Blue Book: A set of cybersecurity roadmaps and challenges for researchers and policymakers

Markatos, Evangelos; Rannenberg, Kai; Athanasopoulos, Elias; Kompara, Marko; Bountakas, Panagiotis; Kotzanikolaou, Panayiotis; Chaudhary, Sunil; Krenn, Stephan; Daoudagh, Said; Lioy, Antonio

Total number of authors:
26

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

BLUE BOOK

A set of cybersecurity roadmaps and challenges for researchers and policymakers

December 2022

EDITORS

Evangelos Markatos
WP4 Leader
Foundation for Research and Technology - Hellas

Kai Rannenberg
CyberSec4Europe Manager
Goethe-University Frankfurt

AUTHORS

Elias Athanasopoulos
University of Cyprus
Panagiotis Bountakas
University of Piraeus
Sunil Chaudhary
NTNU
Said Daoudagh
CNR
Antreas Dionysiou
University of Cyprus
Christos Douligeris
UPRC
Afonso Ferreira
IRIT
Simone Fischer-Hübner
Karlstads Universitet
Vasileios Gkioulos
NTNU
Leonardo Horn Iwaya
Karlstads Universitet
Meiko Jensen
Karlstads Universitet
Wouter Joosen
KU Leuven
Marko Kompara
University of Maribor
Panayiotis Kotzanikolaou
University of Piraeus
Stephan Krenn
AIT
Antonio Lioy
Politecnico di Torino
Alberto Lluch Lafuente
Technical University of Denmark
Harry Manifavas
FORTH
Eda Marchetti
CNR
Constantinos Patsakis
University of Piraeus
Joao Roberto Peres
KOMP & Faculdade Getulio
Laurens Sion
KU Leuven
Silvia Sisinni
Politecnico di Torino
Christos Xenakis
University of Piraeus

CONTRIBUTORS

Oum-El-Kheir Aktouf, *Grenoble INP, France*
Spiros Antonatos, *Aegis Technologies, Singapore*
Antanas Čenys, *Vilnius Gediminas Technical University*
Nabin Chowdhury, *NTNU*
Christoforos Dadoyan, *Ionian University*
Claudia Diaz, *KU Leuven*
Nikolaj Goranin, *Vilnius Gediminas Technical University*
Dimitris Gritzalis, *Athens University of Economics and Business*
Steven Furnell, *University of Nottingham*
Marit Hansen, *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD SH)*
Marko Hőlbl, *University of Maribor*
Thorsten Holz, *CISPA Helmholtz Center for Information Security*
Jaap-Henk Hoepman, *Radboud University*
Alexandros Kapravelos, *North Carolina State University*
Marieke Huisman, *University of Twente, Netherlands*
Marc Juarez, *University of Edinburgh*
Nicolas Kourtellis, *Telefonica I+D*
Giuseppe Lami, *CNR, Italy*
Alexios Lekidis,*Public Power Corporation*
Nicolas Mayer, *Luxembourg Institute of Science and Technology*
Weizhi Meng, *Technical University of Denmark*
Marino Miculan, *Università di Udine*
Panagiotis Papadopoulos, *iProov Limited*
Jason Polakis, *UIC*
Lorenzo Pupillo, *CEPS and LUISS*
Joao Resende, *NOVA university of Lisbon*
Konrad Rieck, *Technische Universität Braunschweig*
Vittorio Rosato, *ENEA and University Campus Biomedico of Rome*
Antonio F. Skarmeta, *University of Murcia*
Thomas Schaberreiter, *CS-AWARE Corporation*
Stefan Schiffner, *University of Münster*
Roberto Settola, *Università Campus Bio-Medico di Roma*
Maurice H. ter Beek, *CNR, Italy*
Denis Trček, *University of Ljubljana*
Andrea Vandin, *SSSUP, Italy*
Luca Viganò, *Kings College London*
Kim Wuyts, *imec-DistriNet, KU Leuven*
Apostolis Zarras, *University of Piraeus*
Alejandro Cabrera Aldaya, *Network and Information Security (NISEC) Group, Tampere University*
Arttu Paju, *Network and Information Security (NISEC) Group, Tampere University*
Juha Nurmi, *Network and Information Security (NISEC) Group, Tampere University*
Muhammad Owais Javed, *Network and Information Security (NISEC) Group, Tampere University*
Nicola Tuveri, *Network and Information Security (NISEC) Group, Tampere University*
Alberto Carelli, *LINKS Foundation, Italy*
Andrea Vesco, *LINKS Foundation, Italy*

List of Acronyms

Acronym	Explanation
3D	Three Dimensions
2FA	Two-Factor Authentication
AI	Artificial Intelligence
API	Application Programming Interface
APT	Advanced Persistent Threat
ATM	Automated Teller Machine
AV	Anti virus
C2	Command and Control
CCC	Confidential Computing Consortium
CI	Critical Infrastructure
CI/CD	Continuous Integration and Continuous Delivery
CIoT	Consumer Internet of Things
CPU	Central Processing Unit
CVE	Common Vulnerabilities and Exposures
CWE	Common Weakness Enumeration
DDoS	Distributed Denial of Service
DevOps	Software development (Dev) and IT operation (Ops)
DL	Deep Learning
DNS	Domain Name Service
DP	Differential Privacy
ECCC	European Cybersecurity Competence Centre
EDR	Endpoint Detection and Response
ENISA	European Union Agency for Cybersecurity
ESG	Environmental, Social, and Governance
EU	European Union
ETSI	European Telecommunications Standards Institute
FIDO	Fast Identity Online
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
HTTP	Hypertext Transfer Protocol
HW	Hardware
KPI	Key Performance Indicator
ICS	Industrial Control System
IIoT	Industrial Internet of Things
IoEdT	Internet of Education Things
IoET	Internet of Energy Things
IoFT	Internet of Farming Things

IoHT	Internet of Healthcare Things
IoMT	Internet of Medical Things
IoP	Internet of People
IoT	Internet of Things
IoTT	Internet of Transportation Things
IoV	Internet of Vehicles
ISP	Internet Service Provider
IP	Internet Protocol
IT	Information Technology
KLoC	Kilo Lines of Code
LOLBin	Living Of the Land Binary
MFA	Multi-Factor Authentication
MIA	Membership Inference Attack
ML	Machine Learning
MLaaS	Machine Learning as a Service
NCC	National Coordination Centres
NFT	Non-Fungible Token
NIS	Network and Information Systems
NIS2	Network and Information Security Directive 2
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
OES	Operators of Essential Services
OS	Operating System
OTP	One-time Password
OWASP	Open Web Application Security Project
OPCC	Organisational Privacy Culture and Climate
PbD	Privacy by Design
PC	Personal Computer
PETs	Privacy-Enhancing Technologies
RAM	Random Access Memory
ROM	Read-only Memory
SCADA	Supervisory Control and Data Acquisition
SDL	Software Development Lifecycle
SDK	Software Development Kit
SEV	Secure Encrypted Virtualization
SGX	Software Guard Extensions
SIM	Subscriber Identification Module
SME	Small and Medium Enterprise
SMS	Short Message Service

SNP	Secure Nested Paging
SoS	Systems of Systems
SSO	Single Sign On
SW	Software
TC	Trusted Computing
TCB	Trusted Computing Base
TCG	Trusted Computing Group
TDX	Trusted Domain Extension
TEE	Trusted Execution Environment
TLS	Transport Layer Security
TPM	Trusted Platform Module
VM	Virtual Machine
VR	Virtual Reality
XR	Extended Reality

Preface

After the completion of its third year of operation in 2022, the CyberSec4Europe[1] pilot project (https://cybersec4europe.eu/) produced this "Blue Book" (and delivered it as Deliverable D4.7) to serve as a Horizon Research Roadmap in the area of cyber security. To make this book a reality, the project put together a "Task Force" of young and senior researchers in the area of cyber security. The Task Force proposed an initial set of topics and referred back to its constituency, which is composed of top cyber security researchers, asking them what the important research problems should be in relation to these topics. The result of this consultation was a description of each topic that contained the following aspects:

- **What is the topic**? Describe the topic and how it interacts with cyber security.
- **Who is going to be affected** by cyberattacks in this area? ordinary people? organisations? the government? who?
- **What is expected to happen** if we are subjected to such cyberattacks? financial loss? loss of productivity? loss of life? what?
- **What is the worst thing** that can happen if things go really wrong? massive loss of life? a war? financial losses in the range of billions of euros? what?
- **What are the main research gaps?** What do we need to do from a research point of view in order to deal with this problem? What are the important research questions that need to be addressed?
- **Example problems**. Provide specific research problems that can be addressed in a single PhD thesis or in a small number of theses.

[1]CyberSec4Europe is funded by the European Union under the H2020 Programme Grant Agreement No. 830929. This publication reflects only the authors' view. The Commission is not responsible for any use that may be made of the information it contains.

After this consultation, the task force also asked the experts what the **Grand Challenges in cyber security** should be. These whould be topics that would need hundreds of people and several years to solve. However, if solved, they would fundamentally change the problem of cyber security.

We hope that this book will provide useful direction to researchers, will give good advice to policy makers and will prove useful to all who read it.

How to Read this Book

Policy makers may want to focus on Chapter 1 (page 1) which provides a short Executive Summary of the book, and on Chapter 16 (page 111) which describes Grand Challenge Research Problems in the area, which can be solved only with the collaboration of several research organisations and the support of leading funding agencies.

Young researchers who are interested in doing a Ph.D. in systems security should read at least the final section of each chapter, which describes problems that are appropriate to be solved within the context of a Ph.D. thesis.

Experienced researchers may want to read all chapters but especially Chapter 16 (page 111), which describes Grand Challenge Research Problems in the area.

Contents

1 Executive Summary and Main Recommendations

1.1 Research Directions

Over the last year of the CyberSec4Europe project, the beneficiaries of the project, taking into account input from the project's associates and external experts, have formulated a number of research directions that will be important for the future. These directions include:

- Privacy and anonymity
- Emerging technologies: metaverses, IoT, machine learning, etc.
- Novel approaches to authentication: beyond passwords, biometrics, etc.
- Defences "by-design": software development, threat modelling, etc.
- Strong technologies: secure communications, testing, trusted execution, etc.

For each direction a number of research priorities have been defined. Such priorities include:

- Provide strong anonymous communication at large scale
- Build trustworthy metaverses
- Improve password-less authentication methods
- Develop early detection approaches for armoured malware
- Provide privacy in IoT environments
- Realise machine-learning models that remain secure under different adversarial scenarios
- Ensure that critical infrastructures are resilient to cyberattacks
- Support "by-design" testing and certification approaches integrating industrial, social and ethical values, sustainability, and trustworthiness needs

1.2 Grand Challenges

Although short-term projects[1] may have an immediate impact on the market, such impact is usually incremental and may not be long-lasting as it focuses on an immediate problem that may not be so relevant, say, five to ten years down the road. To make fundamental breakthroughs in the area of cyber security, we have proposed several long-term "Grand Challenge" problems. To select a small number of "Grand Challenges", the members of the Task Force, along with the members of the broader constituency proposed several such "Grand Challenges", from which the following were selected:

- Give users assurance about the security of their devices
- If it can be done anonymously in the offline world, it can also be done anonymously online
- Make artificial intelligence safe for people
- Make systems resilient under attack
- Enhance the general public's awareness of cybersecurity

[1] When we say short-term projects we mean projects that last two to three years and have a funding of two to three million euros.

2 Introduction

The penetration of cyberspace into our everyday lives has reached unprecedented levels. Although 30 years ago the Internet was a curiosity mostly used among academics, today more than 92% of the households in the European Union have access to the Internet [2]. The Europeans use the Internet for several aspects of everyday lives: more than 50% use it for social media, around 50% use it for Internet banking, around 66% use it to find information about goods and services, and 55% use it to seek health information [81]. The COVID-19 pandemic just increased the use of the Internet, as even more everyday activities moved online. For example, during the pandemic, schooling, shopping, and socialising could only be done online for extended periods of time. Although the pandemic is a thing of the past, the penetration of some of these Internet activities is here to stay.

Although moving activities online has certain advantages, it may also create threats for people. Indeed, as more and more activities move from the physical world to the digital world, this just increases the **attack surface**. That is, cyberattackers have more opportunities to attack. This is simple to understand: if people do their banking online, thieves will try to steal money online. Similarly, if people do their telephone calls using some online video conferencing system, eavesdroppers will try to listen to these conversations online via a wide variety of options: they may offer such a system for use for free; they may compromise one of those systems; they may bug the software with a virus of their own; they may "purchase" such a bug in order to compromise the system. Here the sky is the limit. The most important point is that people have moved their conversations to online platforms. Once this move has been made, attackers will think of a number of different ways to eavesdrop on these conversations. The same applies to all other activities of our everyday lives: once we move an activity to cyberspace, cyberattackers have a wealth of new opportunities to attack.

Having realised this increasing threat in the area of cybersecurity, the partners of the CyberSec4Europe project put together a list of cybersecurity areas that we should focus on over the next few years. They have explained the security threats in these areas and they have elaborated on what kind of

cybersecurity research needs to be done. The areas they have studied are: anonymity, authentication, critical infrastructures, effective threat modelling, IoT security, machine learning, malware, metaverses, privacy by design, security awareness and training, software life cycle, testing and certification, and trusted execution.

Among the most important research areas we see:

- Provide strong anonymous communication at large scale
- Build trustworthy metaverses
- Improve password-less authentication methods
- Develop early detection approaches for armoured malware
- Provide privacy in IoT environments
- Realise machine-learning models that remain secure under different adversarial scenarios
- Ensure that critical infrastructures are resilient to cyberattacks
- Support “by-design" testing and certification approaches integrating industrial, social and ethical values, sustainability, and trustworthiness needs

To make fundamental breakthroughs in the area of cyber security, we have also proposed several long-term “Grand Challenge problems” including:

- Give users assurance about the security of their devices
- If it can be done anonymously in the offline world, it can also be done anonymously online
- Make artificial intelligence safe for people
- Make systems resilient under attack
- Enhance the general public’s awareness of cybersecurity

3 The Erosion of Anonymity

3.1 Introduction

Over the past few years we have increasingly been using cyberspace for most of our everyday activities: shopping, working, watching movies, listening to music, chatting with friends, entertaining, etc. The recent COVID-19 pandemic intensified this effect and forced us to do most of our activities on line: schooling, shopping for groceries, socialising, keeping in touch, almost everything was done online. In some cases, things became so extreme that doing some of these activities off line was completely illegal. Indeed, during those lock-down periods, face-to-face visits to friends were illegal in some countries and incurred heavy fines. Thus, during such periods the only way to visit friends was through some on-line video conferencing tool.

Although such online activities were convenient (or even absolutely necessary during the pandemic), they usually required strong authentication and identification for all parties involved. For example, online shopping was not possible with anonymous cash, but required the use of debit/credit cards and possibly online bank accounts. Delivery of the purchased products required the disclosure of the delivery address, the presentation of some identifying information, possibly the disclosure of a mobile phone number, etc.[1] The situation was no better for other forms of interaction, such as keeping in touch

[1] Although this information is required for such online transactions independent of the pandemic, before the pandemic people had a choice: They could opt out of such transactions. During the pandemic the choice was not there anymore.

with family. Indeed, as a physical "visit" to family was almost impossible, the only way of interaction was through videoconferencing, which usually implied the installation of some videoconferencing software that needed the user's name, their address, and possibly a credit card for payment purposes. And to make matters worse, this software had the ability to track who is talking to whom, and what they say.

The disclosure of all these personal data is in sharp contrast to the pre-COVID era where people could carry out all these kinds of interactions without the need to disclose any kind of personal information. This disclosure of personal information usually leads to a loss of anonymity: people cannot visit their parents without informing several different companies online. The same loss of anonymity happens in other areas of our lives. For example, in the past people could purchase a can of soda from their minimarket, pay cash, and stay relatively anonymous. Today, in order to purchase a can of soda online they need to disclose their name, their address, their credit card details, while they may ultimately be tracked by dozens of cookies, trackers and advertisers, which use their data for all sorts of marketing purposes.

One might be tempted to say: "It is not necessary to carry out these interactions online: we can always go back to physical interactions." Although it is nice to have such optimistic points of view, we are afraid that soon there may be no "back" to go "back to". Online interactions keep increasing and there is no indication that they are going to significantly decrease: online shopping is on the rise, the use of smartphones continues to increase, and people seem to spend ever more time online. As a result, it seems that online interactions are here to stay and we just need to deal with the tracking and the erosion of anonymity that comes with them.

3.2 Who Is Going to Be Affected?

As it is more difficult to stay anonymous online (compared to the offline world), most law-abiding citizens who use the Internet without any special anonymisation software are potentially going to be affected by this erosion of anonymity. It seems, however, that younger people will be affected the most, as they can be expected to spend a longer percentage of their lives online. In addition, people who have some role that is visible to the public (such as actors, politicians, etc.) will also be disproportionately affected, as their (private) lives will be heavily scrutinised. Unlike theft of physical property, erosion of anonymity is much like data theft: once the data are gone there is usually no way of getting them back. It is not like stolen silverware, which the owner will get back if they catch the thief. Stolen data may be copied and gone forever: there is no "back" to go back to. In addition to people,

their contacts will be affected as well. Exposing the personal information of a single person not only harms the person herself, but may potentially harm anyone who interacts online with her: her friends, relatives, etc.

In addition, people who need anonymity for their physical safety will be severely impacted. For example, people in non-democratic countries may face immediate danger. Even people in democratic countries, such as whistle-blowers and journalists, may be severely impacted if they cannot operate anonymously.

Finally, organisations will also be significantly affected. Indeed, information that used to be confidential within a business (such as number of customers, number of sales, peak times, etc.) could now be found (or at least inferred with high accuracy) by trackers and advertisers that are involved in the interaction. One might think that large organisations would be able to scrutinise their web sites and eradicate any tracking done by third parties. This is probably true. It is not clear, however, whether small companies will have the expertise and/or the capability to do something like that.

3.3 What Is Expected to Happen?

In a world where anonymity is not easy to achieve, people will just not be able to act anonymously. All aspects of their activity will be recorded somewhere online by someone they probably do not know: what time they wake up, what time they go to work, what items they purchase, what books they read, what notes they take, what news they are interested in, where they eat, where they spend the night, who they spend the night with—everything is going to be recorded online.[2] People will have little (if any at all) private life any more. In the absence of a strong legal system that heavily penalises unauthorised access to information, we are afraid that this information may eventually reach the wrong people. Indeed, although initially information may be shared with a trusted entity (such as our ISP or our email provider), information, much like any other digital commodity may eventually be sold, acquired, or even stolen. The worst thing of all is that we do not really know if this will happen, or even if it has already happened.

Some people might say "I have nothing to hide", so they may think that it is reasonable to disclose all of their activities online. However, the main point here is that once information is disclosed online it may eventually find its way to the wrong people or may fall into the wrong hands. If it falls into the wrong hands, information may cause major damage to people. Imagine, for example, organised crime syndicates. They would love to know the whereabouts of

[2] Even the time of the day when I am typing these characters and the time of the day the reader reads this text is maybe being recorded somewhere online.

people: who is alone, who is on vacation, which house is empty, which elderly people bought jewellery, etc. Recent studies suggest that 78% of burglars use social media to find their targets [3]. These burglars use social media to find pictures of homes, or even pictures of house keys [33], to see whether potential targets are on vacation, to find their daily routines, and to see whether they have checked in at a restaurant. All this information can be used in order to find the most promising targets and when is the best time to rob them. One might be tempted to think "Oh! I do not post such information online, thus I am safe." We are afraid that this is far from true. Indeed, several of the apps in our smartphones (and especially those that have access to our GPS coordinates) know where we are. They know if we are on vacation, they know which restaurant we are in, they know when we leave home, they know when we return, etc. The fact that we do not post such information in social media does not mean that this information is not recorded online by several different actors who have access to it. And, as we have said, if some information is collected online, it may later be shared, sold, or even stolen.

It seems that most people are not aware of these dangers. As a result, they do not seek anonymity and they expose themselves to malicious actors out there: burglars, robbers, or even killers! For example, recent research on 350 homicides suggests that before murdering their victims killers stalk their victims in social media [198]. These examples suggest that this lack of anonymous interaction, in which several people engage, may lead to serious damage: theft, loss of property, and even loss of life!

3.4 What Is the Worst That Can Happen?

We are afraid that the impact on society will be much greater than what has been described so far. If anonymity is completely lost, it will be like living in a world where each and every activity of ours is being monitored all the time. This will be like living in a "Big Brother-like" dystopian society, where each and every action will be monitored and recorded. And the worst part of all is that we do not really know who is recording it and who has access to this information. Is it an advertiser who wants to know what colour of shoes we like? Is it a crime gang that would like to know which elderly people recently bought jewellery? Is it the government of a hostile country that would like to know the daily routine of the people in our country and possibly bug them when they visit on vacation? We do not really know.

We are afraid that this complete loss of anonymity will not only transform the lives of individuals, but will transform entire societies. People may become extremely conservative and may become afraid of each and every action they take. In such an environment people may refrain from exercising their rights out of fear that doing so may have consequences; this would severely damage democracy itself. The 1984-like dystopian societies that we managed to avoid will come again to haunt us through our own faults and our own negligence.

3.5 Research Gaps

To address the problem we need a combination of legal and technical activities in this area.

3.5.1 Provide Strong Anonymous Communication at Large Scale

Today there are very few opportunities for anonymous communication. The onion router (Tor) is one of the best-known ones [62]. However, less than 1‰ of Internet users use it. We need to provide easy-to-use systems that give strong protection and can resist powerful adversaries under a variety of threat models.

3.5.2 Provide online the same level of anonymity you expect offline

Today, anonymity has been implemented in only a small portion of online interactions, mostly in anonymous web browsing. Indeed, the Tor network mentioned above comes with a browser that makes installation configuration much easier for users. This anonymity should be extended to all kinds of interactions, including anonymous shopping, anonymous entertainment, etc. The rule of thumb here should be: if it can be done anonymously offline, we should try to do it anonymously online as well.

3.5.3 Measure/Monitor the scale of the problem - Achieve Transparency

It is not clear to most people what the scale of this problem is: what is the amount of personal information that is being shared. The web trackers keep inventing new ways to track users online and to deprive them of the ability to operate anonymously [181]. It is basically a “game of cat-and-mouse”, where trackers invent new ways of tracking and researchers try to detect these ways of tracking, possibly via reverse engineering. We need to better understand the scale and mechanisms of tracking and loss of anonymity. We need to develop mechanisms that continuously monitor this erosion of anonymity at all different levels in all possible different ways. These mechanisms should be able to operate frequently without the cooperation of web content providers..

3.5.4 Novel Data Anonymization and De-Anonymization approaches

We need to develop novel data anonymization mechanisms that will allow sharing of data at a larger scale. Although some data anonymization approaches already exist (see [226], and [68]), there is still a long way to go before anonymous data can be shared on a large scale. We need to study attacks to existing data anonymization approaches that aim to de-anonymize the data, and develop defences that will result in better anonymization approaches.

3.5.5 Resist Censorship

Several countries all over the globe censor communications on the Internet. In such settings users have limited access to the Internet or, in some cases, no access at all. We need to develop robust and practical systems that bypass censorship and enable people to safely (and anonymously) publish and access information.

3.5.6 Develop robust anti-fingerprinting methods

To break the anonymity of their users, several web sites use fingerprinting methods. Such methods try to identify various aspects of the user's browser (e.g. browser type, fonts supported), or the user's computer (such as local language, operating system version, screen size, etc.) in order to uniquely identify users as they browse the web. Although each of these features alone (such as screen size) is not enough to uniquely identify a user, the combination of all of them is usually sufficient. We need to develop strong anti-fingerprinting approaches that allow little (or no) information to be collected about the users as they roam around the Internet.

3.6 Example problems

Tangible example problems might include:

Identity leaks. Monitor how web sites use all kinds of mechanisms (such as cookies, URL arguments, URL header fields, etc.) to transfer personal data from one web site to another. Develop defences against such mechanisms.

Make Anonymizing Networks more resistant to attacks. Study possible attacks that may compromise anonymity in anonymizing networks. Explore the magnitude of these attacks and propose possible solutions. Initially focus on website fingerprinting attacks on Tor.

Operate with anonymous personas. Develop fake personas that allow users to use the web without revealing their true identity. Develop a system that will clearly evaluate the trade-off between usability and privacy in providing fake information in different settings. Explore the situations where personas provide added utility.

Understanding of Privacy. Improve users' understanding of their privacy-related decisions, such as the cookie consent forms that they agree to. Develop (semi-)automated tools that improve this understanding and quantify the choices made by the users.

Data Provenance. Develop systems that enable users to detect the provenance of data and thus discover stolen/leaked data. Address the problem for different kinds of data including time series, images, videos, multi-dimensional signals, etc.

4 Machine Learning

4.1 Introduction

Machine Learning (ML) has become the technology powering a wide-range of applications and services. The performance and the generalisability of ML models made them a good candidate for tackling a series of real-life problems that exhibit high complexity. Take for example the recent advances of Generative Adversarial Networks (GANs) that manage to synthesise highly realistic human faces with a small number of real-world samples [127]. Generally speaking, ML-based systems managed to achieve high success rates on problems where the classic rule-based approaches did not perform well.

Nowadays, ML has been deployed in many sectors of our everyday lives. For example, during our online shopping on Ebay or Amazon, an ML-based personalised recommender system, running in the background, proposes products according to different parameters related to the user, e.g. the history of previous purchases and the time spent looking at a specific product. In addition, the automotive industry has incorporated ML technologies into their cars to make them drive themselves without any human supervision whatsoever. Furthermore, ML-based Natural Language Processing (NLP) techniques have been developed for improving the safety of online discussion environments, e.g. to detect toxic, sarcastic, harassing and abusive content [169]. In general, ML technologies have benefited various sectors, some of them being the following: medical diagnosis [131], detection of credit card fraud [146], stock market analysis [41], bioinformatics [63], speech recognition [99], object detection [40], and robot locomotion [129].

To grasp the potential of ML algorithms, it is enough to say that many tech giants, such as Google and Amazon, offer Machine Learning as a Service (MLaaS) platforms, where the users can upload their own data to train their own ML models and solve a specific classification/prediction task. Thus, the

users' data –which in many cases contain sensitive information, such as medical records, photos and other personal descriptors– is used as the training data by the MLaaS platforms. Additionally, some MLaaS operators may give data owners the option to sell access to their trained ML models to the general public.

Despite the massive success of ML in tackling numerous difficult problems, several security and privacy vulnerabilities have been shown to coexist with these models [142, 182]. For example, think of the case where an NLP model misclassifies a movie's review as "excellent" instead of "bad". This (misclassification) error results in a higher score for that particular movie. Thus, users that consult a specific site for movie ratings will be lured to watch that movie because of its high rating. After watching that movie users will realise that it was not as good as the rating site suggested and, as a consequence, avoid using the same site again. On a more serious note, think of the case where an image recognition model is deployed on an autonomous driving vehicle for identifying road signs. If an attacker deliberately perturbs the input (video) to the image recognition model, then the model might wrongly recognise a "stop" sign as a "minimum speed limit" sign and accelerate instead of stopping the car. As you can easily imagine, such attacks can have serious consequences, even causing fatalities. In conclusion, since ML has dominated across many sectors, we need to come up with solutions for ensuring its secure operation.

4.2 Who Is Going to Be Affected?

Since the widespread adoption of ML models into a variety of services and applications, anyone who has access to a modern device (e.g. a smartphone, a personal computer, a vehicle, or even a home appliance) can be affected. In general, any individual who possesses an electronic device can be affected. Nonetheless, youngsters are expected to be affected to a larger degree compared to older individuals, since they often utilise newer technologies and applications that are often powered by ML [124].

A large portion of ML-based applications are often trained on personal (sensitive) data. Leaks of such data may lead to serious consequences for the affected individuals. Think of the case where an ML model is trained to associate a patient's information with a specific disease class. If an adversary knows that a patient's data was included in the model's training dataset, they can draw conclusions about the victim's health status (known as membership inference attacks [211]). In a similar fashion, if an adversary manages to successfully generate inputs resembling the original ones used for training the target model, then this might enable the de-anonymisation of users

and expose personal or sensitive information (known as model inversion attacks [85]). Finally, adversarial image generation attacks, where an adversary introduces slight modifications to an existing image in order to confuse or deceive an image recognition ML model into performing a misclassification, have been proposed in the literature [96].

Finally, companies that provide ML-based solutions may also be impacted, in addition to individuals. In particular, disclosing that the ML services offered by a company are vulnerable to the aforementioned attacks can seriously harm that company's finances and reputation.

4.3 What Is Expected to Happen?

The generalisation ability (performance) of ML-based applications heavily depends on the quantity of available training data. But, as the training data volume grows, so does the chance that sensitive data will be present. Thus, it is realistic to assume that the attention of potential adversaries and malicious groups is going to be focused on attacking systems that utilise ML components.

The number of data breach incidents that exploit ML components will increase in the near future. This is because ML models, running in the background and collecting sensitive personal data, will be deployed within more and more applications and services. Thus, potential adversaries will have access to a wider range of exploitable targets.

ML models can be deployed in sectors where wrong decision making implies serious consequences (e.g. in healthcare). Thus, legislations/regulations will be drawn up in order to explicitly state the liable entities in case something goes wrong or not as expected. In addition, the security and privacy standards that must be met by deployed ML models will be released. These standards will ensure that deployed ML models are robust against specific (known) threats. Finally, guidelines for best practices will be formed in order to help non ML-expert developers who wish to incorporate ML technologies in their applications.

4.4 What Is the Worst That Can Happen?

As already mentioned, it is expected that the number of data breach incidents caused by the exploitation of ML components will increase. In order to prevent potential exploits that could have serious repercussions, relevant authorities, such as the European Union (EU), should keep taking bold steps (e.g. see Pupillo et al. [144] and ENISA press releases on AI/ML-security [73, 74, 76]) to strengthen the security and privacy of ML-based sys-

tems and services. Only with such concrete regulations/policies in place will the community experience the full potential of ML technologies.

Moreover, for companies that offer ML-based solutions, potential attacks on their systems may imply millions of dollars in financial damage and loss of reputation. In addition, attacks such as those described in Sec. 4.2 might make a large portion of ML-based systems unusable.

Last, but not least, the degree to which people trust ML-based systems will be greatly decreased if appropriate security and privacy measures are not considered. This is important, because the traction (usage) of such systems will be decreased as well. People will be hesitant to provide their valuable data for training ML models. Thus, advances in ML, and artificial intelligence (AI) in general, will decline significantly. In fact, people will become so suspicious of technology that they will be hesitant to use it. Much like the 5G case, we may even see uprisings and protest movements against ML technologies.

4.5 Research Gaps

In order to improve the secure operation of ML-based systems several actions can be taken.

4.5.1 Exploring the security and privacy robustness of state-of-the-art ML models under different adversarial scenarios

So far, the scientific community has identified a number of security/privacy vulnerabilities that coexist with state-of-the-art ML models. Nonetheless, exposing those models in different adversarial scenarios might reveal additional vulnerabilities from which they may suffer. Discovering those weaknesses will significantly aid the community in developing generally applicable defences or designing improved architectures in terms of providing specific security/privacy guarantees. In that sense, ML auditing frameworks can be developed that will be solely responsible for evaluating the robustness of ML models against specific security/privacy threats.

4.5.2 Designing architectures and training algorithms for increasing ML models' generalisation and robustness against security/privacy attacks

A number of the attributes of ML models might be related to their vulnerability to specific security/privacy attacks. For example, membership inference attacks (MIAs) have been shown to be more effective on overfitted models (i.e. models that demonstrate low generalisation) rather than well-generalised ones [211]. In addition, the architecture of the model itself has been shown to play an important role in its vulnerability against MIAs. In that sense, researchers have demonstrated that a naive Bayes model is much more resilient to MIAs compared to a decision tree and, therefore, may be the preferred model type for a particular ML service [234]. Thus, ML model architectures that offer increased robustness against security and privacy attacks should be developed.

In a similar fashion, the training of ML models should be optimised towards offering increased security and privacy guarantees. For example, Differential Privacy (DP) [68] offers probabilistic guarantees about the privacy of individual records in a database. DP retains the global statistical distribution of a dataset, and its contribution to an ML model's weights, while at the same time reducing the influence of each training instance. Similarly to k-anonymity [139] and diversification [9], DP can be used to mitigate the risk against various privacy attacks, such as membership inference and re-identification. The application of DP, however, imposes a trade-off between security and utility (usefulness). In other words, the stronger the security guarantees that DP offers, the larger the negative impact on the model's performance. Thus, novel training algorithms and techniques that maximise the security/privacy guarantees, while also sacrificing as little performance as possible, should be developed.

4.5.3 On the transparency and interpretability of deep ML models

ML models are often viewed as black boxes that can make a decision based on any possible input variant. The complex nature of ML models makes their inner workings difficult to comprehend. However, what is difficult to understand is also difficult to audit. And what is difficult to audit is also difficult to trust. Generally speaking, the level of model transparency depends on the knowledge required to understand the internal mechanics of the ML algorithm.

There are quite a few ML algorithms that directly or indirectly produce human comprehensible output, such as a linear model or a decision tree. Suppose that we can trace the chain of reasoning of each decision that an algorithm makes. Can we claim the algorithm is transparent? The answer

is unfortunately no. The chain of reasoning only tells us "how" a decision was made for a given input but not "why". For example, knowing "how" is not sufficient to justify that the decision is made consistently, accurately, reliably, and validly. Thus, for a learning model to be truly transparent we need to know both "how" and "why". Due to the high complexity of deep ML models, which often incorporate hundreds of fully (or partially) interconnected layers, a promising approach for increasing their transparency and interpretability is to provide justifications and insights for the decisions that can be gauged externally.

4.6 Example problems

Tangible example problems might include:

Exploring security and privacy attacks on ML models. An important direction for enhancing the security and privacy of ML algorithms is to reveal additional hidden vulnerabilities. Apart from the already established security/privacy attacks, such as model inversion, membership inference, model extraction and adversarial sample generation, additional effort is required to determine other possible threats. In addition, we need concepts and techniques to measure the vulnerability/robustness of ML.

Proposing generally applicable defence strategies. Another interesting direction is the development of generally applicable defences, more specifically, defences that can be applied to existing trained ML models without the need for retraining, which is a time-consuming process and would require large computational resources, or any modifications to their architecture/training algorithm, which would require significant manual intervention from experts in the field of AI and ML.

Applying vulnerable learning models in a secure way. One might say that perfectly secure ML is probably an illusion. Thus, instead of focusing on increasing their robustness, an alternative direction is focusing on how to apply them in such a secure way that exploiting those ML models becomes significantly harder.

Developing human-friendly interpretability techniques. This angle involves the development of systems and services that are able to provide human-friendly explanations for the decisions of current state-of-the-art ML models. When we refer to "human-friendly explanations", we mean justifications that are preferably simple enough for people who are not experts in the fields of AI and ML to understand.

5 Authentication - Beyond Passwords

5.1 Introduction

Nowadays, the increasing use of cyberspace requires each person to have several accounts in order to access the systems and web applications necessary for everyday activities. One of the oldest protection mechanisms of systems and web applications is the authentication method, where the user is asked to prove his/her identity to gain access. The most common method of authentication by a system or an application is via the so-called username-password method. In this method, the user has to provide the username and the password that were chosen during the account creation process (registration). Despite the fact that username-password is one of the oldest authentication methods, it is still used by almost every system and application (both online and offline). For instance, a doctor in a hospital deploys the username-password method to access her account in both the hospital and an online shop. During the past few years, the number of accounts each user maintains has greatly increased; consequently, users find it difficult to memorize and manage all these passwords. A recent study by NordPass showed that an average person has 100 different passwords to remember, leading to a problem called password overload [199]. Moreover, the username-password paradigm is subject to various cyber-attacks, such as recovering a password from its leaked hash through brute force (password cracking), recovering a password when transmitted through an untrusted channel (eavesdropping), tricking a user into entering his/her password on an untrusted or compromised endpoint (phishing websites, ATM skimmers), or allowing the use of default passwords that can be used by adversaries [28] [174] [8] [83].

Hence in order for a password to be considered strong, as suggested by Microsoft, it should contain at least 12 characters, be complex (i.e. contain alphanumeric characters, numbers, symbols, and non-dictionary words), be different from other passwords the user used in the past, and be difficult for others to guess [152]. All these conditions along with the high number of different accounts have affected users who find it difficult to memorise (Strength of Memorized Secrets [173]) and manage all these passwords. To solve this password overload problem, users have come up with solutions that directly affect the security of their accounts and the privacy of their data; they either simplify their passwords to be easy to remember, reuse the same password on different services, or store their passwords in a "secure" place, for example on paper or using a password manager. But even if the password is strong and the user handles it appropriately, the service providers also have to keep their end of the deal and store their users' passwords securely. NIST provides suggestions on how to properly store passwords on databases (Memorized Secret Verifiers [173]), though many popular open source server software do not offer adequate security by default [170] and a number of data breaches exposed improperly stored passwords [113].

Several methods have been introduced to enhance the robustness of the authentication process, especially on critical systems and applications; with the best known being two-factor authentication (2FA), also recommended by ENISA to improve password security [77] [5]. During a 2FA method, the user has to prove his/her identity based on two factors rather than one. For instance, to access a web banking account, apart from providing the username and password, the user is also asked to provide a one-time password (OTP) that is received via a Short Message Service (SMS) in order to be authenticated. Although this method improves the security of the authentication process, it lacks user-friendliness [148], which is an important factor in the authentication procedure, and can also be exploited through SIM swap attacks (where the adversary manages to clone the SIM card of the victim, allowing him to steal the SMS) or by tricking the user into revealing the OTP code through a fake call, website or email (phishing).

5.2 Who Is Going to Be Affected?

Anyone who uses a computer or smartphone is going to be affected by the weak security of password-based authenticated methods. However, people with more accounts are more likely to be affected, since the attack surface is wider in other words, attackers have greater chances to compromise an account. For example, if Bob has one account (e.g. an email account) and Alice has three accounts (e.g. email, online shop and streaming accounts), then

an attacker can target Bob on one application, while Alice can be targeted on three different applications. Apart from individuals, companies/organisations might also be affected, since if an employee's password-only protected account is compromised, corporate data could be stolen or malicious software may be planted, resulting in jeopardising the reputation of the company/organisation, which will lead to money loss. Last but not least, governments and critical infrastructures will be affected the most, because if an attacker were to gain unauthorised access, their malicious actions might also have a serious effect on European citizens. For example, the compromise of a power grid will significantly affect the public.

5.3 What Is Expected to Happen?

In a case where a system's authentication is compromised, not only will the user's data be at risk, but also the attacker will have access to the system to perform various malicious actions, such as stealing personal information or documents, installing some type of malware, or performing an Advanced Persistent Threat (APT) attack. Thus, the consequences will vary depending on the criticality of the system and the attacker's actions. In most cases, compromise of the authentication process leads to a data breach and money loss. However, when the authentication process of a critical infrastructure is at stake, the consequences might be much more severe than the loss of money. The worst thing is that we cannot know beforehand the malicious actions that an attacker will perform.

5.4 What Is the Worst That Can Happen?

Most of our digital services rely upon secure authentication of the users, and thus we have to make sure that we use adequately secure authentication methods. Assuming that we will continue to base all of our authentication methods on passwords, eventually every system will be compromised at least once. Every company will be affected by incidents and data breaches, resulting in millions of euros lost. Massive amounts of people's leaked data (e.g. email, photos, residence, social number, telephone, credit card numbers, finance status, medical records, etc.) will be available online to the highest bidder, thus affecting privacy significantly.

Critical infrastructures will also be affected deeply. Such infrastructures (like the power grid, water systems, hospitals, telephone communications) still connected to the internet will possibly pose a threat due to the high risk of being compromised. Cyber-attacks will target such systems, creating a high risk of espionage, cyber terrorism or even loss of lives.

5.5 Research Gaps

To ensure we keep authentication modules adequately secure, actions need to be taken in this area.

5.5.1 Improve passwordless authentication methods

Although passwordless authentication methods — such as the fingerprint unlock of our phones — are already available (e.g. FIDO [6] and WebAuthn [110]), there are a number of points that can be improved. To start with, we should make sure passwordless authentication is accessible by everyone (though our personal devices such as our smartphones or personal computers) in a user-friendly yet secure way, by increasing the adoption of passwordless authentication methods (e.g. increase the number of application and website that support passwordless login) and improving the interoperability between authenticator devices and services requiring authentication (e.g. use the fingerprint sensor on your smartphone to authenticate on your laptop). Since several popular passwordless authentication methods rely on biometrics (e.g. iris scan, fingerprint scan, face scan), looking into ensuring the security and trustworthiness of biometric authentication methods (e.g. by reducing the false positives where an unauthorised entity may be falsely be identified as an authorised one) while also respecting user's privacy (e.g. securely storing biometrics' related data locally only for use to authenticate the user) is of high importance, while also looking into how they can be used along with fuzzy cryptography (where biometric data can be used as an input to cryptographic functions). Furthermore, the usage of passwordless authentication in advance authentication scenarios (e.g. multiparty passwordless authentication, where the authentication/authorisation is performed by more than one entities) should be investigated in order to meet specialised needs that existing methods do not cover (e.g. allowing the authorisation of a transaction or the signing of a document by 2 or more people). We should also look into novel passwordless authentication approaches for both online (e.g. logging into an online website) and offline usage (e.g. logging into your laptop). Last but not least, there needs to be investigation into improving the authentication of users by leveraging existing technologies (e.g. Single Sign On) and new digital identity schemes (e.g. Self-sovereign identity, Decentralized Identifiers [219], Verifiable Credentials [220]) in combination with passwordless authentication (when needed), as well as secure recovery or fallback mechanics for use when the main authentication mechanic is not available (e.g. in case you lose your smartphone or your USB security key).

5.5.2 Measure/monitor the use of insecure authentication methods

It is of high importance to monitor the security state of authentication methods in Europe, by measuring both the adoption of passwordless authentication and the use of insecure password-based authentication methods. With better insight into the problem, measures could be taken to reduce the security risk. For instance, we can introduce new regulations or improve existing ones, targeting critical systems affected by the problem, as well as set minimum security requirements (e.g. appropriate certification, security assessments and auditing) to ensure an appropriate level of security to protect European citizens and our society as a whole. Research could also focus on the economic side of the issue and investigate whether better and newer authentication mechanics are affordable by all kinds of organisations or whether such technological solutions do not fit the bill.

5.5.3 Understanding user's psychology related to authentication

An important research gap is related to the human psychology and authentication. Further research into the user's psychology during authentication should provide more information related to deception attacks (social engineering-related attacks) as well as providing valuable information about the user's perception regarding the usability of an authentication mechanism.

5.5.4 Enhancing biometric authentication methods using AI methods

Biometric authentication methods, such as fingerprint, face, and voice recognition, are heavily utilised in smart phones to login users without passwords. Yet those mechanisms come with their own limitations. To name a few, dirty hands will affect fingerprint recognition, weather conditions face recognition, and loud environments voice recognition. Thus, further research is required to alleviate those restrictions. One approach could be to employ AI methods, such as machine learning (ML) and deep learning (DL), in the authentication process to make the best of the incoming data in cases where the conditions are not optimal.

5.5.5 Continuous authentication

The user's unique characteristics can be deployed for authentication without needing his/her interaction. For instance, in the case of a mobile phone, each person holds his/her phone differently, types differently, swipes from different angles, etc. Utilising all this data regarding each person's behaviour and leveraging AI can result in continuous authentication without the use of passwords or biometrics. Research on this topic should focus on increasing the accuracy of the behavioural authentication mechanisms, at the same time

reducing the false positives and false negatives, while also looking into how to preserve user's privacy and user's control.

5.5.6 Training people in authentication related topics

There are several 2FA methods that can be used today in combination with passwords to provide adequate security to systems, but most users opt not to use them. The research community will have to look into the reasons why many users do not enable passwordless or multi-factor authentication (MFA) and develop efficient user training to tackle the issue. Furthermore, although various recommendations on how to handle passwords exist, both users and software engineers still fail to follow them resulting in handling them insecurely (e.g. users continue to share passwords, engineers continue to store passwords insecurely). Conducting related training (or increasing their efficiency) will ensure that everyone has access to and knows how to use correctly and easily strong authentication mechanics, minimising the risk of their accounts being compromised.

5.6 Example problems

Tangible example problems might include:

User friendliness. Research should be conducted on how the user friendliness of passwordless authentication methods could be improved. Apart from making the methods easier to use for the general public, new easy to use methodologies to transfer or backup credentials used by authenticator devices should be tested.

Transition from password-based to passwordless. In many cases the transition to newer passwordless authentication methods is not trivial as many systems do not support them out of the box. Furthermore, users not familiar with passwordless technology may face difficulties in preparing their environment to use the new authentication methods. Further research in the topic may look into how to introduce passwordless authentication in a user-friendly way and as a security layer wrapping legacy system.

Resistance to attacks. To secure our future we should also look into how attacks can be mitigated and how measures could be integrated into our passwordless authentication methods. In many cases such problems may arise as a result of insecure configuration or faulty implementation, while in other cases they are among the disadvantages of the selected method (e.g. some passwordless authentication methods are not phishing-resistant).

Weak authentication on IoT devices. The introduction of IoT devices to our lives and their interconnection and exposure to the internet created a new attack surface for attackers, namely attackers apart from targeting user authentication, attackers can now target the device authentication process. Novel passwordless authentication methods should be introduced for such small smart devices (e.g. remotely accessible IP cameras) that usually feature limited resources.

6 Security Awareness and Training

6.1 Introduction

Organisational cybersecurity is widely acknowledged to rely on three pillars: namely, technologies, processes and people. Additionally, transforming raw data into eligible information, and information into actionable intelligence, is an increasingly significant component of maintaining situational awareness of cybersecurity.

People are often perceived as the weakest link in the cybersecurity chain [32] [164]. Though this negative characterisation of human nature is debatable [123], it is undeniable that the human is a major contributing factor to the majority of cybersecurity breaches [128]. Cybercriminals frequently employ techniques, such as social engineering, that exploit innate human weaknesses to carry out attacks and to improve their chances of success.

Cybersecurity competence development focuses on enabling people to establish technical and operational barriers to cybersecurity threats, and to conduct themselves appropriately, through the vigilant processing of actionable intelligence. It is an iterative process of continuous and incremental improvement [249] targeted toward transforming the human factor from a potential attack vector to a multiplier of organisational preparedness to protect against, detect, respond to and recover from cyber-attacks. Cybersecurity competence development is based on a continuum that expands formal education through added value activities, such as i) hands-on experience, ii) awareness programmes and iii) training programmes, with each of these multipliers serving particular functions in maintaining organisational cyber hygiene.

Leveraging human factors in cybersecurity goes beyond traditional training and awareness methods. It calls for modern approaches that draw on understandings human behaviour and implementing tools that provide targeted cyber training and awareness. Hands-on experience (also known as learning by doing) is an extremely effective approach to teaching and learning cybersecurity [213]. It engages the learners and improves knowledge comprehension and retention, as well as the possibility of translating acquired knowledge into action [90]. Many successful strategies are used for this purpose, including exercising cyber-attacks detection and defence skills in a cyber range envi-

ronment [42], participating in cybersecurity competitions [162], participating in flagship cybersecurity exercises [55], and learning through gameplay (e.g., serious games) [207]. However, integrating cybersecurity awareness and training only reduces, not eliminates, the possibility of human neglect and errors, implying that smart technical interventions to check and regulate employees' mistakes remain vital for an organisation's overall cybersecurity posture [143].

6.2 Who Is Going to Be Affected?

As mentioned earlier, cybersecurity is widely acknowledged to rely on three pillars: namely, technologies, processes and people. Humans can be negligent are prone to errors, and can represent, either intentionally or unintentionally, a weak link [164]. Therefore, technologies and processes aim to reduce the overall burden or responsibility by automating and demarcating procedures, as we see through the ongoing digital transformation [161]. However, it is people who develop, operationalise and maintain technologies and processes. Thus, while technologies and processes constitute essential tools for cybersecurity hardening, the human factor plays the most critical role in ensuring cyber hygiene. Regardless of how many expensive and sophisticated technological security solutions have been deployed, they cannot be considered secure as long as human factors do not work and behave in a secure manner. Moreover, technological security solutions require human input for proper and effective functioning: for example, firewalls must be activated, software must be updated, and security warnings must be acknowledged and acted upon.

Lack of emphasis on security awareness and training has personal, organisational, and even national ramifications, while improved vigilance, or lack thereof, permeates and spills over between the personal and professional spheres. We see the rippling effects of low awareness and knowledge across nearly all cybersecurity topics and sectors [179], from privacy implications to critical infrastructure security [43]. Human behaviour, more often than not, is the soft underbelly of security designs and architectures, presenting to potential attackers a path of least resistance, if not a clear entry point, with a limited technical threshold. Therefore, the challenge is not to determine who will be affected by limited cybersecurity awareness and training, but to identify who may not.

It must also be noted that the overall impact of digital transformation highly depends on the acceptance of the newly developed digital technologies, referring to both those that are developed with a cybersecurity focus and those that are not. Cybersecurity awareness and training can facilitate stakeholder acceptance and adoption of innovative digital technologies, as it

enhances understanding of the related cybersecurity risks and develops active barriers against them.

6.3 What Is Expected to Happen?

Automated and autonomous systems have been developed across several sectors including cybersecurity [161] to assist humans or even to remove them from the loop. Nevertheless, this process is still in its infancy, and even in developing those systems, people are the principal contributors [221]. Additionally, cybersecurity best practices have been developed across all the phases of secure systems engineering, from planning all the way to disposal. However, these processes, whether referring to systems, policies or processes, often include inputs that are biased by qualitative expert knowledge [101], or require compromises to meet requirements and constraints. Limited cybersecurity awareness and training represent the root causes of vulnerabilities introduced within the deployed systems, technologies, processes and policies. This occurs across all the stages of their lifecycle, arising from several factors, such as design flaws, integration mistakes, or operational negligence. The exact impact and consequences can only be estimated on a case-by-case basis. However, it is critical to acknowledge that promoting targeted cybersecurity awareness and training in an iterative process of continuous development is essential for ensuring cyber hygiene, preparedness and resilience.

6.4 What Is the Worst That Can Happen?

The extent of the potential impact and consequences due to limited cybersecurity awareness and training can only be estimated on a case-by-case basis. The major consideration in that respect has to do with the fact that a lack of relevant competences has a knock-on effect on the cybersecurity and resilience of technologies, systems, processes and policies. Therefore, although the impact and consequences depend on the specifics of an incident (e.g. sector, scope, objectives, attacker capabilities), limitations in cybersecurity competences play a critical role in the probability of an incident occurring, and will have an impact on the effectiveness of the response and recovery actions taken. Thus, it is natural to consider cybersecurity awareness and training as a positive or negative multiplier across the overall cyber hygiene.

The benefits of cybersecurity awareness and training extend beyond the detection and mitigation of cybersecurity issues [151]. To begin, with skilled employees who are familiar with cybersecurity principles and understand their role in keeping the business secure, downtime of critical business systems due to security breaches or incidents could be avoided. This will save organisations from the costly and time-consuming process of repairing and reinstating normal business operations. Next, employees who are familiar

with compliance regulations and have a clear understanding of how to handle sensitive data and information can help to minimise regulatory compliance infractions and their negative reputational and financial impact on businesses. Finally, organisations that implement proactive cybersecurity measures and have demonstrated cyber resilience boost customer confidence.

Let us look at the impact of data breaches in organisations to have a better idea of the issues that could arise as a result of a lack of cybersecurity awareness and training. We considered data breaches as the example simply because 82% of data breaches involved a human factor [240]. These breaches occurred because people fell victim to social attacks, and either deliberately (misuse) or inadvertently (errors) acted or failed to act when necessary. More importantly, they could have been avoided to a greater extent if the people involved were properly aware and trained in relation to their security operations and responsibilities. Now let us assess what could potentially happen if there is a data breach in some organisation. When it comes to hospital data, a breach could jeopardise and harm the patient's health and safety, i.e. endanger human life. In the case of financial service data, its breach could result in a huge financial loss. And in the case of government data, a breach could compromise national security. Last but not the least, irrespective of the organisation type, a data breach would cause a loss of customers' and partners' trust, diminished market reputation, loss in business, and penalties levied, which might lead to bankruptcy.

6.5 Research Gaps

6.5.1 Cybersecurity awareness and training needs across levels and fields of study

The ongoing digitalisation of products, services, supply and value chains highlights the need for the increased technical literacy of digital natives. Therefore, in addition to dedicated study programmes for the development of dedicated professional competences (e.g. computer science, network engineering), relevant modules are integrated across most study programmes, and levels and fields of study [246]. However, topics related to cybersecurity are scarcely introduced outside programmes that are particularly targeted towards developing cybersecurity professionals.

Accordingly, it is essential to identify the cybersecurity skills and competences that are needed, as well as suitable delivery methods, starting from primary education all the way to higher education and specialised fields of study. This requires examining the universal cybersecurity-related components that are targeted at enhancing the cybersecurity awareness of the broader public,

as well as specialised topics that are specific to distinct occupations. Furthermore, it requires assessing delivery mechanisms that are adjusted and optimised with respect to the attributes of the relevant target groups.

6.5.2 Cybersecurity awareness and training needs multidisciplinary approach investigations

It was and is appropriate at this time to ask, "Why are cybersecurity awareness and training failing to yield the expected outcomes?" [22] The question has been the subject of numerous investigations, but no clear answer has been found yet. This may be a result of the narrow or limited perspective from which we view the issue.

Cybersecurity awareness and training mostly revolve around comprehending and transforming human thought and behaviour, which are undoubtedly complex topics. Therefore, as long as cybersecurity researchers and professionals attempt to specify and control human thinking and behaviour through a small set of drivers, which most psychologists and social scientists would consider misleading, the likelihood of successful cybersecurity awareness and training will probably remain low [22]. This also implies that addressing the issue would require a more comprehensive and holistic approach that utilises knowledge and expertise from multiple disciplines, including engineering, pedagogy, behavioural economics, marketing, and social, cognitive and organisational psychology, among others.

6.5.3 Computer-based cybersecurity awareness and training need the implementation of AI and ML algorithms for their automation purposes

There are hardly any disciplines that are not utilising the capabilities of artificial intelligence (AI) and ML, or at least attempting to do so. Cybersecurity awareness and training cannot be an exception. In fact, there are numerous ways that AI and ML could be useful to raise the standard and impact of cybersecurity awareness and training [207].

By utilising AI and ML algorithms, many activities of cybersecurity awareness and training could be automated. Automation would help to achieve on-demand cybersecurity awareness and training. Additionally, they could facilitate the design and delivery of a more customised, personalised and optimised awareness and training experience to the audience. For example, AI and ML-assisted computer-based tests could be developed and used to identify vulnerable groups. Furthermore, based on the test results, and once more with the application of AI and ML algorithms, more customised, personalised and optimised awareness and training resources could be prepared for the audience.

6.6 Example Problems

Tangible example problems might include:

IoT cybersecurity awareness and training modules The use of Internet of Things (IoT) technology is expanding daily in all spheres of business and society, from consumer-focused goods and services to industrial IoT. This has also introduced unprecedented safety, security and privacy risks [23]. The majority of IoT security deployments take place at the business unit level, where IT does participate, albeit insufficiently. This also implies that a number of key stakeholders in IoT security are unfamiliar with the IT security side of things. Further exacerbating the situation, IoT-related risks are often not well articulated, resulting in low awareness among users and employees. Thus, IoT security cannot be robust if the people involved do not have a good understanding, and this requires them to have the relevant awareness and training [134].

Awareness of adversarial AI attacks Contrary to the use of AI/ML methods to strengthen cybersecurity, threat actors are leveraging AI/ML methods for malicious purposes, for example, to increase the number of attack surfaces and bolster their attacking capabilities [154].

Adversarial AI methods are used to craft misleading data or behaviours with the intention of manipulating and disrupting critical AI systems. There is growing evidence that adversarial AI methods have been implemented in real-world attacks. In spite of this, the effort to defend AI systems from adversarial AI attacks is generally an afterthought. It is unfortunate that many companies still remain unaware of adversarial AI attacks and the failure of AI systems the attacks can cause. Therefore, it is urgent to raise companies' awareness of adversarial AI attacks and motivate them to be alert and prepared to defend their AI systems, especially those used in crucial sectors, against the attacks.

Cybersecurity awareness and training modules for mobile users The mobile phone has gained widespread acceptance as a commonplace tool for accessing the Internet and doing sensitive jobs. These could be the causes of the daily rise in cyberattacks and crimes aimed at mobile phone users [200] [31]. However, suitable cybersecurity awareness and training for mobile phone users are still rare. There is a common assumption that mobile phone use is similar to using a desktop or laptop, which is only partially correct. Indeed they share a commonality as computing devices; however, at the same time they also have many differences. For example, mobile phones possess a higher risk for theft or loss, authentication used to lock a mobile phone is often weak as a result of

the high frequency of logins to mobile phones, and the smaller screen size of mobile phones often makes it difficult to notice security warnings. Additionally, mobile phone users are far more diverse than those of laptops or desktops. People of various backgrounds, from urban to rural, educated to uneducated, white-collar to blue-collar, and so on, use mobile phones. There have not been many investigations into why and how these diverse individuals use a mobile phone, and what their expectations from cybersecurity awareness and training might be.

Cybersecurity awareness and training evaluation focusing on behavioural change. Evaluations of cybersecurity awareness and training are frequently restricted to gauging security knowledge and self-reported attitude shifts. Indeed, improvement in knowledge and attitude is important, but the evaluation should actually measure the change in cybersecurity behaviour; after all, behaviour change is what the awareness and training programmes are ultimately aiming to achieve [39]. Studies examining actual cybersecurity behaviour are uncommon (most studies are often limited to assessing intention), and those that do so are often incomprehensible and incomplete. Regrettably, while numerous components of cybersecurity awareness and training are being discussed, there is still no proper and reliable method to measure cybersecurity behavioural change.

7 Trusted Execution

7.1 Introduction

In the last two decades, almost every aspect of people's daily lives and all areas of human activity have been pervaded and revolutionised by digital technology. Sectors vital to society and nations, such as the economy, industry, culture, healthcare, social and government activities, nowadays use massive amounts of software to deliver their services, benefiting from indisputable advantages in terms of time, cost and efficiency. However, IT systems are vulnerable to a huge number of cyber-attacks, that are constantly growing in both number and severity, thus trusted software execution is the goal that industry and academia are pursuing to protect IT systems and their sensitive data from cybercrime attacks.

Traditionally, hardware isolation mechanisms have been introduced to provide various protection mechanisms: virtual address spaces and memory control units protect user applications from each other, privileged instructions protect system software from user applications, and hardware virtualisation creates isolated execution environments protected from each other. However, user applications remain unprotected by the privileged software of the operating system and hypervisor, consisting of millions of lines of code that host a very high number of bugs [53, 88], exploitable by attackers to gain privileged access to the platform [187].

This scenario is further complicated by the advent of cloud computing, nowadays increasingly used by companies due to its indisputable economic advantages. In this case, the user applications have to trust the honesty of the infrastructure provider, the employees with privileged accounts or physical access to the cloud nodes, and the other tenants running their workloads on the same platform.

Trusted Execution Environments (TEEs) were introduced to allow security-sensitive user applications, or the most critical portions of them, to trust only

the hardware support for the TEE plus a software layer that runs in isolation and constitutes the *Trusted Computing Base* (TCB) for the application. The smaller the TCB and the better its security, because this reduces the attack surface and the likley number of vulnerabilities. TEEs also protect applications from physical attackers, for example those that could read sensitive data loaded in clear into the RAM of the platform. This protection is achieved by means of cryptographic layers that shield data while they are processed.

In the 2000s the Trusted Computing Group (TCG) proposed the *Trusted Platform Module* (TPM) as a secure co-processor to perform particular services defined by the TCG, mainly aimed at the verification of the platform's integrity status and the protection of private keys from unauthorised access. However, the TPM is not intended to execute arbitrary applications in its isolated environment, nor can it be installed on any type of device. To meet the need to protect arbitrary user code and data, the industry world began to work towards the creation of TEEs solutions based on special secure modes of the main processor, the first of which was *TrustZone* [17], proposed in 2002 by ARM, followed in 2014 by Intel with *Software Guard Extensions* (SGX) [119], and in 2016 by AMD with *Secure Encrypted Virtualization* (SEV) [125]. At the same time, the academy also looked for suitable software solutions to create TEEs, among which we find *AEGIS* [225], proposed in 2003, *Bastion* [37] in 2010, *Sanctum* [54] in 2016, and *Keystone* [138] in 2020.

Despite the improvements introduced by TEE solutions to pursue trusted software execution through smaller TCB and strong isolation, achieving security depends not only on the TEE technology adopted but also on the trustworthiness of the application code that runs inside it. Identifying vulnerabilities present in the code running in the TEE, as well as detecting its compromise at run-time, constitute challenges that current state-of-the-art TEEs do not address but need to be considered by the scientific community in next years [166].

7.2 Who Is Going to Be Affected?

As the IT systems are becoming more pervasive, distributed, and vital in the current world, there is no sector of our society that can live without trust in the execution of its software components and protection of the sensitive data.

Of course, there is a relative scale of importance. If individuals are not offered trusted execution on their personal systems, then the risks are limited to the assets of that specific individual. On another hand, if the IT system of a commercial company or a government body does not support trusted execution, then the stakes are much higher, depending on the application area of the affected system. In particular, such large systems are the preferred targets for ransomware attacks (i.e., a malware that encrypts data and ask a ransom

to decrypt them) and APT injection (Advanced Persistent Threat, i.e., a permanent malicious application that remains hidden to continuously exfiltrate information or waiting a critical time to perform a destructive attack). Recovering from ransomware may take a very long time, from days to months (note that paying the ransom is no guarantee to have all the data back). APT are even more insidious as they can go undetected for years.

7.3 What Is Expected to Happen?

If software components are executed without proper protection, then the results generated cannot be relied upon for any purpose. Hence any kind of damage can be expected.

If used in an industrial control system (ICS) then production can be blocked or products may be manufactured in the wrong way, eventually leading to defects or damage in other systems using these products as components.

If the attacked software element is an application handling (directly or indirectly) money (such as an Internet banking app or a company payment system) then financial loss can be expected.

Trusted execution is particularly important for cyber-physical systems interacting with humans. For example, this is the case of railway or air traffic control systems or autonomous vehicles. Injection of malware or modification of the configuration of these systems may lead to physical harm to persons, up to death.

Another possible scenario concerns the theft of sensitive user data, such as digital identity, bank credentials, or commercial plans. If this information is not properly protected and used within trusted execution environments, it is vulnerable to theft by an attacker, who can use it to impersonate another person to obtain money illegally through unauthorised banking transactions, commit scams, discredit or put a person in a bad light by carrying out illegitimate actions on his behalf.

In the field of commercial espionage, companies can suffer considerable economic and image damage if attackers manage to steal customer data or confidential information, related to production processes or new projects upon which the future development of the company depends, thus bringing illicit advantage to competing companies.

7.4 What Is the Worst That Can Happen?

The worst possible consequences depend on the application controlled by the system targeted by the attacker. Therefore, it is obvious that the more critical the system and the worst the effect of the attack.

The study "Cost of a Data Breach Report 2022" [116] shows that ransomware and destructive attacks represented 28% of breaches against the

critical infrastructures examined, highlighting that attackers aim to interrupt financial services and to damage industrial, transportation and health-care organisations. The criticality of these infrastructures requires the adoption of cutting-edge security techniques, such as the creation of trusted execution environments and the timely detection of any tampering with the code and configuration of the systems.

For example, in the event that a group of attackers succeeds in blocking a nation's electricity grid, millions of families would be left in the dark, companies' production would be blocked, communications would be cut, banks would be offline, hospitals would not be able to guarantee health care, air and train traffic would stop. One such attack happened in December 2015 in Ukraine, when three utility companies were attacked simultaneously by the BlackEnergy malware, leaving hundreds of thousands of homes without electricity for six hours. Another attack of considerable gravity occurred in Iran in 2010, when the Iranian nuclear program was blocked due to sabotage of the Natanz enrichment plant by means of the Stuxnet virus, which caused the destruction of the centrifuges of the plant while preventing the detection of the malfunctioning of the system itself. Running the critical application that supervises the operation of the centrifuges within a TEE would have protected it from a virus that infects the Rich OS.

This last attack is a clear example of what is the worst scenario: the injected malicious application does not limit itself to block the normal behaviour of the system but completely subverts it to perform wrong operations that would directly damage the system itself or persons that use it.

7.5 Research Gaps

Over the past two decades, a lot of work has been done to build execution environments able to guarantee confidentiality and integrity to execution and to allow external entities to assess the trustworthiness level of systems. Nonetheless, the TEEs themselves pose new challenges that need to be addressed by the scientific community.

7.5.1 Attack vectors against TEE security guarantees

A TEE can be exposed to typical software vulnerabilities, with the addition of architectural vulnerabilities native to a particular TEE solution. A TEE should have a small TCB with a narrow interface to minimise the attack surface. Over the years, several software and structural vulnerabilities have been found in specific TEE implementations. However, more experienced teams are developing smaller and more secure TEEs, thanks to the scrupulous adoption of secure software development best practices and rigorous validation of the TEE design and code. This has caused attackers to shift their focus to more

sophisticated attacks at the edge between hardware and software [197]. An important research area for the next years will concern the study of micro-architectural side-channel attacks, that is, attacks that exploit information leakage from the hardware infrastructure to reveal sensitive information, such as private keys.

7.5.2 Protection mechanisms against compromised TEE applications

TEEs represent a valid technological solution to execute security sensitive workloads in a protected environment. However, if the application code deployed within them contains vulnerabilities, they can be exploited by an attacker to compromise the security of the entire TEE. This problem is becoming more and more concrete, and its solution more urgent, because developers begin to use TEEs to run complex applications containing a large code base, thus increasing the likelihood that exploitable bugs are present within the TEE. It has also been observed that the security features of the TEEs themselves can help attackers to install higher level stealthy rootkits that are extremely difficult to detect through current defense mechanisms [166]. For example, antivirus tools running on the operating system can not detect malicious code nested in a TEE because, by design, the OS cannot access the TEE memory, which is often also encrypted.

For what has been said, the security of a TEE cannot be given for granted because it's a complex matter, not guaranteed just by a perfect architectural design and implementation. Therefore the creation of solutions able to detect bugs in the application code developed for a TEE and monitor its trustworthiness at run-time represents an important research area.

7.5.3 TEEs and cloud computing: interoperability and management challenges

Some of the major cloud infrastructure providers have included TEEs in their service offering, since TEEs are able to improve the security and privacy guarantees of cloud computing. However, two conceptually different TEEs models can be adopted for cloud computing [94]: the virtual machine-based model encrypts the entire system memory of a virtual machine; the process-based model selectively encrypts a memory zone of the deployed application, delegating to the developer the decision to choose which section of an application's code to protect. Concrete implementations of these models have been developed for different platforms – SGX and the new *Trusted Domain Extension* (TDX) for Intel platforms, SEV and the forthcoming *Secure Nested Paging* (SNP) for AMD platforms, TrustZone and Realms for ARM – and CPU architectures (x86, RISC-V, ARM). The great variety of proposals fielded by research and industry causes interoperability problems in moving a service from one ar-

chitecture to another, and compatibility problems in deploying applications written for traditional systems within TEEs. An important research area is the study and development of frameworks that offer a level of abstraction capable of making the heterogeneity of TEE solutions transparent to the application developer, while maintaining the same security guarantees offered by the underlying TEE.

Another aspect that is gaining more and more importance is the development of solutions that allow to combine TEE technologies with container technologies, in order to promote the use of TEEs in cloud-native scenarios and facilitate the deployment of TEEs-based applications inside containers, at the same time offering the same user experience as ordinary containers and a smooth integration with the Kubernetes ecosystem.

7.5.4 TEEs cryptographic primitives in the post-quantum era

In recent years we have witnessed remarkable advances in the field of quantum computers, which allow us to push computational capabilities far beyond classical ones. This has inevitably caused important consequences in the field of cryptography, as quantum computers allow the execution of algorithms that offer quantum speed to the solution of the mathematical problems on which classical cryptosystems are based. This threat was highlighted with NIST's call, in 2016, to present new cryptographic algorithms resistant to quantum computer attacks. In July 2022, NIST selected the first four algorithms that will become part of NIST's post-quantum cryptographic standard [168]: CRYSTALS-Kyber for general encryption, CRYSTALS-Dilithium, FALCON and SPHINCS+ for digital signatures.

TEEs base their security on cryptographic primitives implemented in the hardware root of trust of the platform, currently based on classical cryptosystems. An important research area for the next few years will be the design and implementation of hardware root of trust relying on post-quantum cryptography, in order to withstand quantum computation and quantum side-channel attacks.

7.6 Example problems

Tangible example problems might include the following ones:

Detecting a compromised TEE application. An application running in a TEE could be compromised by an attacker due to the presence of vulnerabilities in its code. The strong security and isolation guarantees offered by TEEs can be exploited by attackers to implement and install hard-to-detect advanced rootkits in a platform [166]. Aim of the research is to develop solutions able to detect compromised TEE applications at run-time.

Technology-agnostic TEE solutions in cloud computing. In 2019, a group of companies, including Intel, Microsoft, Google and ARM, founded the *Confidential Computing Consortium* (CCC) with the aim of promoting the adoption of TEE solutions in the Cloud. CCC sponsors several open-source projects that offer solutions to the compatibility and interoperability problems that TEE technologies pose, such as Enarx [70], Gramine [98], and Occlum [172]. The objective of this research is to analyse the effectiveness of current technology-agnostic confidential computing solutions, evaluate their performance, study their possible security shortcomings, apply them to the cloud computing domain.

TEE applications in cloud-native scenarios. Today, many service providers offer technical solutions to facilitate the development and execution of TEE applications in the cloud- Google's Asylo [97] and Azure's OpenEnclave [21] are two important examples of them. However, while they simplify the development of TEE-based applications, they still require the developer to acquire new programming skills and develop the code using the corresponding SDKs. Furthermore, even though the goal of these frameworks is to support heterogeneous TEEs by using the same API, they still rely primarily on Intel SGX technology. The aim of the research is to design and develop solutions that allow users to run their services inside "TEE-based containers", without requiring modifications to the application code, while supporting heterogeneous TEE back-ends and providing easy integration with the Kubernetes orchestrator.

Trusted execution in low-end IoT devices. Nowadays, the security of IoT devices is essential as they are increasingly used in multiple fields (e.g. vehicles, industry, smart cities, healthcare). However, IoT systems present special security challenges due to their heterogeneity, considering not only the embedded devices but also the networks, the management and data analysis services, and the storage. Furthermore, while high-/middle-end devices can benefit from the security guarantees offered by TEEs, low-end devices typically do not have hardware security mechanisms to protect security-sensitive applications. A research area in the IoT field concerns the design of TEE architectures that meet the challenges posed by low-cost and low-power devices, to ensure the trustworthiness of a wider range of IoT applications and the data they produce. This should go along with the development of solutions for a secure remote and automated management of the IoT devices, often installed in uncontrolled environments.

TEE's security functions integration in the network. Goal of the research is the integration of TEEs technologies within the common network op-

erational mechanisms and the enhancement of their security thanks to TEE's hardware and software guarantees. For example, an important aspect concerns the creation of mutually trusted channels between TEE-based applications, extending the TLS protocol with mechanisms that allow the verification of the integrity and authenticity of the end-points of the communication channel, portable for heterogeneous TEEs.

Quantum-resistant roots of trust for TEEs. OpenTitan [177] is an open-source framework that supports the design and integration of vendor- and platform-agnostic silicon roots of trust to integrate into servers, storage devices, peripherals or other types of platforms. The goal of this research is to realise an OpenTitan extension capable of using post-quantum cryptography in silicon design and firmware implementation of a root of trust, in order to support quantum-resistant TEEs.

Runtime detection of manipulation of system conditions Altering the correct configuration in which a chip has to operate can lead to unexpected software behaviour or changes in the execution flow of the code; this is typically accomplished by physically modifying the power of the device, the clock, the electromagnetic field or the physical interfaces [197]. The aim of this research is the creation of runtime mechanisms capable of dynamically sending alerts when a change in system conditions is detected.

8 Privacy by Design

8.1 Introduction

In a world that is increasingly digital, vast amounts of personal data are collected and processed, often ubiquitously and intransparently, and used by governments and/or commercialised among several service providers, data brokers, and advertisers. This commoditisation of personal data has further eroded individuals' rights to privacy. For many decades, researchers have looked into this growing Orwellian trend of profiling and surveillance, attempting to find a balance between the advances in technology and the protection of privacy. Aiming at the very core of the systems' design, Ann Cavoukian coined the term Privacy by Design (PbD) back in the '90s, proposing a series of seven foundational principles, instilling privacy assurance as an organisation's default mode of operation [35] (see Figure 8.1). Behind these principles is also the observation that privacy is best achieved when addressed at the earliest stages of technology development, i.e. in the conceptual design phase.

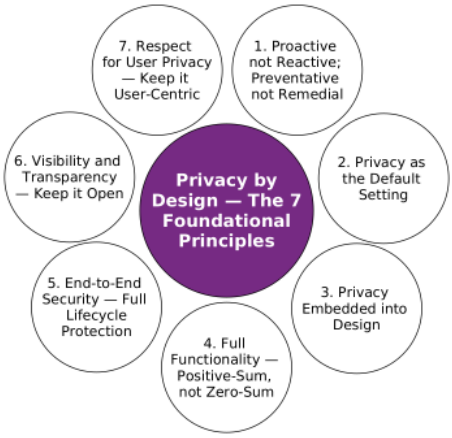


Figure 8.1: Cavoukian's 7 Foundational PbD Principles [35]

Although acclaimed by many researchers and policymakers, PbD is often criticised as being too vague and hard to translate into concrete software engineering practices [239]. In fact, today, there is still a significant gap between research and practice, e.g., translating high-level PbD principles into concrete engineering practices that software practitioners can effectively and efficiently adopt. Aiming to close this gap, the emerging discipline of Privacy

Engineering has been formed that focuses on designing, implementing, adapting, and evaluating theories, methods, techniques, and tools to systematically capture and address privacy issues in the development of socio-technical systems [102]. Therefore, further developing this area of Privacy Engineering is a significant challenge for researchers as well as for organisations that want to integrate and operationalise PbD. For organisations, this challenge is also a matter of regulatory compliance now that the notions of data-protection-by-design and data-protection-by-default have been incorporated as part of the European General Data Protection Regulation (GDPR), in force since 2018.

8.2 Who Is Going to Be Affected?

As mentioned, the GDPR has raised the bar for privacy, including PbD as part of Article 25 "Data protection by design and by default" for protecting the fundamental rights of individuals in Europe. This regulation affects all organisations that collect and process personal data of EU citizens and residents, meaning that it can apply even if an organisation is based outside EU. However, the legislation leaves it open to which exact technical and organisational protective measures are to be taken to fulfil the requirements of PbD. This, of course, creates further challenges to organisations, but more specifically, to software architects and developers, who are ultimately responsible for designing the systems.

Besides that, large technology organisations have started hiring privacy engineers and establishing privacy red team operations, which help to embed privacy in the system's design and proactively test processes and systems to identify privacy risks. However, not many organisations have the resources to hire privacy engineers, let alone to maintain an entire privacy engineering department. From what we see, this is especially the case for small and medium-sized enterprises which comprise the majority of organisations nowadays. Even though the practical effects of GDPR are beneficial to individuals and society as a whole, they pose significant challenges for organisations, and in turn, to the research community that aims to make PbD a reality in an ever-changing technology landscape.

8.3 What Is Expected to Happen?

In our daily lives we are surrounded by technology, with a wide range of data-intensive software systems being used for personal and professional activities. Failing to accomplish PbD in today's working systems can severely jeopardise individuals and the democratic society as a whole [205]. The lack of privacy has negative impacts to individuals ranging from embarrassment and reputation damage to various forms of discrimination that adversely affect individuals' rights and freedoms and physical and mental health. On a

societal level, privacy is also considered as an essential component for a functioning democratic society [26, 69]. If people cannot fully exercise their rights and freedoms, such as freedom of association (e.g., political and religious) and freedom of opinion and expression (including holding back one's views), there are negative impacts on the individuals democratic participation, also harming their human dignity and personal autonomy.

For such reasons, it is expected that organisations will responsibly create and adapt software systems following PbD principles, adhering to privacy rights as enshrined in today's regulations. As a result, people would be able to take privacy for granted, with the full expectation that any data that is collected and processed has been lawfully acquired, that the specific purposes for processing are transparently communicated and adhered to, and that whenever possible users are able to exercise various rights over their data, e.g., access, correction, deletion, object processing, etc.

8.4 What Is the Worst That Can Happen?

The GDPR is sometimes referred as a regulation that has "real teeth", applying massive fines to organisations that violate privacy rights. Non-compliance with the GDPR can lead to fines of up to 20 million euros or up to 4% of an organisation's worldwide annual turnover, whichever is higher. Studies are showing an increasing number of fines based on the GDPR since its publication in 2018 [190,250], with the largest fine so far of 746 million euros imposed by Luxembourg's Data Protection Authority against the tech giant Amazon in July 2021 [141].

However, arguably, the collective societal costs of privacy violations can be much higher than any legal fines. As mentioned, the deterioration of individual privacy rights incurs in the weakening of democracy in itself. If left unchecked, organisations can exploit advanced technologies such as artificial intelligence to carefully craft and target advertisements, generating a scenario of social manipulation [147]. Evidences for such exploitative use of AI-based profiling of users have been seen in social media manipulation, spreading fake news and misinformation, and targeting voters with the intent to tilt the results of elections – the Cambridge Analytica scandal probably as the most widely known case, but there is also evidence of interference in the Brexit referendum and elections in Brazil, Sweden, and India [15].

8.5 Research Gaps

8.5.1 Privacy Goals vs. Other Goals

Solving trade-offs that need to be made between privacy protection and other goals constitutes a major challenge, as illustrated in [104]. Also, our inter-

views conducted within the requirement elicitation phase of CyberSec4Europe [82] conveyed that for the domain of privacy-enhancing identity management systems, research and practical challenges for adequately addressing trade-offs that need to be made between privacy protection, usability and trust need to be addressed. Preferences for privacy trade-off preferences also differs culturally, which also needs to be considered for achieving usable privacy and identity management solutions by design [121].

8.5.2 Building the Theory of Organisational Privacy Culture and Climate

As advocated by [24], organisations can be seen as living human entities, and as a human group structure, they have a culture. This group culture is a reflection of the consciousness of its leaders. Therefore, the values and behaviours of the leaders will largely influence the culture of the entire organisation. If a culture is "toxic" in one or more of its facets, it is important to look closely at the values and behaviours displayed by leaders and top management. And this includes the facet of privacy and how it is perceived inside the organisation.

Many researchers have addressed the topics of organisational privacy culture [57] and climate [16, 103], showing that these components strongly influence the employees' perceptions, attitudes, and behaviours concerning privacy. Such research emphasises that leaders must create a conducive environment to integrate PbD successfully into the organisational processes. However, since this area of Organisational Privacy Culture and Climate (OPCC) is in its embryonic stage [122], there still needs to be more primary research to solidly build the theory around the topic, as well as to define ways to measure and embed privacy in organisations reliably.

8.5.3 Countering Device Fingerprinting

Device fingerprinting attacks, which can recall a device by coincidental data that the device leaves while communicating in a network, become serious threat for location privacy. Network devices become increasingly heterogeneous, which enables a diversity of fingerprints that can be exploited for attacks. Further research is needed for understanding and measuring the accuracy of fingerprinting attacks, e.g., by measuring how much entropy is contained in a specific fingerprinting source for providing guidance on achieving data minimisation in a PbD process.

8.5.4 Data Subject Rights Engineering

According to Art. 15-21 GDPR, European citizens whose personal data is processed at any organisation globally have a set of rights towards these data processing organisations. For instance, the right of access allows them to

be informed about the nature and purpose of processing, as well as about the set of data stored and processed. To some degree, this even spreads to sub-processors that are involved in the data processing as well. The right to erasure, often also dubbed the right to be forgotten, allows for demanding deletion of all (or part of the) personal data stored at an organisation – unless other explicit reasoning restricts this (e.g., concerning personal records at law enforcement agencies). The right to rectification enables the individuals concerned to change their data, e.g., to correct false information in a data record.

However, when it comes to enforcement of these data subject rights granted by the GDPR, a lot of open issues arise. How can the transparency demanded by the right of access be realised in a multi-organisational, distributed workflow? How can restrictions to processing according to Art. 18 GDPR be implemented into such a workflow? How can a request for erasure or rectification be propagated throughout a processing chain, and which part of the processing constitute the same workflow with respect to the specific purpose of processing? When do two processing activities belong to the same workflow, and when do they instantiate a separate data processing instance, with separate needs for user consent and data subject rights enforcement?

8.6 Example Problems

When addressing these and other open research challenges in the domain of Privacy by Design or Privacy Engineering, the following specific problem domains need to be addressed more closely.

Identifying factors and defining constructs in the OPCC area. Organisations can greatly benefit from practical instruments, such as questionnaires [57], that could help them to measure or assess organisational aspects such as "privacy culture" and "privacy climate". To do so, researchers still need to understand and identify the key factors that form OPCC constructs, and test instruments in terms of validity and reliability.

Algorithmic fairness vs. data minimisation. Trade-offs between data minimisation and fairness for machine learning models was recently discussed [38] and is still to a large extend an unsolved issue.

Metering risks in device fingerprinting. Further research is needed on defence mechanisms for device finger printing risks that can be avoided (e.g. based on software-defined behaviour such as APIs) and on remaining risks that will be hard to defend (e.g., finger printing attacks based on physical device properties, such as drift of physical clocks).

Measuring the level of privacy protection offered. It is often unclear what level of protection is actually provided by a certain privacy-enhancing

technology of privacy-aware design decision. Along with this uncertainty comes the question whether a given set of privacy-enhancing measures was sufficient to be considered state of the art in the sense of GDPR, or what other levels of protection would have been adequate. Hence, the selection of Privacy-Enhancing Technologies (PETs) to apply in a given context and scenario, along with the determination of the protection achieved, is an open research question. Early methodologies exist (like the Privacy Design Strategies [111], LINDDUN [60], or the Standard Data Protection Model [51]), but these are not sufficient in detailing the metering of the level of protection provided.

Threat modelling as "by design" enabler. Privacy needs to become integrated into all steps of software development by design. According to GDPR's risk-based approach, it is crucial to first determine the privacy problems that can arise in order to properly resolve them. Threat modelling is a well-known approach in the security domain[1] and has been gaining momentum in the privacy community as well. Privacy threat modelling allows to systematically identify and mitigate privacy issues at the architectural level. By identifying these problems early, they can be tackled at the system's core in a more efficient way. The threat model should inform decisions in subsequent design, development, testing, and post-deployment phases[2] (e.g. determine the key verification targets for software testing). Risk assessment should guide prioritisation. Threat modelling automation is the next step to facilitate a growing adoption. Developments in run-time and adaptive threat modelling will also strengthen the incorporation in Continuous Integration and Continuous Delivery (CI/CD) settings.

Data custodians and delegated data subject rights. Utilising data subject rights against a data controller requires a specific type of interaction according to the rules outlined in the GDPR. Data Controllers have to provide communication means for such requests accordingly, and may have a large incentive to automate or at least structure such requests as far as possible, e.g. to save workforce. At the same time, the burden of utilising one's data subjects rights over time can easily become cumbersome, e.g. if requests for erasure need to be done repetitively due to data collection processes not properly controlled by the data controller, or if right of access requests must be preprocessed to become understandable to human readers. In such cases, the instantiation of a dedicated data custodian that enforces data subject rights on behalf of the data subject

[1] see e.g. OWASP's top 10: `https://owasp.org/Top10/`

[2] see also `www.threatmodelingmanifesto.org`

may become essential. As foreseen in the European Data Governance Act, data intermediaries that enforce data subject rights must fulfil special requirements, and may be tempted to automate their operations as far as possible as well. Here, open research challenges can be identified in these aspects, such as data subject rights engineering, transparency by design, right of access as a service, or data custodian trust delegation models.

9 Critical Infrastructures

9.1 Introduction

Although the protection of critical infrastructures (CIs) has received the attention of the research community for more than a decade, securing CIs from emerging cyber and hybrid (cyber-physical) attacks is still an open challenge. Various definitions of CIs can be found in the scientific literature, international standards and regulatory documents. In simple terms, and in line with the relevant European Council Directive [64], CIs are large-scale systems or systems-of-systems, that are essential for the proper operation of vital societal functions and for people's well-being.

Take for example the *healthcare* sector: this sector is comprised, among other things, of hospitals, healthcare centres, pharmaceutical labs, blood supply facilities, emergency services and research facilities. The disruption or destruction of such facilities, especially if extensive or for a significant duration, may have a severe impact on public health. As another example, consider the *transport* sector: in this case airports, ports, railway infrastructures and road traffic control systems play a significant role in people's mobility, as well as in the proper operation of the supply chain. Other examples of CI sectors include *information and communication technology* infrastructures, such as telecommunication networks and cloud infrastructures; *energy installations* including electrical, gas, oil or nuclear power production, storage, transmission and distribution networks; *water facilities*, including dams, water storage, management and networks; *finance*, such as banking facilities and inter-banking communications; *food management*, including food production, food safety systems, wholesale supply chain, and many more.

One might argue that, “since CIs have been are around for several decades (or even centuries), they must already be mature enough and adequately protected". Unfortunately this is far from being true, for several reasons. The first reason is the increased accessibility of modern CIs and their increased coupling with information and communication systems. A few decades ago, CIs used to be closed systems. Nowadays, Internet connectivity offers CI administrators more efficient, real-time and remote management, without requiring physical proximity to the infrastructure. On top of that, CIs have also become more accessible to end users and closely connected with Internet-of-things (IoT) systems. For example, while some years ago measurement consumption in the electric grid required physical access to the end-user metering systems, nowadays smart metering systems allow not only remote measurement, but also remote control. While changes like the ones described above have increased the efficiency of CI operations, at the same time they have increased their attack surface and have enabled their exposure to remote cyber-attacks.

Last but not least, the increased connectivity of CIs has also increased the *dependencies* between those infrastructures. Different types of infrastructure dependencies exist. For example, an energy provider who receives communication services from a telecommunication operator has a *cyber dependency*. On the other hand, the telecom operator will require electrical power to support its network operations. Any dependency on a physical resource, such as on the energy supply as described above, is a *physical dependency*. Other types of dependencies include *geographical* (when two CIs depend on each other because of their physical location) and *logical* (when some kind of dependency other than those above can be identified).

9.2 Who Is Going to Be Affected?

Anyone who acts as a “consumer" of the services provided by a CI will be affected if a CI is compromised, including people, companies and other organisations. Unfortunately, the dependencies between CI providers increase the significance of potential attacks, as well as the extent of the organisations and people that will eventually be affected. Consider for example a cyber-attack against an electrical distribution network, which supports many other nearby CIs (geographical and physical dependencies), such as telecom providers, traffic-light systems, government services, data facilities, hospitals, data centres or airports. The disruption or the degradation of the electrical supply will concurrently affect to some extent, all the CIs that depend on the specific electrical distribution network under attack. Such types of concurrent dependencies of multiple infrastructures on a single CI may result in what is known as *common-cause failures*. Common-cause failures will obviously affect

multiple organisations, in both the public and in the private sector, as well as many people who “consume" the affected services.

Another type of dependencies that may concurrently affect a considerable number of people, companies and organisations are those dependencies that cascade from one CI to another. One of the most famous and well-studied cases is the California black-out [196], where the failure of a power station caused multiple *cascading failures*, due to a series of CI dependencies. For example, the energy reduction caused a decrease in the amount of petroleum that was channelled to the airport facilities, therefore causing severe problems to the airport services and ultimately to the flight schedules of the airport operators. At the same time, the loss of electrical power led to the degradation of the steam injection units that were used to power oil recovery units. The latter led to a feedback effect, since the produced oil was also used as a fuel by the electrical power operator that was initially affected! Finally the electrical power reduction also affected the water pumps that were used in crop fields.

9.3 What Is Expected to Happen?

Attacks against CIs may lead to all kinds of consequences, such as loss of life, financial loss, public disorder or disruption of business operations [222]. Attacks against hospitals may affect patient treatment. For example, a ransomware attack in a German hospital caused a delay on a patient’s emergency treatment, who eventually lost her life[1]. Although the relevant police investigation concluded that “the delay was of no relevance to the final outcome"[2], it also warned that it’s a matter of time before hacking hospitals leads to tragic results. Attacks against energy infrastructures can directly lead to loss of productivity, and to severe economic loss, especially if cascading effects on other infrastructures are also considered. Examples of cyber attacks of this kind, allegedly caused by nation-state adversaries, include the attacks against Ukraine’s electrical grid in 2015 and in 2016[3]. Telecommunication infrastructures are also very attractive attack targets, since they and energy are the two sectors with the highest level of incoming dependencies from other infrastructures (almost any CI depends on energy and telecommunication services). Attacks against road traffic management infrastructures or direct attacks on

[1]The untold story of a cyberattack, a hospital and a dying woman: `https://www.wired.co.uk/article/ransomware-hospital-death-germany`

[2]Ransomware attack in German hospital. A report on the investigation findings and warnings can be found in: `https://www.technologyreview.com/2020/11/12/1012015/ransomware-did-not-kill-a-german-hospital-patient/`

[3]Hackers trigger yet another power outage in Ukraine: `https://arstechnica.com/information-technology/2017/01/the-new-normal-yet-another-hacker-caused-power-outage-hits-ukraine/`

vehicles of any kind (cars, ships or planes) may lead to disruption of traffic or even to lethal accidents.

9.4 What Is the Worst That Can Happen?

As CIs are vital for people's well-being and their disruption could lead to severe societal, financial and safety consequences, they are very attractive targets for malicious attackers.

The preparation of cyberattacks against CIs requires high motivation, usually high resources and skills and some kind of capability for initial access by the adversaries. Unfortunately, various such adversaries exist in the current threat landscape. For example, nation state adversaries may be sufficiently motivated and may have the required time, resources and skills to deploy *Advanced Persistent Threats* (APT) against targeted CIs, with the intention of causing severe damage to the CIs of an enemy state. This may be used as an asymmetric attack or as part of a hybrid cyber-physical war. Terrorists may also be motivated to cause severe disorder and loss of public confidence. Finally, cybercriminals may be motivated to attack CIs, aiming to a high economic gain, e.g. through ransomware and blackmailing attacks.

The increased network connectivity and the inter-connectivity of CIs is increasing the attack surface, as it may provide adversaries with several initial points of entry. In addition, the lack of security training and awareness may also be exploited by adversaries to gain initial access. For example, spear phishing campaigns aimed at targeted users can be a preparatory action for an APT.

9.5 Research Gaps

Obviously the protection of CIs from attacks such as those described above is not a trivial task. CI protection is a multi-disciplinary process. From a technical point of view, it requires a better understanding of the threats, vulnerabilities and exposures, as well as an efficient and effective protection. From a social and business perspective, it requires a better understanding of the dependencies between CIs and the impact related to the specific attacks, as well as increased training and awareness of the people involved.

9.5.1 Modelling, analysis and simulation of non-trivial threats including APTs, cyber-physical and climate-change related disasters

APTs are expected to become more powerful, even more sophisticated and more frequent. The same is true for other types of man-made hybrid (cyber-physical) attacks or even for natural and climate-change disasters. Because of their complexity [209], CIs are not currently equipped with advanced modelling tools that will allow them to adequately prepare themselves for such

non-trivial threats and to effectively manage any likely attack [242]. There is a need for novel approaches to support the modelling, analysis and simulation of such threats, e.g. [254], in order to better prepare CIs to deal with them in the real world, but also to propose fast, efficient and reliable response tactics.

9.5.2 Develop risk assessment and management methodologies for systemic and supply-chain risks

As was highlighted very clearly through the events triggered by the war in Ukraine, several CI sectors, such as energy and transport, may trigger systemic, cross-sector and cross-border risks for society and the economy (e.g. by disrupting energy or food sufficiency on a European or global scale). Given their special characteristics [256], there is a need to develop novel methods for the early identification and proactive management of such risks, especially in a cross-sector and cross-border context.

9.5.3 Resilience of Critical Infrastructures

Increasing the resilience of CIs is an ongoing research goal, at both a “microscopic" and at a “macroscopic" level [193]. From a component-wise point of view, there is a need for additional research into *resilient and fault-tolerant embedded systems*, which are essential for the proper monitoring and control of critical cyber-physical systems. From a system-wise view, assuring a level of *resilience for critical operations* and services in a cost-efficient way is an open challenge. There is a need for (re)designing resilient-by-design infrastructures, by developing nearly cost-optimal solutions that assure the controlled redundancy, resourcefulness and quick recovery of critical operations [210]. A resilient design should also consider the integration of cost-efficient (semi-)automated response capabilities to effectively minimise the impact of cyber attacks at the earliest possible stage.

9.5.4 Improved AI/ML assisted models for (inter)dependency analysis

Despite past and recent research efforts (e.g. [204, 208, 224, 243]), there is still a need for improved models for the analysis of CI (inter)dependencies, exploiting real-time threat and risk monitoring systems assisted by artificial intelligence (AI) and machine learning (ML). For example, there is a need to develop models for understanding the perturbation flux from one system to another, which encompass non-local effects with large difference in time scales. As model training should be based on real data from actual systems, this requires a more direct involvement of the infrastructure operators.

9.5.5 Event prediction based on all types of dependencies

Although cyber and physical dependencies are generally captured in depth in current event prediction models, geographical dependencies are not adequately captured. Because of this, prediction of disruptive events is not accurately mapped on a specific territory. Efforts should be made in that direction, for example using geographic information systems to capture the maximum possible level of spatial resolution and to map this information onto data concerning dislocation of assets, also considering the most important perturbation types for each asset (e.g. ground-shaking, rain, wind, or temperature).

9.5.6 Collaborative situational awareness for the CI ecosystem

As a continuously increasing amount of cybersecurity data (e.g. emerging threats, zero-day vulnerabilities) becomes available on a daily basis, it becomes more difficult to effectively utilise that data to improve the cybersecurity situation in an organisation [11]. Improving situational awareness for CI operators requires a multi-disciplinary socio-technical approach, which includes people as part of the solution. Methods and tools are needed to facilitate cooperation and collaboration *within* and *between* the CI operators and the relevant sectoral, national and European authorities.

9.6 Example problems

Tangible example problems might include:

Digital twins of CIs. Develop 4-dimensional models to continuously monitor the behaviour of the infrastructure, by constantly receiving input from IoT devices. This may be a priority for infrastructures suffering from ageing problems, in order to continuously analyse their expected state and to perceive possible deviations from their normal structural behaviour.

Supply chain security for CIs. CI operators depend on various supply chains to deliver their services to end-users. Attackers are increasingly using the whole supply-chain to perform attacks. As supply chain security is also mandated by the NIS/NIS2, supply chain security for CIs requires further investigation.

Develop tools to support effective recovery from critical all-hazard events. The increased number of interdependencies requires the development of innovative decision support tools to assist in the early identification of critical events and in the most effective recovery strategy, taking into consideration their dependencies and other constraints.

Develop methods and tools for the early detection of cyber-physical events. Since CIs are cyber-physical systems with strongly tight dimensions, there is a need to develop methods and tools to detect and assess cyber-physical attacks, by concurrently considering both their cyber and physical vulnerabilities from a holistic perspective.

Design improved multi-disciplinary regulatory framework. Competent authorities, as defined in NIS/NIS2, are playing a key role in CI protection. It is necessary to develop sustainable and collaborative regulatory frameworks that can consider all pertinent risks and handle incidents involving various dimensions such as technical, societal or legal aspects.

10 Metaverses

10.1 Introduction

Despite what many readers may think, the metaverse is not a product, not even a brand of some social network company, but a name given to a set of technologies applied in platforms for the Web on the Internet. In fact, the concept of virtual worlds dates back at least to the 19th Century [27]. Still, the term metaverse was used to name a futuristic concept, described in a science fiction book in 1992, which popularised it [223], and was shown in a visual format in a movie 20 years later (i.e. 10 years ago) [159]. Metaverse in practice today, refers to a new type of Web platform, which is supported through a comprehensive set of technologies, some of which already consolidated and others in evolution, which will allow users greater interactivity and socialisation in immersive 3D digital environments, represented by a universe of new digital virtual worlds, preferably mirrored in the physical world.

Metaverse research witnessed a first wave of "hype" between the years 2000 and 2006, with many results and visibility. Currently, in 2022, it is going through a second wave of interest, now brought about by commercial players that started to market their metaverses and events held inside them, but also by a widely publicised metaverse-related public announcement by one of the Western Big-Techs in late 2021.

Today, it is possible to understand what metaverses are, or could be, by browsing through Web platforms such as Second Life, Decentraland, Somnium Space, The Sandbox, Roblox, Horizon Worlds, Avakin Life, Mesh, among others.

The concepts of digital virtual worlds of today's metaverses are typically based on Web 2.0 technologies that include 2D and 3D Virtual Reality spaces, with computer graphics images ranging from low to high resolution, and some platforms using Augmented Reality technologies in various activities. The representation of users is always through avatars, and, as access to platforms depends on an exclusive login, there is a lack of interoperability, as avatars are confined to a single metaverse and its worlds, not being allowed

to move from one metaverse to another, on another platform, without logging in again in the physical world.

The means of accessing the platforms can be done through various devices that include smartphones, tablets, laptops, desktops, workstations, and even head-mounted displays or virtual reality glasses. Some platforms already use monetisation through blockchain and cryptocurrencies, with the adoption of smart contracts and fungible and non-fungible tokens (NFTs) that enable mercantile activities. Note, however, that today there are still few platforms for metaverses that use Web 3.0 technologies, the HTTP/3 protocol, and other more advanced and secure technological resources, but this is clearly the path for the future.

To support these features, the most advanced technologies such as Web 3.0 (latest Internet version), Artificial Intelligence, Brain-Computer Interfaces, IoT (Internet of Things), Blockchains, and Virtual, Augmented, Extended, and Mixed Reality will usher in a large number of opportunities that will probably impact large parts of our societies, just like Social Networks did.

10.2 Who Is Going to Be Affected?

In order to analyse the plausible impact of metaverses in future, let's embrace in this chapter their full vision as digital worlds that are massive, immersive, persistent, open and economically developed, as follows [95].

- **Massive:** They can host an unlimited number, or at least a very high number of concurrent users, as the computing power of the Web platforms and of the users' machines evolves in terms of graphics processing and connectivity.

- **Immersive:** They offer three-dimensional and embodied experiences, based on Virtual Reality (VR) and Extended Reality (XR). Imagine that after work you go to a small room in your house or neighbourhood, dress up in a connected "sensory suit", and tell the computer the metaverse of your choice and, from there, you enter the site, having the sensation of being present and living "inside" a chosen digital virtual world, controlling many things with your thoughts. This is in contrast to the current experience of most game universes, which are two-dimensional, confined to screens, and mediated by clicks, typing, and either screen or mouse.

- **Persistent:** Metaverses will never stop or reset. Or at least that will be the perception of their users. The life and society of a metaverse will continuously evolve, even if some avatars are not present, as it happens to normal life in our world.

- **Open:** Anyone with good Internet connectivity and VR/XR computing power can go into metaverses, move within them as an avatar, interact with other avatars, socialise, trade, build, produce intellectually, and so on.

- **Economically developed:** There will be extensive trade in goods and services within the metaverses, which may or may not have an impact in the physical world outside them. They will likely be supported by Decentralized Finance (DeFi) architectures and digital monetary systems that encompass blockchain technologies, cryptocurrencies, smart contracts, and fungible and non-fungible tokens that will enable property rights assurance practices.

Clearly, such an ambitious vision points to a high likelihood of a renewed collision between Industrial Age Governance and Digital Age Governance, which would affect all layers of the population, from simple metaverse users to policy makers.

In fact, governments are already nervous. In the EU the European Parliament is concerned mainly about Competition, Data Protection, Responsibilities, Financial Transactions, Cybersecurity, Health, Accessibility, and Inclusion [145], while the EU Council's main points of preoccupation are Geopolitics, Economic growth, Jurisdiction, Health, Consumer protection, Civil and Penal codes, and Climate change [19]. We note that massive intellectual investment would be required in order for practical solutions to be found and implemented in each of these areas. Besides, there will be thorny issues around reaching consensus in any of these topics. These are some reasons why the European Commission has just included metaverse policy among its priorities [44, 50].

10.3 What Is Expected to Happen?

An analysis of the evolution of metaverse support technologies, such as those described above, the Internet, and the Web – from the Web 1.0 version and the current Web 2.0, to the new level of Web 3.0, especially when thinking about Web platforms with great interactivity and greater social reach –, brings many question and concerns, especially regarding cybersecurity, privacy and protection of (personal) data, regulations, and various aspects of the governance of such digital worlds [216].

Take governance *inside* of metaverses as an example. The concept of "inside" is highlighted because it is different from the concept of interface between the digital world of a metaverse and our physical world since such an

interface is becoming regulated, at least in the European Union (EU), since 2016.

In the EU, the rule-of-law is dominant and its institutions are mostly fit for purpose. However, in this new technological frontier that are metaverses, it is not clear what will be regulated, who will establish and enforce rules, or how this will be done. But any place, physical or digital, at some point of population density will need some kind of order maintenance, including the notion of fundamental rights.

Indeed, thinking of unregulated parallel digital universes is worrisome. And as commerce will be ubiquitous, products, transactions, property rights, and other businesses will need some kind of protocols for markets to thrive. Then all kinds of conflicting situations will have to be resolved by some form of authorities, police, and courts. As well, there must be rules of trade, taxation, income, etc. But then, if a large set of rules has to be established, another important question is who is going to set them: Are they going to be the owners of the platforms of metaverses, since these universes are privately owned? Will they put users to help set up local rules? Or are public authorities from the physical world starting out and expanding their reach into the digital world as well? Whose public authorities to start with? Or are libertarians thinking about creative technologies to govern the metaverses, promoting the ideology that "code is law"? Likewise, what form does such a body of rules take? Accordingly, we can think of the following forms of regulation.

- Signing of usage contracts. However, they may be as long as constitutions.
- Replication of laws and regulations from the physical world. However, this may hinder innovation, and good justifications would be expected for the choice of one model over another.
- Distributed models, based on digital technologies, like blockchain, bitcoins, NFTs, smart contracts (i.e. , persistent scripts).

In addition, the very technological offer of interactivity and immersion of next-generation metaverses will heavily depend on wearable devices monitoring both biometric (e.g., gait, facial expressions, temperature) and neurometric (e.g., fear, satisfaction, attention) data, which will imply continuous and full surveillance of users. In Western societies, where privacy and protection of personal data are fundamental rights, commercial and public interests will have a very difficult relationship concerning this topic.

To compound such issues, the attack surface for security breaches and privacy invasions can become very large in the metaverse, because it integrates a variety of older, current, as well as untested new technologies and systems

whose intrinsic vulnerabilities and flaws will be inherited by the larger system. As a consequence, existing security threats will be amplified, with more severe effects. They include the following (non exhaustive) [244]:

- Lack of security culture from the part of users in such new environments,
- Mismanagement of massive data streams,
- Widespread user-profiling activities,
- Unfair results from Artificial Intelligence (AI) algorithms,
- Digital twins security,
- Security of metaverse physical infrastructures,
- Personal data involved in the metaverse will be more granular and biometric, including emotional, etc.

Finally, the enlargement of the attack surface brought by metaverses will facilitate existing threats in physical and cyber spaces, like persecution, harassment, and espionage, which may increase in frequency and impact. The use of emerging technologies will make more likely security incidents, like hijacking wearable devices or cloud storage, virtual currency theft, or AI misconduct to produce fake news autonomously within metaverses [67].

10.4 What Is the Worst That Can Happen?

Many things can go wrong if provision and usage of metaverses run amok in future, and most of them are related to the notion of trust in them.

It is certain that the vast majority of metaverse users are and will be law-abiding citizens and people who value civilised behaviour. However, among the users is also certain that there will be cheaters and other less honest persons who will join in just to try and make easy money out of what would be defined in most parts of the physical world as criminal activity. Such an environment would not invite trust from users, and licit commercial returns over investment may plunge as a consequence, while illicit undertakings may flourish.

On another, perhaps more important registry, metaverses place major challenges to privacy and governance and they may have the potential to accelerate the geopolitical shift of power from Nation States to private companies. Remind that already today some social network companies have populations that are larger than that of the largest country on Earth. If national governments cannot trust that metaverses will treat their citizens in a legal manner,

then governments may decide to over-regulate metaverses, hampering innovation and increasing fragmentation.

Accordingly, the lack of trustworthy governance and of security and privacy regulations inside metaverses may turn this high-tech Eldorado into a 21st Century Wild-West, where fortunes will be made and lawlessness will be the rule rather than the exception.

10.5 Research Gaps

As seen above, the field for State regulation of metaverses is vast, ranging from issues at macro levels (e.g., geopolitics) to micro levels (e.g., selling a digital bracelet in the metaverse). In a nutshell, the major current EU legislation and policies governing the digital sphere are as follows.

- Digital Markets Act : Regulation of competition for online markets. It establishes harmonised rules that define and prohibit unfair practices, such as the use of competitors' data and lack of interoperability, on the part of "gatekeepers" of the Web.
- Digital Services Act : Due diligence obligations on all digital services that connect consumers to goods, services, or content, including procedures for faster removal of illegal content as well as comprehensive protection for the fundamental rights of online users [46].
- GDPR : Protection of personal data. Due diligence and cybersecurity [49].
- Data Governance Regulation and Data Act: While the Data Governance Regulation creates the processes and structures to facilitate data, the Data Act clarifies who can create value from data and under which conditions. [48]
- Various in cybersecurity: Cybersecurity Act (eg certification), NIS2, ENISA, ECCC / NCCs , Joint Cyber Unit, Cyber Resilience Act, etc. [45]

However, from a governance and policy viewpoint such existing legislation are probably not sufficient to induce trust in the domain, and perhaps not even suitable for metaverses. Consequently, much research is needed in these areas in the near future. For instance, there will be a need to regulate security and privacy in multiple universes that are being built from scratch. Questions may be simple extensions of existing concerns, like whether metaverses should be subject to existing laws for the physical world and, if so, how not to hinder innovation and creativity. Or they may be turned much more towards future concepts, like whether avatars should be given citizen status.

Likewise, the technologies needed to build metaverses as envisioned here are just emerging, and a great deal of technological research will be required in the next few years. Moreover, one likely result of market forces is that several metaverses will be created, representing parallel universes, not only between them, but also to the physical one we are used to live in.

What is certain is that a new gold rush has already begun. Required research areas can be presented in clusters, as follows.

10.5.1 Building trustworthy metaverses

One governance research area should analyse all aspects within metaverses that would impact individual users. These encompass inter alia Data protection, Liability, Digital Identities, Cybersecurity at the user level, Mental and Physical Health, Accessibility, Inclusion, Financial transactions, and Consumer protection.

10.5.2 Metaverses and the physical world

Another governance research area should propose new societal systems for metaverses and their interrelation with existing forms of governance and government. These would include Cybersecurity at physical infrastructure and at systems levels, Privacy, Competition, Global governance, Jurisdiction, Civil rights, Penal code, Climate change, Innovation.

10.5.3 Compliance by design

The emergence of metaverses raise a wide range of concerns regarding their compatibility with the law, as seen above. Therefore, it will be necessary to go beyond the well-known concepts of security-by-design and privacy-by-design towards an encompassing compliance-by-design paradigm, if at all possible. For instance, research will be required about adapted technical regulations to guide hardware manufacturers and software developers with respect to compliance, including data governance and operational governance rules.

Such governance topics should be addressed together with research in the new technologies and systems integration that will be needed in order to achieve the full metaverse concept described above in this chapter. Some technological and systems research areas are as follows. Note that they are intrinsically transdisciplinary.

10.5.4 Interactivity and immersive technologies

Making the metaverse fully interactive and immersive is an evolutionary research area. it should be focused on the massive capture and fast analysis of data (telemetry, biometric, and neurometric tracking, among others) of

users and their avatars. Data will be collected through "wearable interfaces" (wearable devices) of different types that will gradually bring to metaverse XR platforms more and more sensitive personal information, which will need systemic protection.

10.5.5 Metaverses design

The area of research on the establishment of structured projects and design of digital virtual worlds in a metaverse environment now has great potential to study and establish a minimum necessary architecture. These can be platform infrastructures, usual protocol standards, security systems, or even the constructive and operational aspects of the application of XR in 3D. The establishment of a minimum standard should not make creativity unfeasible, but encourage the effective construction of interoperable metaverses with rules for social coexistence among avatars, which are acceptable in ethical and moral terms, universally, whether in digital or physical worlds.

10.5.6 Interoperability between metaverse platforms

Interoperability of metaverses needs to be intensified, so that it should be possible for avatars (users) who are experiencing a digital virtual world on a particular metaverse platform of a company, to be able to move, without impediments and in a transparent way, into another platform of metaverse, from another company, without the need to identify themselves again in the physical world. Research on interoperability in metaverse environments would touch upon digital identities and allow the establishment of a seamless collection of metaverses, maybe using the concept of self-sovereign digital identities and digital passports [248].

10.5.7 Metaverses and Environmental, Social, and Governance (ESG) issues

One of the key research points concerning metaverses relates to their impact on climate change, because of their need to rely on huge data centres, high performance computing, and even blockchain platforms, all of which necessitate very high electricity consumption. This area of research requires advances in architectures and algorithms, but also in other areas such as cooling techniques, that can enable the use of those technologies without major environmental impact. ESG considerations will play a major role in the provision and adoption of metaverses in future.

It's worth noticing that in the areas mentioned above, isolated and unconsolidated actions are already ongoing, which aim to cover the existing gaps in metaverse research. We can mention the actions of: the World Economic Forum [84], the Metaverses Standards Forum [150], the Open Metaverses In-

teroperability Group [175], and the Metaverses Interoperability Community Group at the W3C [229], among others.

10.6 Example problems

Tangible example problems include:

Data protection inside the metaverse. Personal data collected in the metaverse will be more granular, biometric, and neurometric. The question is then how to reconcile the fundamental need of metaverse immersion technologies to implement widespread user-profiling and the fundamental right to data protection, including bioethics. Note that such a question touches upon protecting the data from both the physical user and the digital avatar. More specifically, it should be investigated how to ensure that metaverses will not make illegal use of such data, for example for sales and monetisation (such as social networks already do), for promoting media influence, or in the effective production of subliminal advertisements, among other aspects of active and interactive persuasion.

Protecting avatars from identity theft. The protection of avatars' identity is a very important issue to be solved. Although there are already several proposals and strategies for applying security in databases, with the use of technologies such as distributed ledgers and scatter or hash tree structures, such as Merkle Trees (which are, by the way, key elements of Blockchain), there is still no consensus on how to keep avatars' digital identities without compromising their Lifelogging (metaverses life history).

Regulation of creation of metaverses. The technologies currently applied by many Web platforms already provide easy-to-use tools that allow users to create their own metaverses. Even if these are simple, they are totally under the users' control. The problem here is centred on the improper creation of metaverses that camouflage digital worlds meant to harbour avatar gangs for criminal practices, social activism, racism, and terrorism, among other unethical and illegal practices.

Equal opportunities in the metaverse. Ensure accessibility and inclusion in the metaverse in order to safeguard equal opportunities. The Web platforms that host metaverses will be able to segregate avatars based on their physical users' hardware characteristics, computing capacity, personal profile, or according to the geographic region of their access, giving more privileges to some than others.

Cryptocurrencies and NFTs usage in the metaverse. Issues of ownership, misuse, interoperability and portability. As the Web platforms are proprietary, they maintain control over the digital assets owned by avatars, as well as, determine the monetary standards used. Some platforms have their own internal cryptocurrencies, a fact that can jeopardise the portability and interoperability of avatars' digital assets between platforms.

11 Malware

11.1 Introduction

Modern malware comes in different forms: viruses, worms, spyware, adware, trojans, backdoors, and ransomware, to name a few. Although, computer viruses were the most frequent form of malware a couple of decades ago, nowadays, it is ransomware that seems to be the most prevalent. This is because ransomware provides a highly profitable and direct way for malicious actors to monetise the infected systems. Indeed, using ransomware these actors infect victim systems, encrypt all data, and then ask for money (ransom) in order to provide the decryption key. Without the decryption key, the legitimate owners of the victim systems can not really use them as all information is encrypted.

To defend against malware, computer security practitioners usually need a way to *detect* it in the first place. Detecting a file containing malware used to be easy: computer security companies computed a hash (a summary) of the malicious file and just tried to find files that matched this hash value. Antivirus systems used to be nothing more than a set of hash values (one hash value for each piece of malware) and just searched for files that matched any of these hash values. To avoid this type of (static) detection, modern malware mutates so that two "copies" of the same malware are not the same. For example, in each "copy" of the malware they introduce small changes that, while not changing the main functionality of the software, do change its appearance, and consequently its hash value. Following this phi-

losophy, malware authors obfuscate their code to deter or at least impede the reverse engineering of their binaries but also to remove possible code patterns that could be used to detect the malware. This may come in the form of packers, programs that try to compress and/or encrypt the code of the malware so that the malicious code is unpacked and executed after several steps that would make the life of a malware author difficult. Finally, modern malware is armoured in the sense that it has anti-analysis functionalities such as anti-debugging, anti-hooking, and anti-VM to name a few.

Setting aside the differences of scope that malware may have, e.g. worms, trojans, miners, it is important to highlight some differences in the sophistication and range of targets. Practically, highly sophisticated malware, from Pegasus to Stuxnet, is mostly focused on attacking a specific individual or crafted for a single organisation. In this case, a chain of exploits is used, many of which may be zero-days yet the attacker is not financially motivated. However, the sophistication of the malware significantly decreases when the attack is targeted at general information systems. This sophistication is what can make malware stay below the radar and increase its impact on its victims.

11.2 Who Is Going to Be Affected?

Common practice proves that malware can infect almost any computing device. Indeed, in the past few years we have witnessed a long list of high-profile organisations being compromised: e.g. the Colonial Pipeline [236], Uber [52], AXA [194] to name a few, however, the ransomware cases are the ones that surface in the news mainly because of the monetisation method that the ransomware groups adopt. In essence, beyond encrypting the data and asking for ransom in return for providing the decryption key, the attackers also exfiltrate the data so that they can still threaten the victim with publication of the sensitive data. Malware may also try to exfiltrate sensitive user information via keyloggers, compromised recording media (e.g., cameras and microphones), etc.

Attackers may also use malware to infect thousands of hosts and use them as an army that obediently carries out all the tasks that it is assigned. The network is called a botnet and may also be used for denial-of-service attacks. One striking example that stands out in this category, not because of its size but because of the devices that comprised it, is Mirai. Mirai [14] is a botnet that mainly infects insecure IoT devices and uses them to perform denial of service attacks. The size of the produced bandwidth targeted a popular DNS provider DYN and as a result, high-profile websites and services such as GitHub, Twitter, and Netflix were inaccessible [34, 251]. Beyond Mirai, there are several botnets which are currently active, e.g. Emotet [185] which

was resurrected [117] after its shutdown [80] after compromising a host used to deliver several other malware such as Trickbot and Ryuk, Mozi [20], and Mantis [252]. Notably, they have also been used by state actors to launch attacks for cyber warfare [237].

11.3 What Is Expected to Happen?

The impact of malware is multifaceted. There are several direct costs that can be relatively easily quantified, such as the amount of ransom requested. However, there are also costs that are more difficult to quantify, such as lost customers, lost productivity, etc. According to Sophos, the average cost to recover from a ransomware attack is on the order of $810,000 for organisations that did not pay ransom and double that for organisations that did pay [218]. These costs cover all the operational costs and downtime costs caused by the ransomware. In fact, the damages caused by ransomware are estimated to reach the staggering amount of $265 billion by 2031 [29].

Even organisations that are not IT-oriented may suffer from malware. Consider, for example, the case of a hotel. While its core business is not delivering IT products and services, hotels that have suffered a ransomware attack [160] have witnessed their guests being locked out of their rooms, and their billing, reservations, check-in/out systems rendered useless effectively blocking any possible business transaction. Similarly, several health organisations have suffered malware attacks, and we have reached a point where it is just a matter of time until there are casualties [176].

Based on the above, there are obvious monetary and reputation loses for organisations and individuals whose systems are compromised by malware. Consider that for individuals, other mechanisms such as sextortion may be used to harm the victim further on the personal level [183].

11.4 What Is the Worst That Can Happen?

As highlighted in the previous paragraph, we are on the verge of having casualties due to malware attacks. However, this is not the only nefarious scenario. Stuxnet [137] was a worm targeted to disrupt Iran's nuclear program. While this may be more than a decade ago, considering the current turmoil in the political landscape, malware attacks are expected to be further utilised as a means to attack a country's digital infrastructure. In this regard, malware attacks to cripple smart cities, critical infrastructures or big service providers are expected to increase, as proved by the recent cyber attacks on HSE [191], the Colonial Pipeline, and the Danish train operator [195]. Unfortunately, this is aligned with the modus operandi of several advanced persistent

threat (APT) groups which are not necessarily financially motivated but are state actors or state-supported. In fact, as recently reported by ENISA, APT groups were responsible for more than half of the supply chain attacks that were investigated [75]. Indeed, this leads to several unprecedented attacks, e.g. the recent Sandworm attack which targeted a Ukrainian agricultural firm's network to disrupt grain production and exports [155]. The presence of APT groups in conjunction with the shifts to IoT and remote working is significantly increasing the potential impact of a cyber-attack. Indeed, using the Zmap network scanner [66], one can easily see that millions of vulnerable devices are connected to the Internet. To make matters worse, a quick search in search engines, such as Shodan(`https://shodan.io/`) and Censys (`https://censys.io/`), reveals similar results. Such techniques have been used by, e.g. APT41 to target U.S. State Governments [30],

All this creates a dangerous mix where strategically motivated threat actors have access to a myriad of vulnerable devices that may access, directly or indirectly, systems that store, exchange and process sensitive and/or critical information. Therefore, in the coming years, cyber warfare as an extension of geopolitical turbulence, and the resulting use of malware, is going to lead to large-scale cyber-attacks on critical infrastructures significantly impacting sectors such as banking, energy, telecommunications to name but a few, or even organizations in the defence industry [92]. The latter implies that we may face unprecedented attacks that may paralyse mission-critical systems and services, and impact organisations, individuals, and the social fabric in both the cyber and the physical layer.

11.5 Research Gaps

11.5.1 Provably secure systems

As discussed, malware often exploits system vulnerabilities. Therefore, an obvious question is how do we build systems free from any vulnerabilities that malware can exploit? While this line of research might be too broad, there is still plenty of security to be had from impervious containers or sandboxes. For example, while some malware may compromise the underlying operating system or firmware the research question is whether we can build microkernel and sandboxing architectures that are provably secure. Of course, this would still leave us vulnerable to malware that compromises application-level software, but containing the adversary in a sandbox would allow us to keep core system functionality secure and also maintain separation between different applications and services running in different sandboxes. seL4 (`https:`

//sel4.systems/) as microkernel and Qubes (https://www.qubes-os.org/) as OS can be considered well-known examples in this direction.

11.5.2 Malware detection

Currently there is an ongoing arms race between malware authors and the "defenders", whether they are malware analysts, digital forensics investigators, SOCs, CERTs, CSIRTs etc. As already discussed, modern malware is armoured to prevent analysis and to be more stealthy. Therefore, malware detection is still a core issue in this research field. Although modern antivirus (AV) software may be far more accurate than in the past, it is not enough to prevent the infection of millions of devices, primarily because AVs are focused on static features. A new stream of anti-malware mechanisms, namely endpoint detection and response systems has emerged during the past few years. These, along with their variants, e.g. extended detection and response systems (XDR), try to exploit behavioural features and AI/ML mechanisms to detect and block malware attacks. While more efficient than AVs as they can detect advanced techniques and lateral movement, EDRs are far from being considered silver bullets [126]. To this end, a critical research question is how to determine that a file is malicious at runtime, and block it once it performs a malicious action without allocating a lot of resources.

This research question also has many more extensions. For instance, when analysing malware we often execute it in sandboxes to record and understand its capabilities in a highly monitored environment. This sandboxed-based execution has two main disadvantages: (i) it consumes a lot of resources and (ii) if the malware realises that it is being executed in a sandbox, it may alter its behaviour to avoid being detected. Therefore, the research here lies in how to perform dynamic malware analysis without wasting precious resources and how this can be performed against evasive malware [132, 133]. Moreover, we need to find methods to automatically trigger the malware appropriately without creating long execution paths and unlock its functionality. To this end, binary emulation and symbolic execution may come to the rescue. Finally, we have to highlight that many system calls performed by malware, if treated individually, do not always differ much from those issued by benign programs; thus, evasive malware can still bypass many classifiers that cannot see the whole picture [171].

11.5.3 Machine learning in malware detection and classification

The continuous use of machine learning and artificial intelligence in cyber security has also paved the way for its application in malware detection and analysis. Nevertheless, we have to consider that it can also be leveraged by malware authors to bypass the detection mechanisms. Hence, it is essential to

consider that malware authors will try to exploit feature selection algorithms to make their malware undetectable by some classifiers. As a result, machine learning and artificial intelligence cannot simply be used and expected to provide excellent results. First, we have to devote major research efforts in order to understand how to fill in the gap in imbalanced datasets where a malware family may be underrepresented. Next, we have to study adversarial machine learning and how to make our mechanisms robust against possible feature injection or blinding [255]. One should also consider the explainability and interpretability of the results of machine learning and artificial intelligence algorithms and how feature engineering can impact them as malware samples may have thousands of sparsely distributed features. Finally, one should also consider the relevance of the datasets and models over time. Using older datasets and models that might be state of the art now may soon be outperformed or perform poorly due to the evolution of both malware and ICT systems.

11.5.4 Extend the platform scope

While most users of personal computers are using Windows and represent one of the biggest targets of malware attacks, they are not the only ones. Similarly, in mobile devices Android may have the biggest share in smart phones, but it is not the only platform for mobile devices. Moreover, we all know that a significant part of the Internet is not running on only these two platforms and that malware has been developed for, e.g., IoT devices, Linux-based hosts, and MacOS, among others, focusing research only on Windows and Android creates a huge gap that is exploited by threat actors who find many of the less-focused platforms unprepared. For instance, the bulk of research focuses on PE32 files, overlooking e.g. ELF files that target Unix and Linux hosts. Even when researchers try to study ELF files the datasets are highly unbalanced as most samples come from a single family, i.e., Mirai, which may severely bias the outcomes. Therefore, there is a definite need to extend the scope of platforms and architectures that are used in malware analysis research and to develop new methods and tools.

11.5.5 Command and control servers

Finally, a rather thorny issue is the gradual integration of decentralised mechanisms by malware to control the botnet but also to deliver payloads. For example, blockchains and decentralised storage (e.g. IPFS) have been proven to be a very robust mechanisms to act as Command and Control servers but also to host payloads [12,184,188]. The crucial point here is that most of these decentralised mechanisms are not regulated (indeed some of them cannot be) and takedown mechanisms may not be possible, for example, once something

is committed in bitcoin's blockchain, it cannot be erased. Thus, there is a lot of research on how to protect against such malware and how to minimise the exploitation of such ecosystems.

11.5.6 Post-infection management

Acknowledging that there is no 100% accuracy in malware detection and prevention means that in practice a given system will be infected with malware at least once. Naturally, one may wonder what should be the next steps when malware is detected. Most of the existing antimalware solutions try to thwart attempts at infection and clean up. Perhaps there are other things we can do post-infection to minimise harm or facilitate digital forensics. This could involve automatically rolling the system back to a state just before infection. While for storage the option of incremental file systems may provide a solution, the same does not apply for memory.

11.6 Example problems

Tangible example problems might include:

Command and Control (C2) servers and defence mechanisms To manage the compromised hosts, many malware authors use C2 servers, some of which are commercial, e.g. Cobalt Strike, whose copies have been leaked but are legitimately used in red team scenarios. Regardless of their origin, C2 servers allow threat actors to coordinate the actions of their bots, issue commands, exfiltrate data and perform other attacks. Currently, there are many C2 servers, many of which are open source, and it would be interesting to study how different security mechanisms, e.g., AVs, EDRs, firewalls, treat these beacons and whether they are detected as malicious. Malicious patterns in memory and system calls can be leveraged through memory scanners and hooking to promptly block their functionalities.

Malware classifiers The sheer amount of malware samples on a daily basis imposes many constraints on resources and timing. Binary classification (benign and malware) is a traditional problem in the field. Going a step further, family classification and clustering are very important. Regardless of whether these analyses are performed based on binary similarity measures, static or dynamic features, it is crucial to determine their accuracy and robustness, especially in an adversarial scenario where the threat actors may want to bypass security mechanisms but if this fails, raise a false flag [25].

Anti-evasion mechanisms and triggering mechanisms Malware may try to evade detection and analysis in various ways. Automating the bypassing of such mechanisms and collecting robust results from malware through the correlation of static and dynamic features is a big challenge. How do we trigger the malware properly to exhibit its behaviour when static analysis indicates that a file is malicious yet the dynamic analysis fails to detect the maliciousness of the file in question?

Covert communication channels and malware Many malware instances would try to hide their communication channels by mixing their interventions with legitimate traffic, e.g. using a social network or another legitimate service to communicate between the C2 server and the compromised host. However, malware may use steganography and other covert channels to exfiltrate data or to disseminate commands. Detecting possible malicious covert communication and stegomalware is a challenging problem.

Abuse of legitimate processes Living Off The Land Binaries and Scripts (and also Libraries)[1], commonly referred to as LOLBin/Script/Lib are files that are shipped from Microsoft in Windows and other tools (e.g. Office, Visual Studio), which bear the signature of Microsoft and can execute additional functionalities to those with which they were initially designed, e.g. download files, execute arbitrary content, etc. Because of their signature, when executed, they do not request any user interaction, are whitelisted by most security mechanisms, and can be found in almost all Windows machines.

Threat actors have repeatedly used these files in malicious campaigns to trojanise Microsoft Office documents to execute malicious payloads. This approach has gradually been abused by other malware, especially fileless malware attacks [230]. Moreover, they are abused by ransomware to delete shadow copies, e.g. `cmd` to launch `vssadmin` and delete the shadow copies [136].

Based on the above, the research problem lies into finding ways, based on, for example API call context, process parents and children, and call arguments to determine whether a call to a legitimate process, API, library, or binary is being abused by malware or whether it is in fact a benign call.

[1] `https://lolbas-project.github.io/`

12 Software Life Cycle

12.1 Introduction

Software is at the foundation of all digital technologies, thus it is at the core of the infrastructures, services and products that drive our societies. The life cycle of software consists of several phases, starting from conception, and going through design, realisation, deployment, operation, maintenance and, eventually, decommissioning. Current software development approaches prioritise fast deployment over security, which often results in insecure, expensive to repair [202], applications. Security concerns are, unfortunately, still not fully and suitably integrated within the life cycle of today's increasingly complex software systems [241]. Moreover, software is usually built by assembling components from third-party sources, which raises trust concerns (e.g. as evidenced in supply chain attacks [72]), makes it hard to comply with security requirements and legislation, and compromises digital sovereignty. The raise of artificially synthesised software is expected to aggravate this. Last, security and privacy regulations such as the GDPR [93] or the Cybersecurity Act [7], as well as citizen expectations change frequently and software is subject to continuous update. As a consequence, software compliance cannot be assessed once and for all and needs to be an inherent part of its life cycle [135]. The field has seen advances since early initiatives to build security in software systems [149] and efforts in this direction have been made, such as NIST's Secure Software Development Framework [167], OWASP's Software Assurance Maturity Model [201], Microsoft's SDL [153], and ETSI's standard 303 645 [78] (see also Chapter 12 of the Cybersecurity Body of Knowledge [56]). Nonetheless, many challenges remain (see the last part of this chapter and also [135], as examples).

12.2 Who Is Going to Be Affected?

Traditionally, ensuring high-quality software was considered to be mainly relevant for critical infrastructures: finance, healthcare, energy, and so on. However, software is becoming more pervasive and intrinsic, up to the point that it can be seen as the circulatory system of our society's body: you may not

notice that it is there, until its quality starts to affect your health. Nowadays, we use software to regulate the indoor climate of our houses, to plan our commute to work and schools, to carry out our daily activities, to communicate with colleagues, family, and friends, to access medical services and treatments, and so on. Ultimately the quality of life of every single citizen will be highly dependent on the quality of the life of the software facilitating his or her activities.

12.3 What Is Expected to Happen?

Software vulnerabilities in critical sectors can have catastrophic consequences for our lives: companies and individuals can lose money because of flaws in financial software, access to treatments can be delayed by malfunction in software platforms used in hospitals, lives can be lost as a result of software bugs in medical devices or car assistance systems. Chapter 13 provides a representative sample of (in)famous cases such as the Ariane 5 disaster, and the loss of the Mars climate orbiter, and many others can be added. Just to mention a recent example, a vulnerability in the Poly Network smart contract lead to the loss of 600M USD [89]. But even vulnerabilities in cases that we traditionally do not consider as critical can have severe consequences for individual citizens: violation of personal privacy is arguably the most archetypal example.

12.4 What Is the Worst That Can Happen?

Software vulnerabilities can have all sorts of catastrophic consequences, and certainly need to be addressed. However, ensuring quality of software is the least thing we can do. Software can also be of high-quality and adhere to the most strict security and privacy regulations such as the highest levels of the Common Criteria [36], but harm can still be obtained if there is lack of trustworthiness in the way it is developed, acquired, used, maintained and dismantled. Consider for example, what can happen if citizens do not trust the software being used in the next democratic elections. Even the entire democratic system of a country can be at risk.

12.5 Research Gaps

Security must be better integrated in the entire life cycle of software, from conception to dismantlement. We consider the following gaps and possible ways forward to address this.[1]

[1] Inspired partly by the VERSEN Manifesto [241]

12.5.1 Verifiable and Auditable Software

A great portion of the software components that constitute a software product or service is obtained from third parties; thus it is potentially untrustworthty as it may not comply with the expected security requirements. To achieve digital sovereignty, there is a need to be able to rely on software that can be verified and audited. The potential security gain of using open-source software amenable to automated analysis should be further explored.

12.5.2 Continuous Software Assessment

Security and privacy regulations and citizen expectations change frequently and software is subject to continuous update. Therefore the compliance of software systems cannot be assessed once and for all, and hence methods and tooling to perform continuous assessments are needed. Given the high cost of security and software assessments, the use of automated procedures is critical to ensure sustainability and scalability. If this is not implemented effectively, the software becomes too complex, and maintenance and evolution become too expensive, until they are no longer sustainable. We must break this vicious cycle, and find new ways to create software that is long-lasting and that can be cost-efficiently upgraded, assessed and migrated to new technologies.

12.5.3 Secure-by-design Agile Software Development

The dominating approaches to development are agile and prioritise fast deployment over security guarantees. More research is needed to effectively and efficiently develop tools and techniques to support secure-by-design techniques within agile approaches, so that competitiveness and fast deployment are not compromised by security requirements and so that changes in those requirements can be efficiently reassessed at any point, even while the software is running.

12.5.4 Lightweight Formal Methods

Many formal methods technologies have been developed to improve software reliability, such as model checking, theorem proving, and monitoring systems, but applying them on a large scale to modern software systems remains a challenge. More efforts are needed to further develop and promote lightweight, accesible formal methods that can be gradually applied to increase the levels of assurances obtained. Methods must be developed to support a spectrum of guarantee levels, each providing greater assurance, in a way more approachable than the common criteria. Eventual enforcements in regulations must be gradual in order not to close the opportunity for SMEs to deliver software products and services, and appropriate tool support is need.

12.5.5 Decentralised Software Governance

Software with decentralised governance such as smart contracts, blockchain technologies, and crypto-assets, pose several challenges to the management of the software life cycle. In those systems, it is unclear whether and how vulnerabilities should be reported and repaired in a way that harmonises consensus and security across the history of the system.

12.5.6 Trustworthy AI-powered Software Life Cycle

Artificial intelligence techniques are already being used to synthesise small pieces of code. One should expect that in the near future all activities of the life cycle of software (requirements elicitation, code synthesis, verification, monitoring, etc.) will be supported by intelligent agents. While this will certainly bring huge advances in terms of scalability and productivity, it is still unclear how software components and methodologies with intelligent components can be rigorously analysed.

12.5.7 Software Supply Chain Security

Nowadays, creation and deployment of software involes the integration of code and components from third parties, whose development is outside our control. These components can be the target of cyberattacks (e.g. the SolarWind incident). We need to define a methodology for reducing supply chain security risks, by means of assessing and guaranteeing the trustworthiness of components. This methodology must be based on formal models of contract-based software line development and integration, in order to enable the implementation of (semi)automatic tools for the verification of security properties. The development of these models and the corresponding formal methods are an important research priority.

12.5.8 Secure Architectures and Platforms

For building safety- and security-critical systems, it is not enough to have a trusted software supply chain: we need to deploy this software on trusted platforms. This includes the hardware level, but in particular the operating system level. Therefore, an important research priority is to develop a verified platform that provides fine-grained access control through capabilities, and controls communication between components of the system. This kind of platforms is highly sought in operational scenarios (e.g. the SCADA of critical infrastructures) but also in datacentres that provide cloud services. This would help to recover data sovereignity in the EU.

12.5.9 Secure Economics

Another interesting research direction is related to security economics, i.e., the study of the incentives facing different players [71]. It is now well established that purely technological solutions will not fit the bill. Accordingly, each alternative mechanisms must be scrutinised against market dynamics.

12.6 Example problems

Tangible example problems might include:

Verification at the scale of public open source code repositories. Formal verification techniques offer the highest level of assurance for software security. The main challenges of current techniques are, arguably, due to scalability issues in terms of the computational and human expertise needed. How can we raise successful verification techniques to the scale of code bases of the size of average popular public code repositories?

Formal methods-powered DevSecOps DevSecOps has been advocated as an ideal approach to combine DevOps and security, in order to provide a security-aware agile and fast-adapting continuous life cycle. On the other hand, formal methods, which provide the highest possible level of assurance in terms of security, safety and performance, have been traditionally conceived in a water-fall mind-set, rooted on formal specifications as the first step. Can we develop agile formal method methodologies in what could be called formal DevSecOps?

Formal Analysis of Socio-Technical and Cyber-Physical Software Systems. Socio-technical systems, whose security depends intrinsically on human users, and cyber-physical systems, where one needs to explicitly consider the underlying physical processes pose several challenges to formal automated modelling, analysis and testing. Can we develop effective and scalable formal and automated tools for the analysis and testing of such systems?

Verification of ML applications. Probabilistic and randomised software components are at the core of many software applications, from cryptography to machine learning (ML), to privacy protection. Recent years have seen advances in probabilistic programming techniques and verification techniques for ML. However, the field is still in its infancy, while, on the other hand, the application of ML has been advancing swiftly. How can we extend probabilistic programming to cope with real-world ML-based applications?

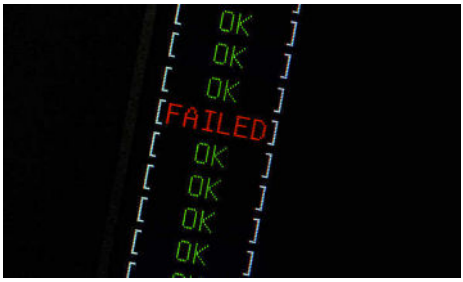
Resilient Smart Contract Repair If a security vulnerability is discovered in a smart contract, reporting it -or trying to repair it- could trigger a race for

its exploitation that is likely to end up with financial gain for malicious agents. Can we design disclosure and repair techniques that a resilient w.r.t. malicious agents trying take profit?

Secure and Privacy-friendly Explainability. Explainable security extends explainable AI with the need to consider security and privacy aspects of the explanation process and of the explanations themselves. How can we adapt approaches to explainability to take into account security and privacy considerations?

13 Testing and Certification

13.1 Introduction



Information technology (IT) is pervasive in both work and social sectors. Home, industries, offices, cars, streets and public buildings are full of IT devices, systems apps, or electronic equipment. In our daily lives, under normal conditions, we are usually not worried about the technology around us. We are reasonably sure that our mobile phone, PC, refrigerator, electronic device, car, or even apps, cannot damage our life, steal data, or cause security or safety issues, because they should have been built according to the required standards, properly tested and fully certified.

However, we have recently witnessed various examples of malfunctioning or issues like the following: Tesla had a failure in a flash memory device, causing a safety risk in more than 135,000 vehicles [163]; the New Jersey hospital vaccine scheduling system bug caused 10 to 11 thousand duplicate appointments [65]; the Zoom app suffered from security issues during the coronavirus pandemic in 2020 [1].

As reported in the recent Cybersecurity act, "Hardware and software products are increasingly subject to successful cyberattacks, leading to an estimated global annual cost of cybercrime of €5.5 trillion by 2021" [79]

Humans and society generally trust industries and the best practices they adopt in testing and certification processes. However, considering that the overall cost of testing is around 40% of the total development costs of a typical software project [91], if not stringent and without concrete safety risks, often verification, validation and assessment procedures are the first to be reduced or skipped to save cost and time. Additionally, pressure from the need to research new products, the time to market, and competition forces in-

dustries and developers towards massive widespread integration and the use of available third-party or open-source components that could surreptitiously increase the cybersecurity risks if not properly tested and certified.

In an IT world that is going to be more human-centric and focused on people's needs (such as the Internet of People [IoP] manifesto [157]), the presence of evidence of the testing and certification activity performed needs to become a common practice. We need to increase our awareness to avoid "poisoned" IT products as well as poisoned food. Therefore, the assessed or certified quality level must be a label for each IT product in order to establish trust and reduce risks to security and privacy.

The quality of digital products (combination of software and hardware) must become a guarantee label, in the same way as the label we find on the food we buy in supermarkets.

13.2 Who Is Going to Be Affected?

Everyone directly or indirectly using products or technologies can be affected by the lack of testing and certification processes. For instance, babies could be damaged by a toy going out of control, Generation Alpha or Zeta could be unconsciously deceived by appealing apps maliciously stealing their pictures, companies can be affected by ransomware hidden in useful plug-ins or libraries, organisations and governments can be subjected to cybersecurity attacks. Of course, testing and certification are not the only means of avoiding such critical situations. Everything has to be executed correctly at every phase of the development process (see Chapter 12 for details). Conceiving and developing (by-design) quality products is crucial, but not sufficient per se to meet the final requirements: building the product right does not guarantee building the right product [217]. Testing and certification remain pivotal activities for trustworthiness and cybersecurity assurance and for guaranteeing that a product is designed and manufactured with quality as a primary objective.

However, as long as stakeholders (ordinary people, companies, organisations, and governments) do not firmly demand transparent, labelled, tested and certified products, the situation will hardly change and cybersecurity risks will still be on the agenda.

13.3 What Is Expected to Happen?

What is the expected damage in the absence of an adequate testing and certification process? Unfortunately, there are many aspects to be considered:

Hardware/software failure: It has been estimated that nearly 80% of unexpected downtime can be ascribed to HW/SW failures and power outages. Proper storage backups can be an ad hoc solution in most cases, but preventing failure would be less costly and risky.

Natural disasters and emergency situations: Lack of testing and certification of the processes and procedures for resuming operations/data and systems in case of (natural) disaster or emergency situations can be extremely costly and cause the loss of business continuity.

Human factor: Even not intentionally, humans may inevitably cause mistakes or execution of unexpected procedures. Testing based on user profiles or exploiting machine learning approaches could avoid or predict possible misbehaviour or accidental situations. User-centred assessment processes and training programmes could be essential for minimising human damage and avoiding permanent losses.

Cybersecurity attack: Because society and organisations increasingly rely on digital information for daily operations, cybersecurity attacks can be more dangerous. Currently, 95% of companies invest in testing and certification activities only after a disaster and then actuate a recovery plan (reactive behaviour). Predicting vulnerabilities beforehand and providing solutions before a cybersecurity attack is, therefore, mandatory (proactive behaviour). The penetration test is pivotal for avoiding and anticipating cyberattacks by hackers who are trying to exploit potential vulnerabilities in order to access company networks and to steal confidential data or to inject malicious codes.

High expectations: : In our hyper-connected world, where IT products need to be available 24h7d without disruptions, failures and loss of services are costly disasters for companies and favour their competitors. Therefore, robust testing and certification processes, which can assure the quality of services and make it possible to establish a suitable recovery plan, are pivotal activities.

Trust or reputation damage: Loss of trust or damage to a reputation is mostly translated into a loss of customers, and hence a loss of revenue: trust and reputation are nearly impossible to regain. Testing and certification are among the most effective means of avoiding this problem.

Compliance requirements: Nowadays, business continuity is not just a mere desire: it is becoming a requirement, especially for Operators of Essential Services (OESs) [47]. All of them must follow specific and strict regulations and standards. That means that adopting certification processes and maintaining their product certification is becoming a legal

obligation and offers a competitive advantage within the reference market.

13.4 What Is the Worst That Can Happen?

Figuring out what could happen without testing and certification should not point to the future but simply to the past. Most worst-case scenarios have already been covered in the newspapers, the default reports and disaster documentation. The worst-case bugs history started as soon as the first computer was massively used and included:

- **The Ariane 5 Disaster**, 4th June, 1996. During the launch of the Ariane 5 spacecraft, 37 seconds after the first rocket ignited it started flipping in the wrong direction, and less than two seconds later the whole world observed its self-destruction. The problem was quickly identified as a software bug in the rocket's inertial reference system and, unfortunately, could have been easily solved with a trivial integration testing procedure [247].
- **The Mars Climate Orbiter**, 23rd September, 1999. During its descent into the Martian atmosphere, the Mars Climate Orbiter was reoriented to pass behind Mars and successfully enter its orbit. Unfortunately, this did not happen: the craft was not on the correct trajectory and it was finally lost without a trace. The root cause analysis of this error yielded a long chain of wrong or unexpected events, which included: the incidental arrangement of solar panels on the craft due to the solar sail effect; the use of two different units in the Ground Control software (data provided using imperial units and pound-seconds on the sender side but expected in metric units on the receiver side); and finally, human errors in communications. Again, proper integration testing procedures and correct use of standards and assessment procedures would have avoided such a critical disaster [105].
- **Therac-25** During the period from 1992 to 1998, the reports about radiation overdoses caused by the 80's computer-controlled radiation therapy were published. In particular, six documented accidents occurred, resulting in deaths or severe injuries. The causes were identified as the application of incorrect procedures by personnel and the weaknesses of the software used for assuring safety. In particular, all accidents involving software had resulted from flawed software requirements. Application of certification processes and a proper system and acceptance testing process would have again avoided significant loss of life [140].
- **Knight Capital Group** On 1st August 2012, during a software update of the production server, an incorrect configuration of an old (2003) system

> caused 97 email notifications and the execution of 4 million unexpected trades. That led to a $460 million loss and the risk of bankruptcy. The post-analysis highlighted that the program believed it was in a test environment and executed trades as quickly as possible without worrying about losing the spread value. As in the previous cases, the testing process would have discovered that misbehaviour and avoided using obsoleted, not aligned software [189].

It is likely that past mistakes have been resolved and lessons learnt, but challenges, vulnerabilities and new scenarios are constantly emerging. Who does not remember the Millennium bug [4]? Or the 2018 cyberattack that interrupted communications on the Midcontinent Independent System Operator? Or even the six/seven hours of the global unavailability of the social network Facebook and its subsidiaries in October 2021 [228]? Or the recent ransomware attacks on the IT network?

The smart and quick discovery and provision of new technologies, programming languages and systems obliges testing and certification to continuously jump **"Back to the Future"** and provide new means, strategies and processes to prevent future worst-case scenarios. Indeed, history teaches that the past can always turn into the future and *vice-versa*.

What Is the Worst That Can Happen? A life without testing and certification, because it means a lack of quality, efficiency and trust in every system and software package.

Indeed, testing and certification seek to mitigate the risks of safety, security and privacy loss or absence for anyone worldwide. Who would use a machine without it being tested? Who would be willing to set up a medical facility without being certified? Who could think to give a child toys that put their life at risk?

Unsafe, not secure or not trustable HW or SW products, elements, components, and libraries make the world dangerous: they can cause environmental disasters; they can play a role in the default or bankruptcy of companies, industries and even nations; they can impact essential services (i.e. energy, transport, financial and banking, healthcare, drinking water supply & distribution, and digital infrastructures); they can compromise health systems or medical devices. The current international situation can also paint even more dramatic scenarios: HW/SW vulnerabilities and security threats could be exploited to allow terrorist attacks on nuclear power plants and military bases.

Luckily, in this catastrophic apocalyptic scenario, learning from the past and focusing on the future, research and industry are starting to understand the importance of strict collaboration in testing and certification to effectively prevent disasters before they happen.

13.5 Research Gaps

Considering that "Program testing can be used to show the presence of bugs, but never to show their absence. (Dijkstra)" [61], exhaustive testing is usually impossible, and issues and problems in testing and certification are far from being exhausted. New challenges are continuously added in parallel with the development of new technologies, features, languages and application domains, and the discovery of new vulnerabilities and threats. In particular, the following areas are recent trends in research activities.

13.5.1 Human-centred Testing and Certification

Supporting human-centred testing and certification approaches that are able to guide, improve and assess technological development in line with social and ethical values, sustainability and trustworthiness. Additionally, increasing inclusiveness by supporting the gender and diversity balance of different stakeholders involved in the testing and certification approach can ensure trustworthy public awareness, the broad adoption of IT methods, and the adoption of standards to increase transparency and openness.

13.5.2 Integrated cybersecurity and functional safety certification

Besides interleaving and overlapping several aspects of cybersecurity and safety, there is still a gap in providing a comprehensive framework and technical standards for their full integration. Indeed, safety assurance/certification cannot be achieved without considering the impact of cybersecurity vulnerabilities and threats on the system. Thus, there is a need to provide a functional safety/cybersecurity assurance risk-based integrated approach.

13.5.3 Quantitative and qualitative testing and certification

Accountability and replicability are essential characteristics of cybersecurity modelling, testing and certification approaches, and require methods and means for quantitative and qualitative collection and the analysis of results and data. Thus, the availability of open-source data sets and conformance test suites as the facilities for the setting up and execution of controlled experiments should be improved. In particular, challenges focus on: (1) Improving formal methods for quantitative security modelling and analysis and their application to risk management, enriching their data-driven aspects, e.g. synthesising and refining models from (possibly underspecified) attack scenarios and validating them concerning data from previous attacks. (2) Realisation of modelling, testing, and certification approaches driven by cybersecurity risks (3) Makingdata collection, quantification approaches/tools, and result analysis more accessible to practitioners and open-access communities. (4) Improving the efficacy and efficiency of the testing and certification processes,

making them more focused on qualitative properties. (5) Making testing and certification by design, guided by user stories, domain-specific needs requirements, and standards. (6) Providing metrics, guidelines, and approaches for securing products and services throughout their lifetime.

13.5.4 Automation of Testing and Certification

Testing and certification are complex, costly and time-consuming activities. Reducing the effort and mitigating the cybersecurity cost and risk is a significant challenge for attainable automation. Important directions are:

1. Developing advanced techniques, finding innovative support procedures to (fully) automate the different activities, or providing metrics, guidelines and approaches applicable throughout the overall process lifetime

2. Providing a holistic methodology that integrates runtime and design-time methods applicable at different specification levels—such as firmware, communication protocols, stacks, operating systems (OSs), and application programming interfaces (APIs)—and that considers the integration of software and hardware.

3. Specifying and developing manageable and human-centric KPIs, metrics, procedures, and tools for dynamic and automatic cybersecurity certification from chip to software and service levels.

13.5.5 Diversity, heterogeneity and flexibility of environments

Diversity, heterogeneity, and flexibility are challenging attributes of testing and certification proposals. In particular, any approaches and solutions provided should move according to vertical and horizontal research levels. Indeed, ecosystems and systems of systems (SoS) rely on the continuous integration of components, apps and devices developed using different languages and operating systems, and on combining and accessing thousands of device-browser-platform combinations simultaneously. To avoid the risk of becoming outdated, testing and certification need highly flexible and modular schemes that rapidly adapt to the changes and updates of the technological environment and elements at each horizontal or vertical level. Additionally, to follow the rapid and pervasive evolution of the different supply chain environments (such as the critical infrastructures described in Chapter 9), and new technologies (like the metaverses described in Chapter 10), holistic, modular proposals are necessary, able to effectively and efficiently validate, verify and certify the different HW/SW elements under real user conditions and considering other interacting systems and application domains.

13.5.6 Including legal aspects inside testing and certification

The interplay between HW and SW elements in current systems promotes a new direction for cybersecurity testing and certification research: to include legal aspects in the verification, validation and assessment procedures. The legal framework and technical standards must be considered necessary parameters during the development life cycle (for more details refer to Chapter 12).Indeed, cybersecurity vulnerabilities may cause legal violations, especially in sensitive applications such as healthcare. The future direction is to ensure that cybersecurity, safety and legal requirements are tested and certified as inseparable aspects of the same process.

13.6 Example problems

Tangible example problems might include:

Testing the unknown. SoSs continuously integrate various new devices and components; some of them could be untested and any intrinsic flaws will be inherited. The research should pave the way to new testing paradigms to achieve self-adaptive testing methodologies aiming at ensuring that unknown and untested components and devices are trustable and have good quality before they join the SoS. In other words, this research should promote "Full Quality – positive-sum, not zero-sum." [1]

Testing of AI/ML/DL. Provide testing methodologies and tools that can be suitable for revealing bugs in artificial intelligence (AI), machine learning (ML) or deep learning (DL) applications. The study should consider the following three main aspects: the required conditions (correctness, robustness, security and privacy); the AI, ML or DL items (e.g. the data, the learning program, or the framework used); and the involved testing activities (test case generation, test oracle identification and definition, and test case adequacy criteria).

Using AI/ML/DL for testing. Provide AI/ML/DL-based methodologies and tools that can help perform most testing tasks, such as test-case generation, test-case classification, oracle derivation or mutation analysis, to cite a few. Therefore, this research aims to leverage state-of-the-art AI/ML/DL technologies to aid software and hardware testers in achieving the desired quality driven by testing data.

Understanding the testability of the metaverse. Improve the understanding of the challenges of testing the metaverse by considering three testing

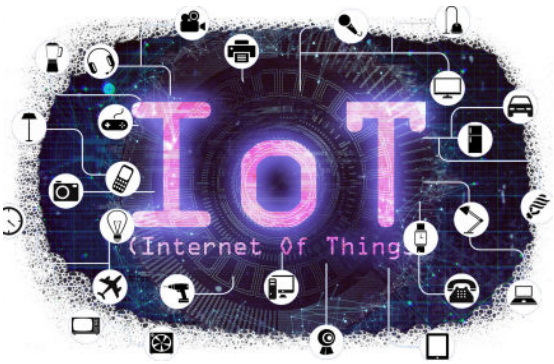
[1] This term is inspired by the well-known privacy by design principle "Full functionality: positive-sum, not zero-sum" [35]. See also Chapter 8.

pillars: cybersecurity, aimed at security testing; API testing, crucial for guaranteeing interoperability, which is a fundamental characteristic of the meta experience; and interactive and immersive testing, which puts the human at the core of testing meta experiences.

We are all testers. Improve the understanding of the role of humans in the testing process. The research should provide theories, insights, and practical solutions for engaging people in the testing and assessment of digital products and services, considering different dimensions of (digital) ethnography. The starting point for this kind of research should be gamification, which aims to convert testing tasks to gameplay components, and crowd-sourced testing (also known as crowdtesting), which is an emerging approach for involving users and experts in testing activities.

14 IoT Security

14.1 Introduction



The Internet of things (IoT) is a collection of devices (i.e. things) that contain sensors and/or actuators, software, and communication capabilities to send and exchange data with other devices on the Internet. The idea of the IoT has been with us for quite some time now, and its security has always been and still is one of its main challenges.

Over time the types and capabilities of everyday devices connected to a network continue to grow quickly. IoT Analytics [106] estimated the number of connected IoT devices for 2021 to have been 12.2 billion globally. Meanwhile, Statista [238] has estimated a more modest 11.3 billion connected IoT devices in 2021. The prediction is for the number of devices to more than double by 2030, with Statista estimating 29.4 billion connected IoT devices. IoT represents one of the biggest security concerns at the present time and in the future, as the number of such devices is projected to grow and permeate all aspects of life even more deeply. This is corroborated by the fact that IoT devices, on average, get attacked within five minutes of being connected to the Internet [158].

While most might think of smart refrigerators, robot vacuums or smart watches when they hear the word IoT, there are many more such devices that support and surround us in our daily lives as well as in industries. IoT can be used in many areas to help optimise and/or automate processes by gathering live data. The most common applications of IoT include smart homes, smart cities, waste management, smart grids and power management,

industry (i.e. Industry 4.0), agriculture, smart healthcare, smart warehouses, smart transport and logistics, etc. From the environment where they are used and their purpose, we can derive different types of IoTs. For example, Internet of Industrial Things or Industrial Internet of Things (IIoT), Consumer Internet of Things (CIoT), Internet of Medical Things (IoMT) or Internet of Healthcare Things (IoHT), Internet of Agricultural Things or Internet of Farming Things (IoFT), Internet of Energy Things (IoET), Internet of Vehicles (IoV), Internet of Transportation Things (IoTT), Internet of Education Things (IoEdT), etc.

There is a considerable list of properties that make IoT devices and networks vulnerable. For example, the ubiquity of IoT devices make it difficult to protect them against physical access. At the same time, the diversity of devices makes it difficult to design "one-size-fits-all" security constructs that could be freely applied to the devices. Even more, the rapid life cycle (of the devices themselves and the development process) also makes it hard to track the devices on the market and apply software patches. Discovered vulnerabilities can go unpatched for an extended amount of time, and even if there is a patch, most of the users fail to regularly update IoT devices.

To make matters worse, IoT devices are often left with their default security configurations (e.g. factory passwords) which leaves them even more vulnerable. And finally, several IoT devices are rather small, with limited power, memory and computational capabilities. This often means they are not capable of running the best security mechanisms and protocols and must instead use less computationally demanding and resource-intense solutions that are generally not as secure.

Addressing common challenges in the IoT ecosystem, of which security is certainly one of the most important, is key to the future of IoT, especially as IoT becomes more and more ingrained in our lives and no longer represents a threat only to our sensitive information, but also to our physical assets and health. For all of the convenience and value that IoT provides, the risks are also unparalleled.

14.2 Who Is Going to Be Affected?

IoT devices affect nearly everybody. For example, individuals can be affected in many ways. If a smart home comes under attack and stops functioning correctly, the inhabitants can lose power, heating, light, entertainment, etc. Web cameras and baby monitors are also very common household IoT devices that regularly get attacked and have previously been used to spy on their owners or to form part of a botnet. Individuals may also be indirectly affected if the attack targets their organisation, their government or any other entity they are part of.

Industries may also be negatively impacted by IoT attacks. Indeed, a successful attack on such an IoT system would cause operations to cease. Any organisation in the supply chain would also suffer consequences, especially if the attack was aimed at postal/transport organisations that manage the transportation of goods. While this is truer for organisations or industries dealing with manufacturing, online services are more vulnerable to things like DDoS (Distributed Denial of Service) attacks, which make online services inaccessible by overloading the service providers with fake requests. Some of the largest such attacks were launched from hijacked IoT devices that formed a botnet (a large collection of devices that were successfully attacked and subverted to do the attacker's bidding: e.g. [178,206]).

For efficiency and transparency, many critical infrastructures and governmental services (e.g. power, water, and waste management) are becoming IoT supported. Any attack that would undermine any of them for any extended amount of time would cause havoc in the population and resentment towards the government. Attacks against IoT could also be used to spy on politicians or, again, using DDoS attacks, to make digitally supported governmental services unavailable (e.g. eHealth).

14.3 What Is Expected to Happen?

As IoT progresses to become part of everything, many things could be affected when something goes wrong. Web cameras, baby monitors, voice assistants, smart toys and similar tools can monitor peoples' activities and conversations (e.g. [108,165,192]). Medical devices collect highly sensitive information, including protected health data. Smart temperature sensors can tell people when somebody is at home, and smart locks can let them in when they are not. IoT devices used in manufacturing could be used for industrial espionage to obtain sensitive information about manufacturing processes and procedures, or the whole manufacturing process could be shut down. Attacks on IoT in smart transport and warehousing will disturb supply chains. Attacks can affect traffic where smart traffic management is used, and attackers can take over smart cars if they can gain access (e.g. [100]). By attacking smart water management and power grids, large regions can be left without power and water, which brings any industry to a stop and causes people there difficulties with cooking and keeping warm in the winter. In agriculture, a successful attack that is not noticed quickly enough can lead to ruined crops or dead livestock. Having IoT devices expands the attack surface, so we can expect more successful attacks by attackers gaining access to protected networks through seemingly inconsequential IoT devices (e.g. [245]). Large amounts of successfully corrupted IoT devices will be merged into botnets

that will then be used for crypto mining or to perform large attacks, such as DDoS, to cripple online services or whole parts of the Internet (e.g. [58]). IoT attacks will regularly come in the form of ransom, where the attackers will demand money to stop an attack or not begin it in the first place.

Given these few examples, the potential damage that could be caused by losing security over IoT systems is immense. Consequences include loss of privacy, identity theft, effect on health or even loss of lives, stealing of intellectual property/competitive advantage, loss of property, goods shortages, decreased food production, unavailability of online services, difficulties with the supply of electricity and other energy sources, etc.

14.4 What Is the Worst That Can Happen?

In the previous section, we tried to show how much could go wrong if IoT systems get compromised. In this section, however, we want to give some worst-case escalations of those examples. In the case of losing privacy, there are two really bad outcomes. The first is the loss of anonymity, which was already covered in a previous book chapter, while the second is identity theft, which is considerably alarming, especially if it happens in large numbers. Malicious medical IoT devices can cause health degradation or even death, but even worse are devices implanted in humans (i.e. pacemakers). Ransomware on such devices is basically remote kidnapping that does not leave the victim with any negotiation options or alternative to paying the ransom. The loss of running water and power is bad, but if a large enough area is affected, that would plunge the inhabitants into a dark age, which in modern times would be catastrophic. An attack on the water supply can not only stop the running water, but it can also make it poisonous by altering the water treatment at the water plant. Any widespread successful attack on critical infrastructure would have devastating consequences for general security (e.g. military), national economic security, national public health or safety. Malicious attacks on manufacturing plants can also cause the production machinery to break, stopping production for a very long time or even causing injuries or deaths among employees. Using large enough botnets to perform DDoS or other types of attack could cripple large sections of the Internet and, with it, everything that relies on that infrastructure (e.g. communications).

14.5 Research Gaps

IoT security is a problem, and it will get worse as the potential attack surface expands with many more devices and with more critical devices (e.g. medical devices).

14.5.1 Artificial Intelligence and Machine Learning

Artificial intelligence (AI) and machine learning (ML) promise to be a huge help in securing and identifying attacks in IoT [10, 114, 203]. The introduction of AI and ML into the IoT environment has some associated difficulties, such as deployment on constrained and distributed devices and the need for updating AI/ML models over time, which can be problematic for reasons of accessibility and general updating practices - as we discuss later in this section. Overcoming these issues, AI and ML can provide a great deal in terms of security for IoT systems. AI and ML can cope with heterogeneous data and go through large volumes of data produced by IoT much more quickly (i.e. in real time) than traditional methods, enabling them to discover attacks as they happen. Such solutions can be utilised for access control, security, malware detection and analysis, risk assessment, threat analysis, privacy, attack detection, and potentially tracing the attack trough the system. AI/ML is also a good foundation for providing additional system resilience. Deep learning has already shown promising results in identifying IP Spoofing and DDoS attacks, and decentralised machine learning could be especially compatible with IoT. We need solutions that are able to identify the subtleties of security breaches and mitigate them while conforming to the limited resources of IoT devices. This includes efficient labelling of input streams and learning with smaller sets of training data. We need methods for such solutions to work not only in enterprises, but potentially also in much smaller environments, regardless of the type of data transmitted through the IoT network.

14.5.2 Strong and Universal Security Standards for IoT Technology

Security standards in IoT and their application, in general, need some work [13]. The quick development of solutions and the heterogeneity of the devices certainly do not make standardising IoT security any easier. Universal standards or guidelines should be set for IoT devices, including data protection at rest and during communication, authentication and authorisation of IoT devices, maintenance and management of IoT devices, auditing and logging, and secure interfaces (web, application API, cloud, and mobile), and IoT security incident response processes. In general, more IoT development should follow the "security and privacy by design" paradigm, especially for devices that collect personal data and/or can have a significant impact on their owners' health or assets (e.g. smart lock).

14.5.3 Develop Strong and Lightweight Cryptography for IoT

Some IoT devices have severely limited resources, and to retain full functionality, security and usability, they require lightweight (cryptographic) protocols. Lightweight solutions must be efficient in their computational, memory and

power consumption. For this purpose, we need (standardised) lightweight IoT solutions for data encryption (at rest and in transit), key management, routing, authentication, and access control. Additionally, malware is also a large problem for IoT systems and for the same reasons of limited resources, malware detection solutions that can be effective in such environments have to be further developed.

14.5.4 Establish Trust and Traceability

Taking into account the security concerns surrounding the IoT, establishing trust in the devices, their processes, and the collected and transmitted data is important. Current IoT systems lack transparency, making it impossible for ordinary users to know what is going on, what data is being collected and what happens to it. This includes live monitoring that can notify users in real time of any malicious behaviour in IoT systems. Monitoring is also very important for self-healing cybersecurity IoT systems that have the potential to automate cybersecurity.

Data traceability and integrity are vital for increasing trust in data and, consequently, the whole IoT system. Distributed ledgers have become the primary solution for data traceability; however, some development, especially in scalability, is still needed before they can be freely applied to larger IoT networks. At the same time, trust is also required amongst IoT devices in a network. This prevents attackers from joining the network or masquerading as one of the devices in the network. For this, we need better secure trust management systems.

14.5.5 IoT Security Awareness and Education

IoT users are currently not well aware of the security risks and especially the available mitigation controls to reduce these risks [115]. This is especially true in personal/home environments and smaller businesses, but it is unfortunately also often true in enterprise environments. The most common problem, and one that has been exploited very successfully even in the recent past, is the use of the default passwords that the devices were shipped with or the use of weak passwords. More effort is required for effective awareness methods and tools for informing the public of the dangers of insecure IoT (either insecure devices or weak configurations). IoT products should come with clearer instructions for the users on how to set up their devices with an emphasis on the importance of security and privacy settings (this could be part of the manual and/or as hardwired policies, e.g. default passwords would have to be changed during the setup to a password of some minimum quality). Reportedly there is also a large shortage of professionals to implement

IoT networks in businesses, including cybersecurity talent [118]. Appropriate training and upskilling programs should be designed and put in place.

14.5.6 Hardware Security

With IoT devices, it is important to remember that they cover a wide range of use cases, and in some of them (e.g. when devices are installed outside protected environments), the physical or hardware security of the device itself is as important as anything else [18]. This aspect often seems to be forgotten, and IoT devices lack hardware security, such as cryptographic coprocessors or anti-tampering technologies. Therefore, we need more low-cost, efficient and well-tested modules, which include hardware security that manufacturers can reliably use in their IoT products, and we must provide incentives for them to be used. In this section, trusted gateways can also be mentioned as a way to minimise the attack surface and the damage to organisations.

14.5.7 Privacy in IoT

Privacy is an important challenge in IoT [107]. Privacy preservation restricts the processing of data to only the strictly necessary, and in a way that prevents additional sensitive data from being inferred throughout the data's lifecycle. It must also strike a balance between data utility and privacy. We need more emphasis on privacy during the design and development of IoT and better privacy-preserving techniques (e.g. anonymisation) that could be widely adopted in IoT.

14.5.8 Life cycle management

A device can be secure today, but this condition could change during its life cycle because of a newly discovered vulnerability. The security management should be scalable and as automatic as possible if we want to deal with a large number of heterogeneous IoT devices [109]. However, this might not always be possible. Since IoT devices are not usually equipped with traditional interfaces, and updates are not pushed to the devices, users do not know there are new updates or patches they should install. We need methods of notifying device owners when there are crucial updates or patches they need to install without them losing any functionality of the systems they have set up (if updating means losing data or device configuration, many will choose not to update). Finally, an additional challenge that needs further research is to develop efficient update procedures for IoT devices with very limited resources (e.g. not enough memory to download an update).

14.5.9 IoT Regulation and Policies

At the end of the day, even if the technology exists that can make IoT secure, it is still important for the technology to be implemented. As is often the case, regulation takes some time to catch up with technological advances, and while we have recently seen some movement on regulating IoT solutions and the expected levels of security they should provide, there should be more. We need some way of imposing minimum security standards for IoT devices (e.g. certification).

One crucial matter that could be alleviated with regulation is the long-term support of IoT devices. Today you can buy a device, and the manufacturer will end its support (if it had any in the first place) at any point in the future, without even notifying the device owners. Given the current policies of sustainable development, minimum critical security support could be prescribed by regulation, or there could be a requirement for products to have a clearly marked support duration on their packaging at the time of sale, which the manufacturer will guarantee.

14.6 Example problems

Tangible example problems might include:

Machine learning-based cybersecurity for IoT. Study IoT attack patterns and develop IoT-friendly raw data-labelling methods for new machine learning solutions to recognise attacks. Create anomaly datasets. Develop new deep learning solutions for detecting attacks and/or malware on IoT networks.

IoT device security classifications. To alleviate the problem of IoT device heterogeneity, develop a classification scheme for IoT devices based on their resource limitations and purpose (i.e. how crucial is security for the device, based on what it is meant to do and what types of data are involved). The classification could be used to determine what are the minimum security features (e.g. security protocols) the device has to support for it to be considered to have an acceptable level of security, given its security class.

Smart honeypots for IoT. Establish emulation of IoT devices on universal computer platforms. Enable monitoring and collection of data from the distributed IoT honeypot network.

Lightweight protocols for IoT. Find or adapt suitable existing protocols or develop new cryptography protocols for IoT (potentially for each IoT device security classification from the previous example). The selection

of protocols (for data encryption, both for data at rest and in transit, key management, routing, mutual authentication of devices in the network, etc.) can be promoted as good practices and/or standardised.

Update and patch notifications for ordinary users. Compile a database of IoT devices and hardware used and any consequent updates or patches released for their software or firmware. Give users options to find their devices in the database and subscribe to be notified if updates or patches are ever available for their devices. Provide instructions on where to get them and how to install them.

Improved authentication. IoT devices suffer from overuse of default and weak passwords. Efforts should be put into developing convenient ways of incorporating multi-factor authentication into IoT devices and developing and implementing passwordless authentication for IoT devices.

15 Effective Threat Modelling

15.1 Introduction

There is growing trend in security of shifting left, that is applying security activities earlier in the software development lifecycle. Threat modelling starts from an architecture-level (or design-level) description of the software system or service that is being developed, and strives for early improvements in terms of security and privacy by (1) identifying threats, (2) prioritising these threats in terms of risk and possible damage, and (3) suggesting/offering possible mitigations at the architectural level. Such an approach is beneficial, as it enables the identification of security flaws early on to reduce the impact of changes [232]. The relevance and usefulness of techniques like security and privacy threat modelling is demonstrated by the growing interest in threat modelling. Indeed, organisations such as Microsoft have made great strides in addressing security in the early phases of the development lifecycle as part of their security push in the early 2000s [112, 130, 227], with the introduction of security threat modelling and the security development lifecycle. The relevance and importance of considering security in these phases continues to be recognised and is confirmed with the 2021 release of the Owasp top 10 [180] which explicitly includes insecure design as a top 10 entry and specifies the need to perform more threat modelling [212]. Furthermore, it has been applied to many systems in practice. For several of these, concrete threat models are available, such as the SecureDrop whistle-blower submission system [86] and Kubernetes [233]. Such a systematic and comprehensive analysis can be an indispensable tool to identify problematic data flows in applications that are later leveraged as part of ransomware attacks to further propagate themselves.

15.2 Who Is Going to Be Affected?

Clearly, the activity of threat modelling involves software architects and security experts. It introduces an additional and possibly costly activity to the development process, yet the yield can be a relatively high level of assurance: many classical security and privacy flaws can be avoided "by Design". If this

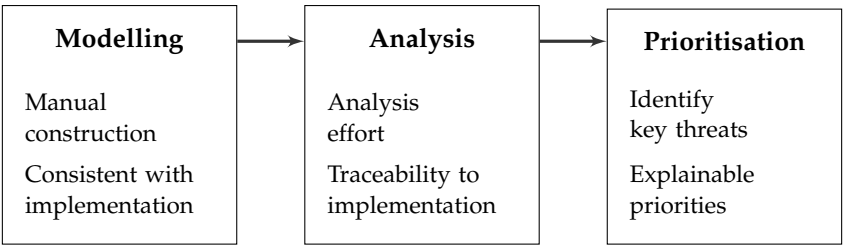


Figure 15.1: Challenges in each threat modelling phase

were not covered, the same flaws could be hidden and not discovered until later, at the implementation level. This would entail much larger investments and efforts to deal with these specific security problems. In effect, the service or software provider (company) remains in need of a cost-effective threat modelling process. Users and user organisations of the corresponding product or service might not be aware of this matter, yet they would still have to pay the bill at the end of the journey of solving structural security problems.

While techniques for threat modelling have already shown great potential in supporting the design and development of secure software systems, the broader application of these techniques as a part of the software development processes introduces a number of challenges (shown in figure 15.1) for practitioners with regard to the cost of applying these techniques in contemporary development processes [214]. First, the application of these techniques is typically an activity that includes the involvement of security experts, a scarce resource in many companies, which hinders the broader and more frequent application of these techniques [253]. Second, the application of these techniques entails some manual effort in creating and maintaining a representation of the system and analysing such a representation to identify security threats. Any manual effort as part of an activity that is, ideally, frequently repeated as a software system is further developed and extended, introduces a non-trivial overhead that impedes its frequent application. Furthermore, the cost of maintaining and re-analysing this representation is exacerbated in the context of contemporary development practices that are characterised by frequent iterations and fast-paced development.

15.3 What Is Expected to Happen?

The drawback of not performing threat modelling has been suggested above. Yet the current cost of threat modelling is high, and the research challenges introduced in this chapter are of utmost importance to increase the cost-effectiveness of current and future threat modelling practices.

As mentioned above, the application of security and privacy threat modelling commonly involves a manual input or assessments by threat modellers, such as the creation of a model representation of the system under consideration, the elicitation of the security and privacy threats, the prioritisation of these threats to determine the most important ones and, finally, suggesting appropriate mitigations to address the identified threats. Practitioners encounter several challenges when applying these threat modelling activities: (i) a comprehensive analysis of a software system entails a significant amount of work, in both constructing the model of the system and the actual threat elicitation; (ii) the analyses can frequently lead to long lists of threats, but these results lacks information on the relevance of these threats, hindering the identification of the most critical ones; (iii) it is essential to ensure that the model used for the analysis remains consistent with the actual implementation of the system under development. Each of these challenges will be explained in more detail below.

15.3.1 Manual work

One of the largest challenges to the cost-effectiveness of security and privacy threat modelling is the reliance on manual effort in both the creation of the models and the analysis for eliciting security and privacy threats. Since the threat modelling relies on using a design representation of the system (typically a data flow diagram [59,112]) to analyse for security and privacy threats, such a representation must be retrieved or constructed before the threat analysis can start. However, frequently such design documentation is not available for the systems that have been built or are being extended. Because of that, the design of the system under analysis will have to be reconstructed by relying on documentation (to the extent it is available) and going through the implementation of the application. This reconstruction effort already imposes additional cost when performing a threat modelling exercise, and this effort may have to be repeated frequently if the model documentation is not kept up to date with the application as it is further developed. A second source of manual effort can be the analysis itself. The amount of effort introduced by this step depends on the extent to which practitioners can rely on tool support for the analysis or instead perform the analysis manually. The more informal the system descriptions are, the more the analysis will have to rely on a manual assessment by a threat modeller, as automated tools require a richer model input, including more information, to enable the tool to make the threat elicitation decisions automatically.

15.3.2 Prioritisation

The second challenge is related to using the results of the threat analysis in subsequent phases to support decisions on applying security and privacy countermeasures in the application under development. As the available resources to address security and privacy threats are limited, practitioners need to be able to determine which threats are the most relevant and important to address. However, the security and privacy threat elicitation only renders a (large) list of threats that are applicable. It does not provide any support in identifying the most relevant threats among them that should be addressed first. These elicited threats commonly lack information needed to prioritise the resulting threats. As a consequence, the prioritisation of the threats involves a manual activity in which each threat has to be manually assessed to determine its relevance. While such an approach may be appropriate for a single-shot analysis, it is ineffective if the threat elicitation is frequently repeated and the resulting list of threats changes as well. Furthermore, support for tracking the priority or importance of threat types is frequently limited to a very coarse grained classification (e.g. low, medium, or high) that does not include any kind of traceability information when such a classification decision will have to be reassessed later on. Because of the lack of information, it is not possible to assess why that particular priority was assigned to that threat at the time. If certain assumptions underlying that decision turn out to be invalid, it is not possible to identify all relevant threats that would require a reassessment.

15.3.3 Ensuring up to date results

A final challenge for practitioners is to ensure that the threat analysis results remain up to date and relevant to the application under development. Especially with contemporary development practices that involve fast-paced development and frequent iterations, the design of the application can change frequently. The result is that the threat analysis results from previous versions of the design are no longer relevant, as some threats may no longer be applicable (for example, due to the removal of certain elements in the design). This introduces a challenge in keeping the design representations of the system up to date with the implementation as it evolves during development leading to additional maintenance costs to ensure the threat results are current.

15.4 What Is the Worst That Can Happen?

The previous section outlined the different challenges and problems experienced by practitioners in the application of threat modelling, especially in terms of overhead and cost-effectiveness of these approaches. while a great number of scenarios can be constructed to illustrate the impact of various se-

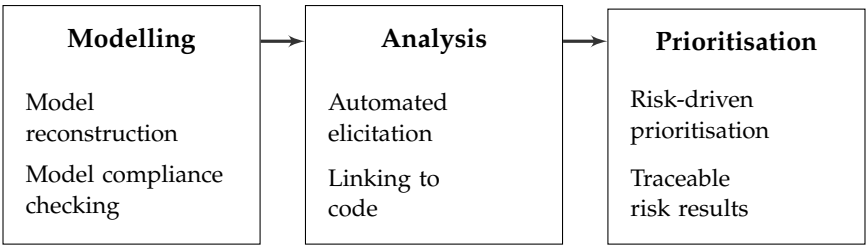


Figure 15.2: Opportunities and improvements in each threat modelling phase

curity flaws being missed in the development of a concrete software product. The worst case is actually unknown, as the actual impact of not performing any of these security analyses cannot be predicted because of the uncertainties in the applications, the organisations, the contexts in which the applications are used, the types of data processed, etc. Hence, the main focus is on the role of automation in reducing overhead and effort as a way to enable the broader use and application of these techniques. The successful enhancement of threat modelling, largely through automation, will be necessary to drive to adoption, which in its turn will enable the avoidance of expensive worst-case scenarios in the reengineering and fixing of complex software systems and services. It is hard to estimate worst-case scenarios in terms of damage.

15.5 Research Gaps

The research agenda that is essential to drive this subdomain of the secure software development lifecycle is relatively straight-forward. It mainly includes activities that relate to reusable knowledge, automation and tool support, etc. The essential research themes and activities are summarised below.

15.5.1 Automation

A key element in the strategy to address these challenges is to strengthen automation and apply it in many threat modelling activities to reduce the manual steps and enable frequent reassessment as part of iterative development practices. Indeed, automation can play a crucial role in reducing the cost of threat modelling by automating many steps that involve costly and manual inputs by developers and experts. We briefly outline each of the phases (shown in figure 15.2) of the threat modelling processes in which automation can significantly improve cost-effectiveness by reducing or eliminating manual labour.

Modelling One of the first steps where automation can be applied is in the construction of the model representation that serves as the input for the threat

modelling activity. This is also one of the most challenging areas to apply automation. There are two main approaches that can be taken that provide different degrees of reduced manual effort. First, after the construction of an initial model representation of the system, model compliance checking [186] or architectural drift analysis [231] can be used to verify whether the model representation actually corresponds with the source code implementation of the system. Such an approach still requires an initial model, but can reduce the cost of keeping the model up to date as the system continuous to be further developed. Second, a more complex and more fully automated approach is to rely on model construction. This approach employs tooling to automatically create a model starting from the source of the application, thus eliminating the initial effort in model construction. These techniques can of course be combined with the compliance checking to verify the accuracy of the reconstructed models.

Threat Elicitation The second step where automation can be leveraged is during threat elicitation. There are two areas in which automation reduces effort and manual input: the elicitation itself and the automated application of expert knowledge. For the threat elicitation itself, the use of automation can ensure a comprehensive, systematic, and repeatable analysis of the system. Many existing threat modelling tools [120, 156] do provide this functionality already, ranging from simple criteria to more complex model patterns [235]. Automated tools can consistently apply complex rulesets to system designs to ensure repeatable threat elicitation. The second benefit of automation in the context of threat elicitation is that it allows expert knowledge about security and privacy threats to be encoded into tool support, enabling the automated application of this knowledge without having to rely on security and privacy experts to assist in the assessments, as these are scarce resources for organisations.

Prioritisation The third step where automation introduces benefits is in the prioritisation of the elicited security and privacy threats. Given the substantial number of security and privacy threats that may be elicited, being able to prioritise them becomes essential. The large number of threats makes it increasingly difficult to review them for prioritisation, especially if the analysis is frequently repeated in response to changes to the system design. Automation provides two key benefits in this context. First, because the automation will rely on additional information in the input models to determine the priorities of the threat, it actually forces the explicit specification of this information in the input models. While this introduces some overhead to provide additional input, it also allows traceability of the results, as the resulting priorities can be explained through the inputs and revisited later. Second, it removes the need for manual assessment and prioritisation of the threats, making it much

more economical to frequently reanalyse a system. Such automation requires the construction of risk models [87,215] that can be systematically applied.

15.5.2 Tool support

As illustrated above, there are many opportunities for automation to reduce manual effort and enable a more cost-effect threat analysis of a system. Tool support is crucial for achieving automation in these different phases of threat modelling. The necessary tool support ranges from: (1) source code analysis tools to perform compliance checking or model reconstruction; (2) automated threat elicitation, leveraging encoded expert knowledge; and (3) automated prioritisation of elicited threats using risk assessment.

15.5.3 Education and training

A final area of improvement is to provide education and training to enable all personnel to participate in threat modelling and further reduce the reliance on security and privacy experts for threat modelling activities. Together with tool support, education and training facilitates the embedding of threat modelling in existing software development processes.

15.6 Example problems

Tangible example problems might include:

Creating and maintaining models. Any threat modelling activity relies on the creation of an initial model of the system to be analysed. The creation and maintenance of these models can introduce significant overhead for threat modellers hindering the frequent application of these techniques during development. There have been several advances [186] that make it easier to determine whether these models are still compliant with the code, thus reducing the effort involved in maintenance. The analysis of source code to construct models that are readily useable in threat modelling analysis is still a challenging problem.

Automating threat knowledge. There are many publicly available resources with information about previously identified security bugs, weaknesses, and flaws (e.g. CVEs, CWEs, etc.). These resources are highly dynamic, as they are frequently updated when new issues are identified. While some of these resources have already been successfully integrated into automated analysis activities, such as the detection of vulnerable dependencies, not all resources are easily translated and applied in a threat modelling context.

Integration in development processes. The application of threat modelling is usually an activity that happens in isolation. This introduces some

additional overhead and complexity in translating the threats identified in the system's design into very concrete and actionable items for developers to work on. There are several challenges in improving the actionability of the results of threat analyses by supporting a tighter integration in development processes and relating threat modelling results to concrete source code artefacts, for example, by guiding towards starting points when mitigating the identified threats.

16 Grand Challenges

In this section we describe some "grand challenges" that we will need to face in the next few years. These challenges require the collaboration of hundreds of people from several different realms of science. Most of these challenges not only involve novel research, but also need appropriate regulation and possibly legal frameworks in place. We hope that the funding agencies will provide support to these areas and that the research community will start working towards these challenges.

16.1 Give users assurance about the security of their devices

Most computing devices today offer little, if any, assurance about the level of security they provide. Although some of them (such as medical devices) may adhere to **safety standards**, most of them do not adhere to any **security standards** at all. As a result, they provide no guarantees to their users: they may crash at any time; they may get compromised at any time; they may turn hostile at any time. We believe that we should provide users with (i) better transparency and (ii) better guarantees about the security of their devices. Although this sounds like a task that can be achieved through regulation, it has significant research and development dimensions including continuous monitoring, aggressive penetration testing, and continuous bug detection to name a few.

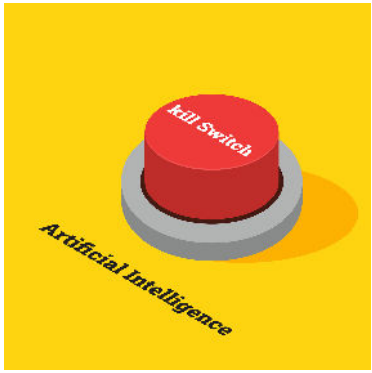
16.2 If it can be done anonymously in the offline world, it can also be done anonymously online

Over the past years we have moved several of our everyday activities to cyberspace. The COVID-19 pandemic intensified this trend so that at the peak of the pandemic the only ways to interact with other people involved the digital world at some level. As a result, we started doing all our shopping online,

our visits moved to teleconferences, our schooling was done via Zoom, several aspects of work also moved online, etc. What we did not easily realise, though, was that in order to carry out these activities online we had to provide a great deal of personal information, and in this way sacrifice our privacy. For example, in the past it was possible to do most of our shopping practically anonymously. We could visit stores anonymously, browse for various products anonymously, we could even pay anonymously using cash. At no point in this process did we have to reveal our name, our address, our telephone number etc. We could reveal this information (if we wanted to), but we did not have to. Today it is almost impossible to do any shopping online without revealing a great deal of personal information such as our name, our telephone number, our address, etc. Such personal information is revealed to a wide range of different actors including the merchant, online advertisers, the courier company, etc. We believe that it is now time to reclaim our privacy and reveal as little information as possible. The **guiding principle** here is that **if it can be done anonymously offline, it can also be done anonymously online**. This is not an easy task and it may involve several aspects besides research including, for example, awareness and deployment. It may not even be possible in some cases and with some providers. However, having this as a guiding principle will help us trim down all the cases where privacy has been unnecessarily sacrificed.

16.3 Make AI Safe for People

AI is spreading widely and rapidly. For example, a recent whitepaper by Deloitte showed that the world will see AI-driven GDP growth of $15.7 trillion by 2030. The capability of AI, and ML models in particular, to extract/learn complex features from massive volumes of (often) unstructured data is what makes them a popular choice for tackling various problems. Yet, as discussed in Chapter 4, ML-powered applications offer a whole new spectrum of security and privacy exploits for potential adversaries.

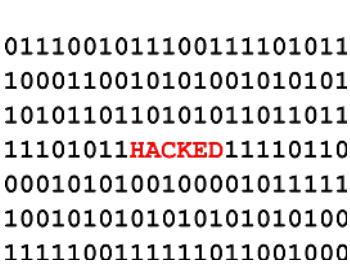


First, ML models are often applied to sectors where wrong decision making can have serious implications. Yet it may often not possible to offer formal security guarantees, given those models' non-deterministic nature. Second, ML models are often trained on personal/sensitive data, especially models deployed in the healthcare field. Thus,

revealing training instances constitutes a serious violation of individuals' privacy.

As a result, we need to develop techniques and mechanisms for **making AI safe for people**. Note that doing so is not an easy task and involves bringing together researchers and practitioners from a wide range of fields, such as mathematics, linguistics, informatics, etc. In fact, for specific use cases, it may not even be possible to provide the desired guarantees without sacrificing the model's performance. However, working towards this direction will surely lead to significant improvements and novel techniques offering acceptable trade-offs.

16.4 Make systems resilient under attack



Computer systems can be remarkably fragile. Indeed, a wrong `if statement`, a wrong `assignment statement`, or an undefined `global variable` is all it takes to crash an application or even to compromise a computer. To make matters worse, if a program with the wrong `if statement` runs on millions of computers, all these computers may be compromised in a matter of hours or even minutes! The grand challenge here is to develop computer systems that are able to tolerate cyberattacks. We would like to have systems that fail gracefully when are attacked by cyberattackers. We cannot avoid having millions (or even billions) of copies of a program running on various devices. Indeed, there are billions of people and tens of billions of devices running a small number of ultra-popular applications. The challenge in this environment is to make these ultra-popular programs (and all computers in general) resilient to cyberattacks. There are several different paths one can explore in order to achieve this resilience. Although the paths may be different, most of them agree that an application should fail gracefully under attack. This graceful failure may mean that only a small fraction of the computers will be compromised, or that only a tiny part of the functionality will be compromised, or something else. The unifying point, however, is to make systems more resilient to cyberattacks; one wrong `if statement` should not be able to compromise millions of computers. We should do much better than that.

16.5 Enhance General Public Awareness of Cybersecurity

People are often perceived as the weakest link in the cybersecurity chain. They are a major contributing factor to the majority of cybersecurity breaches, as cybercriminals frequently employ techniques that exploit innate human

weaknesses to carry out attacks. Enhancing cybersecurity competence development through training and awareness initiatives focuses on enabling people to establish technical and operational barriers to cybersecurity threats, and to operate themselves as such, through the vigilant processing of actionable intelligence. Boosting the potential impact of such initiatives requires the personalisation and tailoring of the awareness or training experience. This must take into account, among other things, personnel roles, knowledge foundations, competences, and experiences. It should also include the operational context of the involved organisations, including policies, processes, and applicable regulatory frameworks.

The first grand challenge here has to do with creating a mapping of the competence benchmarks that are to be achieved, depending on the distinct organisational contexts and the corresponding personnel roles. This also reflects on the personal sphere when referring to societal hardening and awareness. The second grand challenge has to do with the delivery of competence development programs, which means structuring the appropriate message to achieve the targeted learning objectives, selecting a suitable medium of communication, and determining the time intervals and other parameters that are dependent on the participants and can significantly affect participation and retention. Addressing these challenges requires a multidisciplinary approach, involving expertise not only in pedagogical sciences and cybersecurity, but also psychology, domain (i.e. sector specific) experience, and other areas. Furthermore, social sciences and data analytics can be contributing factors that can enhance and facilitate the aforementioned mapping, while also contributing to tailored delivery.

Bibliography

[1] 8 zoom security issues you need to know about. https://www.sigmundsoftware.com/blog/zoom-security-issues-coronavirus/. Accessed: 2022-11-119.

[2] Digital economy and society statistics - households and individuals. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Digitaleconomyandsocietystatistics-householdsandindividuals. Accessed: 2022-11-119.

[3] How do burglars use social media to find targets? https://www.homewatchgroup.com/how-do-burglars-use-social-media-to-find-targets/. Accessed: 2022-11-119.

[4] Year 2000 problem. https://en.wikipedia.org/wiki/Year_2000_problem. Accessed: 2022-11-119.

[5] Securesme: Cyber tips for passwords, Sep 2021.

[6] Open authentication standards more secure than passwords, Nov 2022.

[7] N. Achiaga and M. D. Mar. The NIS2 Directive: A high common level of cybersecurity in the EU. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)689333. [Accessed: 07 November 2022].

[8] A. Adams and M. A. Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, dec 1999.

[9] R. Agrawal and R. Srikant. Privacy-preserving data mining. *SIGMOD Rec.*, 29(2):439–450, may 2000.

[10] R. Ahmad and I. Alsmadi. Machine learning approaches to iot security: A systematic literature review. *Internet of Things*, 14:100365, 2021.

[11] C. Alcaraz and J. Lopez. Wide-area situational awareness for critical infrastructure protection. *Computer*, 46(4):30–37, 2013.

[12] S. T. Ali, P. McCorry, P. H.-J. Lee, and F. Hao. Zombiecoin 2.0: managing next-generation botnets using bitcoin. *International Journal of Information Security*, 17(4):411–422, 2018.

[13] E. Andrukiewicz, S. Cadzow, and S. Górniak. Iot security standards gap analysis. = https://www.enisa.europa.eu/publications/iot-security-standards-gap-analysis, 1 2019. [Accessed: 07 November 2022].

[14] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou. Understanding the mirai botnet. In *Proceedings of the 26th USENIX Conference on Security Symposium*, SEC'17, page 1093–1110, USA, 2017. USENIX Association.

[15] S. Aral and D. Eckles. Protecting elections from social media manipulation. *Science*, 365(6456):858–861, 2019.

[16] R. Arizon-Peretz, I. Hadar, G. Luria, and S. Sherman. Understanding developers' privacy and security mindsets via climate theory. *Empirical Softw. Engg.*, 26(6), nov 2021.

[17] ARM. Building a secure system using trustzone technology. In *ARM Security Technology*. ARM, April 2009. [Accessed: 16 November 2022].

[18] Arrow. Understanding the increased importance of hardware security in iot technologies. = https://www.arrow.com/en/research-and-events/articles/understanding-the-importance-of-hardware-security, 5 2020. [Accessed: 07 November 2022].

[19] ART. Metaverse : Virtual world, real challenges. Technical report, Analysis and Research Team of the Council of the European Union, Mar. 2022. [Accessed: 07 November 2022].

[20] D. Atch, G. Regev, and R. Bevington. `https://www.microsoft.com/en-us/security/blog/2021/08/19/how-to-proactively-defend-against-mozi-iot-botnet/`, 2021.

[21] M. Azure. Openenclave sdk. `https://openenclave.io/sdk/`. [Accessed: 17 November 2022].

[22] M. Bada, A. Sasse, and J. Nurse. Cyber security awareness campaigns: Why do they fail to change behaviour? In *International Conference on Cyber Security for Sustainable Society*, pages 118–131, 01 2015.

[23] O. Barajas. How the internet of things (iot) is changing the cybersecurity landscape. `https://securityintelligence.com/how-the-internet-of-things-iot-is-changing-the-cybersecurity-landscape/`, 09 2014. [Accessed: 07 November 2022].

[24] R. Barrett. *Building a Values-driven Organization: A Whole System Approach to Cultural Transformation*. Butterworth-Heinemann, 2006.

[25] B. Bartholomew and J. A. Guerrero-Saade. Wave your false flags! deception tactics muddying attribution in targeted attacks. In *Virus Bulletin Conference*, pages 1–9, 2016.

[26] V. Boehme-Neßler. Privacy: a matter of democracy. why democracy needs privacy and data protection. *International Data Privacy Law*, 6(3):222–229, 2016.

[27] T. Boellstorff. The metaverse isn't here yet, but it already has a long history. Technical Report 186083, The Conversation, Aug. 2022. [Accessed: 07 November 2022].

[28] J. Bonneau, C. Herley, P. C. v. Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *2012 IEEE Symposium on Security and Privacy*, pages 553–567, 2012.

[29] D. Braue. Global ransomware damage costs predicted to exceed $265 billion by 2031. `https://cybersecurityventures.com/global-ransomware-damage-costs-predicted-to-reach-250-billion-usd-by-2031`, Jun 2022. [Accessed: 07 November 2022].

[30] R. Brown, V. Ta, D. Bienstock, G. Ackerman, and J. Wolfram. Does this look infected? A summary of APT41 targeting U.S. state governments. `https://www.mandiant.com/resources/blog/apt41-us-state-governments`, 2022. [Accessed: 07 November 2022].

[31] M. A. S. Bubukayr and M. A. Almaiah. Cybersecurity concerns in smart-phones and applications: A survey. In *2021 International Conference on Information Technology (ICIT)*, pages 725–731, 2021.

[32] B. Bulgurcu, H. Cavusoglu, and I. Benbasat. Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Q.*, 34(3):523–548, sep 2010.

[33] M. BULL. Ex-burglars warn homeowners of social media posts that put property at risk of a break-in. `https://www.express.co.uk/life-style/property/1559309/burglar-tips-hacks-social-media-posts-break-ins-property`. Accessed: 2022-11-119.

[34] E. Bursztein. Inside the infamous Mirai IOT Botnet: A retrospective analysis. `https://blog.cloudflare.com/inside-mirai-the-infamous-iot-botnet-a-retrospective-analysis`, Sep 2021. [Accessed: 07 November 2022].

[35] A. Cavoukian et al. Privacy by design: The 7 foundational principles. *Information and privacy commissioner of Ontario, Canada*, 5:2009, 2009.

[36] CC. Common criteria for information technology security evaluation. https://www.google.com/search?client=safari&rls=en&q=common+criteria&ie=UTF-8&oe=UTF-8. [Accessed: 07 November 2022].

[37] D. Champagne and R. B. Lee. Scalable architectural support for trusted software. In *HPCA - 16 2010 The Sixteenth International Symposium on High-Performance Computer Architecture*, pages 1–12, 2010.

[38] H. Chang and R. Shokri. On the privacy risks of algorithmic fairness. In *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 292–303, 2021.

[39] S. Chaudhary, V. Gkioulos, and S. Katsikas. Developing metrics to assess the effectiveness of cybersecurity awareness program. *Journal of Cybersecurity*, 8(1), 05 2022. tyac006.

[40] G. Cheng, P. Zhou, and J. Han. Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7405–7415, 2016.

[41] R. Choudhry and K. Garg. A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, 39, 01 2008.

[42] N. Chouliaras, G. Kittes, I. Kantzavelou, L. Maglaras, G. Pantziou, and M. A. Ferrag. Cyber ranges and testbeds for education, training, and research. *Applied Sciences*, 11(4), 2021.

[43] N. Chowdhury and V. Gkioulos. Cyber security training for critical infrastructure protection: A literature review. *Computer Science Review*, 40:100361, 2021.

[44] E. Commission. Commission work programme 2023. https://ec.europa.eu/info/sites/default/files/cwp_2023.pdf. [Accessed: 07 November 2022].

[45] E. Commission. Cybersecurity policies. https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-policies. [Accessed: 07 November 2022].

[46] E. Commission. The digital services act package. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package. [Accessed: 07 November 2022].

[47] E. Commission. Directive (eu) 2016/1148 of the european parliament and of the council of 6 july 2016 concerning measures for a high common level of security of network and information systems across the union (NIS). = https://eur-lex.europa.eu/eli/dir/2016/1148/oj. [Accessed: 07 November 2022].

[48] E. Commission. European data governance act. https://digital-strategy.ec.europa.eu/en/policies/data-act. [Accessed: 07 November 2022].

[49] E. Commission. General data protection regulation. https://eur-lex.europa.eu/eli/reg/2016/679/oj. [Accessed: 07 November 2022].

[50] E. Commission. People, technologies & infrastructure – europe's plan to thrive in the metaverse. https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_22_5525. [Accessed: 07 November 2022].

[51] Conference of the Independent Data Protection Supervisory Authorities of the Federation and the Länder. The standard data protection model – a method for data protection advising and controlling on the basis of uniform protection goals, version 2.0b (english version). https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methodology_V2.0b.pdf, 2020. [Accessed: 07 November 2022].

[52] K. Conger and K. Roose. Uber investigating breach of its computer systems. https://www.nytimes.com/2022/09/15/technology/uber-hacking-breach.html, Sep 2022. [Accessed: 07 November 2022].

[53] M. Corporation. Common vulnerabilities and exposures (cve) details: The ultimate security vulnerability datasource. https://www.cvedetails.com/browse-by-date.php. [Accessed: 20 November 2022].

[54] V. Costan, I. Lebedev, and S. Devadas. Sanctum: Minimal hardware extensions for strong software isolation. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 857–874, Austin, TX, Aug. 2016. USENIX Association.

[55] CyberSec4Europe. Flagship 2: The successful second cybersecurity exercise hosted by cybersec4europe. `https://cybersec4europe.eu/flagship-2-the-successful-second-cybersecurity-exercise-hosted-by-cybersec4europe/`, 03 2022. [Accessed: 07 November 2022].

[56] CyBOK. The cyber security body of knowledge. `https://www.cybok.org`. [Accessed: 07 November 2022].

[57] A. Da Veiga and N. Martins. Information security culture and information protection culture: A validated assessment instrument. *Computer Law & Security Review*, 31(2):243–256, 2015.

[58] S. Dange and M. Chatterjee. Iot botnet: The largest threat to the iot network. *Advances in Intelligent Systems and Computing*, 1049:137–157, 2020.

[59] T. DeMarco. *Structured Analysis and System Specification*. Yourdon Press, 1979.

[60] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen. A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements. *Requir. Eng.*, 16(1):3–32, mar 2011.

[61] E. W. Dijkstra et al. Notes on structured programming. *Section 3 On The Reliability of Mechanisms, corollary at the end*, 1970.

[62] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, SSYM'04, page 21, USA, 2004. USENIX Association.

[63] A. Dionysiou, M. Agathocleous, C. Christodoulou, and V. Promponas. Convolutional neural networks in combination with support vector machines for complex sequential data classification. In V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis, editors, *Artificial Neural Networks and Machine Learning – ICANN 2018*, pages 444–455, Cham, 2018. Springer International Publishing.

[64] C. Directive. Council directive 2008/114/ec of 8 december 2008–on the identification and designation of european critical infrastructures and the assessment of the need to improve their protection. *Official Journal of the European Union. L*, 345:75–82, 2008.

[65] J. Drees. Software bug in new jersey hospital's vaccine scheduling system causes thousands of duplicate appointments. `https://www.beckershospitalreview.com/healthcare-information-technology/software-bug-in-new-jersey-hospital-s-vaccine-scheduling-system-causes-thousands-of-duplicate-appointments.html`. Accessed: 2022-11-119.

[66] Z. Durumeric, E. Wustrow, and J. A. Halderman. {ZMap}: Fast internet-wide scanning and its security applications. In *22nd USENIX Security Symposium (USENIX Security 13)*, pages 605–620, Washington, D.C., 2013. USENIX Association.

[67] Y. K. Dwivedi, L. Hughes, A. M. Baabdullah, S. Ribeiro-Navarrete, M. Giannakis, M. M. Al-Debei, D. Dennehy, B. Metri, D. Buhalis, C. M. Cheung, K. Conboy, R. Doyle, R. Dubey, V. Dutot, R. Felix, D. Goyal, A. Gustafsson, C. Hinsch, I. Jebabli, M. Janssen, Y.-G. Kim, J. Kim, S. Koos, D. Kreps, N. Kshetri, V. Kumar, K.-B. Ooi, S. Papagiannidis, I. O. Pappas, A. Polyviou, S.-M. Park, N. Pandey, M. M. Queiroz, R. Raman, P. A. Rauschnabel, A. Shirish, M. Sigala, K. Spanaki, G. Wei-Han Tan, M. K. Tiwari, G. Viglia, and S. F. Wamba. Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 66:102542, 2022.

[68] C. Dwork. Differential privacy: A survey of results. In M. Agrawal, D. Du, Z. Duan, and A. Li, editors, *Theory and Applications of Models of Computation*, pages 1–19, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[69] C. Dwork and D. K. Mulligan. It's not privacy, and it's not fair. *Stan. L. Rev. Online*, 66:35, 2013.

[70] Enarx. `https://enarx.dev/`. [Accessed: 07 November 2022].

[71] ENISA. Security economics and the internal market. `https://www.enisa.europa.eu/publications/archive/economics-sec`. [Accessed: 07 November 2022].

[72] ENISA. Understanding the increase in supply chain security attacks. `https://www.enisa.europa.eu/news/enisa-news/understanding-the-increase-in-supply-chain-security-attacks`. [Accessed: 07 November 2022].

[73] ENISA. Artificial intelligence cybersecurity challenges. *European Union Agency for Cybersecurity (ENISA)*, Aug 2021.

[74] ENISA. Cybersecurity challenges in the uptake of artificial intelligence in autonomous driving. *European Union Agency for Cybersecurity (ENISA)*, Aug 2021.

[75] ENISA. ENISA threat landscape for supply chain attacks. *European Union Agency for Cybersecurity (ENISA)*, 2021.

[76] ENISA. Securing machine learning algorithms. *European Union Agency for Cybersecurity (ENISA)*, Dec 2021.

[77] ENISA. Tips for secure user authentication, Aug 2021.

[78] ETSI. Etsi en 303 645. `https://www.etsi.org/deliver/etsi_en/303600_303699/303645/02.01.00_30/en_303645v020100v.pdf`. [Accessed: 07 November 2022].

[79] European Commission. Cyber resilience act. `https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act`. Accessed: 2022-11-119.

[80] Europol. World's most dangerous malware emotet disrupted through global action. `https://www.europol.europa.eu/media-press/newsroom/news/world%E2%80%99s-most-dangerous-malware-emotet-disrupted-through-global-action`, 2021.

[81] Eurostat. Individuals - internet activities. `https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_ac_i/default/table?lang=en`. Accessed: 2022-11-119.

[82] S. Fischer-Hübner, C. Alcaraz, A. Ferreira, C. Fernandez-Gago, J. Lopez, E. Markatos, L. Islami, and M. Akil. Stakeholder perspectives and requirements on cybersecurity in europe. *Journal of Information Security and Applications*, 61:102916, 2021.

[83] D. Florencio and C. Herley. A large-scale study of web password habits. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, page 657–666, New York, NY, USA, 2007. Association for Computing Machinery.

[84] T. W. E. Forum. Defining and building the metaverse. Technical report, weforum.org, Jan. 2022. [Accessed 28-Sep-2022].

[85] M. Fredrikson, S. Jha, and T. Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, page 1322–1333, New York, NY, USA, 2015. Association for Computing Machinery.

[86] Freedom of the Press Foundation. SecureDrop Threat Model. `https://docs.securedrop.org/en/stable/threat_model/threat_model.html`, 2022. [Accessed: 07 November 2022].

[87] J. Freund and J. Jones. *Measuring and Managing Information Risk: A FAIR Approach*. Butterworth-Heinemann, 2014.

[88] J. E. Gaffney. Estimating the number of faults in code. *IEEE Transactions on Software Engineering*, SE-10(4):459–464, 1984.

[89] T. Gagliardoni. The poly network hack explained. `https://research.kudelskisecurity.com/2021/08/12/the-poly-network-hack-explained`. [Accessed: 07 November 2022].

[90] B. Gardner and V. Thomas. *Building an Information Security Awareness Program: Defending Against Social Engineering and Technical Threats*. Syngress Publishing, 1st edition, 2014.

[91] V. Garousi, A. Rainer, P. Lauvås, and A. Arcuri. Software-testing education: A systematic literature mapping. *Journal of Systems and Software*, 165:110570, 2020.

[92] S. Gatlan. Chinese hackers use new windows malware to backdoor govt, defense orgs. `https://www.bleepingcomputer.com/news/security/chinese-hackers-use-new-windows-malware-to-backdoor-govt-defense-orgs/`, Aug 2022. [Accessed: 07 November 2022].

[93] GDPR. General Data Protection Regulation. `https://gdpr-info.eu`. [Accessed: 07 November 2022].

[94] T. Geppert, S. Deml, D. Sturzenegger, and N. Ebert. Trusted execution environments: Applications and organizational challenges. *Frontiers in Computer Science*, 4, 2022.

[95] S. Gilbert. The political economy of the metaverse. Technical report, Briefings de l'IFRI, IFRI, June 2022. [Accessed: 07 November 2022].

[96] I. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. *arXiv 1412.6572*, 12 2014.

[97] Google. Asylo. `https://asylo.dev/`. [Accessed: 17 November 2022].

[98] gramine. Gramine. `https://gramineproject.io/`. [Accessed: 07 November 2022].

[99] A. Graves, A.-r. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649, 2013.

[100] A. Greenberg. Hackers remotely kill a jeep on the highway—with me in it. `https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway`, 7 2015. [Accessed: 07 November 2022].

[101] L. Grindstaff. Through your mind's eye: What biases are impacting your security posture? `https://www.mcafee.com/blogs/other-blogs/executive-perspectives/through-your-minds-eye-what-biases-are-impacting-your-security-posture/`, 05 2021. [Accessed: 07 November 2022].

[102] S. Gürses and J. M. Del Alamo. Privacy engineering: Shaping an emerging field of research and practice. *IEEE Security & Privacy*, 14(2):40–46, 2016.

[103] I. Hadar, T. Hasson, O. Ayalon, E. Toch, M. Birnhack, S. Sherman, and A. Balissa. Privacy by designers: Software developers' privacy mindset. *Empirical Softw. Engg.*, 23(1):259–289, feb 2018.

[104] M. Hansen, M. Jensen, and M. Rost. Protection goals for privacy engineering. In *Proceedings of the 2015 IEEE Security and Privacy Workshops*, SPW '15, page 159–166, USA, 2015. IEEE Computer Society.

[105] A. Harish. When nasa lost a spacecraft due to a metric math mistake. `https://www.simscale.com/blog/nasa-mars-climate-orbiter-metric/`. Accessed: 2022-11-119.

[106] M. Hasan. Number of connected iot devices growing 18% to 14.4 billion globally. `https://iot-analytics.com/number-connected-iot-devices/`, 5 2022. [Accessed: 07 November 2022].

[107] N. Hasan, A. Chamoli, and M. Alam. Privacy challenges and their solutions in iot. *Internet of Things (IoT): Concepts and Applications*, pages 219–231, 1 2020.

[108] J. Haworth. Zero-day flaws in iot baby monitors could give attackers access to camera feeds. `https://portswigger.net/daily-swig/zero-day-flaws-in-iot-baby-monitors-could-give-attackers-access-to-camera-feeds`, 9 2021. [Accessed: 07 November 2022].

[109] J. L. Hernández-Ramos, G. Baldini, S. N. Matheu, and A. Skarmeta. Updating iot devices: challenges and potential approaches. *GIoTS 2020 - Global Internet of Things Summit, Proceedings*, pages 1–5, 2020.

[110] J. Hodges, J. Jones, M. B. Jones, A. Kumar, and E. Lundberg. Web authentication: An api for accessing public key credentials level 2.

[111] J.-H. Hoepman. Privacy Design Strategies. In N. Cuppens-Boulahia, F. Cuppens, S. Jajodia, A. Abou El Kalam, and T. Sans, editors, *ICT Systems Security and Privacy Protection*, pages 446–459, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[112] M. Howard and S. Lipner. *The Security Development Lifecycle*. Microsoft Press, 2006.

[113] T. Hunt. Pwned websites.

[114] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain. Machine learning in iot security: Current solutions and future challenges. *IEEE Communications Surveys & Tutorials*, 22(3):1686–1721, 2020.

[115] i SCOOP. Iot security and the consumer: the challenges and education question. = https://www.i-scoop.eu/iot-security-consumer-education/. [Accessed: 07 November 2022].

[116] IBM. Cost of a data breach report 2022. `https://newsroom.ibm.com/2022-07-27-IBM-Report-Consumers-Pay-the-Price-as-Data-Breach-Costs-Reach-All-Time-High`, July 2022. [Accessed: 07 November 2022].

[117] J. Inclan. Emotet exposed: A look inside the cybercriminal supply chain. `https://blogs.vmware.com/security/2022/10/emotet-exposed-a-look-inside-the-cybercriminal-supply-chain.html`, 2022.

[118] Inmarsat. Industrial iot in the time of covid-19. `https://www.inmarsat.com/en/insights/enterprise/2021/research-programme-2021-industrial-iot-covid-19.html`, 2021. [Accessed: 07 November 2022].

[119] Intel. In *Intel Software Guard Extensions Programming Reference*. ARM, October 2014. [Accessed: 16 November 2022].

[120] IriusRisk. IriusRisk. `https://www.iriusrisk.com/`, 2022. [Accessed: 07 November 2022].

[121] L. Islami, S. Fischer-Hübner, and P. Papadimitratos. Capturing drivers' privacy preferences for intelligent transportation systems: An intercultural perspective. *Computers & Security*, 123:102913, 2022.

[122] L. H. Iwaya, G. H. Iwaya, S. Fischer-Hübner, and A. V. Steil. Organisational privacy culture and climate: A scoping review. *IEEE Access*, 10:73907–73930, 2022.

[123] J. Jalkanen. *Is Human the Weakest Link in Information Security? Systematic Literature Review*. University of Jyväskylä, Jyväskylä , Finland, 2019.

[124] S. V. Joshi, D. Stubbe, S.-T. T. Li, and D. M. Hilty. The use of technology by youth: Implications for psychiatric educators. *Academic Psychiatry*, 43(1):101–109, 2019.

[125] D. Kaplan, J. Powell, and T. Woller. In *AMD Memory Encryption*. AMD, April 2016. [Accessed: 16 November 2022].

[126] G. Karantzas and C. Patsakis. An empirical assessment of endpoint detection and response systems against advanced persistent threats attack vectors. *Journal of Cybersecurity and Privacy*, 1(3):387–421, 2021.

[127] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4396–4405, 2019.

[128] Kaspersky. The human factor in it security: How employees are making businesses vulnerable from within. `https://www.kaspersky.com/blog/the-human-factor-in-it-security/`, 11 2022. [Accessed: 07 November 2022].

[129] N. Kohl and P. Stone. Policy gradient reinforcement learning for fast quadrupedal locomotion. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, volume 3, pages 2619–2624 Vol.3, 2004.

[130] L. Kohnfelder and P. Garg. The threats to our products. *Microsoft Interface, Microsoft Corporation*, 33, 1999.

[131] I. Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1):89–109, 2001.

[132] V. Koutsokostas and C. Patsakis. Python and malware: Developing stealth and evasive malware without obfuscation. In S. D. C. di Vimercati and P. Samarati, editors, *Proceedings of the 18th International Conference on Security and Cryptography, SECRYPT 2021, July 6-8, 2021*, pages 125–136. SCITEPRESS, 2021.

[133] A. Küchler, A. Mantovani, Y. Han, L. Bilge, and D. Balzarotti. Does every second count? time-based evolution of malware behavior in sandboxes. In *28th Annual Network and Distributed System Security Symposium, NDSS 2021, virtually, February 21-25, 2021*. The Internet Society, 2021.

[134] I. Kuzminykh, B. Ghita, and J. M. Such. The challenges with internet of things for business. `https://arxiv.org/abs/2012.03589`, 12 2020. [Accessed: 07 November 2022].

[135] A. L. Lafuente, F. Nielson, S. Mödersheim, A. Schlichtkrull, A. Sforzin, C. Sorientea, L. Kamm, R. Martins, J. Soares, L. Antunes, L. Durante, M. Cheminod, E. Athanasopoulos, B. Hamid, A. Omerovic, K. Bernsmed, and R. S. Per H Meland. Research challenges and requirements for secure software development. `https://cybersec4europe.eu/wp-content/uploads/2020/09/CS4E-D3.9-Research-challenges-and-requirements-for-secure-software-development-v1.1-Submitted.pdf`, 2020. [Accessed: 07 November 2022].

[136] T. LAMBERT and B. DONOHUE. It's all fun and games until ransomware deletes the shadow copies. `https://redcanary.com/blog/its-all-fun-and-games-until-ransomware-deletes-the-shadow-copies`, 2022. [Accessed: 07 November 2022].

[137] R. Langner. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, 9(3):49–51, 2011.

[138] D. Lee, D. Kohlbrenner, S. Shinde, K. Asanović, and D. Song. Keystone: An open framework for architecting trusted execution environments. In *Proceedings of the Fifteenth European Conference on Computer Systems*, EuroSys '20, New York, NY, USA, 2020. Association for Computing Machinery.

[139] K. LeFevre, D. DeWitt, and R. Ramakrishnan. Mondrian multidimensional k-anonymity. In *22nd International Conference on Data Engineering (ICDE'06)*, pages 25–25, 2006.

[140] N. G. Leveson. The therac-25: 30 years later. *Computer*, 50(11):8–11, 2017.

[141] M. N. Lintvedt. Putting a price on data protection infringement. *International Data Privacy Law*, 12(1):1–15, 12 2021.

[142] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin. When machine learning meets privacy: A survey and outlook. *ACM Comput. Surv.*, 54(2), mar 2021.

[143] J. Lopez, C. Alcaraz, and R. Roman. Smart control of operational threats in control substations. *Computers & Security*, 38:14–27, 2013. Cybercrime in the Digital Economy.

[144] P. Lorenzo, F. Stefano, A. Ferreira, and P. Carolina. Artificial intelligence and cybersecurity: Technology, governance and policy challenges. `https://www.ceps.eu/wp-content/uploads/2021/05/CEPS-TFR-Artificial-Intelligence-and-Cybersecurity.pdf`, 2021. [Accessed: 07 November 2022].

[145] T. Madiega, P. Car, and M. N. with Louise Van de Pol. Metaverse: Opportunities, risks and policy implications. Technical Report PE 733.557, European Parliamentary Research Service, June 2022. [Accessed: 07 November 2022].

[146] S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick. Credit card fraud detection using bayesian and neural networks. In *Proceedings of the 1st international naiso congress on neuro fuzzy technologies*, pages 261–270, 08 2002.

[147] K. Manheim and L. Kaplan. Artificial intelligence: Risks to privacy and democracy. *Yale JL & Tech.*, 21:106, 2019.

[148] K. Marky, K. Ragozin, G. Chernyshov, A. Matviienko, M. Schmitz, M. Mühlhäuser, C. Eghtebas, and K. Kunze. "nah, it's just annoying!" a deep dive into user perceptions of two-factor authentication. *ACM Transactions on Computer-Human Interaction*, 29(5), oct 2022.

[149] G. McGraw. *Software Security: Building Security In*. Addison-Wesley Professional, 2006.

[150] metaverse standards.org. The metaverse standards forum. Technical report, metaverse-standards.org, June 2022. [Accessed: 07 November 2022].

[151] MicroAge. The benefits of cybersecurity awareness training. `https://microage.ca/the-benefits-of-cybersecurity-awareness-training/`, 10 2022. [Accessed: 07 November 2022].

[152] Microsoft. Create and use strong passwords — support.microsoft.com. `https://support.microsoft.com/en-us/windows/create-and-use-strong-passwords-c5cebb49-8c53-4f5e-2bc4-fe357ca048eb`. [Accessed: 07 November 2022].

[153] Microsoft. Microsoft sdl. `https://www.microsoft.com/en-us/securityengineering/sdl/practices`. [Accessed: 07 November 2022].

[154] Microsoft. Applications for artificial intelligence in department of defense cyber missions. `https://blogs.microsoft.com/on-the-issues/2022/05/03/artificial-intelligence-department-of-defense-cyber-missions/`, 05 2022. [Accessed: 07 November 2022].

[155] Microsoft. Special Report: Ukraine: An overview of Russia's cyberattack activity in Ukraine. `https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4Vwwd`, 2022. [Accessed: 07 November 2022].

[156] Microsoft Corporation. Microsoft Threat Modeling Tool 7. `https://aka.ms/threatmodelingtool`, 2022. [Accessed: 07 November 2022].

[157] J. Miranda, N. Mäkitalo, J. Garcia-Alonso, J. Berrocal, T. Mikkonen, C. Canal, and J. M. Murillo. From the internet of things to the internet of people. *IEEE Internet Computing*, 19(2):40–47, 2015.

[158] H. Modi. Netscout threat intelligence report. `https://www.netscout.com/sites/default/files/2019-02/SECR_001_EN-1901%20-%20NETSCOUT%20Threat%20Intelligence%20Report%202H%202018.pdf`, 2018. [Accessed: 07 November 2022].

[159] Movie. Ready player one, 2018. [Accessed: 07 November 2022].

[160] P. Muncaster. Hotel guests locked out of rooms after ransomware attack. `https://www.infosecurity-magazine.com/news/hotel-guests-locked-out-rooms`, Dec 2021. [Accessed: 07 November 2022].

[161] S. Muppidi, L. Fisher, and G. Parham. Ai and automation for cybersecurity. Technical report, IBM Corporation, June 2022. [Accessed 07-November-2022].

[162] A. S. Namin, Z. Aguirre-Muñoz, and K. S. Jones. Teaching cyber security through competition: An experience report about a participatory training workshop. In *International Conference on Computer Science Education Innovation & Technology (CSEIT). Proceedings*, page 98. Global Science and Technology Forum, 2016.

[163] National Highway Traffic Safety Administration. Part 573 safety recall report. `https://static.nhtsa.gov/odi/rcl/2021/RCLRPT-21V035-4682.PDF`. Accessed: 2022-11-119.

[164] T. Ncubukezi. Human errors: A cybersecurity concern and the weakest link to small businesses. *International Conference on Cyber Warfare and Security*, 17:395–403, 03 2022.

[165] L. H. Newman. Millions of web camera and baby monitor feeds are exposed. `https://www.wired.com/story/kalay-iot-bug-video-feeds/`, 2017. [Accessed: 07 November 2022].

[166] Z. Ning, F. Zhang, W. Shi, and W. Shi. Position paper: Challenges towards securing hardware-assisted execution environments. In *Proceedings of the Hardware and Architectural Support for Security and Privacy*, HASP '17, New York, NY, USA, 2017. Association for Computing Machinery.

[167] NIST. Secure software development framework. `https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-218.pdf`. [Accessed: 07 November 2022].

[168] NIST. Nist announces first four quantum-resistant cryptographic algorithms. `https://www.nist.gov/news-events/news/2022/07/nist-announces-first-four-quantum-resistant-cryptographic-algorithms`, July 2022. [Accessed: 07 November 2022].

[169] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang. Abusive language detection in online user content. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, page 145–153, Republic and Canton of Geneva, CHE, 2016. International World Wide Web Conferences Steering Committee.

[170] C. Ntantogian, S. Malliaros, and C. Xenakis. Evaluation of password hashing schemes in open source web platforms. *Computers & Security*, 84:206–224, 2019.

[171] M. Nunes, P. Burnap, P. Reinecke, and K. Lloyd. Bane or boon: Measuring the effect of evasive malware on system call classifiers. *J. Inf. Secur. Appl.*, 67(C), jun 2022.

[172] occlum. Occlum. `https://occlum.io/`. [Accessed: 07 November 2022].

[173] N. I. of Standards and Technology. Digital identity guidelines: Authentication and lifecycle management. Technical report, U.S. Department of Commerce, Washington, D.C., 2017.

[174] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.

[175] OMIGroup. Open metaverse interoperability group. Technical report, OMI Group, Sept. 2022. [Accessed: 07 November 2022].

[176] P. H. O'Neill. Ransomware did not kill a german hospital patient. `https://www.technologyreview.com/2020/11/12/1012015/ransomware-did-not-kill-a-german-hospital-patient`, Nov 2020. [Accessed: 07 November 2022].

[177] opentitan. Opentitan. `https://opentitan.org/`. [Accessed: 07 November 2022].

[178] C. Osborne. Mirai splinter botnets dominate iot attack scene. = https://www.zdnet.com/article/mirai-splinter-botnets-dominate-iot-attack-scene/, 1 2022. [Accessed: 07 November 2022].

[179] M. Ovelgönne, T. Dumitraş, B. A. Prakash, V. S. Subrahmanian, and B. Wang. Understanding the relationship between human behavior and susceptibility to cyber attacks: A data-driven approach. *ACM Trans. Intell. Syst. Technol.*, 8(4), mar 2017.

[180] OWASP. OWASP top 10 - 2021. `https://owasp.org/Top10/`, 2021. [Accessed: 07 November 2022].

[181] E. Papadogiannakis, P. Papadopoulos, N. Kourtellis, and E. P. Markatos. User tracking in the post-cookie era: How websites bypass gdpr consent to track users. In *Proceedings of the Web Conference 2021*, WWW '21, page 2130–2141, New York, NY, USA, 2021. Association for Computing Machinery.

[182] N. Papernot, P. McDaniel, A. Sinha, and M. P. Wellman. Sok: Security and privacy in machine learning. In *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 399–414, 2018.

[183] M. Paquet-Clouston, M. Romiti, B. Haslhofer, and T. Charvat. Spams meet cryptocurrencies: Sextortion in the bitcoin ecosystem. In *Proceedings of the 1st ACM Conference on Advances in Financial Technologies*, AFT '19, page 76–88, New York, NY, USA, 2019. Association for Computing Machinery.

[184] C. Patsakis and F. Casino. Hydras and ipfs: a decentralised playground for malware. *International Journal of Information Security*, 18(6):787–799, 2019.

[185] C. Patsakis and A. Chrysanthou. Analysing the fall 2020 emotet campaign. *arXiv preprint arXiv:2011.06479*, 2020.

[186] S. Peldszus, K. Tuma, D. Strüber, J. Jürjens, and R. Scandariato. Secure data-flow compliance checks between models and code based on automated mappings. In *2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems (MODELS)*, pages 23–33, 2019.

[187] R. P. Pires. Distributed systems and trusted execution environments: Trade-offs and challenges. `https://arxiv.org/pdf/2001.09670.pdf`, December 2019. [Accessed: 20 November 2022].

[188] S. Pletinckx, C. Trap, and C. Doerr. Malware coordination using the blockchain: An analysis of the cerber ransomware. In *2018 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9, 2018.

[189] N. Popper. Knight capital says trading glitch cost it $ 440 million. `https://archive.nytimes.com/dealbook.nytimes.com/2012/08/02/knight-capital-says-trading-mishap-cost-it-440-million/`. Accessed: 2022-11-119.

[190] W. Presthus and K. F. Sønslien. An analysis of violations and sanctions following the gdpr. *International Journal of Information Systems and Project Management*, 9(1):38–53, Sep 2021.

[191] pwc. Conti cyber attack on the HSE. `https://www.hse.ie/eng/services/publications/conti-cyber-attack-on-the-hse-full-report.pdf`, 2021. [Accessed: 07 November 2022].

[192] D. Reading. Popular iot cameras need patching to fend off catastrophic attacks. `https://www.darkreading.com/attacks-breaches/popular-iot-cameras-patching-catastrophic-attacks`, 9 2022. [Accessed: 07 November 2022].

[193] D. Rehak, P. Senovsky, M. Hromada, and T. Lovecek. Complex approach to assessing resilience of critical infrastructure elements. *International journal of critical infrastructure protection*, 25:125–138, 2019.

[194] Reuters. AXA division in asia hit by ransomware cyber attack. `https://www.reuters.com/article/us-axa-cyber-idUSKCN2CX0B0`, May 2021. [Accessed: 07 November 2022].

[195] Reuters. Danish train standstill on saturday caused by cyber attack. `https://www.reuters.com/technology/danish-train-standstill-saturday-caused-by-cyber-attack-2022-11-03/`, Nov 2022. [Accessed: 07 November 2022].

[196] S. Rinaldi, J. Peerenboom, and T. Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine*, 21(6):11–25, 2001.

[197] riscure. Security pitfalls in tee development. `https://www.riscure.com/publication/security-pitfalls-in-tee-development/`. [Accessed: 07 November 2022].

[198] A. ROBERTSON. Most killers stalk their victims on social media before murdering them, say criminologists. `https://www.dailymail.co.uk/news/article-4439130/Most-killers-stalk-victims-social-media-murder.html`. Accessed: 2022-11-119.

[199] A. Rowe. Study Reveals Average Person Has 100 Passwords | Tech.co — tech.co. https://tech.co/password-managers/how-many-passwords-average-person, 2021. [Accessed: 07 November 2022].

[200] P. Ruggiero and J. Foote. Cyber threats to mobile phones. https://www.cisa.gov/uscert/sites/default/files/publications/cyber_threats_to_mobile_phones.pdf, 2011. [Accessed: 07 November 2022].

[201] O. SAMM. Software assurance maturity model. https://owaspsamm.org/model/. [Accessed: 07 November 2022].

[202] Sanket. The exponential cost of fixing bugs. https://deepsource.io/blog/exponential-cost-of-fixing-bugs/. [Accessed: 07 November 2022].

[203] I. H. Sarker, A. I. Khan, Y. B. Abushark, and F. Alsolami. Internet of things (iot) security intelligence: A comprehensive overview, machine learning solutions and research directions. *Mobile Networks and Applications*, 1:1–17, 3 2022.

[204] T. Schaberreiter, K. Kittilä, K. Halunen, J. Röning, and D. Khadraoui. Risk assessment in critical infrastructure security modelling based on dependency analysis. In *International Workshop on Critical Information Infrastructures Security*, pages 213–217. Springer, 2011.

[205] P. M. Schwartz. Privacy and democracy in cyberspace. *Vand. L. Rev.*, 52:1607, 1999.

[206] T. Seals. Mozi botnet accounts for majority of iot traffic. = https://threatpost.com/mozi-botnet-majority-iot-traffic/159337/, 9 2020. [Accessed: 07 November 2022].

[207] F. T. Security. Boost engagement with serious game training. https://terranovasecurity.com/serious-game/, 10 2022. [Accessed: 07 November 2022].

[208] R. Setola, S. De Porcellinis, and M. Sforna. Critical infrastructure dependency assessment using the input–output inoperability model. *International Journal of Critical Infrastructure Protection*, 2(4):170–178, 2009.

[209] R. Setola, V. Rosato, E. Kyriakides, and E. Rome. *Managing the complexity of critical infrastructures: A modelling and simulation approach*. Springer Nature, 2016.

[210] M. Shahraeini, P. Kotzanikolaou, and M. Nasrolahi. Communication resilience for smart grids based on dependence graphs and eigenspectral analysis. *IEEE Systems Journal*, pages 1–11, 2022.

[211] R. Shokri, M. Stronati, C. Song, and V. Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3–18, 2017.

[212] A. Shostack. *Threat Modeling: Designing for Security*. John Wiley & Sons, Indianapolis, Indiana, 2014.

[213] J. Sigholm, G. Falco, and A. Viswanathan. Enhancing cybersecurity education through high-fidelity live exercises (hiflix). In *HICSS*, 01 2019.

[214] L. Sion, D. Van Landuyt, K. Yskout, S. Verreydt, and W. Joosen. Automated threat analysis and management in a continuous integration pipeline. In *2021 IEEE Secure Development Conference (SecDev)*, pages 30–37, 2021.

[215] L. Sion, K. Yskout, D. Van Landuyt, and W. Joosen. Risk-based Design Security Analysis. In *Proceedings - 2018 IEEE/ACM First International Workshop on Security Awareness from Design to Deployment, SEAD 2018*, page 11–18, New York, NY, USA, 2018. Association for Computing Machinery.

[216] J. Smart, N. Cascio, and J. Paffendorf. Metaverse roadmap – pathways to the 3d web: A cross - industry public foresight project. https://metaverseroadmap.org/MetaverseRoadmapOverview.pdf. [Accessed: 07 November 2022].

[217] I. Sommerville. Software engineering 10. *Harlow: Pearson Education Limited*, 2016.

[218] Sophos. The state of ransomware 2022. https://assets.sophos.com/X24WTUEQ/at/4zpw59pnkpxxnhfhgj9bxgj9/sophos-state-of-ransomware-2022-wp.pdf, 2022. [Accessed: 07 November 2022].

[219] M. Sporny, D. Longley, M. Sabadello, D. Reed, O. Steele, and C. Allen. Decentralized identifiers (dids).

[220] M. Sporny, G. Noble, D. Longley, D. C. Burnett, B. Zundel, and K. D. Hartog. Verifiable credentials data model.

[221] R. Steen. 5 reasons automation can't take over cybersecurity. https://www.securitymagazine.com/articles/98396-5-reasons-automation-cant-take-over-cybersecurity, 9 2022. [Accessed: 07 November 2022].

[222] I. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz, and J. Lopez. A survey of iot-enabled cyberattacks: Assessing attack paths to critical infrastructures and services. *IEEE Communications Surveys & Tutorials*, 20(4):3453–3495, 2018.

[223] N. Stephenson. *Snow Crash*. Bantam Books, United States of America, 1992.

[224] G. Stergiopoulos, P. Kotzanikolaou, M. Theocharidou, G. Lykou, and D. Gritzalis. Time-based critical infrastructure dependency analysis for large-scale and cross-sectoral failures. *International Journal of Critical Infrastructure Protection*, 12:46–60, 2016.

[225] G. E. Suh, D. Clarke, B. Gassend, M. van Dijk, and S. Devadas. Aegis: Architecture for tamper-evident and tamper-resistant processing. In *Proceedings of the 17th Annual International Conference on Supercomputing*, ICS '03, page 160–171, New York, NY, USA, 2003. Association for Computing Machinery.

[226] L. Sweeney. K-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, oct 2002.

[227] F. Swiderski and W. Snyder. *Threat modeling*. Microsoft Press, 2004.

[228] J. Taylor. Facebook outage: what went wrong and why did it take so long to fix after social platform went down? https://www.theguardian.com/technology/2021/oct/05/facebook-outage-what-went-wrong-and-why-did-it-take-so-long-to-fix. Accessed: 2022-11-119.

[229] W. C. D. Team. Metaverse interoperability community group. Technical report, w3c.org, Sept. 2022. [Accessed: 07 November 2022].

[230] W. Technologies. New research: Fileless malware attacks surge by 900% and cryptominers make a comeback, while ransomware attacks decline. https://www.watchguard.com/wgrd-about/press-releases/new-research-fileless-malware-attacks-surge-900-and-cryptominers-make, 2021. [Accessed: 07 November 2022].

[231] B. Tekinerdogan. Architectural drift analysis using architecture reflexion viewpoint and design structure reflexion matrices. In *Software Quality Assurance*, pages 221–236. Elsevier, 2016.

[232] P. Torr. Demystifying the threat modeling process. *IEEE Security & Privacy*, 3(5):66–70, 2005.

[233] Trail of Bits. Kubernetes Threat Model. https://github.com/kubernetes/community/raw/683ec8f8a392522933b8950a052dfdce6da6a812/sig-security/security-audit-2019/findings/Kubernetes%20Threat%20Model.pdf, 2019. [Accessed: 07 November 2022].

[234] S. Truex, L. Liu, M. E. Gursoy, L. Yu, and W. Wei. Demystifying membership inference attacks in machine learning as a service. *IEEE Transactions on Services Computing*, 14(6):2073–2089, 2021.

[235] K. Tuma, L. Sion, R. Scandariato, and K. Yskout. Automating the early detection of security design flaws. In *Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*, MODELS '20, page 332–342, New York, NY, USA, 2020. Association for Computing Machinery.

[236] W. Turton and K. Mehrotra. Colonial pipeline cyber attack: Hackers used compromised password. `https://www.bloomberg.com/news/articles/2021-06-04/hackers-breached-colonial-pipeline-using-compromised-password`, Jun 2021. [Accessed: 07 November 2022].

[237] U.S. Department of Justice. Justice Department Announces Court-Authorized Disruption of Botnet Controlled by the Russian Federation's Main Intelligence Directorate (GRU). `https://www.justice.gov/opa/pr/justice-department-announces-court-authorized-disruption-botnet-controlled-russian-federation`, 2022.

[238] L. S. Vailshery. Number of internet of things (iot) connected devices worldwide from 2019 to 2021, with forecasts from 2022 to 2030. `https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/`, 8 2022. [Accessed: 07 November 2022].

[239] J. van Rest, D. Boonstra, M. Everts, M. van Rijn, and R. van Paassen. Designing privacy-by-design. In B. Preneel and D. Ikonomou, editors, *Privacy Technologies and Policy*, pages 55–72, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[240] Verizon. Data breach investigations report. `https://www.verizon.com/business/resources/reports/dbir/`, 06 2022. [Accessed: 07 November 2022].

[241] Versen. Manifesto on software research and education in the netherlands. `https://www.versen.nl/assets/manifesto/digitalfolder.pdf`, 2020. [Accessed: 07 November 2022].

[242] N. Virvilis, D. Gritzalis, and T. Apostolopoulos. Trusted computing vs. advanced persistent threats: Can a defender win this game? In *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*, pages 396–403, 2013.

[243] S. Wang, X. Gu, S. Luan, and M. Zhao. Resilience analysis of interdependent critical infrastructure systems considering deep learning and network theory. *International Journal of Critical Infrastructure Protection*, 35:100459, 2021.

[244] Y. Wang, Z. Su, N. Zhang, R. Xing, D. Liu, T. H. Luan, and X. Shen. A survey on metaverse: Fundamentals, security, and privacy. *IEEE Communications Surveys and Tutorials*, pages 1–1, 2022.

[245] W. Wei. Casino gets hacked through its internet-connected fish tank thermometer. `https://thehackernews.com/2018/04/iot-hacking-thermometer.html`, 4 2018. [Accessed: 07 November 2022].

[246] R. Weiss, X. Mountrouidou, S. Watson, J. Mache, E. Hawthorne, and A. Chattopadhyay. Cybersecurity across all disciplines in 2020. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE '20, page 1404, New York, NY, USA, 2020. Association for Computing Machinery.

[247] E. Weyuker. Testing component-based software: a cautionary tale. *IEEE Software*, 15(5):54–59, 1998.

[248] Wikipedia. Self-sovereign identity. `https://en.wikipedia.org/wiki/Self-sovereign_identity`. [Accessed: 07 November 2022].

[249] M. Wilson and J. Hash. Sp 800-50. building an information technology security awareness and training program. Technical report, National Institute of Standards and Technology, Gaithersburg, MD, USA, 2003.

[250] J. Wolff and N. Atallah. Early gdpr penalties: Analysis of implementation and fines through may 2020. *Journal of Information Policy*, 11(1):63–103, 2021.

[251] N. Woolf. DDoS attack that disrupted internet was largest of its kind in history, experts say. https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet, Oct 2016. [Accessed: 07 November 2022].

[252] O. Yoachimik. https://blog.cloudflare.com/mantis-botnet/, 2022.

[253] K. Yskout, T. Heyman, D. Van Landuyt, L. Sion, K. Wuyts, and W. Joosen. Threat modeling: from infancy to maturity. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: New Ideas and Emerging Results*, pages 9–12. ACM, jun 2020.

[254] K. Yu, L. Tan, S. Mumtaz, S. Al-Rubaye, A. Al-Dulaimi, A. K. Bashir, and F. A. Khan. Securing critical infrastructures: Deep-learning-based threat detection in iiot. *IEEE Communications Magazine*, 59(10):76–82, 2021.

[255] F. Zhang, P. P. K. Chan, B. Biggio, D. S. Yeung, and F. Roli. Adversarial feature selection against evasion attacks. *IEEE Transactions on Cybernetics*, 46(3):766–777, 2016.

[256] E. Zio. Challenges in the vulnerability and risk analysis of critical infrastructures. *Reliability Engineering & System Safety*, 152:137–150, 2016.

Cyber
Security
for Europe

CyberSec4Europe is a research and innovation pilot project for the European Cybersecurity Competence Centre in Bucharest and the network of National Coordination Centres.

As a research project, CyberSec4Europe is working towards harmonising the journey from the development of software components that fit the requirements identified by a set of short- and long-term roadmaps, leading to a series of consequent recommendations. These are tied to the project's real-world demonstration use cases that address cybersecurity challenges within the vertical sectors of digital infrastructure, finance, government and smart cities, healthcare and transportation.

CyberSec4Europe's main objective is piloting the cybersecurity capabilities required to secure and maintain European democracy and the integrity of the Digital Single Market. CyberSec4Europe has translated this broad objective into measurable, concrete steps through a set of policy, technical and innovation objectives.

CyberSec4Europe is funded by the European Union
under the H2020 Programme Grant Agreement No. 830929