May 21th, 2019

SECOND EDITION OF THE FRANCO-GERMAN
COMMON SITUATIONAL PICTURE

AGENCE NATIONALE DE LA SÉCURITÉ DES SYSTÈMES D'INFORMATION
ANSSI

Bundesamt
für Sicherheit in der
Informationstechnik

Foreword by Arne Schönbohm – President of the Federal Office for Information Security (BSI)

Digitalization is an issue that cannot be considered on a national level alone, as it means progress for the people and for the worldwide economy. However, digitalization also means to face a multitude of challenges in the field of cybersecurity which are transnational by their nature. Therefore, it is increasingly necessary for all national cybersecurity agencies to cooperate very closely.

In this regard, the relationship between ANSSI and BSI is based on mutual trust, respect and a longstanding partnership which extends to all technical challenges for a secure and successful digitalization. Regular exchanges between our experts on topics related to cybersecurity, standardization, certification as well as cryptography are the daily basis for our cooperation.

In July 2018, ANSSI and BSI published their first joint report on cybersecurity, highlighting the challenges of ransomware and cryptocurrency crime as well as presenting solutions to cope with these particular issues. The second volume now offers an update on incidents concerning cryptocurrencies but also sheds light on the field of Artificial Intelligence which has growing, but often underestimated intersections with cybersecurity. The analysis of these quickly developing fields outlines the necessity of synchronizing our efforts even closer, in order to face both current and future challenges successfully.

Foreword of Guillaume Poupard - Director General of the French National Cybersecurity Agency (ANSSI)

The Agency celebrates this year its tenth anniversary. Over the last decade BSI has been amongst the oldest and closest partners we have had. This second volume of the Franco-German Common Situation Picture is both a testament and a living testimony of the tight cooperation between ANSSI and BSI. A cooperation furthermore mirrored on a greater scale by the signature on the 22nd of January 2019 of a bilateral treaty on a level of the Élysée Treaty signed 55 years ago.

This production could not have been achieved without the strong foundation of cooperation that can be witnessed between both agencies on a daily basis being it on the certification, technical, operational or research level. More than a relation between peers this cooperation has furthermore been emulating, helping both partners overachieve their targets and expectations.

The two topics addressed in this Common Situational Picture are a good sample of what is ahead of us in terms of both threat and opportunities. Cryptocurrency crimes are indeed a result of in-depth digitalization of the society, the latter being as much a paradigm changing opportunity as it is already actively exploited on a grand scale for nefarious purposes. Artificial Intelligence will represent just such a tremendous shift as well and we are looking forward with our BSI colleagues to transform this in an opportunity that is safe and secure for all society.

CONTENTS

INTRODUCTION TO THE SECOND EDITION OF THE FRANCO-GERMAN COMMON SITUATIONAL PICTURE

For the first edition of the Franco-German Common Situational Picture (CSP), ANSSI and BSI chose to focus on malicious activities associated to ransomware and cryptocurrency mining, which represented a growing threat to citizens, major companies providing critical services as well as small and medium-sized enterprises with direct impact on their activities.

This year's second edition picks up the developments of cryptocurrencies, explaining some methods of criminal exploitation of the underlying block-chain technology. Furthermore, the CSP 2019 provides an introduction to some core principles and concepts of Artificial Intelligence (AI) and especially their possible future implications for IT-security.

The State of Cryptocurrency and Its Criminal Exploitation

Cryptocurrency crime was a growing trend in 2017. The enormous rise in stock prices gave way to numerous technical and non-technical criminal attacks on the underlying block-chain technology and the handling of the currencies themselves. In the following update to the assessment of 2018 on the state of cryptocurrency, this volume will present a few examples for their criminal exploitation.

The volatility of those new pseudo-currencies became ubiquitous in 2018, e.g. visible in the very high volatility of their overall market capitalization. In the beginning of 2019, the capitalization of the largest cryptocurrencies dropped to just a fraction of what it used to be at their heights in November 2017. Hence, the criminal interest in exploiting them decreased significantly. Nonetheless, the given examples shed light on some security concerns towards the use of block-chain technology in general and provide some insights into how a technology, especially designed to be transparent and secure, could become a victim of its own success.

Artificial Intelligence as a Likewise Upcoming IT-Security Concern and Hope

Artificial Intelligence is not only an extraordinary interesting and fast developing part of information technology, it also has tremendous impact on a vast amount of actually running applications which themselves play important roles in global economic development.

The growing number of success stories published in scientific literature and even in mainstream daily news even accelerates the pace of its introduction into everyday life. There are manifold examples of AI applications: Internet users are presented search results generated by Machine Learning (ML) algorithms, insurances use it to determine their fees, banks to decide on buying and selling stocks and even the police uses it for predictive analytics to assess in which areas burglary is most likely to happen in the near future.

Vulnerabilities of AI could therefore result in a large attack surface. To clarify the extent of impacts that could be expected, the CSP 2019 presents basic aspects of AI's general vulnerability and shows some examples in this regard. Another spot is thrown on the use of AI as a weapon for attack as well as for defensive purposes.

The assessment of this CSP is that AI already has impact on cybersecurity, has many applications in private and economic life and hence will definitely be a corner stone technology in future information security. Therefore, high awareness of the ongoing development is advised. Crucial factors are high expertise and continuous monitoring of related hardware, software and supply chain developments.

THE STATE OF CRYPTOCURRENCY CRIME: FOLLOW-UP

In the first Common Situational Picture (CSP), BSI and ANSSI focused on ransomware and cryptocurrency cybercrime. The enormous stock price increase of many cryptocurrencies in 2017 and early 2018 fed the development of new trends for lucrative tactics of cybercrime, such as cryptojacking.

In the period from January 2017 to January 2018, Bitcoin stock prices rose from 900 to 13.000 EUR (+1440%). However, this trend has not lasted, as the burst of the speculative bubble resulted in a massive drop of all cryptocurrencies prices in the first quarter of 2018. For instance, the Bitcoin stock price dropped from around 16.000 EUR (January 2018) to 3.500 EUR (March 2019), which denotes an approximate loss of 80%. The possible causes for the global cryptocurrencies drop and high volatility in 2018 were the media attention bubble – which attracted inexperienced investors –, international efforts for new cryptocurrencies legislation such as profit taxes or money laundering, diversification of cryptocurrency types, confusion about block-chain forks, energy consumption critics and reports of security incidents. These factors are likely to have influenced the investors who began to withdraw their funds that caused a domino effect leading to the drop.

A few months after the drop, ANSSI and BSI noted a significant decrease of cryptocurrency crime, especially those based on cryptojacking techniques. As observed in the first edition of the CSP, Monero cryptocurrency is mainly used in cryptojacking. ANSSI and BSI noticed that web cryptojacking[1] became non-profitable and most of the malware web miners stopped their activity due to the decrease of Monero stock prices[2].

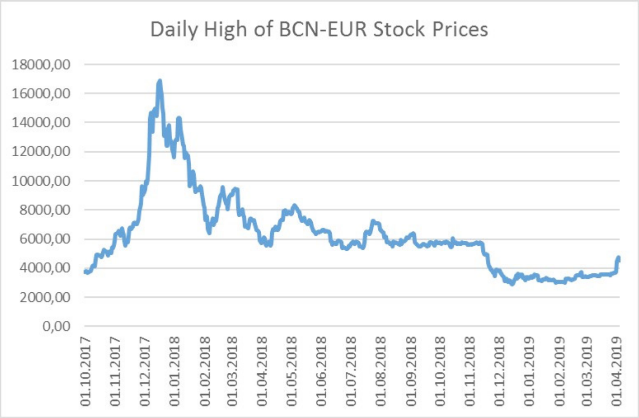


***Figure 1:** Development of the exchange rate for Bitcoin-Euro 01.10.2017-04.04.2019*

Although some cryptocurrency crime trends, such as cryptojacking, decreased in 2018 and early 2019, other tendencies or types of attack are still used or have even increased. Some of them are described in the following part[3].

Types of attacks

The block-chain integrity is theoretically secured and guaranteed by cryptography protocols. However, there are techniques for bypassing the mechanism of transaction validation, such as the use of third-party software or services (mining pool, stock exchanges, "mixers", wallet software, private key storage) exposing the cryptocurrencies to various attacks.

▶ 51% - Attacks – the limits of block-chain integrity

The consensus mechanism, called "proof of work", is fundamental to all the miners creating new blocks of the block-chain. But in the case of a miner providing more than 50% of the overall mining power (total hash rate), it can create new blocks with convenient contents for the attacker and

[1] Technique which targets legitimate and popular websites, on which users are expected to stay several time. The attackers often use a JavaScript insertion to include malicious tools that force the visitors' Internet browser to mine cryptocurrencies.

[2] Marius Musch, Web-based Cryptojacking in the Wild. Chaos Communication Congress, Leipzig 2018.

[3] SIX Financial Information via https://www.finanzen.net

attach it to preceding blocks. This is called a 51%-attack. Because of its computing power majority, the attacker's chain would be longer than other node's chains and would thus be accepted as the valid chain. A 51% attack offers different profitable malicious possibilities, such as "double spending" using the controlled chain to return already accepted transactions to the pool of non-validated transactions which can consequently be spent a second time.

Example. *16.-19.05.2018: The currency Bitcoin Gold which was forked from Bitcoin in October 2017, was hit several times by a 51%-attack. The unknown attackers focused their double spends on cryptocurrency exchanges with high volumes. The potential damage amounts to 18 Mio. USD. Consequently, the Bitcoin Gold hash algorithm was updated.*

▶ Selfish mining – a threat on "minor" cryptocurrencies

A selfish mining attack requires significant mining capabilities of a block-chain, but not necessarily 51%. Selfish mining focuses on the reward provided by the proof-of-work protocol. The attacker publishes new blocks with exceptional delay and is free to work on the next block based on its own. Meanwhile, the other miners waste their mining power on the first. By doing so, the attacker gains an asymmetric advantage generating the most attractive branch of the chain. Referring to Eyal and Sirer [4], only a 25% ownership of the overall computing power is necessary for a successful selfish mining attack. On "minor" cryptocurrencies with a low total "hashrate" - the mining capacity allowed by all the miners - 25% or 51% can be acquired through cloud mining.

Example. *13.-15.05.2018: A special attack scenario threatened owners of the Japanese cryptocurrency Monacoin. The selfish mining attack caused damage of 90.000 USD, as the attacker sent coins to other exchanges receiving a purchase, but invalidating the transaction afterwards.*

▶ Cryptocurrency scams – one of the most important trends in cryptocurrency crime

Beside the attack on the consensus protocol, many attackers find a way to scam by using third-party services or software. For example, a scam occurred for the cryptocurrency IOTA. Over a period of several months, a free service for private key creation was provided on the unofficial website iotaseed.io. During this time, the credentials were saved by the conductors. Finally, the attackers used the collected information to plunder all IOTA wallets of their former customers [5]. This is one of various scams on cryptocurrencies by which private persons or businesses can be affected.

▶ Attacks on trading exchanges – the most profitable trend

Trading platforms are a third-party service which allow cryptocurrency users to trade between crypto- and conventional fiat currencies. Those platforms are comparable to online-banking. These services are currently the main target for advanced criminal individuals or groups. In 2017 and 2018, at least five platforms were partly compromised and their funds were stolen. The record loss occurred at the Japanese trading exchange Coincheck. The attackers stole 470 Mio. EUR worth of the cryptocurrency NEM.

According to various research groups[6, 7] , targeting trading exchange became a major trend in 2018 with a total loss of about 1 billion USD worth of cryptocurrencies.

[4] I. Eyal and E. Sirer: Majority Is Not Enough: Bitcoin Mining Is Vulnerable. Lecture Notes in Computer Science 8437, S.436–454, 2014, https://arxiv.org/pdf/1311.0243.pdf

[5] McAffee: Blockchain-Threats-Report. https://www.mcafee.com/enterprise/en-us/assets/reports/rp-blockchain-security-risks.pdf. 2018.

[6] Ciphertrace, Cryptocurrency Anti-Money Laudering Reports, 2018 Q3.

[7] ChainAnalysis, Crypto Crime Report, Decoding increasingly sophisticated hacks, darkent markets, and scams. January 2019.

Threat evolution in the near future

In order to anticipate new cyber crime trends or cyber threats, it is recommended to have a look at trends in the cryptocurrency sphere that can be expected in the near future.

In the first CSP, ANSSI and BSI explained cryptojacking as an aspect of cryptocurrency crime consisting of stealing computing power – and energy cost – from a victim in order to mine. Monero is massively used for this type of attacks, because of the mining protocol and its anonymity-guaranteed transactions. The privacy coins (Monero, ZedCash, Verge, and a few other minor currencies) are increasingly used for cybercriminal activities, such as illicit business on darknet markets. Indeed, new major markets like Alphabay and Hansa (both closed in 2018) offered payment with Bitcoin and Monero. Until 2018, major cryptocurrencies or illegal activities were "pseudonymous", that means they were public and traceable. Currently, security services begin to rely on block-chain analysis tools, which forces cyber criminals to use only privacy coin as Monero, since it is fully encrypted and untraceable.

In response to mining concentration issues (51% attack, selfish mining) and to energy consumption critics, new cryptocurrencies with different mining protocols have been created. Proof-of-stake (PoS) "mining" could replace Proof-of-Work (PoW) in the coming years. New PoS cryptocurrencies (e.g. Tezos, EOS) rely on a mining system, where the user is required to prove the ownership of a certain quantity of cryptocurrency in order to be authorized to validate additional blocks. If a user gets detected by validating a false block, he loses his stored cryptocurrency. In the future it can be assessed that cyber criminals may develop news techniques based on the PoS protocol or PoS cryptocurrencies third-party services and software.

Following multiple thefts targeting trading exchanges in 2018, the use of distributed trading exchange became standard. Instead of storing all private keys in an easily targeted storage, it is recommended to rather use a connection protocol allowing the users to keep their private key. Should distributed exchange become a standard, it is likely that cyber criminals and other groups would pivot to other attack techniques.

ARTIFICIAL INTELLIGENCE AFFECTS CYBER-SECURITY

The emergence of ubiquitous Artificial Intelligence (AI) applications is today one of the main driving forces in the digital transformation of societies. It is therefore necessary to try to identify and understand the potential implications of AI for cybersecurity, as both fields will intersect and present a cornerstone in the technical foundations of future societies. In this regard, several initiatives both on national and European level are already addressing this intersection of AI and cybersecurity, such as:

- In France, an important program has been announced recently on "**How to secure, certify and make reliable the systems involving AI?**" [8]. A Program Director (recruited since early 2019 in the Services of the Prime Minister) will be responsible for the advancement of this challenge, whose thematic focus was chosen as a follow-up to the consultations lead during the preparation of the Villani report.[9]

- The German Ministry of Education and Research launched a sponsorship for "**Künstliche Intelligenz für IT-Sicherheit**".[10]

- SPARTA is a novel Cybersecurity Competence Network, where both ANSSI and BSI are involved, supported by the EU's Horizon2020 program, with the objective to develop and implement top-tier research and innovation collaborative actions. Among the concrete challenges at the core of SPARTA research Roadmap is the SAFAIR program which will devise approaches to make systems using AI more reliable and resilient through enhanced explainability and better threat understanding[11].

- During the 6th Forum of Franco-German Research Cooperation, privacy-preserving AI (most notably the design of privacy-preserving machine learning algorithms) and reliable architectures have been identified as essential issues for European digital sovereignty[12]. This is only indirectly a cybersecurity topic, but it represents an important side-effect when combined with the fact that AI algorithms are widely used for biometric and continuous identification. Already today AI techniques can be used to bypass anonymization.

The key question is therefore how the ubiquitous introduction of AI is likely to shape the equilibrium in the domain of cybersecurity between offensive and defensive aspects. On the one hand, AI systems are imperfect and their current weaknesses could be exploited as vulnerabilities by cyber-malicious actors. AI techniques could also be exploited to enhance cyber-offensive capabilities. On the other hand, AI presents an unprecedented opportunity to generate many powerful applications.

As the drive for AI deployment is extremely strong today, it has to be understood how to act in order to collectively reinforce cybersecurity. In the following part, the CSP 2019 will therefore elaborate on some examples to show various influences AI techniques have on the security of present digital life and the process of digital transformation. Before doing so, a very general overview of what is actually meant by AI will be given.

[8] https://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2018/09/certification_ia.pdf (for the description of this challenge see page 3).

[9] https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

[10] https://www.bmbf.de/foerderungen/bekanntmachung-2187.html

[11] https://ssi.gouv.fr/uploads/2019/02/press-release-sparta.pdf

[12] p. 38 of https://www.bmbf.de/upload_filestore/pub/BMBF_DF_FF_Dokumentation.pdf

AI

The ongoing tremendous hype surrounding AI was mainly induced by recent success stories in a sub-field of AI called Machine Learning (ML) and especially in Deep Learning (DL). In DL, learning is performed by Neural Networks (NN) containing many layers of artificial neurons and trained on a huge amount of data. This is why DL could only emerge once large data sets and foremost enough computing power (notably GPUs) were available.

In this context, two striking examples are the following: the first case was in 2012 in the domain of image recognition for which DL techniques allowed to significantly diminish the error rate and allegedly even surpassed human capabilities[13]. The second example refers to the domain of games, when in 2016 the first victory of the program AlphaGo[14] against the champion Lee Sedol took place.

AI is obviously not limited to DL and has in fact a long and cyclic history, dating back to pioneering works in the field of cybernetics in the 1940s. Since then, several approaches have prevailed in AI, some of them based on symbolic rules, some of them on machine learning, partially being bio-inspired. This background information should be kept in mind in order to put the following parts in perspective ans grasp thus a bigger picture in the context of AI.

The CSP 2019 will mainly address topics related to ML, as the dependence of these algorithms on data make them particularly interesting to study them from the cybersecurity point of view.

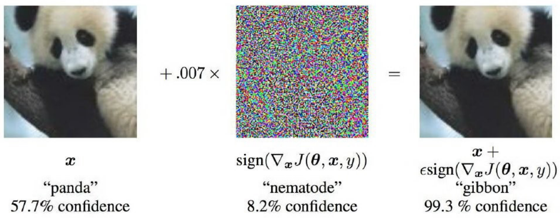
NN trained on vast amount of data can be assessed as a paradigm change in many contexts. In fact, the knowledge of those trained NN is encoded in a myriad of parameters (i.e. numbers) representing implicitly the statistical structure of the data. Only implicit knowledge can be produced this way and it will prove extremely difficult to try to understand explicitly the main determinants of a decision (e.g. by looking at all these parameters), which can be interpreted as lack of transparency and a lack of semantics.

Eventually, despite the huge generalization power of NN, it cannot be known in advance what will be the behavior of NN under new circumstance and especially in rare cases. Even if their application is operationally tested for a certain time, it can only be observed that the process works successfully for the presented amount of data, which is usually similar to the test and training data set. This can be interpreted as a reliability issue as a reliable AI system must be able to work properly with a large enough range of inputs and situations.

The efforts should be put on developing "trustworthy AI", as emphasized by the European Union[15], underlining in particular the importance of transparency and reliability.

Some vulnerabilities of AI

■ AI systems are vulnerable to adversarial attacks, i.e. imperceptible (for a human observer) variations of legitimate examples crafted to deliberately mislead a ML algorithm. A well-known example for this vulnerability is the example of a panda bear, that is detected as gibbon with high confidence after the application of "designed noise", shown in the following image[16]:



+ .007 ×
=
x
"panda"
57.7% confidence
$\text{sign}(\nabla_x J(\theta, x, y))$
"nematode"
8.2% confidence
$x + \epsilon\text{sign}(\nabla_x J(\theta, x, y))$
"gibbon"
99.3 % confidence

[13] https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-networks

[14] AlphaGo is based on a deep convolutional network to guide its exploration of the game, but also on Monte-Carlo methods and reinforcement learning.

[15] Cf. The "Ethics Guidelines for Trustworthy AI" prepared by the High-Level Group on Artificial Intelligence set up by the European Commission https://bdi.eu/media/themenfelder/digitalisierung/publikationen/20190201_Stellungnahme_BDI_Draft_Ethics_Guidelines_for_Trustworthy_AI.pdf: it includes notably seven key requirements for the realization of Trustworthy AI among those technical robustness and safety and transparency.

[16] Goodfellow IJ, Shlens J, Szegedy C (2014) Explaining and harnessing adversarial examples. arXiv. Available online at: https://arxiv.org/pdf/1412.6572.pdf).

These attacks have been demonstrated for a variety of input signals (image, video, sound)[17]. Hereby the principle always remains the same: intentionally disturbed input data cause the activation of areas of the input space of the neural network, which leads to mis-classification of the content or (even worse) a targeted and allegedly reliable (but false) classification. This should be harmless for e.g. an application that labels images for a textual search in a large image database, because failure would only mean that images are not presented in the most useful order. But as soon as security and safety critical applications like autonomous car control, border control or analysis of medical data are addressed, the qualitatively new attack vectors due to AI usage have to be taken into account and proper mitigations to those attacks must be designed [18]and implemented.

Vulnerabilities of AI-Systems, especially qualitatively new and AI-specific attack vectors, necessitate special care when deploying them. Especially for security and safety critical AI-applications effective defense strategies and evaluation methods have to be developed.

- Moreover, there is a danger of **accidental failure**. In another example, an algorithm was trained with many images of huskies and wolves. It allegedly could distinguish between both kinds of "dogs". However, a detailed analysis showed that differentiation was mainly influenced by the presence of snow in the background instead of the characteristics of the animal. In security contexts even more hidden correlations cannot remain undetected.

This underlines the importance to develop interpretability and explainability of AI systems.

- Widely spread ML methods are currently shared in a large developer community. Not all of these developers have access to those large amounts of data that are available to big players like Google and Facebook. To avoid a broad distribution of data and to reduce the time for training, the deployment of pretrained generic models has been established. These models are trained for example with a large text corpus provided by the owner of this corpus. A user subsequently has to train only with a small corpus to achieve his specific aims. Exchange of these pretrained NN is known as Transfer Learning (TL) and presents a way to implicitly disseminate knowledge to customers. Like other types of information, this transferred knowledge may be manipulated to aim at causing malicious effects. This process is known as **data poisoning** and may be performed by all participants in the supply chain including the producer of transferred data sets. It should also be mentioned that the producer of transferred data may leak internal information remaining in the nontransparent data. Intruders may find a handful of ways to intermittently squeeze corrupted data. Although exchange of knowledge is a valuable asset in this fast evolving area, a negative aspect is, from an information security point of view, that much of this data is widely shared on freely accessible developer platforms (like Github) and therefore many players are able to change it in open supply chains. Another scenario is the continuous training of a NN by input from user interactions in the internet. If the same NN produces transfer data used by other ML processes, manipulation of input data with appropriate chat bots is possible as well. These robots may use AI themselves in a malicious manner (see below).

Open supply chains and Transfer Learning may be misused by adversaries poisoning data or trying to manipulate available models.

- The proliferation of AI systems may open

[17] It seems a priori more difficult to design adversarial examples adapted to data with other types of structure, like the ones encountered in classical cyber-security context (net traffic data, logs, etc.).

[18] For a recent paper on robustification against adversarial attacks, see: "Robust Neural networks using Randomized Adversarial Training" (https://arxiv.org/abs/1903.10219). However, as robustification is mostly based on randomization, this can be to the detriment to the NN's accuracy.

new channels for malicious use. Indeed, as soon as ML methods are considered strong enough to allow basic natural communication via channels like images or sounds (speech recognition), those media could be established as interface to control computers or IoT devices (e.g in home automation solutions). While software had been vulnerable in the past due to aspects like stack overflow issues, functions with exploitable source code or hidden and forgotten features, these new channels can affect the information level and are harder to explore, detect and mitigate. One of the exploits of this kind has been published as the "dolphin attack"[19] and consists in the emission of voice commands on ultrasonic wave length (e.g. frequencies > 20 kHz) for personal assistant systems like Siri and Alexa. These commands may be sent by other IoT-devices like smart TVs and cannot be detected by a listening human user due to the high frequencies. The same may be possible for optical devices in the infrared area. In the case when many operating systems work together unforeseen effects may appear like voice squatting, i.e. phonetically similar commands that are added by third party companies.

These are not vulnerabilities of AI systems per se, but this shows an extension of the global attack surface when those systems are used as tools to control other information systems. The risk is that a wide usage of such coupling may lead to a broad (illusory) trust.

AI as weapon / malware

Some applications of Adversarial Machine Learning that appeared in recent years have the potential to produce diverse threats to cybersecurity.

- It has been shown that image recognition can be fooled to detect things, that have not been presented to it. Indeed, there is a research field called **Generative Adversarial Networks (GAN)** representing another specialized trending topic for the developer community within the general field of AI. Although not designed for attacks (as the name may suggest), GANs are used to find strategies encountering the failure of a counterpart NN. This technique may be abused like other technical instruments. Therefore, they may be used to find images that intentionally fool implemented recognition systems, or to generate artificial fingerprints or morphed facial images to pass as authentication for many users. Although GANs serve as tools to exploit weaknesses of Machine Learning methods, they unfortunately cannot avoid the lacking transparency of such methods due to their own lack of transparency[20].

GANs are a good example of the inherent dual nature of most AI techniques.

- Similar effects may be achieved using more classical machine learning methods. In a well known example, a face recognition algorithm is fooled by eyeglasses with an adversarial printed glasses frame which leads to the fact that in this specific targeted attack, one person is recognized as another one.[21]

AI can be attacked by AI and many people work on that field although not interested in malicious use.

- A widely spread element in attack vectors is social engineering. In the near future AI-generated "Deep Fake Videos" may play a stronger role in that game. They present known and unknown people in videos to witness actions that never have taken place. To do this, AI makes it possible to present other actors with faces of the targeted person. This

[19] https://www.heise.de/forum/heise-online/News-Kommentare/Amazon-will-Alexa-das-unkontrollierte-Lachen-austreiben/DolphinAttack-gegen-Alexa/posting-31999888/show/

[20] Karras, T.; Laine, S. & Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks.

[21] Sharif, Bhagavatula, Bauer and Reiter: Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16). ACM, 2016

could be exploited by attackers in video identification processes or to discredit people. Even when it can be shown subsequently that this video material is not authentic, it can have disastrous effects in public processes like elections, dishonor publicly known persons or simply serve for extortion of ransom.

Attacks by social engineering, where false documents are created by AI in order to give a false sense of confidence to the target, and more broadly disinformation is supported by AI methods and serve as examples of how security is more often attacked on the semantic instead on the technical layer.

■ Another approach to leverage AI as malicious tool is to circumvent security features like HIP (human interaction proof). They use the complex task of image recognition to verify interactions of human origin in the internet. For example, this test is used to avoid automatic account creation for internet applications that need a personal counterpart. They often use an image of arbitrary characters that are disturbed, so that an ordinary Optical Character Recognition (OCR) is not capable to "read the text". With growing capabilities of AI, this kind of protection gets weaker. Another more simple example may be the sorting of exposed password databases according the features of the application it is used for, e.g. the password "tre$or123" may be used more likely for a banking application than for a video streaming portal. The possibility of such a usage of ML gets more probable with the growing number of breaches.

Classical and widespread security methods may be weakened by AI.

■ AI may also be used to de-anonymize users and devices in combination with big data and/or physical fingerprinting, i.e. AI may be used to extract and combine data from multiple sources for identification purposes. The later mentioned side-channel attacks can be used to compromise existing hardware to encrypt and decrypt data.

■ Often malware has the problem to run on systems with a great diversity of properties. Some systems have intrusion detection systems, some have AV-software and others do not support the commands that usually work or have very restrictive policies. To adapt to those different circumstances, learning software may be more appropriate to meet these requirements. A back channel to the developer of the malware may provide the training data for newer versions. These data do not necessarily feed a ML procedure but there is more to AI than ML. Even sophisticated knowledge bases that are provided with more logical input are a part of modern AI. AI may therefore serve to act more autonomously.

Malware may be less dependent on external support (C&C server) and therefore detection might be more difficult.

AI as tool of defense

On the other side, several AI applications have been introduced that support the fight against cyberthreats.

AI methods, combined with existing approaches in cybersecurity, have the potential to enhance the defense in all its different phases: in the development and evaluation of products, in the detection of attacks, and in the remediation phase (at least with decision support tools).

Due to the huge amount of data cyber analysts are confronted with and the growing number and sophistication of attacks, AI methods could prove very useful, whenever they are able to partially automate the work of defenders.

■ Modern AV-software detects malware not only by signatures, but also by Machine Learning algorithms that have been trained by thousands of known malicious samples. Many parameters of the software may serve as indicator, partly even the structure of the code. Therefore, threat actors are forced to be highly innovative and must vary their code more rigorously to avoid detection.

AI aides thus in malware detection.

■ The observation of net traffic produces large amounts of data. These can be used to find anomalies that are characteristic to attacks. Although success is not easily demonstrated, many companies work in that field. These techniques may be applied in a local (enterprise, organization) or global context(Internet Service Providers (ISPs) as well as in Content Delivery Networks (CDN)).

AI finds anomalies in net traffic.

■ One of the oldest applications of AI is the detection of spam. When in earlier times the existence of keywords was enough to find spam mails, younger waves of spam and more dangerous mails with attachments and URLs are more elaborated. Nevertheless, modern filters are able to find them as long as the algorithms are constantly trained.

AI helps to avoid certain threat vectors.

■ AI may be used to detect attacks on biometric identification systems such as face and fingerprint recognition systems. Robust detection of morphed facial images is, for example, a challenging task, but employing AI combined with sufficient training and test data results in robust morphing attack detection.

With AI it is possible to detect fraud.

■ Modern cryptographic hardware uses complex algorithms that are difficult to examine by the observation of side-channels. Nevertheless, AI methods have been successfully used to show the weakness of such devices. Therein the analyst uses additional information like power-consumption or execution time to extract the secret key used by the device. Classical methods are prone to inaccuracies in the model assumptions such as alignment and the noise distribution. Side-channel attacks based on ML techniques are less sensitive against deviations from the model assumptions. Therefore, they are more robust and need less preprocessing. All AI tools used as weapons (see chapter above) are helpful to examine software and hardware systems to evaluate their security.

AI aides to examine hardware and software and to harden them.

■ Last but not least, the enormous amount of information found in the Web can be exploited to enhance situational awareness. Currently, the interpretation of text by ML algorithms shows ongoing progress. In this regard, the automatic extraction of entities and enrichment of the text corpus is for example possible. This is also a subject of current research.

Enhanced awareness by text processing is possible though AI.

Many more future applications are conceivable, e.g. continuous authentication of users scanning the behavior, automatic detection of data breaches by observing login frequencies, prediction of future attack waves and more.

AI is a tool for many future applications.

Conclusion

Concerning the topic of AI both agencies already started discussing its challenges and opportunities with regard to cybersecurity. Furthermore, both ANSSI and BSI are together engaged, directly or indirectly, in AI projects on European level such as SPARTA which is part of the Horizon2020 strategic research and development fund. This topic is just one of numerous projects both agencies are leveraging to deepen even further their close cooperation such as on R&D topics as well as on the technical and operational level.

In reflection of the width and depth of this bilateral, but not exclusive, Franco-German collaboration, the Common Situational Picture will pursue its goals of presenting a tangible and high visible output, raising awareness among readership on ongoing topics related to the cybersphere, and sweep the cyber landscape looking for current or upcoming trends, threats and opportunities.

Agence nationale de la sécurité
des systèmes d'information
51, boulevard de La Tour-Maubourg
75700 Paris 07 SP

Bundesamt für Sicherheit in
der Informationstechnik
Postfach 200363
D 53133 Bonn