candidates, and the other (discriminator) evaluates them. Thus, the first network learns to build an image that would allow it to pass the criteria for selecting a second network. In addition, the discriminator is allowed to influence the criteria for selecting the source data for the generator, which makes it a full participant in the learning process [2]. In practice, such a system can achieve significant results.

Further development of neural network technology allowed such generation to be performed in real time, such process is shown on picture no. 2.

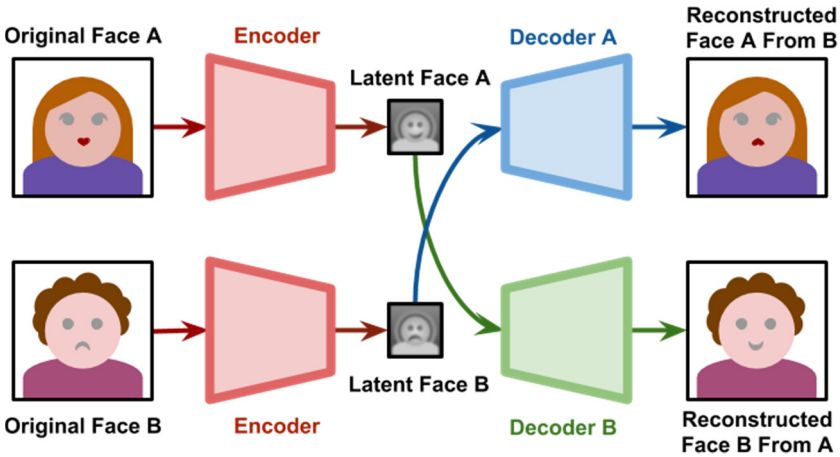

Figure 1. The process of creating a real-time imitation of a human neural network in DeepFake technology

Until recently, DeepFake technology was mainly used for entertainment purposes such as FaceApp application, etc. However, recently a malicious use of this technology was recorded for the first time.

According to the Wall Street Journal [3], in March 2019, the CEO of a British energy company was robbed of € 220,000. He sent the money to a supplier from Hungary because his boss, the head of a parent company in Germany, repeatedly confirmed to him this instruction. Nevertheless, the attacker simply used AI-enabled software to replace his face and voice with the face and voice of the supervisor in real-time and demanded the payment within an hour.
The program used by the thief was able to completely simulate the person's voice: tone, punctuation, even a German accent. The message came from the head of the company in Germany, additionally the British director was sent an email with his contacts.
As a result, all the money was stolen. They were transferred from a Hungarian account to Mexico and then dispersed worldwide. However, the attacker did not stop there and asked for a second urgent transfer, in order to "speed up the deliveries from Hungary". Then the CEO realized that something was wrong and called his real boss. The name of the company and its employees have not been disclosed yet as the case is under investigation and the attacker has not yet been found.
This proves that DeepFake technology has reached a level of development that allows its use in real cyberattacks.

2. Main part

DeepFake technology can be used in a wide range of cyberattacks, both as an auxiliary and as a major tool. It allows an attacker to simulate both the image and the voice of the victim, but enough data to train the model is need to be collected.
Due to the widespread use of social networks, by both celebrities and ordinary people, the problem of gathering the necessary data is a matter of time. Moreover, scientists have recently announced a neural network that is effective enough to train a model using only one photo of a person [4]. Thus, even the person who does not use social networks can become a potential victim since it is enough to take only one photo.
Voice simulation still requires collection of a large amount of audio data for the model to learn to imitate the voice timbre and the victim's accent. However, if the victim is a public figure, then the issue of collecting media generally turns into a search for a several interviews. If the victim is a non-public person, it is enough for the attacker to have a conversation with her on the street and record her on a hidden voice recorder. Therefore, the process of preparing a model for a successful attack does not take too much time. Thus, DeepFake technology has great potential for use in cyberattacks.
Now let us look at the main attack vectors in which DeepFake technology can be used as a tool.

2.1. Fraudulent Attacks

Using DeepFake, attackers can greatly simplify the tasks of social engineering, which in this case boils down to gathering the necessary data about the person whose image will be imitated. Previously needed to gather a lot of information about the victim, his

or her habits and personal life, it is now simplified to choosing the person the victim trusts and to use his or her imitation to persuade the victim to do what the attacker wants. On picture no.2 two photos are shown – a real one, and a one generated by DeepFake [5]. As we can see, the difference is almost impossible to see.

Figure 2. Left photo is real, right photo was generated using DeepFake technology

2.2. Identity theft

If, in the previous vector of attacks, the person who knew and trusted the victim of the attack was a person to imitate, then in this case, the victim itself is imitated.

Identity theft is a common cybercrime in developed countries [6]. It involves the theft and forgery of the victim's personal data, which allows an attacker to make purchases, book tickets or engage in unlawful activities on behalf of the victim. Thus, the attacker uses the victim as an anonymizer, hiding his own identity.

Successful conduct of this type of attack requires access to the victim's confidential data - social security number, passport, etc. [6]. DeepFake in this case helps to extend the limits of the attack by allowing the attacker to act on behalf of the victim in the virtual space - to make videos and skype calls, post photos and videos on social networks, which will be very difficult to distinguish from the original. Because model's training data is much more accessible than the victim's confidential ID data, DeepFake can be a unique complement to this type of attack if an attacker needs to expand its scope.

2.3. DeepFake in politics

DeepFake technology opens up new opportunities for manipulation in politics and media.

Today, world leaders are increasingly using social networks, such as Facebook or Twitter, to advertise or express their public position. There have already been several cases of hacking famous people's accounts and posting messages on their behalf [7]. With DeepFake, such attacks can have a much greater impact. A successful imitation of a political statement by a world leader, distributed on his behalf from his account,

can potentially provoke a big political conflict. In addition, not only well-known politicians who have their own verified accounts can be victimized - attackers can imitate a message from the leader of a terrorist or other illegal organization and upload it to the Web - exposing such a fake is much more difficult, and the consequences can be much worse.
Another way to use DeepFake is to spread fake news - it can present any statement from any politician. Of course, such a fake will eventually be exposed, but the social impact resulting from its publication may have far-reaching consequences.

2.4. DeepFake for tampering with media evidence and personal compromise

DeepFake can also be used to falsify evidence and compromise a person. With the help of this technology, an attacker potentially can fake evidences of a particular person's whereabouts at a certain point in time.
If the fake is not revealed by expertise, this fraudulent evidence can even be used in court.
Another option for such a vector of attack is the use of the technology by totalitarian regimes in order to compromise a person in the media or justify special measures against their citizens.
In this way, DeepFake technology can be used as an effective primary or secondary tool in a range of cyberattacks, delivering a new level of cybercrime.

2.5. Anti-attack measures with DeepFake technology

Since the main purpose of the technology is to create fake information about an individual or to disseminate false information on his behalf, it is important to be able to use an additional verified source of information to refute the fake information.
Possible security measures are different for each of the above attack vectors.
If an attacker uses DeepFake to engage in fraudulent attacks, the victim must have an additional hidden channel of communication with the individual, or request that the information provided be confirmed.
An example of a hidden link may be an additional phone number, or a social network account created by a different name so that abusers do not know it.
Confirmation of the information provided is possible by using special keywords agreed in advance, which will not be known to the attacker.
When it comes to financial transactions, it is a good option to use EDS to confirm instructions for conducting financial transactions. Such a defense system could prevent the attack described in [3].
In order to protect against DeepFake in an identity theft attack, the victim should identify the suspicious activity on his or her behalf and alerted all her entourage in advance, as well as the relevant government agencies. This will make it much more difficult for the attacker to take further actions, because all interested persons will check all the information coming from the victim more closely.
Protection against policy and media attacks is to collate and verify the information. It is important to know that political negotiations take place through secure channels of

communication and controversial publications on behalf of officials should be immediately interpreted as a possible fake.
The protection against compromise of the person and the evidence should be done by the introduction of advanced DeepFake recognition technologies and the adoption of relevant regulatory acts by government bodies. This will allow to identify false evidence in a timely manner and prevent the compromise.

3. Conclusion

The further development of DeepFake technology initiated its use by cyber-attackers. Now, in the absence of software that can detect fake audio and video data quickly and efficiently, we can predict that DeepFake will quickly enter the arsenal of attackers as an effective social engineering tool.

REFERENCES:

1. ROOSE K.: Here Come the Fake Videos, Too // The New York Times. – 2018 ISSN 0362-4331. URL: *https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html*
2. GOODFELLOW I.J., POUGET-ABADIE J., MIRZA M. and others: Generative Adversarial Nets // Proceedings of the International Conference on Neural Information Processing Systems (NIPS) – 2014. pp. 2672–2680.
3. STUPP C.: Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case // The Wall Street Journal - Aug. 30, 2019. - *https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402*
4. QUACH K.: Deepfake 3.0 (beta), the bad news: This AI can turn ONE photo of you into a talking head. Good news: There is none // The Register. – Jun 2019. – URL: https://www.theregister.co.uk/2019/06/19/deepfake_ai_single_photo/
5. DVORSKY G.: Deepfake Videos Are Getting Impossibly Good. // Gizmodo. – 2018. – URL: *https://gizmodo.com/deepfake-videos-are-getting-impossibly-good-1826759848*
6. What Is Identity Theft // Symantec Corporation. – 2019. – URL: *https://www.lifelock.com/how-it-works/what-is-identity-theft/*
7. Twitter closes thousands of fake news accounts worldwide // CAN-World. – Sep 2019. – URL: *https://www.channelnewsasia.com/news/world/twitter-closes-thousands-of-fake-news-accounts-worldwide-11928230*