Proposal No. 830929
Call H2020-SU-ICT-03-2018

Project start: February 1, 2019
Project duration: 42 months

Cyber Security for Europe

# D3.10

# Cybersecurity outlook 1

| Document Identification | |
|---|---|
| Due date | 30 September 2020 |
| Submission date | 29 September 2020 |
| Revision | 1.0 |

| | | | |
|---|---|---|---|
| Related WP | WP3 | Dissemination Level | PU |
| Lead Participant | POLITO | Lead Author | Daniele Canavese (POLITO) |
| Contributing Beneficiaries | GUF, POLITO, UMU | Related Deliverables | - |

**Abstract:** This deliverable describes the state-of-the-art, the current trends that are currently emerging in the cybersecurity field. This document is the first outcome of Task 3.9 'Continuous scouting', whose goal is to constantly look for new outgoing research challenges and developments that can provide interesting ideas not only to the academic partners, but also to the use case owners. Henceforth, this document reports a variety of new advancements, and emerging security technologies in various branches, such as artificial intelligence, 5G applications and trusted execution environments, that, in the near future, could play a pivotal role in our daily lives, but, at least for now, provide us with new hints and food for thoughts.

This document is issued within the CyberSec4Europe project. This project has received funding from the European Union's Horizon 2020 Programme under grant agreement no. 830929. 

## Executive Summary

This deliverable reports the outcomes of Task 3.9 on Continuous Scouting. The goal of this task is to constantly analyze emerging technologies and new trends in the cybersecurity field in order to give both academic and industrial partners interesting food for thoughts for the future.

This document describes four IT branches that have seen several interesting developments and applications in the cybersecurity area in the last couple of years:

- *intelligence techniques*, mostly based on machine learning models, can be used to handle threats in a timely and privacy-preserving manner with the least amount of human interaction possible;
- *artificial intelligence* is frequently used to secure digital assets, but an attacker can also use them for a variety of purposes such as to confuse a machine learning system in order perform the wrong action when an input is received (e.g. misclassify a cyberattack) or help an evildoer to automatically gather information about a victim (i.e. for social engineering purposes);
- *zero-trust security systems* can be used to create a more secure environments and the last few years have seen the developments of new technologies and the discovery of state-of-the-art attacks;
- *5G*, the new wireless standard, promises to make the world more connected than ever since it is being adopted in various embedded and Internet-of-Things devices.

This document reports not only new trends, attacks and emerging developments in the research world, but it also discusses how these state-of-the-art technologies are deployed and used by the current industries.

# Document information

## Contributors

| Name | Partner |
|---|---|
| Daniele Canavese | POLITO |
| Antonio Lioy | POLITO |
| Ignazio Pedone | POLITO |
| Leonardo Regano | POLITO |
| Majid Hatamian | GUF |
| Sascha Löbner | GUF |
| Sebastian Pape | GUF |
| Narges Arastouei | GUF |
| Antonio Skarmeta | UMU |
| Alba Hita | UMU |
| Jorge Bernal | UMU |

## Reviewers

| Name | Partner |
|---|---|
| Kimmo Halunen | VTT |
| João Resende | C3P |

## History

| | | | |
|---|---|---|---|
| 0.01 | 2020-07-03 | Daniele Canavese, Antonio Lioy | 1st draft |
| 0.02 | 2020-07-23 | Antonio Skarmeta, Alba Hita, Jorge Bernal | Added the 5G chapter |
| 0.03 | 2020-07-27 | Sascha Löbner | Added the intelligence GDPR section |
| 0.04 | 2020-07-27 | Ignazio Pedone | Added the ZTS systems chapter |
| 0.05 | 2020-07-29 | Leonardo Regano | Added the threat detection section |
| 0.06 | 2020-07-29 | Majid Hatamian, Sebastian Pape | Added the adversarial AI chapter |
| 0.07 | 2020-08-05 | Daniele Canavese | Minor cleanup, added the abstract and the executive summary |
| 0.08 | 2020-08-06 | Daniele Canavese | Added the introduction |
| 0.09 | 2020-08-07 | Daniele Canavese | Major restructuring of the deliverable |
| 0.10 | 2020-08-07 | Daniele Canavese | Cleanup of Chapters 1 and 2 |
| 0.11 | 2020-08-08 | Daniele Canavese | Cleanup of Chapters 3, 4 and 5 |
| 0.12 | 2020-08-08 | Daniele Canavese | Added the conclusion chapter |
| 0.13 | 2020-08-28 | Ignazio Pedone, Daniele Canavese | Major refactoring of Chapter 4 |
| 0.14 | 2020-09-02 | Daniele Canavese | Major cleanups here and there |
| 0.15 | 2020-09-04 | Daniele Canavese | Minor fixes and bibliography cleanup |
| 0.16 | 2020-09-18 | Daniele Canavese | Fixes after internal review |
| 0.17 | 2020-09-18 | Sebastian Pape | Updates to Section 3.2.2 |
| 0.18 | 2020-09-21 | Sascha Löbner | Updates to Section 3.1.3 |

| | | | |
|---|---|---|---|
| 0.19 | 2020-09-24 | Daniele Canavese | Minor cleanup |
| 1.0 | 2020-09-29 | Narges Arastouei, Daniele Canavese | Final check before submission |

# List of contents

# List of figures

# List of tables

# List of acronyms

| | |
|---|---|
| 3GPP | 3rd Generation Partnership Project |
| 5GPPP | 5G infrastructure Public Private Partnership |
| AAA | Authentication, Authorization and Accounting |
| ACM | Authenticated Code Modules |
| AI | Artificial Intelligence |
| AIK | Attestation Identity Key |
| ANN | Artificial Neural Network |
| API | Application Programming Interface |
| ARPF | Authentication credential Repository and Processing Function |
| AUC | Area Under Curve |
| AUSF | AUthentication Server Function |
| BIOS | Basic Input-Output System |
| C&C | Command & Control |
| CoT | Chain of Trust |
| CPU | Central Processing Unit |
| CSRF | Cross-Site Request Forgery |
| DDoS | Distributed Denial-of-Service |
| DFDC | Deepfake Detection Challenge |
| DGA | Domain Generation Algorithm |
| DNN | Deep Neuronal Network |
| DSM | Digital Single Market |
| DT | Decision Tree |
| EA | Enhanced Authorization |
| EAP | Extensible Authentication Protocol |
| EC | European Commission |
| EK | Endorsement Key |
| eMBB | enhanced Mobile BroadBand |
| ENISA | European union agency for cybersecurity |
| ETSI | European Telecommunications Standard Institute |
| EU | European Union |
| FCNN | Fully Connected Neural Network |
| GAN | Generative Adversarial Network |
| GDPR | General Data Protection Regulation |
| HIPAA | Health Insurance Portability and Accountability Act |
| HMEE | Hardware Mediated Execution Enclave |
| HTTP | HyperText Transfer Protocol |
| ICS | Industrial Control Systems |
| ICT | Information and Communications Technology |
| IoT | Internet-of-Things |
| IP | Internet Protocol |
| ISA | Instruction Set Architecture |

| | |
|---|---|
| k-NN | k-Nearest Neighbors |
| LIME | Local Interpretable Model-agnostic Explanations |
| ME | Mobile Equipment |
| MIMO | Multiple-Input Multiple-Output |
| MitM | Man in the Middle |
| ML | Machine Learning |
| mMTC | Massive Machine Type Communication |
| MPC | Secure MultiParty Computation |
| MS | Member States |
| MTM | Mobile Trusted Module |
| N3IWF | Non-3gpp InterWorking Function |
| NFV | Network Function Virtualisation |
| NGMN | Next Generation Mobile Networks |
| NIS | Network and Information Security (directive) |
| NIST | National Institute of Standards and Technology |
| OS | Operating System |
| OT | Operational Technology |
| PMP | Physical Memory Protection |
| PPML | Privacy Preserving Machine Learning |
| RA | Remote Attestation |
| RAM | Random Access Memory |
| REE | Rich Execution Environment |
| RF | Random Forest |
| RL | Reinforcement Learning |
| RNN | Recurrent Neural Network |
| ROC | Receiver Operating Characteristic |
| ROP | Return Oriented Programming |
| RT | RunTime |
| SDN | Software-Defined Networking |
| SEAF | SEcurity Anchor Function |
| SGX | Software Guard Extensions |
| SIDF | Subscription Identifier De-concealing Function |
| SIM | Subscriber Identity Module |
| SL | Supervised Learning |
| SME | Small and Medium-sized Enterprise |
| SQL | Structured Query Language |
| SVM | Support Vector Machine |
| SM | Security Monitor |
| TA | Trusted Application |
| TC | Trusted Computing |
| TCG | Trusted Computing Group |
| TCP | Transmission Control Protocol |
| TEE | Trusted Execution Environment |

| | |
|---|---|
| TPM | Trusted Platform Module |
| TSG-SA | Technical Specification Group of Services and system Aspects |
| TXT | Trusted Execution Technology |
| UDP | User Datagram Protocol |
| UDM | Unified Data Management |
| UE | User Equipment |
| UL | Unsupervised Learning |
| URLLC | Ultra-Reliable Low Latency Communications |
| USIM | Universal Subscriber Identity Module |
| VANET | Vehicular Ad-hoc NETwork |
| VM | Virtual Machine |
| VNF | Virtualised Network Function |
| VPN | Virtual Private Network |
| vRoT | virtual Root of Trust |
| vTPM | virtual Trusted Platform Module |
| WAF | Web Application Firewall |
| XSS | Cross-Site Scripting |
| ZTS | Zero-Trust Security |

# 1 Introduction

In our fast-moving world, life quality is largely determined by the technological level that an individual can directly and indirectly access. Computers, Internet-of-things and embedded devices are ubiquitous. They surround us in our working place, in our homes, on the streets and even in our cars. With so much hardware and software close to our personal lives, a cyberattack can threaten us closer than ever. The first and last line of defense against these threats is given by many new cybersecurity techniques that are constantly emerging and evolving.

Due to the ever-changing quality of technology it is important to be constantly prepared against new form of menaces. In this document we report a variety of new cybersecurity technologies, trends and issues that have emerged in the last couple of years and that can become game changers soon. We report the current state-of-the-art in various cybersecurity fields and, where applicable, we also discuss what are the challenges for their deployment.

The investigation documented in this report is useful not only for researchers, but also for the people more directly involved in the industry as it can yield interesting food for thought for real world applications.

This document contains several chapters, each one devoted to a specific cybersecurity field:

- in Chapter 2 *Threat intelligence* we analyze what are the most recent trends and machine-learning techniques that can be used to detect network and malware attacks, discussing also their applicability and their privacy issues according to the recent GDPR [GDPR 2016];
- in Chapter 3 *AI for adversarial purposes* we investigate how various artificial intelligence techniques can be exploited by an evildoer to craft ad-hoc samples to purposely confuse various types of detection systems and how they can use AI-based approaches to automate various aspects of social engineering attacks (e.g. phishing);
- in Chapter 4 *Zero-trust security systems* we discuss what are the most recent zero-trust security technologies that can be used to create tamper-free execution environments, and, in addition, we also report some new (worrying) attacks against these technologies;
- in Chapter 5 *5G applications* we analyze the security of 5G networks, what are their weaknesses, and how artificial intelligence, hardware protections and various other approaches can be used to strengthen this soon-to-be pervasive technology.

Finally, in Chapter 6 we present our conclusions and in Chapter 7 we list all our references in alphabetical order.

# 2 Threat intelligence

*Threat intelligence* is a broad field in the heterogeneous realm of cybersecurity where information about attacks and attackers is gathered in a manual or automated fashion. The standard approach over the last years has been the manual analysis of security related events. However, the exponential growth of data produced by organizations makes the adoption of artificial intelligence and automatic techniques for cybersecurity purposes a necessity. Machine and deep learning techniques have been recently applied in a plethora of real-life scenarios.

## 2.1 Threat detection

The NIST (National Institute of Standards and Technology) defines a *threat*[1] as "any circumstance or event with the potential to adversely impact organizational operations (including mission, functions, image, or reputation), organizational assets, or individuals through an information system via unauthorized access, destruction, disclosure, modification of information, and/or denial of service". Identifying threats in a timely fashion is of paramount importance for organizations, since it is the only way to appropriately respond to such menaces and thus limit the related damages. In this section we present a selection of these use cases, explaining the use of threat intelligence techniques for the automatic configuration of web application firewalls, the protection of Internet-of-things appliances against botnet infections and the detection of malware threats.

### 2.1.1 Web application firewalls

*Web Application Firewalls* (WAF) are security appliances employed on web servers to monitor the HTTP (HyperText Transfer Protocol) traffic exchanged between the hosted applications and the clients requesting the offered services, in order to detect and consequently block attacks, typically mounted by malicious actors by leveraging vulnerabilities in the code of web applications. WAFs are typically configured manually by network administrators, specifying rules on the payload of HTTP packets to identify specific attacks. However, writing such rules is typically cumbersome and error prone. Thus, a lot of research is being conducted in using machine learning algorithms to automate this task.

Mereani et al. [Mereani 2018] investigated the use of an algorithm to identify *Cross-Site Scripting* (XSS) attacks, which occur when an attacker successfully injects a malicious script in a web application's code. The script is therefore downloaded and executed in the browser of the victim browsing the infected page. It should be noted that typically attackers employ strong forms of obfuscation to render the scope of such script unintelligible by manual analysis (the same is done by legitimate web application coders to preserve their intellectual rights). The authors analyzed scripts from the XSSed archive[2] to build a dataset, extracting 59 features both structural (presence of non-alphanumeric characters, employed extensively by obfuscation techniques) and behavioral (including objects, events, methods and tags used in the script code). The authors used the dataset to train SVM, k-NN and random forest models able to identify malicious XSS scripts. All

[1] See https://csrc.nist.gov/glossary/term/threat.
[2] See http://www.xssed.com/.

models performed well, with an accuracy over 95%, with the best one being the k-NN based one with an accuracy over 99%.

Another typical attack against web applications is the *SQL injection*. Attackers can try to exfiltrate sensitive data (e.g. login credentials, e-mails, credit card numbers) hosted in databases used by such applications, by abusing user input (e.g. login forms) to execute special SQL queries. Ross et al. [Ross 2018] presented a comprehensive comparison of different machine learning techniques to identify injection attacks against a MySQL server, acting as a backend for a custom messaging web application. They propose a concurrent analysis of traffic on both the web application server interface over the Internet and on the connection between the server and the MySQL database, in order to increase the attack detection performances. Three different datasets are obtained generating and consequently capturing traffic using a set of custom-made Python scripts. The first two are obtained by capturing traffic respectively on the web application and MySQL servers, and the third by correlating samples captured on both sites originating from the same SQL query. The authors trained various algorithms using different machine learning techniques (RF, SVM and ANN) on the three datasets. Using the dataset containing the correlated samples, the average accuracy among different machine learning techniques increased from 95% to 97%, with the best algorithm being based on ANNs (97,2% accuracy).

Finally, Calzavara et al. presented Mitch [Calzavara 2019], an approach to detect *Cross-Site Request Forgery* (CSRF) vulnerabilities on web applications. In this kind of attack, a user, logged in a legitimate but vulnerable website, visits (unknowingly) a malicious website, which uses the logged state of the user to forge requests to the legitimate website on the user's behalf. The authors have built a dataset by manually browsing 60 websites obtained from the Alexa ranking[3], capturing the HTTP traffic, and marking the requests that can be potentially subject to CSRF attacks (e.g. functionalities of the web application that are available only after a login by the user). The authors have subsequently built a dataset by extracting from the requests 49 features, originating from structural (e.g. number of parameters), textual (e.g. occurrences of specific words) and functional (e.g. HTTP request method) characteristics of the analyzed requests. Using this dataset, they have tested various machine learning techniques (comprising SVM, DT and RF), the best performing being the RF-based one with a ROC AUC score of 93.2%.

### 2.1.2 Internet-of-things

The introduction of the Internet-Of-Things (IoT) paradigm is one of the last big revolutions in the IT scenario. The IoT term encompasses the introduction of computational intelligence in objects used in many objects of everyday life, including, for example, wearable devices (e.g. smartwatches), automobiles, domestic and video surveillance appliances. Typically, these objects do not need a high amount of computational power to carry out their tasks. Employing powerful hardware would lead therefore to unnecessary costs, and in many cases would be impossible (e.g. battery capacity limitations on wearable devices). However, from a cybersecurity point of view, this limitation renders the application of typical security solutions (e.g. antiviruses and firewalls) impossible. This in turn renders such devices a perfect target for attacks by malicious actors. Typically, they try to infect such unprotected devices with some kind

[3] See https://www.alexa.com/topsites.

of malware, taking control of them in order to create botnets, which are typically used to carry on Distributed Denial-of-Service (DDoS) attacks, unknowingly from the legitimate proprietors of the involved devices. A notable example has been the Mirai malware [Manos 2017], which in 2016 has been used to create botnets, which in turn have been employed to undertake successful DDoS attacks against a wide range of well-known websites and web hosting platforms (including GitHub, Twitter, Reddit and OVH).

A report by Kaspersky[4] [Kupreev 2020] indicates a staggering increase of the occurrence of DDoS attacks, with the number of such attack duplicating in Q1 2020 with respect to the same quarter of the previous year.

Two main strategies are found in literature to defend against this kind of attacks, both employing various machine learning techniques. The first one is the identification of TCP/UDP connections established by the controlled botnet devices (or *zombies*) with the target (e.g. a web server hosting the attacked web page). Such controls may be executed both on the access point of the network comprising the zombies (e.g. the home gateway), or on the targeted devices (e.g. a web server). The latter can close these connections as soon as they are identified as DDoS-related ones, thus preserving computational resources and consequently mitigating the attack effects (i.e. preserving the availability of the service offered by the attack target). A recent work by Doshi et al. [Doshi 2018] tackles this problem, presenting the performances of five different machine learning algorithms (k-nearest neighbors, support vector machines, decision trees, random forests and fully connected neural networks) trained to identify three typical DoS attacks (employed for example by the Mirai botnets): TCP SYN flood, UDP flood and HTTP GET flood. They used a combination of stateless (e.g. packet size, inter-packet interval and protocol) and stateful (e.g. bandwidth, IP destination address cardinality and novelty) features evaluated on the analyzed traffic to train the machine learning algorithms, with the one performing better being the FCNN with an accuracy of 99%. Another work by Aamir et al. [Aamir 2019] presents a semi-supervised approach for the same problem. The work is particularly interesting, since the traffic constituting the dataset must not be completely labelled (i.e. only a portion of the dataset must be manually labelled as DDoS traffic). The authors first use two different clustering algorithms (agglomerative clustering and k-means with feature extraction via principal component analysis) to analyze the traffic of unknown origin, thus obtaining a fully labelled dataset. Then, they evaluate the performance of three different supervised machine learning algorithms (KNN, SVM and RF), with the best performing one being the random forest with an accuracy of 96.6%.

The other strategy typically adopted to counter DDoS attacks consists in the identification of the Command and Control (C&C) channel, which the owner of the botnet (or *botmaster*) establishes with the *zombies* to control them, for example to coordinate an attack against a specific target. These types of controls are especially interesting on the edge network appliances. Occurrences of this kind of traffic clearly indicate the presence of infected devices on the network, so the network administrator can take the appropriate actions (e.g. detect and remove malware from the infected devices). Gadelrab et al. introduced BotCap [Gadelrab 2018], an approach to identify HTTP-based botnet C&C channels with machine learning algorithms trained on statistical features of the traffic. In particular, the authors created a dataset by running six different botware families (all employing HTTP C&C channels) in a controlled scenario, capturing the

[4] See https://securelist.com/ddos-attacks-in-q1-2020/96837/.

generated traffic and consequently analyzing it to extract 55 features, comprising, for example, statistics about duration of connections and the amount of exchanged traffic. The resulting dataset has been used to train three different SVM-based algorithms, with the best one obtaining a 95% F1-score in identifying HTTP C&C channels. Hoang and Nguyen [Hoang 2018] have proposed a different approach to tackle the same problem, based on the identification of the DNS requests submitted by zombies to obtain the IP address of the C&C server. Such requests are necessary, since botmasters typically change over time their IP in order to avoid detection. In particular, they analyzed both benign (top 30000 domain names ranked by Alexa) and botnet domain names (created by the Conficker and DGA botnets), then they extracted 18 features obtained from various text characteristics of the domain names (e.g. bi-gram and tri-gram clusters, vowel distribution) to generate three datasets, with a varying ratio between benign and botnet samples. The authors used these datasets to train four different machine learning algorithms (kNN, DT, RF and Naïve Bayes), with the best performing one being the fandom forest model, with an average accuracy of about 90% on the three datasets.

### 2.1.3 Malware

*Malware* (malicious software) is any kind of software designed with the objective of harming the devices on which it is installed (i.e. *infected* by it). Indeed, malware applications are one of the most longstanding menaces in the IT scenario. In fact, the first known malware dates to 1988 [Orman 2003]. While in the early days such programs were written mainly for fun or as experiments, nowadays malwares are typically either financially motivated or engineered for political and industrial espionage. The most common typologies of malware are:

- *ransomware* (or cryptolockers): after infecting a device, these viruses encrypt all or some data present on a victim and then ask the device owner to pay a ransom (to be paid with hard-to-trace cryptocurrencies such as Bitcoin), in order to obtain the encryption key and regain access to his data. A notable example is the WannaCry worm, which encrypted data on more than 200000 devices in 150 countries [Mackenzie 2019].
- *cryptojackers*: this kind of malware hides a cryptomining process on the infected device, using the computational power of the latter to mine new cryptocurrency units for the malware owner. This may take a high toll on the performances of the infected device, may cause an early obsolescence and a reduced battery life in the case of mobile devices. Cryptojacking infections started in 2017 and these kinds of malwares are among the ones spreading with the highest rate [Liebenberg 2018].
- *trojans* (short for *trojan horse*): these viruses are typically downloaded and executed on the target machine unknowingly by the user, for example by misleading him into opening an apparently legitimate e-mail attachment. While the payload of this kind of malware can be virtually anything, usually they are used as vectors to gain complete remote control of the infected machine (remote access trojans), either for financial motives (e.g. Metamorfo [Szeles 2020], a trojan written to obtain access to the user's bank account) or to gain illicitly information (e.g. industrial espionage or state surveillance).

In general, given the dominant position in the market of operating systems, malware have traditionally targeted devices running Microsoft Windows (in 2018 still half of the new malware targeted this OS). Still, there is an increase in diffusion of malware targeting Android mobile devices and for the Apple MacOS operating system [Kujawa 2020].

Anti-virus software usually resorts to two main strategies to identify the presence of malware on the protected devices:

- *signature* (or *definition*): a unique identifier (typically obtained with an hash or a custom algorithm executed on the malware's binary) of a specific malware or of a specific portion of it (e.g. a function or method); anti-virus programs using this kind of strategy analyze continuously the protected system, searching processes in execution that match signatures of known malware;
- *behavioral analysis*: potentially malicious software is executed in a virtual and controlled environment (*sandbox*), to analyze the possible outcomes of its execution on the main OS of the device, trying to identify typical behaviors of malware (e.g. replication in different locations of the file system, attempts of spreading via network interfaces, encryption of the files on the device).

Given the importance of this menace, a lot of effort has been put in researching new and more effective ways of identifying malware. In particular, the application of machine learning techniques has been found to be promising in this area. Regarding the signature-based approach, a recent survey by Shalaginov et al. [Shalaginov 2018] show that the most commonly used features, employed to train machine learning algorithm for this purpose, are the raw representation of bytes of the executable (e.g. for randomness analysis, which may indicate the presence of encrypted malicious content), the disassembled binary code, and, regarding the Windows portable executable format (an executable format frequently used to distribute malwares), the information contained in the file header.

Another interesting work using the same approach is the one by Ma et al. [Ma 2019]. The authors focused on Android malware, using a collection of more than 10000 Android malwares (obtained from the Android malware dataset [Jang 2017]). They built a malware detection framework based on three machine learning algorithms. They extract API calls from the malware artifacts and correlated them to create three datasets, based on the presence of certain API calls known to be commonly used by malware, the frequency of these calls and their sequences. The first two datasets are used to build classical decision tree algorithms. The third dataset is definitely more interesting, since the API (Application Programming Interface) sequences are modeled as time-series, and used to train a long-short term memory RNN (Recurrent Neural Network), a neural network able to classify sequences of events (used for example for financial predictions). The framework proves to be very effective, with a detection accuracy of more than 98%.

### 2.1.4 Conclusion

Machine learning approaches can be successfully used to automatize various critical tasks. This trend seems to be exponentially growing in the last few years and it will most likely continue in the foreseeable future.

Machine Learning (ML) models are continuously proving to be very effective for coping with cyberthreats in a timely and accurate manner. Given the astonishing results obtained in the last few years it seems sensible that various industries will integrate even more ML algorithms in their software and hardware appliances. In fact, several existing anti-virus frameworks started to adopt these methods. For example, Microsoft launched recently Advanced threat protection, an extension of its Defender anti-virus software that uses a combination of big-data and machine learning techniques to improve its malware detection capabilities.

Apart from direct threat detection, these approaches can also be used to (semi-)automatically configure WAFs, with less space for human error. Many industry leaders in the cybersecurity fields are starting to integrate such approaches in their firewalls, such as Fortinet[5] and Alibaba Cloud[6].

## 2.2 Machine learning under the restriction of GDPR

While the applications of machine learning seem to be endless, many countries have started to restrict and regulate the handling and usage of data and therefore also the field of application by data protection regulations such as the EU GDPR [GDPR 2016]. Nowadays, many companies still struggle with the implementation and maintenance of GDPR-conform data handling, not only for cybersecurity related tasks but also for many other purposes. Especially in conservative markets such as for many finance applications, the usage of complex models or even the storage of related data is obviated [Guidotti 2018]. To fulfill the requirements of the GDPR on the one hand and to enter new fields of application on the other hand, a variety of new technologies that enable privacy protecting machine learning have emerged during the recent years. Before the potentially game changing technologies are presented, a brief look is taken into the GDPR and four major requirements are elicited:

- *explainability*: the first requirement in designing ML models is to make the models explainable [Goodman 2017]. Following Article 13 and 14 of the GDPR a user has the right to explanation if decision making or profiling is involved. Then the controller must provide information in an understandable way that the user can assess a fair and transparent processing of the model's logic. This leverages the tradeoff between simple easy understandable so called "white-box models" that often generalize better with a lower accuracy and more complex so called "black-box models" that can achieve a higher accuracy but are much more difficult to explain.
- *non-discrimination*: the second requirement, the right to non-discrimination, can be defined as the absence of unfair treatment of a natural person based on the belonging to a specific group, such as religion, gender or race [Goodman 2017]. In relation to machine learning, this stands in opposition to the concept of allocating individuals in different classes on basis of a huge amount of data, collected from society. Reasoned by the fact that the society exhibits per definition exclusion, discrimination or inequality, bigdata can only be treated as fair, if these differences in treatment are detected and balanced during the development of the ML model.
- *the right to be forgotten*: the third requirement is the user's right to be forgotten what implies the deletion of instances from an existing dataset [Yang 2019]. This leads to the question whether it can use a ML model that has been trained on deleted or withdrawn data. Therefore, ML model should be somehow adoptable during their lifetime.
- *data security*: the fourth requirement is the data security. Personal data that is "any information relating to an identified or identifiable natural person" (GDPR article 4) has to be protected against loss, damage and unauthorized processing [GDPR 2016]. Data from different locations or

[5] See https://www.fortinet.com/products/web-application-firewall/fortiweb.
[6] See https://www.alibabacloud.com/product/waf.

companies cannot easily combined to one big dataset because collected data must be restricted to a defined purpose [Yang 2019].

While the above-mentioned requirements restrict the application of classic ML approaches although they aim to protect the data of a user, new technologies have evolved that try to guarantee and fulfill the above-mentioned challenges.

### 2.2.1 Privacy preserving machine learning

To protect ML models from a variety of attacks that try to reveal the data, training features or the algorithm itself, a variety of countermeasures have evolved during the recent years. These techniques in general, can be summarized under the term of *Privacy Preserving Machine Learning* (PPML). While most techniques, were not invented in the recent years, their application to the field of machine learning is new and most of them are not well established or applied. Examples for this are cryptographic protocols to encrypt data that is submitted from multiple parties to one single database, or homomorphic encryption that enables simple computation tasks with encrypted data [Al-Rubaie 2019].

Distributed learning was designed for problems that a single machine cannot handle in enough time, due to the size of the data or the complexity of the model [Li 2014]. While distributed machine learning is widely used nowadays to address performance issues, current methods such as federated learning focus also on privacy protection. Generally, federated learning is expected to break the barriers between data sources while the leakage of data is prevented. The idea of *federated learning* was first proposed 2017 by Google and aims to build a ML model based on datasets that are distributed across multiple devices whereby the data is not merged to an overall dataset. There are different approaches, such as horizontal, vertical federated learning and federated transfer learning [Yang 2019].

One of the potentially most relevant fields in the future is the domain of medicine. While noticeable success in the training of Deep Neuronal Networks (DNN) has been achieved, AI for reliable clinical decision support, requires larger amounts of imaging and clinical data. These data cannot be achieved in voluntary clinical studies among a small number of institutions that are not well geographically diversified. As described above, this problem is leveraged by regulations such as the GDPR or the United States Health Insurance Portability and Accountability Act (HIPAA) that strictly regulate the exchange and storage of personal data [Kaissis 2020]. This is where federated learning comes in, ensuring data protection on the one hand and the usage of data among institutions on the other hand.

Another potential field is the domain of banking. In the prediction of credit risk, a large variety of factors is combined. Although efficient intra-bank machine learning systems exist, a huge gain of efficiency can be expected for inter-bank models. Federated learning can enable banks to share information about the credit risk of their customers while the privacy relevant data remains locally stored and is not visible for other banks [Kawa 2019].

Although a lot of advantages can be expected, it is not easy to distinguish between federated learning as privacy protecting or privacy vanishing. e.g. in smart retail federated learning can be used to break the barrier between different data holders. This enables companies to combine financial data from banks disclosing the willingness to pay, personal preferences from social products characteristics from e-shops

[Yang 2019]. Instead of using this information for personalized advertising, this could be also used for perfect price discrimination.

While the principle of differential privacy is not new, its relevance has significantly increased during the recent years. In general, *differential privacy* tries to disguise certain rare attributes by adding noise or using generalization [Yang 19]. The aim of this is that a single user cannot be identified by using other studies, additional datasets or further information sources [Dwork 2014]. What makes this technology relevant for the future is its usage in combination with other new emerging technologies. E.g. Yang et al. [Yang 2018] propose a multifunctional data aggregation method with differential privacy based on a fog computing architecture. They point out that one advantage of differential privacy compared to homomorphic encryption or cryptographic protocols are the support of a variety of statistical aggregation functions and the low computational costs. Chamikara et al. [Chamikara 2019] claim that current privacy-preserving deep learning approaches are build up on server centric approaches that cause high processing costs. To overcome this, they propose their new local differentially private algorithm LATENT that is based on adding a randomization layer before the data of a user is used to train a ML model. Another approach comes from Lecuyer et al. [Lecuyer 2019] who present the defense PixelDP that is a connection of robustness against adversarial examples and differential privacy that scales to large networks and datasets.

Like differential privacy, the idea of *secure MultiParty Computation* (MPC) is not new. By involving multiple parties that each only know their input and output this concept aims for zero knowledge [Yang 2019]. Reich et al. [Reich 19] have identified a research gap in the privacy preserving solutions for text classification. Their proposed method is based on MPC and is used to classify hate speech against women and immigrants. They claim that their model has the advantage that the ML model does not learn anything about the text and the author does not learn anything of the model. In the past, a tradeoff in MPC between efficiency and security existed that caused less secure models so called “semi honest” models. Therefore, Chen et al. [Chen 2019] tested the SPDZ[7] framework with simple applications such as linear and logistic regression, with increased security and performance. They claim that future research topics in this technology are the application to more complex models while limitations in computational power must be overcomed.

### 2.2.2 Explainable machine learning

As mentioned above, when designing a machine learning model, good scores in an evaluation metric are not enough to evaluate the performance of an algorithm. With the data protection regulations, algorithms have also to be designed in a way that they are explainable and non-discriminatory [Hall 2018]. Therefore, methods for explainable machine learning will be important in the near and long future. Model specific methods give insights in how a specific model makes decisions and often try to explain the black-box. These methods can often not be compared over different models. In the opposite to this, model agnostic methods give insights into a model without understanding how the model works. The model is treated as a black-box and the relation between input and output is analyzed. One of the most common methods are Local

[7] SPDZ is a curious acronym for speedz.

Interpretable Model-agnostic Explanations (LIME) [Ribeiro 16] and is constantly improved, e.g. by Visany et al. [Visani 2020] who aim to increase the stability to make the explanations more reliable.

To be precise, explainability has significant relevance for the future because nowadays many models lack in providing enough explainability to the user of an application in the real world [Bhatt 2020]. This demand is leverage by an increasing complexity and distribution of ML models such as with adversarial and federated learning. Besides this, the relevance of explainable ML is also increasing in the field of natural sciences where ML is utilized to achieve insights in observational or simulated data [Roscher 2020].

### 2.2.3 Conclusion

The GDPR has started to affect not only our daily lives, but also how machine learning and artificial intelligence models are trained and used by the industries. The security of the data and the ML model itself as well as its explainability are most relevant. While trends such as edge computing in combination with federated learning approaches or differential privacy are likely to increase the security of users but also come with a significant increase in complexity. These technologies are most likely to help in future to overcome the tradeoff between complex models with high accuracy and secure models that can be explained easily.

# 3 AI for adversarial purposes

This chapter provides some insights into the use of Artificial intelligence (AI) for adversarial purposes. While Section 3.1 investigates the overview of existing adversarial attacks, threat models, applications, and future trends in AI itself, Section 3.2 provides insights into the application of AI for social engineering. Besides these two selected topics, many further possibilities exist how AI can support attacks on cyberphysical systems. For further reading, we refer to a recent survey of Kaloudi and Li [Kaloudi 2020].

## 3.1 Adversarial AI

AI has nowadays become quite prevalent, thanks to its intrinsic automation features. While the adoption of artificial intelligence and corresponding deep learning techniques can help to handle a diverse number of complex tasks such as image processing [Savadjiev 2019], natural language processing [Wen 2019], autonomous cars routing [Lazar], etc., it is of particular importance to ensure the security and robustness of the deployed algorithms [Wanga 2019]. The aim of this topic is to provide a high-level introductory overview of adversarial AI. We also inspect the existing threat models, examine the adversarial applications and provide some insights into the future trends of challenges associated with adversarial AI.

Machine learning models – as a subset of AI – are constantly evolving and have had a significant impact on various application domains. However, there are serious concerns regarding the reliability, trustworthiness, and security aspects of these models [Wanga 2019]. More specifically, previous research has shown that most of the advanced machine learning models are vulnerable to adversarial attacks [Ma 2020, Finlayson 2019, Zhou 2019]. To put it simply, as shown in Figure 1, in a simple adversarial attack the adversary perturbates the original samples in such a way that the changes are almost undetectable to the human eye. The modified samples are then called *adversarial samples*, and when submitted to a classifier they are misclassified, while the original one is correctly classified. As an example, a given supervised machine learning classification algorithm might be attacked through discovering the minimum changes that should be applied to the input data leading to a different classification outcome. A typical example is the computer vision systems deployed within autonomous cars where a negligible change in a stop signal that is completely unnoticeable to the human eyes may cause such cars to identify stop signals as 45 mph signals [Eykholt 2016].

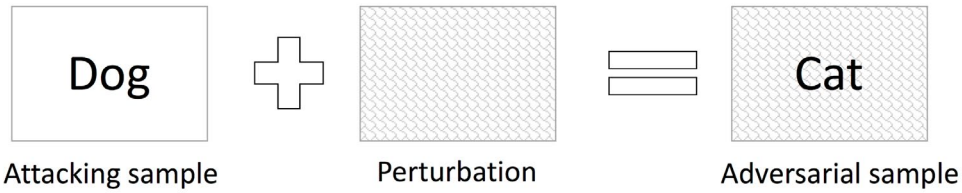


Figure 1: An overview of adversarial AI.

There are three main threat models for adversarial attacks, namely the black-box, grey-box, and white-box models (see Figure 2). In the *black-box model*, an adversary does not know the structure of the target machine learning algorithm or its parameters. However, it can communicate with the algorithm to query the predictions for specific inputs. Such queries may impose a considerable burden on the overall performance of the algorithm. As such, the number of queries is considered as an important criterion for the efficiency of the attack. As for the *grey-box model*, it is assumed that the adversary has a kind of understanding of the overall architecture of the algorithm. However, it does not have any information about the detailed parameters such as the internal weights in a neural network. Like black-box model, in the grey-box model the adversary can communicate with the algorithm. Compared to the black-box model, a grey-box adversary always performs better than the black-box adversary. The *white-box model*, on the other hand, is usually referred to as the strongest adversary as it has a comprehensive understanding of the target model, including its respective parameters. To be more precise, the adversary can adapt the attacks and directly craft adversarial samples on the target model.

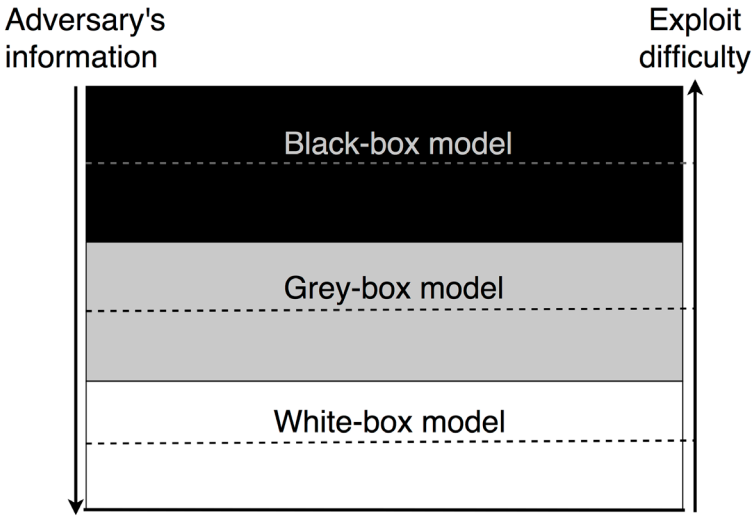


Figure 2: Threat models within adversarial AI.

### 3.1.1 Network security

The use and development of machine learning techniques in cybersecurity related areas has become quite prevalent these days. As an example, machine learning techniques are nowadays widely used in malware classification and intrusion detection systems due to the ever-evolving nature of threats within such systems. In [Grosse 2016] the authors demonstrated the applicability of efficient adversarial sample crafting attacks for neural networks used for classifying malware. The experimental results indicated that the conducted attacks can lead to a misclassification rate as much as 80%. This highlights that adversarial crafting is a real threat in security-critical domains such as malware detection. Similarly, the authors in [Huang 2018] focused on the same problem but with a focus on intrusion detection systems. They analyzed software defined networking-based intrusion detection systems and showed how adversarial attacks can exploit the vulnerability of several deep learning classifiers in this area. Through experimental results, the authors concluded that the conducted attack can provide an accuracy drop of about 35%.

### 3.1.2 Natural language processing

Researchers in [Liang 2017] demonstrated that text classification algorithms can be easily fooled through crafting text adversarial samples. Similarly, the authors in [Hosseini 2017] showed the applicability of crafting adversarial text samples by modification of the original samples. The experimental results obtained from movie reviews on IMDB and tweets available on Twitter showed the efficiency of conducted adversarial attacks. The authors in [Samanta 2018] examined the vulnerability of Google's Perspective API (Google's Perspective API uses machine learning models to score the perceived impact a comment might have on a conversation) against the adversarial examples. Through multidimensional experiments, they showed that an adversary can deceive the system by misspelling the abusive words or by adding punctuations between the letters.

### 3.1.3 Computer vision

The authors in [Sharif 2016] proposed techniques to generate adversarial attacks in facial biometric systems allowing an attacker to evade recognition or impersonate individuals. Using pair of eyeglass frames worn by the attacker, the eyeglasses allow the attacker to evade being recognized or to impersonate another individual. The authors in [Zhou 2018] focused on the same problem. They proposed new attack scenarios against face recognition systems showing that face recognition systems can be bypassed or misled. It is shown that such attacks can not only deceive surveillance cameras, but they can also impersonate the victim and bypass the face authentication system, using only the victim's photo.

According to [Dolhansky 2019], from Facebook's AI Red Team, *deepfakes*, that is artificially generated images and media of a person, can lead to intimidation, harassment, or manipulation of financial systems or elections. Up to them, key driving methods for deepfakes such as face swapping and image manipulation are state-of-the-art techniques from computer vision and deep learning. In order to deal with this massive problem and to find adequate countermeasures, the Deepfake Detection Challenge (DFDC) was announced in 2019. The power of state-of-the art Generative Adversarial Networks (GAN) that produce photo-realistic images from randomly sampled codes is shown by [Shen 2020]. Shen et al. trained a model called InterFaceGAN that is capable to interprete semantics enabling them to manipulate facial attributes precisely with any fixed GAN model.

### 3.1.4 Conclusion

AI-driven technologies have become an indispensable part of our lives. Thanks to the ever-evolving nature of machine learning techniques, they are now increasingly applied in various applications. However, serious concerns have been raised about the security and reliability issues of machine learning models. As discussed throughout this report, a great number of advanced machine learning models are vulnerable to adversarial attacks. Previous studies have shown that such attacks can be efficiently applied to many application domains ranging from computer vision to natural language processing. As such, it is of importance to initiate calls for action considering the evolving nature, likelihood, criticality, and impact of such attacks through providing a comprehensive roadmap considering the future challenges associated with adversarial artificial intelligence.

There exist multiple challenges associated with the future trend of adversarial AI. To avoid or at least minimize the negative effect of adversarial attacks on available AI-driven technologies, it is highly crucial to provide a holistic roadmap that will tackle the following challenges.

Explainability refers to a set of methods and techniques in the application of AI such that the generated results and the decisions made by the solution can be understood by humans. Although explainable AI (see also Section 2.2.2) can help to explain the knowledge within an AI model and reason about what the model acts upon, the information that it reveals by explainability techniques can be used by adversaries to conduct adversarial attacks [Arrietaa 2020].

It has been shown that adversarial examples generated against an ANN can fool the same networks trained by different data sets, with different architectures, and even other classifiers trained by different machine learning algorithms [Papernot 2016]. Transferability can even be more critical when it comes to black-box models since in black-box models the attacker does not have access to the architectural details of the model and the training data set [Yuan 2019].

In the report "Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation" [Brundage 2018], the authors analyzed various scenarios where AI techniques can be used by an attacker and provided some suggestions to mitigate such potential issues. The researchers have also shown that most of the existing defensive mechanisms are suffering from robust, granular, and thorough evaluation. This is because, although such defensive mechanisms can defend against a particular attack, they are extremely vulnerable to slight changes in the attack, and therefore, they become inefficient. As such, a holistic evaluation of defensive mechanisms can help to make the defensive mechanisms not only defending a certain attack, but also similar attacks resulted from slight changes within the attack's architecture.

## 3.2 Social engineering attacks

Social engineering is the clever manipulation of the human tendency to trust. If the technical security of most critical systems remains high or even increases, attacking the systems by social engineering is getting more and more attractive. While the amount of social engineering attacks and the damage they cause rise every year, the defenses against social engineering do not evolve accordingly [Schaab 2016, Schaab 2017]. We sketch in this section how AI will further improve attacks on human beings.

### 3.2.1 AI-supported vishing

Recent efforts in automating tasks such as bookings via phone led to systems like Google Duplex[8], which conduct natural conversations, allowing people to speak like they would with another person and not with a computer. While the main goal of Google Duplex is to automatize bookings for e.g. a hairdresser or a restaurant, naturally these enhanced capabilities can also be used to mislead people. Spam over Internet telephony (SPIT) is an already existing phenomena which is regulated, e.g. by the federal Telephone Consumer Protection Act of 1991 (TCPA) [Telephone 1991]. But legal actions are not enough and lead to some effort to technically counter robocalls [Gallagher 2020].

Furthermore, with additional information, i.e. voice samples, the attacker can make use of AI to imitate someone's voice, improving the setup for fraud [Bendel 2019] and vishing (voice phishing) attacks.

[8] See https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

### 3.2.2 AI-supported phishing

AI is, unfortunately, also beneficial to phishing attacks in a couple of different ways. Artificial intelligence can support large scale or spear-phishing attacks. Besides trying to overcome spam filters by varying the accompanying text or malware, AI has helped in the creation of new phishing URLs based on the patterns of the most effective URLs in previous attacks [Bahnsen 2018], the creation of fake reviews to improve the credibility of a site or an account [Yao 2017], and to personalize phishing message to improve success of spear-phishing attacks [Seymour 2016]. Furthermore, recent work on language models, such as GPT-3, shows that they can achieve good results on tasks as translation, question-answering, and even on several tasks that require on-the-fly reasoning or domain adaptation [Brown 2020]. GPT-3 is a text-generating program from OpenAI, which recently gained huge attention in the social networks and on media[9,10]. Thus, it is easy to imagine that high quality text generation AI can be used for the content of phishing attacks.

With humans being unable to evolve as fast as the machine learning discipline, this seems to be an arm race that humans can only lose, particularly given the amount of spam mails already circulating.

### 3.2.3 Automated information gathering

Numerous tools exist for intelligence gathering for social engineering attacks [Beckers 2017]. Attackers already have a wide range of tools available and most of them do not require high technical skills. With the rise of AI, it can be expected that AI can not only assist in the attacks itself by vishing (see Section 3.2.1) or phishing (see Section 3.2.2), but also on the level of information gathering. While it is still a challenge to connect existing data meaningful with other databases [Pape 2017] or to detect sensitive information in unstructured text [Tesfay 2016], the use of AI in this area is on the rise. But AI can not only support automatic data gathering on the fine-grained data collection itself. AI has also been successfully used at higher information levels, e.g. to select the most valuable targets for phishing messages [Seymour 2016] or to interpret feelings and emotions [McStay 2018]. The exploitation of the information collection processes can also benefit from techniques developed by digital markets [Darmody 2020] to manipulate customers.

### 3.2.4 Conclusion

While AI and ML are usually seen, at least from the researchers point-of-view, as powerful tools for the defenders, they can also simplify the attacker's life.

In the future, we can expect more automated phishing and in general social engineering calls, because the described machine learning and AI capabilities not only lead to an increased quality of the attacks, are harder to spot by countermeasures and for the humans, but also reduce the effort for the attacker. AI trained to gather some information instead of assisting humans in making appointments for the hairdresser or restaurants might also be a severe threat in the future.

[9] See https://www.wired.com/story/ai-text-generator-gpt-3-learning-language-fitfully/.
[10] See https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

# 4 Zero-trust security systems

For years, defining a network perimeter and adding defense layers to keep threats out has served companies well. Nowadays, the evolution of cyberattacks and the borderless nature of companies' networks lead to boundless points of exposure that could not be confined with a perimeter-based approach. With the introduction of new paradigms and technologies such as cloud computing, IoT devices, it has become less and less clear where the edge of a network ends.

Because of these factors, perimeter-less security is gaining momentum starting with the *Zero-Trust Security* (ZTS) model. In a ZTS model, any device, user or application, regardless of their position within the network perimeter, must be verified and authenticated. The integrity of the actors accessing the infrastructure as well as their identity is pivotal in this model. Micro-segmentation is another keyword within the scope of ZTS. This means that it is possible to create different network segments (not related to the network perimeter) or "trust zones" in which we could enforce different access policies to get the resources. The granularity of the micro-segmentation is arbitrary and depends on the sensitivity of the data.

This evolution in terms of security model as well as the requirements for new device integrity check and authentication mechanism, lead to the adoption of new emerging security technologies to create ZTS systems that are safe and, hopefully, tamper-free.

## 4.1 Trusted computing

*Trusted Computing* (TC) has been introduced by the Trusted Computing Group[11] (TCG) to enhance trust and security of modern information and communication technologies. The core component of the TC idea is the *Trusted Platform module* (TPM), a tamper-resistant chip designed to resist against software attack and mitigate the hardware ones. TPM in combination with some ad-hoc software components can provide a set of security mechanisms such as memory curtaining, protected execution, secure I/O, sealed storage, platform measurement and remote attestation. As pointed out by [Yan 2020], in recent years the TC has evolved thanks to efforts coming from both academia and industry.

In its latest version, the TPM 2.0 introduces new features that make its adoption and usage more effective with respect to its predecessors. It supports a wide range of hash and asymmetric algorithms as well as various block ciphers and also a flexible mechanism of authorization that allows to grant access to hardware and software resources based on the state of some components within the TPM (i.e. specific values of the internal TPM registers).

Intel TXT, namely LaGrande technology, in conjunction with a TPM provides a hardware root of trust available on Intel server and client platforms that activate the measured launch and protected execution capabilities[12]. By using these two paired technologies, starting from the BIOS itself, at each stage of the launch process, cryptographic hashes are measured and compared with the ones stored in the TPM to verify

[11] See https://trustedcomputinggroup.org.
[12] See https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04282020-draft.pdf.

the system integrity [Shepherd 2016] from its very boot. Deviations from the expected configurations result in an untrusted platform state.

### 4.1.1 Adoption in cloud environments

The greatest impact so far of the TPM technology is over cloud computing, especially thanks to its ability to provide a (partially) hardware-based remote attestation (RA) procedure, which allows to authenticate remote entities (e.g. infrastructure nodes) and measure their integrity. When an integrity breach is detected, several reactions can be performed such as stopping the compromised virtual machine and/or sending alert messages to the system administrators. Apart from verifying the genuineness of the physical host, each VM also have its own *virtual TPM* (vTPM), a virtual instance of the TPM providing all its functionalities (i.e. integrity measurement of components inside the VM).

An excellent report regarding the most recent applications of the TPM, its challenges and research question in the cloud environments is provided by [Hosseinzadeh 2020].

### 4.1.2 Criticalities

Despite the consistent adoption of TPM-related technologies in cloud environments, there is still some progress to make concerning the aim to mitigate physical attacks and the certification process of these devices.

A recent study [Moghimi 2020] depicts how it is possible to exploit secret-dependent execution times during digital signature generation to recover private keys protected by the TPM. The attack involves the generation of signature based on elliptic curves (e.g. ECDSA and ECSchnorr). The chips affected by this vulnerability were both firmware and hardware based. The study not only shows that the TPM in this case is vulnerable to system adversaries but also to remote attacks, making the applications using the TPM less secure than the ones without it. This have been demonstrated as a result of an attack performed against a StrongSwan IPsec VPN which uses the TPM to generate the digital signature for authentication.

The continuous progress related to quantum computing technologies also represent a future threat to public-key cryptography used extensively by the TPM. The FutureTPM[13] project aims to design a Quantum-Resistant TPM. This involves designing and developing algorithms that could be integrated within a TPM and do not suffer from quantum-related vulnerabilities. As reported in the paper [Fiolhais 2020], some efforts in exploring the integration of new cryptographic primitives within the TPM have been already made. This work discusses a new software TPM architecture which exploits post-quantum cryptography[14] algorithms (e.g. Dilithium and NTTRU) and maintains much of the same infrastructure of the TPM 2.0.

### 4.1.3 Conclusion

In order to make simpler and even wider the adoption of TPM as a solution to enhance security of the IaaS infrastructure, more flexible solutions must be adopted. For instance, a middleware capable of interacting

[13] See https://cordis.europa.eu/project/id/779391.
[14] See https://csrc.nist.gov/projects/post-quantum-cryptography.

with both hypervisors and VMs and creating a software layer that allows a simpler integration of the TPM in a cloud environment is required. Remote attestation of virtual machines is still not fully supported. Different mechanisms exist, such as the vTPM (see Section4.1.1), but they still lack a proper method to bind the vTPMs to the physical TPM.

Furthermore, the use of this technology has a heavy toll on the performance of a machine. The RA operations are particularly expensive in terms of execution time. This situation is exacerbated when dozens of VMs are deployed on a single node and all of them may require the use of these operations periodically. This challenge still has no real solution and needs to be addressed in order to deploy the TPM technology at full scale.

In addition, due some new side-channel attacks, issues in the validation and certification process of devices and the advent of quantum technologies, the security of the current TPM-enabled devices is at risk.

The TPM is a promising technology, but, as we discussed, it has several limitations and drawback. It is foreseeable that soon we will see a new version of this flexible chip that will increase even more its adoption, especially in cloud environments.

## 4.2 Trusted execution environments

A *Trusted Execution Environment* (TEE) is a secure area or enclave protected by the system processor. It has been introduced as a security technology in mobile execution environments. TEE holds sensitive data such as cryptographic keys, authentication strings, or data with intellectual property (i.e. digital right management) and privacy concerns. It is also possible to execute code and operations within TEE concerning this information. Thus, it is not necessary to let this sensitive data leave the TEE.

A typical TEE architecture[15] (see Figure 3) is composed of two different environments: a rich execution environment (REE) and a trust execution environment (TEE). The implementation of the two environments (REE and TEE) depends on the specific technology and thus on the TEE provider, but the GlobalPlatform specifications[16] require a hardware-based separation between REE and TEE. The REE contains the *rich OS* (e.g. a traditional Android) which allows the user to run a plethora of (rich) applications in an "untrusted environment". The TEE is instead an environment in which it is possible to run *Trusted Applications* (TAs) that leverage hardware secure resources provided by the platform (e.g. storage, keys and biometric sensors) and, as the specifications suggest, are completely separated from the other environment. When an application in the REE needs to use a trusted application service, it could send commands and requests to the specific TA through TEE client API.

[15] See https://www.securetechalliance.org/publications-trusted-execution-environment-101-a-primer/.
[16] See https://globalplatform.org/specs-library/?filter-committee=tee.

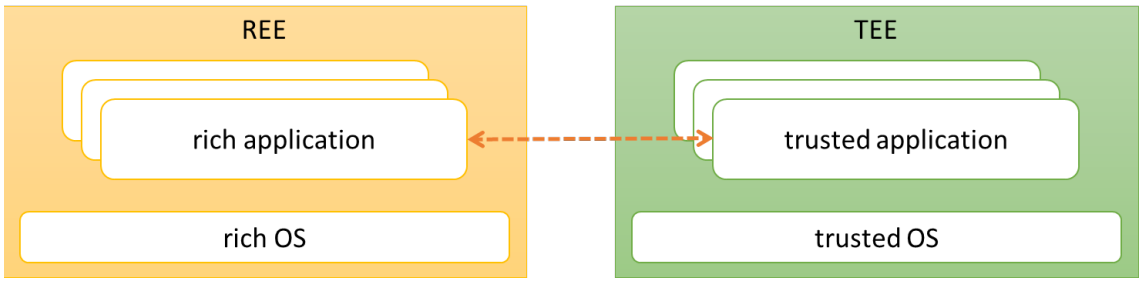


Figure 3: TEE architecture.

A TEE must be compliant to the following security principles[17]:

- code integrity must be verified at each stage of the boot process;
- execution of sensitive code must be isolated from the rich OS by means of hardware-based techniques;
- TAs must be isolated from each other;
- secure data storage must be provided by the TEE;
- a secure and privileged access to peripherals must be provided.

TEEs are not confined to mobile devices. Different uses cases have been proposed regarding diverse execution environments such as IoT, fog computing and cloud computing.

### 4.2.1 Intel SGX

Intel SGX[18] is a set of instructions which provides integrity and confidentiality for secure computations on systems where privileged software is potentially insecure. Security-sensitive code and data could be executed and operated into SGX enclaves. These enclaves are isolated and executed in protected region of the CPU. This technology provides also mechanisms to perform the remote attestation, where a remote provider could verify that an enclave is running on a real Intel processor and has not been tampered with.

SGX is, however, vulnerable to a variety of attacks, mostly side-channel based. An extensive survey of SGX threats is available in [Nilsson 2020].

An attack, which depicts the improvements still required to meet the SGX objectives, is Plundervolt [Murdock 2020]. Plundervolt exploits the ability of modern Intel processors to dynamically scale their frequency and operating voltage to corrupt the integrity of Intel SGX enclave computations. More in detail, the authors observed that in many processors (e.g. Intel Core series) a privileged software interface is exposed to control these parameters. Because of this, a privileged adversary could exploit the CPU supply voltage and induce some targeted faults within the processor. In order to mitigate this attack, microcode

[17] See https://www.securetechalliance.org/publications-trusted-execution-environment-101-a-primer/.
[18] See https://software.intel.com/en-us/sgx.

updates are required. In the past, there have been some other high-profile attacks with the same impact on SGX such as Foreshadow[19] and in that case microcode patches were required to mitigate the attack.

### 4.2.2 Keystone

Beyond Intel SGX, a plethora of TEE technologies have been proposed and developed so far, such as ARM TrustZone[20], AMD SEV[21] and RISC-V Sanctum[22]. All of them suffer from considerable design limitations since they are tied up to the hardware platform. Keystone [Lee 2020] is an open framework carrying out the idea of customizable TEEs. It is built on RISC-V[23] which is an open Instruction Set Architecture (ISA) based on RISC.

RISC-V operates in 4 different modes (see Figure 4):

- U-mode (user): for user-space processes;
- S-mode (supervisor): for the kernel;
- M-mode (machine): for accessing the physical resources (e.g. memory and devices);
- H-mode (hypervisor): hypervisor-level isolation (not yet used in Keystone).

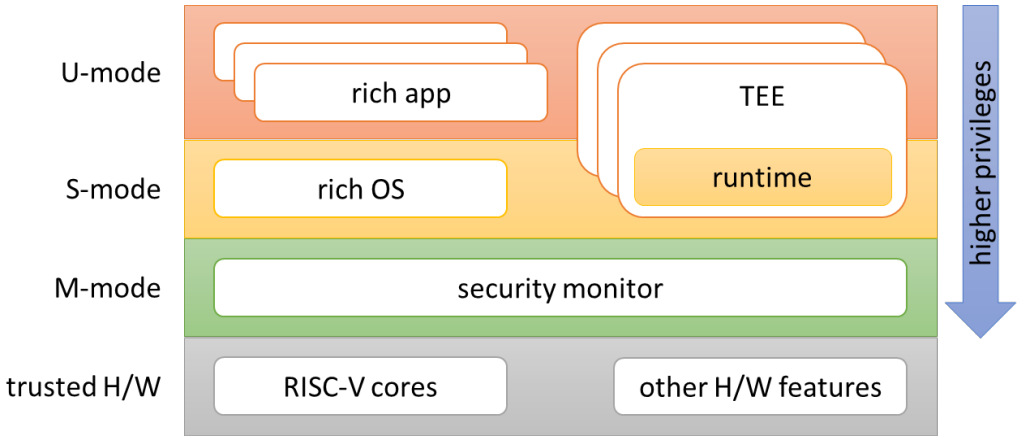


Figure 4: Keystone architecture.

According to [Lee 2020], a customizable TEE could be realized creating a trusted layer in-between the trusted hardware (i.e. RISC-V primitives such as memory isolation) and the untrusted OS (e.g. rich OS). As depicted in figure 4, this new layer is composed of two elements in Keystone's Architecture: the Security Monitor (SM) and Runtime (RT). The SM leverages RISC-V Physical Memory Protection (PMP) to associate specific protections to different physical memory regions. This process creates security boundaries making each enclave able to operate its own isolated physical memory region. At this point the RT is isolated from the untrusted OS as well as the other enclaves. The RT manages the lifecycle of the code inside the

[19] See https://www.usenix.org/system/files/conference/usenixsecurity18/sec18-van_bulck.pdf.
[20] See https://developer.arm.com/ip-products/security-ip/trustzone.
[21] See https://developer.amd.com/sev/.
[22] See https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/costan.
[23] See https://riscv.org.

enclave, manages the memory, the services system calls and contacts the SM on behalf of the application. This modular approach enhances also the flexibility from the more traditional TEEs because the enclave programmer could optimize the applications independently from the other enclaves and not considering any a priori hardware design limitation.

### 4.2.3 Conclusion

TEE technologies are pivotal for implementing trusted systems in emerging execution environments (e.g. IoT, cloud and edge computing). Soon we will most like see a wider adoption of this kind of hardware-based security controls.

However, the current trusted execution environments are still suffering from several side-channel attacks, most notable the diffused Intel SGX technology (see Section 4.2.1). In addition, there it does not yet exists a proper standard specifying how a TEE should be implemented, thus tying the software solutions to a specific platform, thus, in turn, limiting their diffusion. A new technology, Keystone (see Section 4.2.2), promises a more flexible approach regarding the design and the adaptation of TEE in specific case scenarios. In this direction, both industry and academia are expending a significant amount of effort, especially towards research of new attack strategies and methodologies to validate existing solutions.

# 5 5G applications

5G is not only the evolution of a current technology, but a disruption both in capacity and in the potential of the services that will be supported by it. The main characteristics of this technology are its low latency, its higher speed, between 10 and 250 times higher than the current 4G, its capacity for real-time data transmission and hyper connectivity. The potential of the 5G service, as well as the multiple applications that it can offer, are strategic enough to pose a medium-term scenario in which the European Union and member states can go from being in tow to leading the adoption of this technology, as well as the models of smart cities of the future. The European Commission (EC) has launched ambitious initiatives to support the cooperation among stakeholders in different Member States (MSs) for the development of 5G-enabled services. These initiatives include the 5G Action Plan[24], which represents a strategic effort to align roadmaps and priorities for a coordinated 5G deployment across the EU. Furthermore, the 5G Infrastructure Public Private Partnership (5GPPP)[25] is a joint initiative between the EC and EU industry (including telecommunications operators, SMEs or research institutes) to foster a common vision about 5G developments in the EU. Indeed, the development of 5G is widely considered crucial to ensure the EU's strategic autonomy.

5G is meant to be the backbone over a wide range of essential services such as energy, transport, banking and eHealth. Therefore, we are facing a scenario, in which sensible infrastructures could be built on technologies outside the EU, about which the lack of control or guarantees begins to raise suspicions. In this context, previous initiatives consider cybersecurity as a critical aspect for the deployment of 5G in the EU. In fact, it is expected that 5G technologies will play a key role in the Digital Single Market (DSM) with a strong impact in several scenarios, such as energy, transport, or health services. Furthermore, 5G will enable a more interconnected world, where vulnerabilities of 5G systems in a single Member State could affect the EU as a whole. Therefore, there is a need to promote collaboration and cooperation among countries to support a coordinated and secure deployment of 5G. To address such aspects, the EC launched the Recommendation "Cybersecurity of 5G networks"[26] in 2019 to propose a set of concrete actions for ensuring cybersecurity of 5G networks, including the development of national risk assessment strategies of 5G infrastructures. The main goal is to leverage national efforts to develop a coordinated EU risk assessment, in order to create a common toolbox of best risk management measures. As part of these efforts, the "EU coordinated risk assessment of the cybersecurity of 5G networks" report[27] identifies the main threats, threat actors, sensitive assets, vulnerabilities and associated risks of 5G networks. This report was used together with a recent ENISA report on 5G threats[28] to create the initial version of the mentioned toolbox[29]. By October 2020, the idea is that Member States should assess the effects of the recommendation report to determine whether there is a need for further action.

[24] See https://ec.europa.eu/digital-single-market/en/5g-europe-action-plan.
[25] See https://5g-ppp.eu/.
[26] See https://ec.europa.eu/digital-single-market/en/news/cybersecurity-5g-networks.
[27] See https://ec.europa.eu/commission/presscorner/detail/en/ip_19_6049.
[28] See https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-5g-networks.
[29] See https://ec.europa.eu/commission/presscorner/detail/en/IP_20_123.

The ever-changing communications ecosystem is one of the drivers for 5G security. The current networks employ monolithic designs, where a single mobile operator controls all the infrastructure hardware and services. But, the 5G will be composed by many different specialized stakeholders that will provide end-user network services [Chandramouli 2019]. For this reason, flexibility is one of the major 5G security requirements in order to support any efficient use-case, even those not conceived as of today. For example, security mechanisms for ultra-low latency critical verticals may not be appropriate for massive IoT applications with constrained end-devices that transmit delay-tolerant application data.

## 5.1 Certification

To ensure the development of secure 5G deployments, cybersecurity certification is essential to promote a transparent and trustworthy ecosystem of 5G devices and systems. The new regulation "Cybersecurity Act" entered into force in 2019 to create a cybersecurity certification framework for any ICT product, service or process. It complements the existing GDPR and NIS Directive to strengthen the cybersecurity in the EU. Indeed, it is expected that the Cybersecurity Act plays a key role in the development of 5G technologies. As described in the already mentioned Recommendation "Cybersecurity of 5G networks", the realization of such framework is an essential tool to promote consistent levels of security and the creation of certification schemes adapted to 5G related equipment. Furthermore, the mentioned toolbox identifies the EU certification for 5G network components, customer equipment and/or suppliers' processes as one of the main technical measures to strengthen the security of 5G networks. Toward this end, the realization of a cybersecurity certification framework is established as an essential issue to promote consistent security levels and the creation of certification schemes adapted to 5G systems across all the Member States.

### 5.1.1 Certification schemes

5G technology poses several challenges regarding cybersecurity certification. On the one hand, the high degree of heterogeneity of devices and technologies is conflicting with the need for objective comparisons regarding security aspects. On the other hand, due to the security dynamism, the certification approach must consider these changing conditions, in which the system will be operating. Therefore, agile self-assessment schemes and test automation environments will need to be created and evolved to ensure the security level is updated throughout the lifecycle. The borderless nature of the infrastructures and threats involved also mean that any vulnerability or security incident in one country can have disastrous consequences in the whole European Union. Therefore, the certification should consider interdependencies between components and cascade effects that may be produced consequently. Certain components of the 5G architecture are especially sensitive such as base stations or key technical management functions of the networks, representing a critical point for the security that must be highly considered during the certification. Furthermore, the 5G threat landscape combines traditional IP-based threats with the all-5G network, insecure legacy 2/3/4G generations and threats introduced by the IoT paradigm and the virtualization technology, creating security dependencies between different technologies.

In this sense, technologies associated to the 5G such as the virtualization or the cloud are also on the focus of security assessment. In the US, the NIST published a guide of security for full virtualization technologies with security recommendations [Scarfone 2011]. ENISA also published a document with the current status of virtualization, regarding risks and gaps [ENISA 2017]. Recently, ENISA has received the role of working towards an EU cloud security certifications scheme. In this sense, the list of cloud certification schemes

developed by Cloud Selected Industry Group on Certification[30] supported by the European Commission is being considered[31]. This list considers the TüV Rheinland Cloud Security Certification[32], whose audit checks if the requirements and criteria for cloud services, based on standards and regulations, have been implemented and the quality of the processes. The Cloud Security Alliance Cloud Controls Matrix[33] is also in the list. It provides different levels of security certification against a list of requirements for security assurance in the cloud. It considers self-assessment for organizations with a low/moderate risk profile (level 1) and third-party attestation, certification and monitoring (levels 2, 3 and 4) for organizations with high risk profiles. EuroCloud StarAudit Certification[34], from the list, provides certification to cloud services, providing trust to both the customer and the end user. The certification can vary depending on the requirements against the audit has been performed (3, 4 or 5-star Trusted Cloud Service). The list also includes the Code of Practice for Cloud Service Providers[35] for remote hosted IT services (e.g., multi-tenant). The certification can be based on self-assessment, with randomly auditing of the certificates or independent assessment performed by a third party.

### 5.1.2 Conclusion

Although currently there are many well-known cybersecurity certification schemes (e.g. Common Criteria[36], Certification de Sécurité de Premier Niveau[37] and Commercial Product Assurance[38]), none of them explicitly considers 5G and the challenges associated to this technology. Furthermore, the different schemes and standards regarding the 5G associated technologies are not enough to capture the complexity of the 5G and the security vulnerabilities that can arise because of the interconnectivity of the different components, technologies and paradigms. In this direction, a common understanding of the threats, assets, attacks and risks of 5G systems is essential to create a certification scheme that could help to recognize the cybersecurity level of a certain 5G system across all the Member States.

## 5.2 Security

Due to the expected pervasiveness of 5G-related devices (e.g. embedded and IoT appliances), there is a strong need for a secure 5G applications.

The 3GPP technical specification group of services and system aspects (TSG-SA) compiled a list of security requirements for 5G systems in [3GPP 2018]. These requirements emphasize the flexibility aspect of the system, designed to hold a wide range of different service requirements, from mobile broad-band services with data-rates up to several Gbps, to massive IoT deployments. For the latter, security is of utmost importance due to the relatively large life cycle of IoT devices, up to 10 years with one battery charge,

[30] See https://ec.europa.eu/digital-agenda/en/cloud-select-industry-group-certification-schemes.
[31] See https://resilience.enisa.europa.eu/cloud-computing-certification.
[32] See https://www.cloudwatchhub.eu/t%C3%BCv-rheinland-certified-cloud-service.
[33] See https://cloudsecurityalliance.org/artifacts/cloud-controls-matrix-v3-0-1/.
[34] See https://eurocloud.org/streams/staraudit/.
[35] See https://www.cloudindustryforum.org/content/code-practice-cloud-service-providers.
[36] See https://www.commoncriteriaportal.org/.
[37] See https://www.ssi.gouv.fr/administration/produits-certifies/cspn/.
[38] See https://www.ncsc.gov.uk/information/commercial-product-assurance-cpa.

operating without human supervision, at hard to reach locations, and lacking keypads or displays. Also, IoT devices may change owners several times. This is an issue when the user cannot update the security keys or firmware in the device, e.g., inherited IoT devices or consumer smart-home products. This only exacerbates the need for dynamically establishing or refreshing subscription data.

Additionally, 5G aims to support massive IoT deployments of heterogeneous constrained devices. These devices also support a major challenge to the leveraging platform. Billions of IoT devices require cloud services to store, process, and share the information they gather. At the same time, due to their constraints, IoT devices suppose an easy cyberattack target, posing the risk of creating DDoS botnets [Cheng 2017]. For this reason, massive Machine Type Communication (mMTC) scenarios in 5G will impose a major challenge for the signaling plane and core networks.

### 5.2.1 Software attacks

Software in 5G environments must be secure and tamper resistant. Due to its highly pervasive and connected nature, a breach in a 5G-enabled application can have dire consequences (e.g. exploiting a vulnerability in a car while it is being used).

Tampering RAM memory, while the execution is being carried out and not only when the code is being loaded into the machine is an attack that needs to be prevented. Since the code is not changed, these types of attacks cannot be detected by code integrity checks. Another vector of attack is possible by taking control of the processor and injecting code into sensitive memory areas like the stack or heap.

Return Oriented programming (ROP) consists of using some parts of the code itself in arbitrary order so that the effect is that of a malicious program. To inflict this kind of attack, the attacker needs to gain access to software memory layout which in turn is obtained by introspection techniques.

Another attack vector is exploited by accessing memory thanks to smart sequences of operating system instructions that allow the extraction of timing information, power consumption, cache access, etcetera. This side channel collected data can be then used to recover secret keys or any other valuable information.

Some remediation techniques are [Lefebvre 2018]:

- *obfuscation*: re-writes an iso-functional software more complex to analyze at the cost of slowing the software execution but providing software confidentiality;
- *anti-tampering*: performing self-integrity checks based on processing few selected memory bytes. These bytes, however, can be overtaken by a knowledgeable attacker and, on the other hand, their footprint is rather low.
- *remote execution*: placing the software on a remote trusted location not accessible to the attacker. It is one of the most secure approaches but at the cost of depending on a remote Internet connection. This technique provides both software confidentiality and integrity;
- *randomization*: mapping the memory to different locations during the execution to avoid code reutilization attacks. This technique is contrary to the TEE concept in which the memory layout signatures are used;

- *TEE*: placing the code into a trusted memory area not traceable nor modifiable with minimal impact on performance since specific hardware capabilities are used (see also Section 4.2). There are different approaches per vendor, as shown in Table 1.

| requirements | TPM+TXT | ARM Trustzone | Intel SGX | AMD SEV |
|---|---|---|---|---|
| isolation | no | yes | yes | yes |
| remote attestation | yes | no | yes | yes |
| sealing | yes | no[39] | yes | yes |
| dynamic root of trust | yes (single) | no | yes | yes |
| multiple containers | no | no | yes | yes |
| S/W+data confidentiality | yes | yes | yes | yes |
| S/W+data integrity | yes | yes | yes | no |

Table 1: TEE vs requirements.

### 5.2.2 TEE applications

Trusted execution environments (see also Section 4.2) are hardware powered techniques that fight against introspection risks in which an attacker breaks confidentiality and integrity on processed data or executed software. Data is usually encrypted and decrypted by software before sending it to the network, an attacker may achieve software modification therefore breaking data confidentiality and integrity.

Within the 5GPPP project 5G-City the use of TEE [5G-CITY 2018] is envisioned to protect Virtual Infrastructure Managers (VIM), in particular OpenStack, by means of OpenStack's trusted compute pools [Weis 2014] feature consisting of computer nodes with Intel TXT verified by a remote attestation server. With Intel TXT enabled in the compute nodes, the measured data is sent to the attestation server in the form of a TCG-standard TPM quote, a signed report of the current internal TPM registers. Then this data is compared with well-known behavioral data, determining the trustworthiness of the executed code.

Also, within 5G-City, the use of *Unikraft kernels*[40] development of tools that will enable lightweight VM development to be as easy as compiling an app for an existing OS. These light VMs contain only the necessary functionality to achieve the objective of the VM. These unikernels do not run a complete operating system but rather against small pieces of OS functionality needed. There are already unikernel systems available in the market such as: ClickOS, MiniCache, Mirage, Minipython, Solo5, OSv, Erlang on Xen, HalVM. The objective of Unikraft is to ease the development and maintenance of unikernels.

Within ENSURE [5G-ENSURE 2016] the protection of virtual switches is proposed avoiding the impersonation of the devices in SDN networks, allowing only the attested devices to connect to the central controller.

In addition, several researchers proposed a number of interesting TEE applications that can be useful in 5G environments:

[39] Secure key inside Trustzone is possible.
[40] See https://xenproject.org/developers/teams/unikraft/.

- Lazard et al. [Lazard 2018] propose TeeShift to automatically strip functions from an existing binary and to run them within a given TEE;
- in [Shih 2016] S-NFV is proposed where original NFV application is split into two parts: S-NFV enclave and host. Authors provide a performance evaluation using OpenSGX to secure Snort;
- a completely different approach [Felsen 2019] is the mapping of security functions to Boolean circuits that can be then evaluated within an Intel SGX enclave – this approach eliminates the need of a per-application design allowing reusability;
- MicroTEE [Zhang 2019] proposes a TEE Operating system based on the microkernel for ARM TrustZone and SEL4 Microkernel[41]. This approach is applicable to mobile phones which is interesting for fog environments.

Finally, some open source projects provide with abstractions and tools to develop TEE like solutions, like asylo[42], OpenSGX[43] and OpenEnclave[44].

### 5.2.3 Conclusion

5G goals revolve around connecting every aspect of life and society. For instance, sectors like e-Health, Intelligent Transportation Systems, Industry 4.0, Smart Phones, Wearables, etcetera. However, if critical systems connected suffer an attack, the resulting consequences may be catastrophic [Ahmad 2019]. For this reason, there is a lot of attention by industry and academia to identify and solve these challenges during the 5G standardization process. The NGMN Alliance[45] identified a set of highlighted security challenges at the early stages of 5G standardization. Some of these are highly discussed in the literature [Ahmad 2018], specifically: (i) flash or surge network traffic, due to massive public events like sports or music concerts, (ii) security at radio interfaces, prone to passive attacks, (iii) user plane confidentiality and integrity protection, (iv) roaming security due to long-term security keys not being updated when roaming from one operator to another, (v) Denial-of-Service (DoS) Attacks on end-devices where the device's operative system implements the security, and (vi) signaling storms due to distributed control systems.

5G introduced several novel and disruptive networking technologies. In pre-established networking technologies, security achieved maturity over a long period of time, but the recent paradigm changes of cellular network brought several challenges to be further researched. These challenges can be found in all the different 5G networks. The key technologies to obtain the service requirements in the access network are identified as (i) ultra-densification and offloading by use of pico-cells and femto-cells, (ii) the use of millimeter Wave (mmWave), (iii) employing unlicensed radio bands, and (iv) several advances in MIMO technologies [Andrews 2014]. However, all these access network enablers must be further investigated from the security point of view. The new security challenges of these technologies are way different to those presented in last-generation networking technologies. For instance, while 5G heavily leverages on MIMO technologies, full secrecy protection in resource allocation needs further research [Ng 2015].

[41] See https://sel4.systems/.
[42] See https://asylo.dev/.
[43] See https://github.com/sslab-gatech/opensgx.
[44] See https://openenclave.io/sdk/.
[45] See https://www.ngmn.org/work-programme/5g-white-paper.html.

Novel technologies like NFV and SDN are a requirement for 5G [Chandramouli 2019]. In order to achieve an efficient and flexible overall performance in 5G, these new networking paradigms have been considered as a necessity. However, several security risks and attacks have been identified when employing them. For instance, the integrity and confidentiality of the data circulating through Virtual Network Functions (VNFs) depends on the hypervisor and cloud stack implementation. Vulnerabilities in NFV and SDN implementations have been frequently found. Hence, it is still a major security challenge to achieve a secure NFV environment. In SDN there are still severe risks that a control application bug may wreak havoc in a central network controller.

## 5.3 AI-based security

The dependence that exists nowadays in computer networks and services brings the necessity of adequate and reliable security mechanisms to protect the individuals and their data. Security procedures in 5G networks need to be reshaped in order to cope with the new requirements of this paradigm, as the traditional solutions adopted in the ossified networks are now outdated. The 5G ecosystem is bringing together many technologies expected to coexist in the same infrastructure. Besides, the virtualization of the resources in the form of VNFs implies the sharing of physical resources among different service providers. Machine learning arises as the potential leader to lead this change, by supporting the design of new mechanisms that can dynamically adjust themselves to the volatile nature of the envisioned 5G networks. ML is a field of study that gives the computers the ability to learn without being explicitly programmed for this task. ML models receive a dataset containing the information from which the algorithm is going to learn. ML algorithms are usually categorized in three different categories: Supervised Learning (SL), Unsupervised Learning (UL) and Reinforcement Learning (RL).

### 5.3.1 Machine learning techniques

The application of ML techniques in 5G networks security is being deeply studied by the scientific community. Numerous surveys covering the state of the art in this area have been published. Works in [HaddadPajouh 2019, da Costa 2019, Mohanta 2020, Al-Garadi 2020] explored the requirements, challenges and existing solutions regarding learning-based security procedures in Internet of Things networks, envisioned to be a fundamental pillar of the 5G ecosystem. In [Huang 2020], the authors examined ML techniques against hardware trojan attacks from four perspectives, i.e. detection, design for security, bus security, and secure architectures. Alrehan et al. [Alrehan 2019] studied how to detect Distributed Denial of Service (DDoS) on Vehicular Ad-hoc Networks (VANETs) using ML algorithms. Authors in [Sagar 2019] introduced the necessity of cybersecurity solutions based on artificial neural networks, in order to prevent attacks before they occur. Finally, in [Tang 2019], the application of AI techniques to communications, networking and security in vehicular networks was reviewed.

AI-based security is a hot topic, and, in consequence, there are several works exploring the synergies between ML models and their application in security mechanisms. Work in [Tang 2019] proposed a network intrusion detection method based on semantic re-encoding to increase the distinguishing ability of traffic and enhanced by using deep learning to improve the generalization ability of the algorithm. In [Wu 2020], authors presented a first approach in designing a deep learning model to predict the pattern of the next sequence of cyberattacks in certain areas. Sarker et al. [Sarker 2020] developed an intrusion detection model, based on snort, that uses ML to learn from a ranking of security features and builds a tree-based generalized

intrusion detection model. The designed solution is effective in terms of prediction accuracy and, besides, it is also able of minimizing the computational complexity by reducing the features dimension. Authors in [Mulinka 2019] proposed a continuous and adaptive learning framework for network security that builds dynamic models to detect network attacks in real time. Bagaal et al. [Bagaa 2020] presented a ML-based security framework that handles the security aspects in IoT systems. The framework includes a monitoring agent and a reaction agent that use ML models, i.e. supervised learning, to perform traffic packets analysis and anomaly-based intrusion detection. In [Mamolar 2019] authors define a self-protection cognitive framework to protect dynamically against DDoS attacks in 5G networks. implemented and validated in real 5G testbed 5gppp SELFNet. In [Maimó 2018] authors proposed a deep learning-based system for botnet detection in 5G networks. They used an existing botnet database to test the performance of the solution.

### 5.3.2 Conclusion

The envisioned adoption of ML-based management and orchestration systems in 5G has opened a new way for attack vectors in heterogeneous networks. Both the hosting frameworks and the ML algorithms themselves can be corrupted in many different forms, impacting the network performance and the management and orchestration capabilities. The system availability is the first aspect to take into account, as the ML modules can be directly attacked, provoking malfunctions. The integrity of the data can also be in danger because the attackers can alter the information collected by the system, altering the later used ML decision making process. Finally, data or user privacy can also be violated by the interception of sensitive information. The security and privacy concerns of the ML components of the network are a fundamental pillar in the development of 5G and must be scrupulously studied.

## 5.4 Authentication, authorization and accounting

Among the highlighted requirements for authentication services in 5G [3GPP 2018] there are: (i) support for efficient mechanisms for a wide range of devices, (ii) a suitable framework that allows alternative authentication and key agreement methods to non-public networks employing subscriber data not defined by the 3GPP itself, (iii) support for third-party operator controlled authentication methods and protocols with custom credential schemes for isolated environments — like industrial automated factories. Another characteristic in the 5G security design is that there is no trust in the roaming partner, i.e. the home network does not trust the serving network employed by the UE in roaming scenarios. For this reason, 5G grants the home network total control over authentication and key derivation strategies. This way, the home network and UE can authenticate the serving network and protect the subscriber against serving network impersonating attacks [Kunz 2018]. Simultaneously, the 5G system must facilitate resource-efficient technologies that minimize signaling overhead, and support multicast downlink transmissions to authenticate groups of devices in massive IoT scenarios.

### 5.4.1 Unified authentication framework architecture

The 5G system authentication features include: (i) a unified authentication framework that will achieve support for more use-cases, (ii) UE identity protection, (iii) enhanced home network control, (iv) and key separation in key derivation elements. Authentication and key management are fundamental to the secure operation of cellular networks due to the need for mutual authentication between the network and the user. The derived cryptographic keys are employed mainly for user plane integrity and confidentiality and signaling messages.

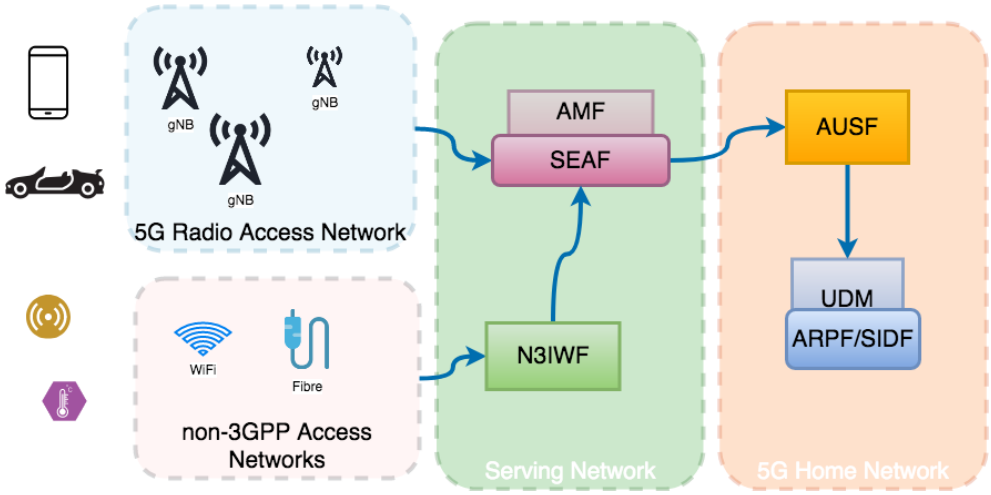


Figure 5: Unified authentication architecture.

The different elements of the 5G unified authentication (see also Figure 5) and key derivation framework [3GPP 2019] are:

- *User Equipment* (UE): it is the device owned by the subscriber itself. It contains the Mobile Equipment (ME) — the device hardware — and the (U)SIM card;
- *Security Anchor Function* (SEAF): a component located in the serving network acting as a transparent relay during authentication between UE and the home network. Under normal circumstances, it relies on the home network decision to accept or reject the registration attempt by the UE.
- *Authentication Server Function* (AUSF): a module in the home network performing the authentication together with the UE. This is the architecture element that eventually decides to accept or reject the UE access to the system. It may rely on back-end services depending of the kind of authentication method employed. The AUSF handles authentication requests at the home network side coming from either 3GPP or non-3GPP access networks.
- *Unified Data Management* (UDM): a component containing the functions related to data management such as the Authentication Credential Repository and Processing Function (*ARPF*). The latter chooses the authentication method based on the subscriber data, i.e., the subscriber ID and the configured policy. The ARPF also computes the keying materials for the AUSF.
- *Subscription Identifier De-Concealing Function* (SIDF): this module de-encrypts the long-term subscriber identity that the UE sends as part of the registration request. The subscriber identity is encrypted with the home network public key and decrypted by the SIDF.
- *Non-3GPP Interworking Function* (N3IWF): this needed only when the authentication process is performed over an untrusted access network (e.g. Wi-Fi or fiber), this module acts as an VPN server that establishes a security link between the UE and the serving network. This way, the UE can securely access the core, regardless of the intermediary access network security.

The goal of the 5G Unified Authentication Framework is to achieve authentication mechanisms that are open — through technologies like EAP — and access network agnostic, i.e., devices can access the core services through 3GPP and non-3GPP access networks — like Wi-Fi or fiber. Due to EAP being compatible

with several methods, the network operator may choose the preferred method for deployment characteristics. This is how 5G offers authentication flexibility. Additionally, EAP enables AAA framework that are easily extensible, further improving heterogeneous deployments.

### 5.4.2 Conclusion

User authentication is still an open challenge in 5G systems for a variety of reasons. First, different stakeholders may have different needs and security requirements. Second, end device heterogeneity is a key factor for supporting as many use-cases as possible by 5G systems. Not only devices are different in form and factor, but also in network and authentication requirements. In the first phase of 5G standardization, use cases were grouped in three major categories, namely: enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC), and massive Machine Type Communications (mMTC). These category groups focus on the service needs with regards to data-rate, low-latency and how important the data is. As an illustration, the periodic packets sent from a light-post may not be as important as an automatic fire extinguisher valve. And finally, 5G aims also at integrating emerging communication access technologies, including those that employ licensed and unlicensed radio bands. Also, different integration mechanisms have been put in place in order to integrate technologies by non-3GPP standard developing organizations. This is, devices employing non-3GPP technologies, e.g., Wi-Fi or fiber, can access the 5G core services and data networks through intermediate architecture elements. This pushes the need for a seamless and access-independent security infrastructure in 5G systems.

In the future, with the adoption of a unified authentication architecture (see Section 5.4.1) 5G-enabled devices and applications will be able to create a “world of 5G”, where everything is connected, hopefully, in a safe manner. However, for the present, a full AAA for 5G networks is not yet available, due to the astonishing variety of involved partners, technologies and security requirements.

# 6 Conclusion

This deliverable reports the outcomes of the investigations of Task 3.9 on Continuous Scouting. It presents a collection of novel technologies and state-of-the-art models in the cybersecurity world that may significantly impact the near future. In addition, this document also discusses what are the challenges and the most recent trends that are affecting the current research.

Machine and deep learning techniques (see Chapter 2) are key technologies for detecting new cyberthreats in a timely manner. Each new year this type of AI techniques are proving themselves more and more effective. However, the application of these models is critical when the privacy of users is at stake. Their application in several sectors (e.g. the health industry) must be rigorously performed in order to be compliant with the new GDPR.

Artificial intelligence techniques (see Chapter 3) can help the defender but also the attacker. An attacker can introduce some ad-hoc noise in the training process of a classifier in order to misclassify some samples (e.g. classify as benign events an attack, or vice-verssa). Furthermore automatic AI-based techniques are starting to surface as new powerful tools in the social engineering area.

Trusted systems (Chapter 4) are not a new idea, however, their diffusion is starting to take place, especially in cloud environments. Several new attacks have been discovered, making these hardware components less secure than though. In addition, the lack of proper standards (especially for TEEs) is causing the production of very specialized software components, thus limiting their portability on different hardware platforms.

5G (Chapter 5) is the new promised land where every device is connected. The adoption of this family of technologies, however, will also bring new problems. An attack on a 5G-enabled appliance may have severe repercussions on its connected devices, thus exacerbating the attack's effects. The wide use of both hardware and software protections will most likely be a decisive factor to make 5G networks more secure, especially with the growth of the SDN paradigm. Furthermore, user authentication will play a pivotal role. Different stakeholders and the vast heterogeneity of devices supporting 5G will be a challenge in the near future.

While it is not possible to predict the future and hence how the IT field will evolve, our investigations can, hopefully, provide some interesting food for thought about how the current technologies and their applications are starting to shape our world of tomorrow.

# 7 References

[3GPP 2018] 3GPP, TS 22.261, *Service requirements for the 5G system,* Stage 1 Release 16

[3GPP 2019] 3GPP, *Security Architecture and Procedures for 5G System. Technical Specification (TS)*, 3rd Generation Partnership Project (3GPP), Release 16.0, http://www.3gpp.org/ftp//Specs/archive/33_

[5G-CITY 2018] 5G-CITY (2nd Phase), Deliverable 3.1, *5GCity Edge Virtualization Infrastructure Design*

[5G-ENSURE 2016] 5G-ENSURE, Deliverable 3.4, *5G-PPP Security Enablers Documentation*

[Aamir 2019] Aamir M., Ali Zaidi S. M., *Clustering based semi-supervised machine learning for DDoS attack classification*, Journal of King Saud University, Computer and Information Sciences, 2019, https://doi.org/10.1016/j.jksuci.2019.02.003.

[Ahmad 2018] Ahmad I., Kumar T., Liyanage M., Okwuibe J., Ylianttila M., Gurtov A., *Overview of 5G Security Challenges and Solutions,* IEEE Communications Standards Magazine, 2018, vol. 2, pp. 36–43, https://doi.org/10.1109/MCOMSTD.2018.1700063

[Ahmad 2019] Ahmad I., Shahabuddin S., Kumar T., Okwuibe J., Gurtov A., Ylianttila M., *Security for 5G and Beyond*, IEEE Communications Surveys & Tutorials, 2019, https://doi.org/10.1109/comst.2019.2916180

[Al-Garadi 2020] Al-Garadi M. A., Mohamed A., Al-Ali A., Du X., Ali I., Guizani M., *A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security*, IEEE Communications Surveys & Tutorials, 2020, https://doi.org/10.1109/comst.2020.2988293

[Alrehan 2019] Alrehan A. M., Alhaidari F. A., *Machine Learning Techniques to Detect DDoS Attacks on VANET System: A Survey*, 2nd International Conference on Computer Applications and Information Security, 2019, https://doi.org/10.1109/CAIS.2019.8769454

[Al-Rubaie 2019] Al-Rubaie, M., Chang, J. M., *Privacy-preserving machine learning: Threats and solutions*, IEEE Security & Privacy, IEEE, vol. 17, pp. 49-58

[Andrews 2014] Andrews J. G., Buzzi S., Choi W., Hanly S. V., Lozano A., Soong A. C. K., Zhang J. C., *What Will 5G Be?,* IEEE Journal on Selected Areas in Communications, 2014, vol. 32, pp. 1065–1082. https://doi.org/10.1109/JSAC.2014.2328098

[Arrietaa 2020] Arrietaa A. B., Díaz-Rodríguez N., Del Sera J., Bennetotb A., Tabikg S., Barbado A., Garciag S., Gil-Lopez S., Molina D., Benjamins R., Chatila R., Herrera F., *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*, Information Fusion, 2020, vol. 58

[Bagaa 2020] Bagaa M., Taleb T., Bernabe J. B. and Skarmeta A., *A Machine Learning Security Framework for Iot Systems*, IEEE Access, 2020, vol. 8, pp. 114066-114077, https://doi.org/10.1109/ACCESS.2020.2996214

[Bahnsen 2018] Bahnsen A. C., Torroledo I., Camacho D., Villegas S., *DeepPhish: Simulating malicious AI,* Proceedings of the 2018 APWG Symposium on Electronic Crime Research, 2018, pp. 1–8

[Beckers 2017] Beckers K., Schosser D., Pape S., Schaab P., *A Structured Comparison of Social Engineering Intelligence Gathering Tools*, Trust, Privacy and Security in Digital Business - 14th International Conference, 2017, pp. 232-246

[Bendel 2019] Bendel O, *The synthetization of human voices*, AI & Society, 2019, vol. 34, pp. 83–89

[Bhatt 2020] Bhatt U., Xiang A., Sharma S., Weller A., Taly A., Jia Y., Eckersley P., *Explainable machine learning in deployment*, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, pp. 648-657

[Brown 2020] Brown T. B. et al., *Language Models are Few-Shot Learners*, arXiv preprint, 2020, https://arxiv.org/abs/2005.14165

[Brundage 2018] Brundage M. et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, 2018, https://maliciousaireport.com/

[Calzavara 2019] Calzavara S., Conti M., Focardi R., Rabitti A. and Tolomei G., *Mitch: A Machine Learning Approach to the Black-Box Detection of CSRF Vulnerabilities*, IEEE European Symposium on Security and Privacy (EuroS&P2019), 2019, pp. 528–543, https://doi-org/10.1109/EuroSP.2019.00045

[Chamikara 2019] Chamikara M. A. P., Bertok P., Khalil I., Liu D., Camtepe S., *Local differential privacy for deep learning,* arXiv preprint, 2019, https://arxiv.org/abs/1908.02997

[Chandramouli 2019] Chandramouli D., Liebhart R., Pirskanen J., Choudhary G., Kim J., Sharma V., *5G for the Connected World*, Wiley, 2019, vol. 9, https://doi.org/10.1002/9781119247111

[Chen 2019] Chen V., Pastro V., Raykova M., *Secure computation for machine learning with SPDZ*, arXiv preprint, 2019, https://arxiv.org/abs/1901.00329

[Cheng 2017] Cheng C., Lu R., Petzoldt A., Takagi T., *Securing the Internet of Things in a Quantum World*, IEEE Communications Magazine, 2017, vol. 55, pp. 116–120, https://doi.org/10.1109/MCOM.2017.1600522CM

[da Costa 2019] da Costa K. A. P., Papa J. P., Lisboa C. O., Munoz R., de Albuquerque V. H. C., *Internet of Things: A survey on machine learning-based intrusion detection approaches,* Computer Networks, 2019, vol. 151, pp. 147–157, https://doi.org/10.1016/j.comnet.2019.01.023

[Darmody 2020] Darmody A., Zwick, D., *Manipulate to empower: Hyper-relevance and the contradictions of marketing in the age of surveillance capitalism,* Big Data & Society, vol. 7

[Dolhansky 2019] Dolhansky B., Howes R., Pflaum B., Baram N., Ferrer C. C., *The Deepfake Detection Challenge (DFDC) Dataset,* arXiv preprint, 2019, https://arxiv.org/abs/2006.07397

[Doshi 2018] Doshi R., Apthorpe N. and Feamster N., *Machine Learning DDoS Detection for Consumer Internet of Things Devices*, IEEE Security and Privacy Workshops (SPW), 2018, pp. 29–35, https://doi-org/10.1109/SPW.2018.00013

[Dwork 2014] Dwork C., Roth A., *The algorithmic foundations of differential privacy*, Foundations and Trends in Theoretical Computer Science, 2014, vol. 9, pp. 211-407

[ENISA 2017] ENISA, *Security aspects of virtualization*, 2017, https://www.enisa.europa.eu/publications/security-aspects-of-virtualization/at_download/fullReport

[Eykholt 2016] Eykholt K., Evtimov I., Fernandes E., Li B., Rahmati A., Xiao C., Prakash A., Kohno , Song D., *Robust physical-world attacks on deep learning models*, cleverhans v0.1: an adversarial machine learning library, 2016

[Felsen 2019] Felsen S., Kiss Á., Schneider T., Weinert C., *Secure and Private Function Evaluation with Intel SGX*, Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop, 2019, pp. 165–181, https://doi.org/10.1145/3338466.3358919

[Finlayson 2019] Finlayson S. G., Bowers J. D., Ito J., Zittrain J. L., Beam A. L., Kohane I. S., *Adversarial attacks on medical machine learning*, Science, 2019, vol. 363

[Fiolhais 2020] Fiolhais L., Martins P. Sousa L., *Software Emulation of Quantum Resistant Trusted Platform Modules*, pp. 477-484, https://doi.org/10.5220/0009886004770484

[Gadelrab 2018] Gadelrab M., Elsheikh M., Ghoneim M., Rashwan M., *BotCap: Machine Learning Approach for Botnet Detection Based on Statistical Features*, International Journal of Communication Networks and Information Security, 2018, pp. 563-579

[Gallagher 2020] Gallagher D. E., Sonnick J. J., Thorpe, M. G., *Robocall detection*, U.S. Patent No. 10,582,041, U.S. Patent and Trademark Office

[GDPR 2016] Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (*General Data Protection Regulation*) OJ L119/1

[Goodman 2017] Goodman, B., Flaxman, S., *European Union regulations on algorithmic decision-making and a "right to explanation"*, AI magazine, vol. 38, pp. 50-57

[Grosse 2016] Grosse K., Papernot N., Manoharan P., Backes M., McDaniel P., *Adversarial perturbations against deep neural networks for malware classification*. arXiv preprint, 2016, https://arxiv.org/abs/1606.04435

[Guidotti 2018] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., *A survey of methods for explaining black box models*, ACM computing surveys, 2018, ACM, vol. 51, pp. 1-42

[HaddadPajouh 2019] HaddadPajouh H., Dehghantanha A. Parizi, R. M., Aledhari M., Karimipour H., *A survey on internet of things security: Requirements, challenges, and solutions*, Internet of Things, https://doi.org/10.1016/j.iot.2019.100129

[Hall18] Hall P., *An introduction to machine learning interpretability,* O'Reilly Media Incorporated

[Hoang 2018] Hoang X.D., Nguyen Q.C., *Botnet Detection Based On Machine Learning Techniques Using DNS Query Data*, 2018, Future Internet, https://doi.org/10.3390/fi10050043

[Hosseini 2017] Hosseini H, Kannan S, Zhang B, Poovendran R. *Deceiving Google's Perspective API Built for Detecting Toxic Comments*, arXiv preprint, 2017, https://arxiv.org/abs/1702.08138

[Hosseinzadeh 2020] Hosseinzadeh S., et al., *Recent trends in applying TPM to cloud computing*, Security and Privacy, 2020, https://doi.org/10.1002/spy2.93

[Huang 2018] Huang C.H, Lee T.H, Chang L.H, Lin J. R, Horng G., *Adversarial Attacks on SDN-Based Deep Learning IDS System*, International Conference on Mobile and Wireless Technology, 2018, pp. 181–191

[Huang 2020] Huang Z., Wang Q., Chen Y., Jiang X., *A Survey on Machine Learning against Hardware Trojan Attacks: Recent Advances and Challenges*, IEEE Access, 2020, vol. 8, pp. 10796–10826, https://doi.org/10.1109/ACCESS.2020.2965016

[Lazar 2019] Lazar D. A., Bıyık E., Sadigh D., Pedarsani R., *Learning How to Dynamically Route Autonomous Vehicles on Shared Roads*, arXiv preprint 2019, https://arxiv.org/abs/1909.03664

[Liang 2017] Liang B, Li H, Su M., Bian P, Li X, Shi W., *Deep text classification can be fooled*, arXiv preprint, 2017, https://arxiv.org/abs/1704.08006

[Jang 2017] Li Y., Jang J., Hu X. and Ou X., *Android malware clustering through malicious payload mining,* Research in Attacks Intrusions and Defenses, 2017, Springer, pp. 192-214

[Kaissis 2020] Kaissis G. A., Makowski M. R., Rückert D., Braren R. F., *Secure, privacy-preserving and federated machine learning in medical imaging*, Nature Machine Intelligence, 2020, pp. 1-7

[Kaloudi 2020] Kaloudi N., Li J., *The AI-based cyber threat landscape: A survey,* ACM Computing Surveys ACM, 2020, vol. 53, pp. 1-34

[Kawa 2019] Kawa D., Punyani S., Nayak P., Karkera A., Jyotinagar V., *Credit Risk Assessment from Combined Bank Records using Federated Learning*, 2019

[Kujawa 2020] Kujawa A. et al., *2020 State of Malware Report*, Malware Bytes, 2020, https://resources.malwarebytes.com/files/2020/02/2020_State-of-Malware-Report.pdf

[Kunz 2018] Kunz A., Zhang X., *New 3GPP Security Features in 5G Phase 1. 2018 IEEE Conference on Standards for Communications and Networking*, IEEE Conference on Standards for Communications and Networking, 2018, https://doi.org/10.1109/CSCN.2018.8581763

[Lazard 2018] Lazard T., Götzfried J., Müller T., Santinelli G., Lefebvre V., *TEEshift: Protecting code confidentiality by selectively shifting functions into TEEs,* Proceedings of the ACM Conference on Computer and Communications Security, 2018, pp. 14–19, https://doi.org/10.1145/3268935.3268938

[Lecuyer 2019] Lecuyer M., Atlidakis V., Geambasu R., Hsu D., Jana S., *Certified robustness to adversarial examples with differential privacy,* IEEE Symposium on Security and Privacy, 2019, IEEE, pp. 656-672

[Lefebvre 2018] Lefebvre V., Santinelli G., Müller T., Götzfried J., *Universal trusted execution environments for securing SDN/NFV operations*, ACM International Conference Proceeding Series, https://doi.org/10.1145/3230833.3233256

[Lee 2020] Lee D., et al., *Keystone: An open framework for architecting trusted execution environments*, Proceedings of the Fifteenth European Conference on Computer Systems, 2020

[Li 2014] Li, M., Andersen, D. G., Park, J. W., Smola, A. J., Ahmed, A., Josifovski, V., Su, B. Y., *Scaling distributed machine learning with the parameter server*, 11th USENIX Symposium on Operating Systems Design and Implementation, pp. 583-598

[Liebenberg 2018] Liebenberg D. et al., *The Illicit Cryptocurrency Mining Threat*, Cyberthreat Allliance, 2018, https://www.cyberthreatalliance.org/wp-content/uploads/2018/09/CTA-Illicit-CryptoMining-Whitepaper.pdf

[Ma 2019] Ma Z., Ge H., Liu Y., Zhao M. and Ma J., *A Combination Method for Android Malware Detection Based on Control Flow Graphs and Machine Learning Algorithms*, IEEE Access, 2019, vol. 7, pp. 21235-21245, 2019, https://doi:.org/10.1109/ACCESS.2019.2896003

[Ma 2020] Ma Y., Xie T., Li J., Maciejewski R., *Explaining Vulnerabilities to Adversarial Machine Learning through Visual Analytics*, IEEE Transactions on Visualization and Computer Graphics, 2020, vol. 26

[Mackenzie 2019] Mackenzie P., *WannaCry Aftershock*, Sophos Ltd., 2019, https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/WannaCry-Aftershock.pdf

[Maimó 2018] Maimó L. F., et al. *A self-adaptive deep learning-based system for anomaly detection in 5G networks*, IEEE Access, 2018, vol. 6, pp. 7700-7712.

[Mamolar 2019] Mamolar A. S., Salvá-García P., Chirivella-Perez E., Pervez Z., Alcaraz Calero J. M., Wang Q., *Autonomic protection of multi-tenant 5G mobile networks against UDP flooding DDoS attacks*, Journal of Network and Computer Applications, 2019, vol. 145, https://doi.org/10.1016/j.jnca.2019.102416

[Manos 2017] Manos A. et al., *Understanding the mirai botnet*, proceedings of the 26th USENIX security symposium, 2017, pp. 1093–1110, August 16-18, 2017

[McStay 2018] McStay A., *Emotional AI: The rise of empathic media*, Sage Publications, 2018, ISBN: 978-1473971103

[Mereani 2018] Mereani F. A. and Howe J. M., *Detecting Cross-Site Scripting Attacks Using Machine Learning*, The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018), 2018, Advances in Intelligent Systems and Computing, vol 723, Springer, https://doi-org/10.1007/978-3-319-74690-6_20

[Moghimi 2020] Moghimi D., et al., *TPM-FAIL:TPM meets Timing and Lattice Attacks*, 29th USENIX Security Symposium , 2020

[Mohanta 2020] Mohanta B. K., Jena D., Satapathy U., Patnaik S., *Survey on IoT security: Challenges and solution using machine learning, artificial intelligence and blockchain technology,* Internet of Things, 2020, vol. 11, https://doi.org/10.1016/j.iot.2020.100227

[Mulinka 2019] Mulinka P., Casas P., Vanerio J., *Continuous and Adaptive Learning over Big Streaming Data for Network Security,* Proceeding of the 2019 IEEE 8th International Conference on Cloud Networking, 2019, https://doi.org/10.1109/CloudNet47604.2019.9064134

[Murdock 2020] Murdock Kit, et al., *Plundervolt: Software-based Fault Injection Attacks against Intel SGX*, IEEE Symposium on Security and Privacy, 2020, pp. 1466-1482

[Nilsson 2020] Nilsson A., et al., *A Survey of Published Attacks on Intel SGX*, arXiv preprint, 2020, https://arxiv.org/abs/2006.13598

[Ng 2015] Ng D. W. K., Schober, R., *Secure and Green SWIPT in Distributed Antenna Networks With Limited Backhaul Capacity*, IEEE Transactions on Wireless Communications, vol. 14, pp. 5082–5097. https://doi.org/10.1109/TWC.2015.2432753

[Orman 2003] Orman H., *The Morris Worm: A Fifteen-Year Perspective*, Security & Privacy, IEEE, 2003,, https://doi.org/10.1109/MSECP.2003.1236233

[Pape 2017] Pape S., Serna-Olvera J., Tesfay W., *Why Open Data May Threaten Your Privacy*, Workshop on Privacy and Inference, 2017

[Papernot 2016] Papernot N., McDaniel P. D., Goodfellow I. J., *Transferability in machine learning: From phenomena to black- box attacks using adversarial samples*, arXiv preprint, 2016, https://arxiv.org/abs/1605.07277

[Reich19] Reich D., Todoki A., Dowsley R., De Cock M., *Privacy-preserving classification of personal text messages with secure multi-party computation*, Advances in Neural Information Processing Systems, pp. 3757-3769

[Ribeiro 2016] Ribeiro M. T., Singh S., Guestrin C., *" Why should I trust you?" Explaining the predictions of any classifier*, Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1135-1144

[Roscher 2020] Roscher R., Bohn B., Duarte M. F., Garcke J., *Explainable machine learning for scientific insights and discoveries*, IEEE Access, 2020, vol. 8, pp. 42200-42216

[Ross 2018] Ross K., Moh M., Moh T. and Yao J., *Multi-source data analysis and evaluation of machine learning techniques for SQL injection detection*, proceedings of the ACMSE 2018 Conference (ACMSE 2018), 2018, Association for Computing Machinery, pp. 1–8, https://doi.org/10.1145/3190645.3190670

[Sagar 2019] Sagar B. S., Niranjan S., Kashyap N., Sachin D. N., *Providing cyber security using artificial intelligence - A survey,* Proceedings of the 3rd International Conference on Computing Methodologies and Communication, 2019, pp. 717–720, https://doi.org/10.1109/ICCMC.2019.8819719

[Samanta 2018] Samanta S, Mehta S., *Generating Adversarial Text Samples,* Proceedings of the 40th European Conference on Information Retrieval Research, 2018

[Sarker 2020] Sarker I. H., Abushark Y. B., Alsolami F., Khan A. I., *IntruDTree: A machine learning based cyber security intrusion detection model*, Symmetry, 2020, vol. 12, https://doi.org/10.3390/SYM12050754

[Savadjiev 2019] Savadjiev P., Chong, J., Dohan A., Vakalopoulou M., Reinhold C., Paragios N., Gallix B., *Demystification of AI-driven medical image interpretation: past, present and future*, European Radiology, 2019, vol. 29, pp. 1616–1624

[Scarfone 2011] Scarfone K., Souppaya M., Hoffman P., *Guide to Security for Full Virtualization Technologies*, NIST Special Publication 800-125, 2011

[Schaab 2016] Schaab P., Beckers K., Pape S., *A systematic Gap Analysis of Social Engineering Defence Mechanisms considering Social Psychology,* 10th International Symposium on Human Aspects of Information Security & Assurance, 2016, pp. 19-21

[Schaab 2017] Schaab P., Beckers K., Pape S., *Social engineering defence mechanisms and counteracting training strategies,* Information and Computer Security, 2017, vol. 25, pp. 206-222

[Shalaginov 2018] Shalaginov A., Banin S., Dehghantanha A., Franke K., *Machine Learning Aided Static Malware Analysis: A Survey and Tutorial*, Cyber Threat Intelligence. Advances in Information Security, Springer, vol 70, https://doi.org/10.1007/978-3-319-73951-9_2

[Shepherd 2016] Shepherd C., et al., *Secure and Trusted Execution: Past, Present, and Future - A Critical Review in the Context of the Internet of Things and Cyber-Physical Systems*, IEEE Trustcom, 2016, pp. 168-177.

[Shih 2016] Shih M. W., Kumar M., Kim T., Gavrilovska A., *S-NFV: Securing NFV states by using SGX*, SDN-NFV Security 2016 - Proceedings of the 2016 ACM International Workshop on Security in Software Defined Networks and Network Function, 2016

[Seymour 2016] Seymour J. and Tully P., *Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter*, Black Hat USA, 2016, pp. 1–39

[Sharif 2016] Sharif M, Bhagavatula S, Bauer L, Reiter M. K., *Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition*, Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016

[Shen 2020] Shen Y., Gu J., Tang X., Zhou B., *Interpreting the latent space of gans for semantic face editing*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9243–9252

[Szeles 2020] Szeles J.G., Condor R.A., *Loading DLLs for illicit profit. A story about a Metamorfo distribution campaign*, Bitdefender, 2020, https://www.bitdefender.com/files/News/CaseStudies/study/333/Bitdefender-PR-Whitepaper-Metamorfo-creat4500-en-EN-GenericUse.pdf

[Tang 2019] Tang F., Kawamoto Y., Kato N., Liu J., *Future Intelligent and Secure Vehicular Network Toward 6G: Machine-Learning Approaches*, Proceedings of the IEEE, vol. 108, pp. 292–307, https://doi.org/10.1109/JPROC.2019.2954595

[Telephone 1991] *Telephone Consumer Protection Act of 1991* (TCPA), 47 U.S.C. § 227

[Tesfay 2016] Tesfay W. B., Serna, J., Pape S., *Challenges in Detecting Privacy Revealing Information in Unstructured Text*, Workshop on Society, Privacy and the Semantic Web - Policy and Technology PrivOn 2016 at the International Semantic Web Conference, 2016

[Visani 2020] Visani G., Bagli E., Chesani F., Poluzzi A., Capuzzo D., *Statistical stability indices for LIME: obtaining reliable explanations for Machine Learning models*, arXiv preprint, 2020, https://arxiv.org/abs/2001.11757

[Wanga 2019] Wanga X., Li J., Kuang X., Tan Y., Li J., *The security of machine learning in an adversarial setting: A survey*, Journal of Parallel and Distributed Computing, 2019, vol. 30

[Weis 2014] Weis S., *Trusted Computing & OpenStack*, PrivateCore, 2014, https://pdfs.semanticscholar.org/presentation/762c/447a11ee258d313fb87c24dab0b5939cd14c.pdf

[Wen 2019] Wen A., Fu S., Moon S., El Wazir M., Rosenbaum A., Kaggal V. C., Liu S., Sohn S., Liu H., Fannpj J., *Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation*, Digital Medicine, 2019, vol. 2

[Wu 2020] Wu Z., Wang J., Hu L., Zhang Z., Wu H., *A network intrusion detection method based on semantic Re-encoding and deep learning,* Journal of Network and Computer Applications, 2020, vol. 164, https://doi.org/10.1016/j.jnca.2020.102688

[Yan 2020] Yan Z., Govindaraju V., Zheng Q., Wang Y., *IEEE Access Special Section Editorial: Trusted Computing*, IEEE Access, 2020, vol. 8, pp. 25722-25726, https://doi.org/10.1109/ACCESS.2020.2969768

[Yang 2018] Yang M., Zhu T., Liu, B., Xiang Y., Zhou W., *Machine learning differential privacy with multifunctional aggregation in a fog computing architecture,* IEEE Access, 2018, vol. 6, pp. 17119-17129

[Yang 2019] Yang, Q., Liu, Y., Chen, T., & Tong, Y., *Federated machine learning: Concept and applications*, ACM Transactions on Intelligent Systems and Technology, 2019, ACM, vol. 10, pp. 1-19

[Yao 2017] Yao Y., Viswanath B., Cryan J., Zheng H., Zhao B. Y., *Automated crowdturfing attacks and defenses in online review systems,* Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1143–1158

[Yuan 2019] Yuan X., He P., Zhu Q., Li X., *Adversarial Examples: Attacks and Defenses for Deep Learning*, IEEE Transactions on Neural Networks and Learning Systems, 2019, vol. 30

[Zhang 2019] Ji D., Zhang Q., Zhao S., Shi Z., Guan Y., *MicroTEE: Designing TEE OS Based on the Microkernel Architecture*, https://doi.org/10.1109/TrustCom/BigDataSE.2019.00014

[Zhou 2018] Zhou Z., Tang D., Wang X., Han W., Liu X., Zhang K., *Invisible Mask: Practical Attacks on Face Recognition with Infrared*. arXiv preprint, 2018, https://arxiv.org/abs/1803.04683

[Zhou 2019] Zhou Y., Kantarciog M., Xi B., *A survey of game theoretic approach for adversarial machine learning*, Data Mining and Knowledge Discovery, 2019