Artificial intelligence in cyber security

Originally published at https://www.itransition.com/blog/artificial-intelligence-in-cyber-security

Published: October 18, 2021 | By Martin Anderson Independent AI expert

The unexpected growth of online systems and correspondingly higher traffic levels caused by the advent of COVID-19 has led to an unprecedented increase in malicious network activity. In a climate where face-to-face discussion has been transitioned to VoIP environments, and where an increasing volume of information is forced onto network channels, the available network attack surface has grown notably since the start of 2020.

The growth of opportunist attacks

It's a long-term situation that's seems set to survive the crisis: according to the International Energy Agency (IEA) 2020 Tracking Report, web traffic is set to double by 2022, compared to rates at the start of 2021, with mobile net users surging from 3.8 billion in 2020 to 5 billion by 2025.

The same machine learning technologies which are increasingly being used to improve systems security are being used to attack them. It's a new cold war for which the old weapons are unsuitable; simple techniques such as heuristics — which look for minor changes in network behavior — are inadequate to a new generation of AI-enabled attackers.

Venerable security protocols and practices, some dating back decades, are unprepared for imaginative new approaches to exfiltration, phishing, identity theft, network incursion and password cracking, among many other attack surfaces that are proving vulnerable to more imaginative approaches.

Where off-the-shelf solutions are outdated, this turbulent period is better addressed by AI consultants. Off-the-shelf SaaS solutions designed in 2014 are not prepared for the cyber attacks of 2022 and beyond.

Here, we'll take a look at some of the key sectors in cyber security, at the extent to which attacks of all types are growing, at the way that AI is increasingly being leveraged to launch devastatingly effective social engineering attacks, at new AI methods that can help bolster systems against more traditional attacks, and at the necessity to adopt a new, proactive mindset in addressing enterprise cyber security.

Key sectors in cyber security incursions

Ransomware

Ransomware has become the poster-child for the COVID-inspired increase in systematic criminal cyber security incidents. A 2021 report from SonicWall found that between 2019 and 2020, ransomware attacks rose by 62% worldwide and by 158% in North America, with headline-capturing events such as the July 2021 attack against US information technology firm Kaseya, which affected 800-1,500 global businesses, emphasizing the power of automated attack and the need for improved recognition capabilities and automated response systems.

Besides hospitals, a traditional target for ransomware, high value industries that are critically dependent on transport infrastructure are most vulnerable to cyber attacks. These include the mining, energy and construction industries, besides potential attacks against transport infrastructure itself, which has secondary effects on local economies.

Phishing

As ever, social engineering remains a signature feature of costly cyber security statistics: the F5 Labs 2020 Phishing and Fraud Report estimated that the frequency of phishing attacks exploded during the lockdown peak by 220%, with many attacks now occurring in real time in order to exploit vulnerabilities in time-critical multi-factor authentication (MFA) systems.

Network attacks

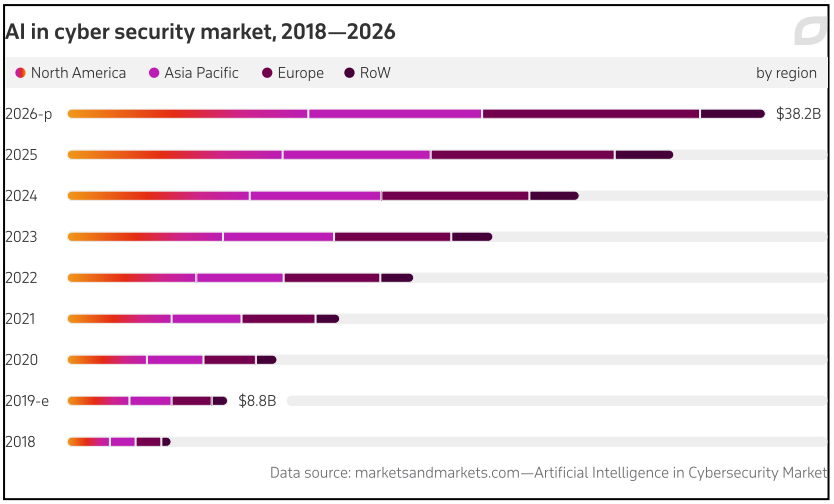
VMWare's 7th annual global report has found that cyber attacks that seek to manipulate or destroy data (so-called 'destructive' or 'integrity' attacks) target organizations 50% of the time, while incidents of attempted network incursion have risen in similar numbers.

Even discounting the influence of the pandemic, the 2021 *Nation States, Cyberconflict and the Web of Profit* study by HP Wolf Security concluded that nation state attackers are 'stockpiling' zero-day vulnerabilities (otherwise unknown weaknesses in network systems), and that over 40% of incidents involved attacks on physical as well as digital assets — an approach called 'hybridization'.

The growth of AI in cyber security

In the meantime, the cyber security market is responding to the challenge with investment: according to Mordor Intelligence's 2021 report, the cyber security market is set to grow at a CAGR of 14.5% to over $350 billion by 2026; Gartner forecasts that enterprise investment in cloud security solutions will rise from $636 million in 2020 to $1.63 billion by 2023; and Forrester predicts a steep rise in spending on cloud security tools, from $5.6 billion in 2018 to $12.6 billion by 2023.

Artificial intelligence is the core aspect behind much of this growth, with the software side of machine learning for anomaly detection and cyber security assurance overall estimated to grow from $8.8 billion in 2019 to $38.2 billion by 2026.



The need for agile and responsive security systems

Machine learning systems can infer patterns of behavior from vast amounts of historical data across a range of applications and processes in the cyber security sector. However, the data can be difficult to obtain, outdated by the time it is processed, or so specific to certain cyber security scenarios that it becomes 'over-fitted' and does not generalize well to new trends in the data that will later be fed to an AI-led algorithm.

For this reason, there is little prospect of a 'set-and-forget' solution to an ever-evolving threat landscape, where a rarely-changing installed solution simply updates definitions periodically from a remote source. Those days are over — now, new threats may come from completely unexpected channels, from a telephone call to a VoIP chat, through novel architectures, and even embedded into machine learning systems themselves.

This signifies the need for proactive systems designed and maintained by specialists in cyber security consulting, or the will and resources to establish protection systems in-house. Because the criminal incursion sector is innovative and resourceful, the response will need equal commitment and intensity of purpose.

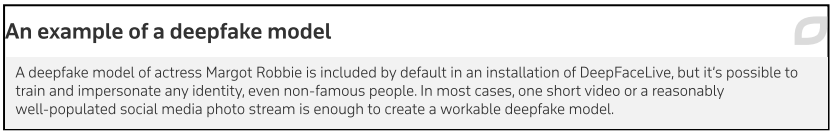
Declining trust in traditional data

Verizon's 2021 Data Breach Investigation report observes that 85% of cyber security attacks are now social engineering attacks that target human credibility. While authentication systems have migrated towards the verification of biometric characteristics such as video or voice data, as well as fingerprints and movement recognition, the same research that underpinned these advances is constantly pushing forward new methods to falsify the data.

This is a unique era — an extraordinarily narrow window of opportunity for cybercriminals, where legacy security architectures and individuals alike still trust what they see and hear, and where biometric data systems (which were difficult to enforce on customers) are being undermined at the very moment of widespread adoption. It's certain that the cyber-incursion factions won't waste the opportunity.

Deepfake manipulation

It's now possible to impose a new deepfake identity onto a participant in real time, thanks to DeepFaceLive — a streaming implementation of DeepFaceLab, the most popular open-source project for superimposing photorealistic identities onto individuals in video footage.



Data source: github.com—DeepFaceLive. Margot Robbie, 2021

In truth, the real victim is the recipient of deepfaked communications, with fraud, social engineering and network incursion (i.e. obtaining of access) emerging as potentially lucrative new avenues for cyber-criminals to explore. In their 2021 Cyber Threat Analysis, Insikt Group's security consultants forecast a drastic coming rise in deepfake attacks of these, and various other types. The report stated:

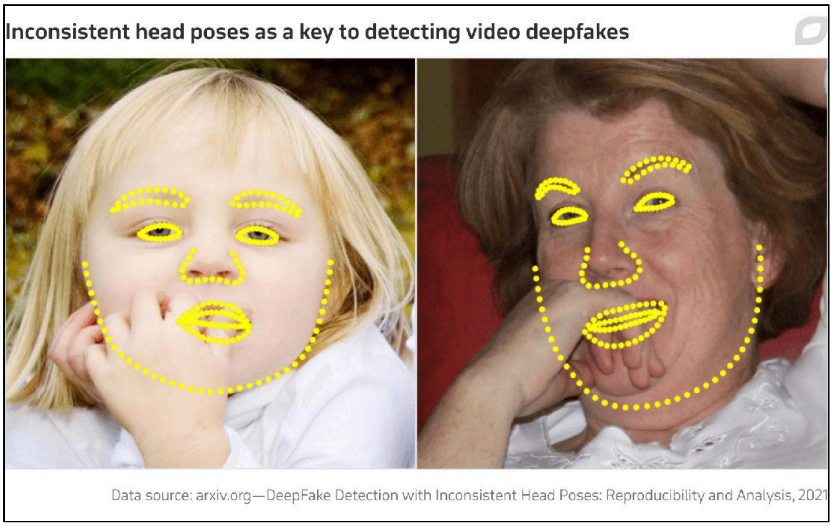
The most famous deepfake crime to date was the defrauding of $243,000 from a company CEO, in a case where an audio deepfake impersonation of a chief executive was used to obtain a wire transfer. Now that it's possible to add video to audio deepfakes in real time, the attack space is wider than ever.

Fighting back against fakes

The civil and state sector have been actively engaged in methods to detect deepfakes since the initial emergence of the technology in 2018. It's an ongoing game of whack-a-mole, as deepfake software creators use publicity around the discovery of perceived 'tells' (such as 'unblinking' faces) as a free 'bug-list', systematically closing most of the loopholes shortly after discovery.

The latest innovation in video deepfake detection involves examining the shape of pupils. Convincing eye detail is one of the very last components to resolve when training a deepfake model, and obtaining a convincing internal eye shape is a high-effort and expensive business. Previous approaches to eye-based deepfake detection also used a lack of reflection in eyes.

A more recent study identifies deepfake video based on the pose of the head on the faked identity:

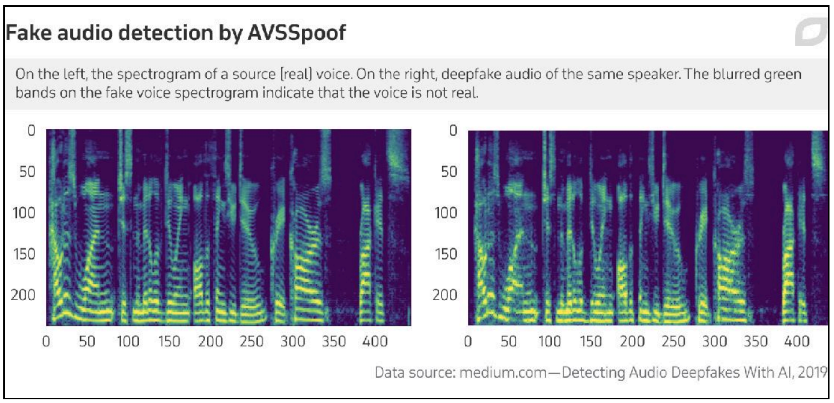


Data source: arxiv.org—DeepFake Detection with Inconsistent Head Poses: Reproducibility and Analysis, 2021

Combatting deepfake voice imitation

Though deepfake audio is a more recent innovation, it is now, as noted earlier, the most successful attack vector for deepfakes. Interest in this area was notably boosted when the voice of late celebrity chef Anthony Bourdain was deepfaked into a recent documentary without notifying the viewer.

The academic and industrial sector are fighting back against fake voices: in September of 2021, researchers from the University of South California announced a series of countermeasures and defense strategies against systems designed to compromise deep speaker recognition systems; in 2019 AI company Square developed a deepfake voice detector based on Google's 2019 AVSSpoof fake audio detection dataset (see image below); and an Asian research consortium (including the Alibaba group) has created a dedicated fake voice detection framework called DeepSonar, which searches artificial dialogue for key characteristics of audio fakes, even where notable background noise is present.



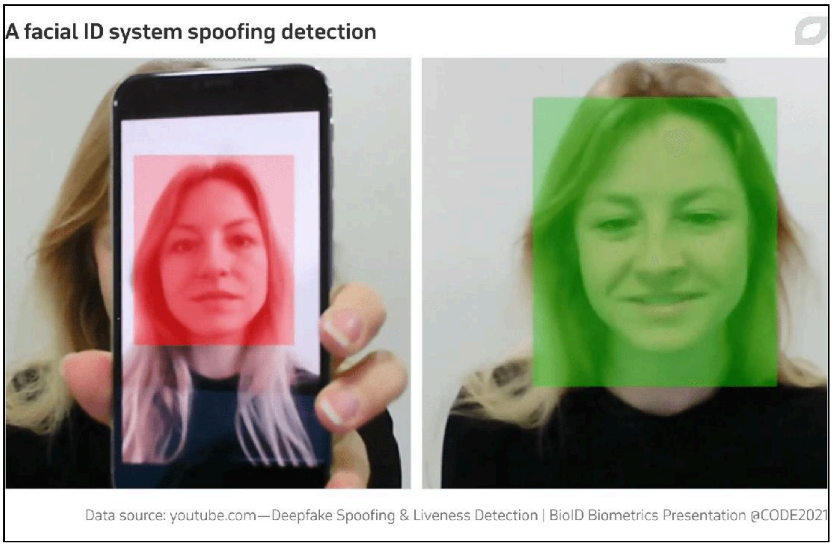
It's possible to implement the latest face/voice fake detection algorithms into an organization's communications channels, and companies that start early down this road are likely to have an advantage in avoiding the worst effects of the emerging wave of video and audio deepfake incursions.

Liveness detection

Emerging attack architectures are quite resistant to challenge, with many incursion attempts anticipating the negotiation of multi-factor authentication systems, like those commonly implemented for mobile banking security. In cases where biometric data is being faked (such as the use of masks, 'master faces' and even neurally-crafted physical make-up to defeat facial ID systems), authentication systems that detect the subject's 'liveness' are becoming an emerging front in bolstering biometric systems.

The topic of liveness detection has inspired LivDet, a bi-annual hackathon begun in 2009, which harvests the latest AI-based techniques designed to combat deceptions based on iris and fingerprint spoofing.

One recent system developed by the researchers from the University of Bridgeport, USA, uses anisotropic diffusion (the way light reacts with skin) to confirm an authentic face; others have used blinking as an index of liveness (though this is now correctible in deepfake work flows); and in June of 2021 a new liveness detection method was proposed that discerns 'unforgeable' lip motion patterns.



Organizations seeking to protect their biometric systems from real-world spoofing and digital attack should consider adding measures to detect liveness based on a number of 'tells' that leverage special equipment sensors (such as infrared scanning), and/or characteristics that distinguish fake and real biometric qualities, such as movement behavior and interaction with the environment.

New types of deepfakes

It's important to view security systems development as an ongoing process, particularly in regard to visual fake attacks. Deepfakes are not necessarily limited to audio and facial data; new detection systems are being devised to discern faked road footage and even satellite maps.

Combatting network incursion with machine learning

What of the remaining 15% of attacks that don't rely on human susceptibility? Effective cyber security systems must now take an anticipatory AI-based approach to detecting more 'traditional' types of attack, such as botnets, malware traffic, and other types of network assaults that may fall outside of recognized and protected attack vectors.

Research into AI-based Intrusion Detection Systems (IDSs) has advanced notably over the last 11 years of GPU-accelerated machine learning. Machine learning-enabled systems are capable of ingesting historical data about attacks and incorporating them into active defense frameworks.

Since the base channels through which most network attacks occur are based on some of the internet's oldest architecture, traditional DOS-based attacks and other types of network incursion operate in a much more limited environment and set of parameters than the new wave of human-centered incursion campaigns.

Protecting software-defined networks with AI

The 2021 research from the Department of Computer Engineering at the King Saud University in Saudi Arabia was able to obtain 'outstanding' results against a gamut of incursion techniques with a new architecture designed to incorporate an AI-centric defense system into the core functioning of software-defined networks (SDNs).

To accomplish this, the researchers developed a comprehensive database of attack type characteristics, which also serve as a list of some of the likeliest routes into a network. These include:

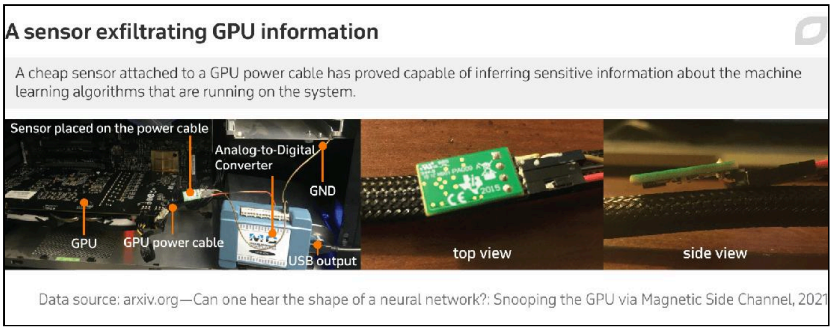
- **DoS attacks** (the flooding of networks with bogus traffic designed to overload the system).
- **Probes** (which hunt out vulnerable or exposed ports in security systems).
- **UR2** (buffer overflow attacks that seek to 'collapse' security safeguards through software vulnerabilities).
- **Remote to Local Attacks** (which send malicious network packets designed to obtain write access to unprotected parts of the target system).

Attacks against machine learning systems

Besides the growing need to fight AI incursions with AI-based countermeasures, it's also necessary to consider that attackers are likely to attempt to compromise these systems in both the virtual and the real world.

Hardware exploits and solutions

One recent exploit demonstrated the possibility of 'listening in' on GPUs that are performing machine learning routines by attaching a $3 sensor to the component's power cable (see image below); another vulnerability recently discovered uses voltage spikes to compromise the guidance software in a self-driving vehicle; and a new graphene-based innovation in 'unclonable' hardware security keys by PennState promises an end to the 'cut-and-paste' era of security incursions.



Defeating AI-based command and control systems

When a computer becomes infected with ransomware, the people who initiated the attack need some way to understand what's happening at the attack site. This means that the malware has to send and receive signals to secret internet locations periodically — a traditional weakness of ransomware.

Now it's possible for attackers to use AI to manipulate apparently random posts in fake social media accounts to keep these channels of communication open. However, a well-maintained network security framework can seek out the patterns of change in the very same information that the malefactors are using and potentially block all communications, buying time for analysts to retrieve the compromised system.

Conclusion

AI-based cyber security attacks are evolving into industrialized, generic attack packages that leverage machine learning technologies, and are increasingly common in illicit markets on the dark web. The rate and volume of cyber-assaults are no longer linked directly to the most advanced hacking groups, but rather to their customers.

Though systematic attack is still susceptible to systematic defense, the new wave of incursions requires a vanguard approach to local and cloud-based cyber security systems. The objective is now to anticipate rather than respond.

In most cases, this will entail custom cyber security solutions that are developed with the same avidity and obsessive detail as is evidenced in the work of a well-motivated and well-equipped new generation of attackers. It may be a long time before the attack vectors consolidate again into so

narrow a channel as a mere TCP/IP switch. In the meantime, we're living in an era where vigilance and creativity are prerequisites for effective protection of organizations.