CANVAS – Constructing an Alliance for Value-driven Cybersecurity

White Paper 4

Technological Challenges in Cybersecurity

Josep Domingo-Ferrer, *Universitat Rovira i Virgili**
Alberto Blanco-Justicia, *Universitat Rovira i Virgili**
Javier Parra Arnau, Oriol Farràs, *Universitat Rovira i Virgili**

Dominik Herrmann, *Universität Hamburg*
Alexey Kirichenko, *F-Secure*
Sean Sullivan, *F-Secure*
Andrew Patel, *F-Secure*
Endre Bangerter, *Berner Fachhochschule*
Reto Inversini, *Berner Fachhochschule*

This report consolidates the findings of Work Package 2 of the CANVAS Support and Coordination Action; * Work Package Leader

The CANVAS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700540.

This work was supported (in part) by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 16.0052-1. The opinions expressed and arguments employed therein do not necessarily reflect the official views of the Swiss Government.

Electronic copy available at: https://ssrn.com/abstract=3091942

Content

Executive Summary..... 4

CANVAS White Papers – Overview..... 5

1. Introduction..... 6
   - 1.1 Considered Values 6
     - 1.1.1 Privacy 6
     - 1.1.2 Fairness 8
     - 1.1.3 Autonomy 8
   - 1.2 Organization of the Document 10

2. Cybersecurity Threats..... 11
   - 2.1 Malware 11
   - 2.2 Distributed Denial of Service 13
   - 2.3 Emerging Threats 13

3. Countermeasures..... 15
   - 3.1 Botnet Tracking 15
   - 3.2 Botnet Takedowns 15
   - 3.3 Lawful Interception 15
   - 3.4 Active Countermeasures 15
   - 3.5 Regulatory Measures 16
   - 3.6 Defence against DDoS Attacks 16
   - 3.7 Incident Handling Capabilities 16
   - 3.8 Countermeasures at Organisational or Individual Level 16
     - 3.8.1 Traditional AV Protection 16
     - 3.8.2 Segmentation at Various Levels 16
     - 3.8.3 Micro Virtualization 17
     - 3.8.4 Sinkholing 17
     - 3.8.5 DDoS Protection and Mitigation 17
     - 3.8.6 Incident Handling 17
   - 3.9 Ethical Concerns Raised by Cybersecurity Activities 17
     - 3.9.1 Metadata Collection 17
     - 3.9.2 Malware Handling 18
     - 3.9.3 Customer Support 18
     - 3.9.4 Malware Investigations 18
     - 3.9.5 Open Source Intelligence 19
     - 3.9.6 Penetration Testing 19
     - 3.9.7 Detection Capability vs. Precision 19
     - 3.9.8 Investigation of Nation-State Operations 19

4. Cryptographic Techniques..... 20
   - 4.1 Secret-key Encryption 20
   - 4.2 Public-key Encryption 20
     - 4.2.1 Homomorphic Encryption 21
     - 4.2.2 End-to-end Encryption 21
   - 4.3 Hash Functions 22
   - 4.4 Digital Signatures 22
     - 4.4.1 Threshold Signatures 23
     - 4.4.2 Group Signatures 23

Electronic copy available at: https://ssrn.com/abstract=3091942

4.5 Secure Multiparty Computation 23
4.6 Functional Cryptographic Schemes 24
4.7 Ethical Considerations Raised by Cryptography 24

5. Data Anonymization and Processing..... 25
5.1 Database Anonymization 25
5.1.1 Non-Perturbative Masking 25
5.1.2 Perturbative Masking 26
5.1.3 Synthetic Microdata Generation 26
5.1.4 Privacy Models 26
5.1.5 Permutation Model for Anonymization 27
5.2 Redaction and Sanitization of Documents 28
5.3 Data Stream Anonymization 29
5.4 Discrimination Prevention in Data Mining 30
5.5 Ethical Considerations on Anonymization applied to Cybersecurity Countermeasures 30

6. Conclusions..... 31

Appendix..... 33
A.1 Glossary 33
A.2 Further reading 34
A.3 References 34

Executive Summary

This White Paper summarizes the current state of discussion regarding the main technological challenges in cybersecurity and impact of those, including ways and approaches to addressing them, on key fundamental values. It will serve as a planning instrument for the CANVAS workshops as well as provide a content base for the three main CANVAS output deliverables. Furthermore, it will inform the review papers that should emerge out of the workshops.

The main contribution of this White Paper is an enumeration of current cybersecurity threats, focusing mostly on malware, denial of service attacks and advanced persistent threats. The countermeasures to these threats clearly show tensions with some values, and therefore the application of such countermeasures provoke some ethical dilemmas. Cybersecurity experts find themselves in many of such dilemmas daily, and have to decide on such issues without a clear guidance. This White Paper aims to help identify such ethical dilemmas when counteracting cybersecurity threats, and help the CANVAS consortium to develop such guides. Section 3.9 explores some of these challenges. One other important finding is that the cybersecurity community relies much more on interpersonal relations when sharing intelligence and data than in explicit national or supranational regulations.

Some of the dilemmas could be solved or minimized using advanced cryptographic techniques and data anonymization techniques. The cryptography community has not limited its efforts in providing only standard security assurances (confidentiality, integrity and autonomy), but have provided several techniques to improve the privacy of users against service providers, in the form of privacy enhancing techniques, some other methods, like secure multiparty computation, emphasize fairness between participants in said protocols. On the other hand, anonymization techniques could improve the performance of cyber threat detection mechanisms by enabling the collection and processing of more data from users (which directly impacts on the detection capabilities) while providing privacy guarantees to these users. One of the main conclusions of this work is that the cybersecurity, cryptography and anonymization communities should improve their collaboration to produce better and most respectful with the citizen values cybersecurity measures.

CANVAS White Papers – Overview

In order to summarize the existing literature on the topics and issues that are relevant for the CANVAS project, the CANVAS consortium has created four White Papers as follows:

- White Paper 1 – Cybersecurity and Ethics: This White Paper outlines how the ethical discourse on cybersecurity has developed in the scientific literature, which ethical issues gained interest, which value conflicts are discussed, and where the "blind spots" in the current ethical discourse on cybersecurity are located. The White Paper is based on an extensive literature with a focus on three reference domains with unique types of value conflicts: health, business/finance and national security. For each domain, a systematic literature search has been performed and the identified papers have been analysed using qualitative and quantitative methods. An important observation is that the ethics of cybersecurity not an established subject. In all domains, cybersecurity is recognized as being an instrumental value, not an end in itself, which opens up the possibility of trade-offs with different values in different spheres. The most prominent common theme is the existence of trade-offs and even conflicts between reasonable goals, for example between usability and security, accessibility and security, privacy and convenience. Other prominent common themes are the importance of cybersecurity to sustain trust (in institutions), and the harmful effect of any loss of control over data.
- White Paper 2 – Cybersecurity and Law: This White Paper explores the legal dimensions of the European Union (EU)'s value-driven cybersecurity. It identifies main critical challenges in this area and discusses specific controversies concerning cybersecurity regulation. The White Paper recognises that legislative and policy measures within the cybersecurity domain challenge EU fundamental rights and principles, stemming from EU values. Annexes provide a review on EU soft-law measures, EU legislative measures, cybersecurity and criminal justice affairs, the relation of cybersecurity to privacy and data protection, cybersecurity definitions in national cybersecurity strategies, and brief descriptions of EU values.
- White Paper 3 – Attitudes and Opinions regarding Cybersecurity: This White Paper summarises currently available empirical data about attitudes and opinions of citizens and state actors regarding cybersecurity. The data emerges from reports of EU projects, Eurobarometer surveys, policy documents of state actors and additional scientific papers. It describes what these stakeholders generally think, what they feel, and what they do about cyber threats and security (counter)measures. For citizens' perspectives, three social spheres of particular interest are examined: 1) health, 2) business, 3) police and national security.
- White Paper 4 – Technological Challenges in Cybersecurity: This White Paper summarizes the current state of discussion regarding the main technological challenges in cybersecurity and impact of those, including ways and approaches to addressing them, on key fundamental values. It provides an overview on current cybersecurity threads and countermeasures and focuses on ethical dilemmas that emerge when counteracting those threads. It also points to the fact that the cybersecurity community relies much more on interpersonal relations when sharing intelligence and data than in explicit national or supranational regulations. Furthermore, the White Paper presents advanced cryptographic techniques and data anonymization techniques that may help to solve or minimize some of the ethical dilemmas.

All White Papers and additional material are available at the Website of the CANVAS project: www.canvas-project.eu

1. Introduction

Our society depends increasingly on digital technologies. Almost every aspect of economy and society in the developed countries has transitioned to a digital environment. Governance, including elections, tax services, justice administration and communication with the citizens; companies managing their stock, logistics, personnel and almost all day-to-day activities; banking, transactions and the stock market; electronic commerce and all kinds of communication among people, such as email, instant messaging and social media depend, at different levels, on information and communication technologies. Therefore, it is of paramount importance that all actors in society can trust the technologies involved in the digital transformation.

"...trust is an expectation about a future behaviour of another person..."[1]

What the public can expect from a software system, service provider or governmental agency is that it behaves according to an established (sometimes implicit) agreement and that the actions by the trusted party do not cause the truster any harm, be it physical, economical or moral. Thus, cybersecurity is essential to build and sustain the trust relationship. However, it is important that the measures employed to keep us safe and secure do not bypass other rights like autonomy, fairness or privacy.

The 2016 Scoping Paper on Cybersecurity defines cybersecurity as the measures put in action to protect the ICT infrastructure, networks, and stored data, both in the civilian and military spaces. The goal of cybersecurity measures is to ensure the availability and integrity of resources, as well as the confidentiality of stored data in the presence of threats such as natural disasters, human errors, corporate espionage, criminality, government-driven attacks, surveillance, terrorism, hacktivism, etc. [2]

While the application of cybersecurity measures can be considered as mandatory, the specific techniques and the way they are applied can put at risk other important values, such as privacy, fairness and autonomy. If these values are not taken into account, the trust of the public on ICT and online services can be damaged, potentially causing a negative impact on economy and society.

In this document, we analyse the current state of the art of cybersecurity and data protection measures, how these measures impact on the privacy and autonomy of the public and whether they can cause social inequalities.

1.1 Considered Values

This section enumerates the values that we consider in this work. These values are privacy, fairness and autonomy. While we might be tempted to take into account more values, we believe that these three are a good representation of the values at stake and also represent the values that are usually considered in the cybersecurity community with a technological focus; and even considering only these three, we find how interrelated they are.

1.1.1 Privacy

Privacy is the ability of an individual or a group of individuals to freely and selectively disclose personal, sensitive or confidential information about themselves. Privacy has been defined in many different ways throughout history: it can be seen as the right to be in seclusion, in solitude, or to keep secrets; the state

[1] From the definition of interpersonal trust by Walter Bamberger from the Technical University of Munich http://www.ldv.ei.tum.de/en/research/fidens/interpersonal-trust/. We use the definition of interpersonal trust in an open way, i.e. a trust relationship in our case will typically occur between a person and an organization, such as a company, a governmental agency, etc.

[2] Scientific Advice Mechanism High Level Group. *Scientific advice mechanism scoping paper: Cybersecurity*. European Commission, 2016: https://ec.europa.eu/research/sam/pdf/meetings/hlg_sam_012016_scoping_paper_cybersecurity.pdf

of being free from unwanted or undue intrusion or disturbance in one's private life or affairs; freedom from damaging publicity, public scrutiny, secret surveillance, or unauthorized disclosure of one's personal data or information.

The ENISA report on Privacy by design,[3] aimed at giving technical guidance to deploy the General Data Protection Regulation of the EU, enumerates a series of design strategies and patterns to ensure that services comply with data protection regulations. While these principles are not directed specifically towards cybersecurity services, they could be useful to design not only concrete mechanisms, but how these mechanisms ought to be applied. These design strategies are *minimize*, *hide*, *separate*, *aggregate*, *inform*, *control*, *enforce* and *demonstrate*.

Next, we survey the particular cases in which cybersecurity measures interact with privacy, including cases in which security and privacy are in conflict, cases in which security is aided by privacy and cases in which security is needed to achieve privacy.

One major point in this topic consists of network monitoring activities conducted either by governmental agencies or by private security providers. Monitoring by the state (e.g. police or agencies) is often considered as the most harmful conflict between security and privacy. However, it should be noted that it strongly depends on what is exactly monitored and under which conditions. Such cases are: i) full cable monitoring by an agency such as it is being discussed in Germany or is practiced by large agencies such as NSA or in China, ii) selective monitoring of offenders, and iii) monitoring of metadata that are not directly related to people, such as monitoring of IP addresses, domain names that are related to criminal activities, etc. Whereas the first case can be harmful to the privacy of many people, especially of innocent people, the second and third cases can be controlled more easily by a system of checks and balances. It is crucial that every measure by the state is balanced between the desired effect of increasing the security of citizens and their right to privacy. Such checks and balances must not only be analysed in the present but for the future as well; that is, data collected under a more rigorous law should not be used for other purposes if the law is changed. It is important to note as well that the third case can be potentially beneficial for citizens in some cases. For example, if an infected device is detected by a security agency and the user is warned, the state has ensured the privacy of this user, as otherwise the attacker could have had access to all data of the victim.

On the opposite side, we acknowledge that complete anonymity and secrecy of communications can be exploited by malicious entities to attack services without being discovered. One example is anonymization services like TOR, which offer the possibility to access websites and online services without disclosing one's IP address. These services protect the privacy of users. However, they pose a threat to the security and trustiness of online services, for instance because malicious activities cannot be traced back to the perpetrator and perpetrators may use the anonymizer to act with multiple identities (Sybil attacks).

No matter whether privacy aids or hampers cybersecurity efforts, it is clear that cybersecurity and data protection are mandatory to achieve any level of privacy. For example, privacy is endangered whenever integrity and confidentiality are violated. This is important to note as integrity and confidentiality are also basic security goals that are often analysed when determining the risk of an application or technology.

Finally, we would also like to point how privacy and research on privacy help cybersecurity. Spear phishing attacks are phishing attacks directed at specific individuals. The availability of private information about individuals makes it easier for attackers to perform these kinds of attacks, which may later lead to more critical attacks. Also, strong and applicable privacy laws lead to better products also in terms of security. A good example is the discussion about electronic health dossiers that currently take place in

[3] George Danezis, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Le Métayer, Rodica Tirtea and Stefan Schiffner. *Privacy and data protection by design-from policy to engineering*. European Network and Information Security Agency-ENISA, 2015. preprint arXiv:1501.03726, 2015.

Switzerland (and many other European countries as well). The fact that privacy is needed and undisputed in such a sensitive environment leads to products that are engineered with better security.

Encrypting and anonymizing collected data is a good way to prevent excessive privacy invasion of the subjects, but the application of these techniques in the cybersecurity field is still not very well understood nor typically applied.

1.1.2 Fairness

Fairness (or non-discrimination) is the absence of bias, the assurance of equal treatment of individuals or groups of individuals from institutions, companies, etc. It stems from the premise that equals should be treated equally, and as per the Universal Declaration of Human Rights, all human beings are born free and equal in dignity and rights.[4] While there might be justified reasons to treat individuals differently (justice means giving each person what he/she deserves), other reasons, such as gender, race, religion, sexual orientation, etc., are not justifiable reasons to discriminate among people.

When dealing with online services and the digital ecosystem, the discussion on fairness and non-discrimination is very closely related to privacy. For example, profiling services are justifiably against the privacy of online users, and the typical consequences of such profiling activities can be most often related to an unfair treatment of the users.

In the cybersecurity space, we acknowledge that the States are generally not yet ready to provide a security level that is comparable to the security provided in the physical space. Therefore, good security measures require a private investment by companies (and individual citizens). The lack of financial resources leaves part of the population and small companies vulnerable to attacks such as DDoS, extortion (e.g. ransomware), theft of confidential information, etc. It is important to note that the security community is very active in this matter and that large enterprises have taken as their responsibility to protect those who cannot protect themselves. A good example of this is the Google Shield Project, which offers a free service to protect news sites and people who stand up for freedom of speech from DDoS attacks.

Another interesting topic is net neutrality, which is the principle by which Internet regulators treat all transmissions (or packets) the same way, independently of their origin, destination and content. This principle is the basis for a fair Internet to everybody. Not only regulators, but also attackers can put net neutrality at risk. Interested parties can dominate big parts of the Internet, for example through the use of malware (botnets) to carry out illegal activities. Security measures are required to avoid these practices.

Possible solutions to unfairness caused by some protocols, defences and data processing activities may stem from the cryptography and statistical disclosure control fields. As an example, in a cryptographic commitment scheme one party commits to a message at present, which is going to be revealed and verified in the future. On the one hand, others cannot determine the message until the initiating party releases the cryptographic key that was used to create the commitment. On the other hand, the initiating party cannot alter the message after having committed to it. Such schemes allow the construction of fair protocols where no party can cheat. Thanks to fair protocols, we can build more complex secure protocols, such as e-voting protocols, certified mail exchange, secure contract signatures without notaries, etc. Anti-discrimination mechanisms allow suppressing biases in the training data of classification systems, or to modify the resulting rules to limit discriminating decisions.

1.1.3 Autonomy

Autonomy or self-determination is the capacity of an individual to make an informed, un-coerced decision. Autonomy may be understood as the boundaries on the actions individuals can make. In the physical space, but also in the cybersecurity space, though, the actions of ones affect not only themselves, but also the well-being, and in our case security, of others. For example, for the general security of

[4] http://www.un.org/en/universal-declaration-human-rights/

interconnected machines and networks, the security level of the whole network is typically considered that of its weakest link. Therefore, a responsible behaviour with respect to security (that is, some restriction on these boundaries) is expected to be followed by all.

As a matter of fact, if you want to be secure on the Internet you have to stick to a set of best practices that infringe your autonomy. These best practices make the life of users more complicated due to higher complexity and poorer usability. For instance, you have to disable plugins, use a secure browser, use long passwords, multiple authentication factors, etc.

People in general have taken responsible attitudes to fight cybercrime. The Internet, referred to as the decentralized collection of autonomous interconnected networks, is managed by many individuals and organizations that act autonomously following their own interests. These autonomous parties have typically had good reactions to new security threats which would not have been possible without individuals taking the right steps. A good example of this is the rise of spam that led to very effective initiatives such as Spamhaus.[5] Therefore, autonomous entities protecting themselves, and finding countermeasures to different threats, increase the security of the whole Internet. Open-source communities and non-profit organizations, on their side, work also to increase the security of all. A notable example is OWASP. OWASP is an online community that creates and publishes freely-available articles, methodologies, tools, technologies and documentation for the protection of web applications. One of their most notable publications is the OWASP Top Ten, which is released regularly, and includes the ten most critical risks of web services and possible countermeasures. OWASP has become a *de facto* standard for web security.[6]

Some measures taken by organizations may impact the ability of autonomous entities to operate on the Internet. For example, it is perfectly legal to operate a mail server at home; no RFC or other standards forbids it. However, the reality is that it is near impossible to do so because most organizations block SMTP traffic from home user IP ranges. Governments also actively impose some not so well justified measures which impact on the autonomy of companies. A good example are the Letters of National Security issued by the FBI, which force ISPs to implement surveillance but hinder them from implementing any transparency.

The disclosure of vulnerabilities as early as possible facilitates autonomous decisions and actions by individuals. However, for the sake of security, vulnerabilities are typically not disclosed until a patch is available. While full vendor disclosure was a typical practice in the early days, the slowness and unwillingness of certain vendors to promptly fix their bugs and publish patches caused some users to adopt immediate public disclosure, which may punish unfairly other vendors that would make an honest effort in handling their vulnerabilities. Hybrid disclosure, where the benign user does not announce the vulnerability knowledge to the public immediately, but instead allows the vendor some time to develop a patch, is a relatively common practice today. If the vendor does not release its patch before the deadline, the public is informed about the vulnerability.

The zero-day vulnerability and exploit trade is currently legal, i.e. the autonomy of such dealers is considered to be more important than security concerns. On the other hand, it is not clear whether security would improve or decrease if vulnerability trade would be regulated. We believe that this is an interesting conflict, whose resolution is not obvious.

[5] The Spamhaus project is an international non-profit organization that compiles and distributes several widely-used anti-spam and related cyber threats lists. https://www.spamhaus.org/
[6] https://www.owasp.org/index.php/Main_Page

1.2 Organization of the Document

The rest of the document is organized as follows. First, Sections 2 and 3 explores common and advanced cybersecurity threats and attacks, and describes some of the applicable countermeasures, indicating how these interact with the presented values. Sections 4 and 5 provide a state of the art on cryptographic and anonymization techniques, respectively. These techniques are powerful tools to reduce the impact of security countermeasures on values. Finally, Section 6 provides some concluding remarks.

2. Cybersecurity Threats

This section enumerates some of the most common threats and attacks currently taking place in the cyberspace, namely, it is focused on malware, denial-of-service attacks and advanced persistent threats. While this list is not exhaustive, we believe that these are the threats that need countermeasures which can incur in ethical dilemmas. Then, we provide a list of available countermeasures and the impact of those countermeasures on the three values described in the previous section.

We note that the cybersecurity domain is highly dynamic, as attackers constantly adapt existing and invent new techniques and tactics to avoid detection and maximise their financial gain, acquired intelligence, harmful effects on their targets, etc. Among the trends observed in 2017, there was a jump in the volume of traffic scanning port 1900, the standard port for the SSDP protocol, which enables discovery of UPnP devices, an indication that the practice of targeting IoT devices has only accelerated since the emergence of the Mirai botnet last autumn. Given the fact that the recent outbreaks of WannaCry and NotPetya leveraged vulnerabilities in Windows SMB using exploits stolen from the NSA, it is no surprise that traffic probing SMB port 445 also experienced a jump. The Eternal exploits leaked from the NSA provided malware authors with powerful worm functionality and illustrated a very alarming trend of commoditisation of attack tools and processes: sophisticated tactics, techniques and procedures developed by nation states and other well-funded organisations rapidly fall into the hands of common cyber criminals, who use them for their own purposes. This means more and more attacks get elevated to the advanced category. PowerShell and the command line were used for a variety of malicious purposes, including launching of new processes, lateral movement, shutting down of “defences” (such as Windows firewall), and downloading of additional payloads. Other common tools native to Windows that were leveraged in attacks include net.exe, netsh.exe, explorer.exe, regsrv32.exe, wmic, and rundll32.exe, showing the attackers’ appreciation of compromising legitimate processes and applications as one of the best ways to avoid detection nowadays.

Recognising that the rapid evolution of the attacking and defending sides influences the contents and depth of ethical challenges in cybersecurity, our goal was to present in this document (Section 3.9) those, which are considered the most relevant by cybersecurity practitioners at present.

2.1 Malware

Malware, short for malicious software, is a broad term for software that, when executed in a target machine, causes unwanted or harmful consequences. These consequences range from simply bothersome effects, such as showing excessive adverts to users, to the destruction or leakage of data, infiltration and control of the target machine, among others. Table 1 provides a classification of malware, according to their target and the actors that typically employ them.

The menace and the visibility of malware strongly depends on the actor group and their intentions. APT actors try to stay as hidden as possible and do a very slow lateral movement (*i.e.*, infect computers in the same network) once they infiltrated a network. Criminals often use speed as their main tactic. They attack a lot of targets at the same time and try to gain as much money as possible until the malware is detected. An example of this is the WannaCry ransomware campaign,[7] which affected at least 74 organizations, from which attackers demanded relatively small amounts of money to recover the organizations’ files. However, there are also criminal actors that have specialized on attacking high value targets

[7] WannaCry Ransomware Attack: https://en.wikipedia.org/wiki/WannaCry_ransomware_attack

and try to gain a large sum from a single victim such as a bank,[8] such as in the case of the Bangladesh Bank robbery in 2016.[9] Attackers were able to obtain the credentials of the bank and issued transfers for a total of $951 million.

Type	Targets	Description	Actors
Banking Trojan	Confidentiality and integrity of financial transactions	Malicious program used to attempt to obtain confidential information about clients using banking and/or payment systems.	Mostly Cybercrime
Ransomware	Availability and integrity of data	Malicious program which encrypts the information stored in the victims' machine. The attackers may demand compensation for the decryption keys.	Mostly Cybercrime (Extortion)
Worm	Networked systems	Malware which can replicate and transmit itself through the network.	Mostly Cybercrime and activists
RAT (Remote Access Trojan)	Control over systems and data, data theft	Malicious program which includes a backdoor providing the attacker with administrative control over the target computer.	Cybercrime and state-sponsored actors
APT (Advanced Persistent Threat)	Control over systems and data, data theft	Sophisticated attack, including several tools and techniques, which aim to take control over the target system or network for long periods of time, to the exfiltration of confidential information.	State-sponsored actors. Cybercriminals may use similar modus operandi.
Surveillance Tools ("GovWare")	Criminals, dissidents, activists	Software and/or Hardware components, possibly installed by manufacturers on orders from governmental agencies, which monitor the targets' network traffic.	Law Enforcement private organizations selling this as a service.
Specialized malware	Special platforms such as ICS platforms	Malware specifically designed to affect specific platforms, such as power plants, industrial control system, IoT devices, power grids, etc.	State-sponsored actors.
Wiper	Availability of data.	Malware which deletes the contents of the storage devices (hard disks or others) of the target computers.	State Sponsored actors
Fileless Malware	Control over systems and data, data theft	Malware which runs entirely from memory. This hampers detection by antivirus software and forensic activities.	State Sponsored actors, organized Cybercrime

Table 1: Classification of malware

Attackers infect their targets by various means, as shown in Table 2.

Attack Vector	Explanation
Waterhole	A website often visited by the target is hacked and used for distribution of malware
Exploit Kit	An exploit kit is used to infect the device by delivering a suitable exploit for the target.
Malspam	The malware is distributed by email. Malspam waves are often used by criminals to target as many victims as possible in a short time.
Spear Phishing	Spear Phishing is a targeted way of delivering the malware to its victim, typically based on specific knowledge about the target. It is often used by state-sponsored actors.
USB Stick	USB sticks can be used to deliver malware as well. It is an interesting vector as it circumvents many of the filters in place and targets the endpoint device directly.
Direct Access	Having physical access to a device enables an attacker to directly implant the malware on the target system. It is mostly used by state-sponsored actors.

Table 2: Attack vectors.

[8] Study of several attacks against banks using the Carbanak backdoor and the network of ATMs; see: https://securelist.com/files/2015/02/Carbanak_APT_eng.pdf

[9] Bangladesh Bank robbery: https://en.wikipedia.org/wiki/Bangladesh_Bank_robbery

Malware communicates with its Command and Control Servers (C&C servers) to receive new commands or to exfiltrate data. Malware uses nearly all channels that are available, such as HTTP, SMTP, Twitter, DNS, TOR, IRC or custom-made protocols.

2.2 Distributed Denial of Service

Distributed denial of service (DDoS) attacks have significantly increased in their intensity and their frequency. The actor groups differ significantly; criminal actors try to extort money by menacing organizations with DDoS attacks, state-sponsored actors try to silence dissidents and activist blogs and, in times of big political tensions, even try to disrupt critical systems in a country. State-sponsored actors partly team up with "patriotic" criminals and/or develop their own sophisticated methods such as the "Great Cannon of China".[10]

The objective of DoS and DDoS attacks is to interfere with the capabilities of internet-connected hosts to properly respond to legitimate requests, by flooding them with malicious requests. The difference between DoS and DDoS attacks is that in the latter kind of attacks, the malicious requests come from several different sources. Successful attacks cause consequences that range from delays to serve legitimate users' requests to complete server crashes. DoS and DDoS attacks can be classified as depicted in Table 3.

Type	Description
Reflection and Amplification	Abuses UDP protocols with faked sender addresses. Small queries trigger large responses (amplification factor). Targets the bandwidth of the victim.
Protocol attacks	Protocol-level attacks target system resources and try to exhaust them (Memory, CPU).
Application level attacks	Attacks directly the application logic (*e.g.,* resource exhaustion due to complex searches on a webpage).

Table 3: Classification of DDoS attacks.

2.3 Emerging Threats

Some of the emerging threats and technologies that may require potentially more invasive protection techniques:

- Fileless or Memory-Resident Malware inject malicious code into memory while no executable files are stored to the disk.[11] Most of the existing protection technologies focus on the detection and analysis of executable files stored in the hard drive, and thus may fail to recognize this kind of attacks. The use of fileless malware also significantly complicates digital forensics, since all evidence may disappear after the system reboots.
- More generally, attacks of new types compromise legitimate processes and applications to carry out malicious activity while avoiding detection by the traditional techniques, as there is no malware involved. In particular, such attacks utilise Java Script, Windows Management Instrumentation (WMI), PowerShell, and Microsoft Office macros.

[10] The Great Cannon of China, co-located with the Big Firewall of China, is a state-sponsored attack tool used to launch DDoS attacks, by intercepting and redirecting legitimate network traffic originally directed at Internet services in China to their targets. https://citizenlab.org/2015/04/chinas-great-cannon/

[11] Heimdal Security introduction to fileless malware. https://heimdalsecurity.com/blog/fileless-malware-infections-guide/

- Use of HTTPS by malware for protecting their C&C traffic.[12] Since encryption limits visibility of end-point protection, some security products attempt to become a trusted man-in-the-middle, intercepting and decrypting all outgoing traffic.
- Security weaknesses of IoT devices such as wearable medical devices, platforms, and applications, and significant increase in the number of access points to networks that can be exploited are leading to a greater need of security monitoring, often involving sensitive data and communications.
- Use of Machine Learning algorithms by attackers as a weapon to enhance social engineering attacks or perform vulnerability scanning.

[12] http://nymag.com/selectall/2017/03/phishing-and-malware-sites-can-use-https-and-ssl-against-you.html

3. Countermeasures

Many states have defined National Cyber Defence Strategies that give a general direction on how Critical Infrastructures should be made safe on a National Level, e.g. in the UK and the Netherlands.[13,14] However, these strategies cannot go in depth when it comes to concrete countermeasures. In the following we try to fill out this gap by giving a short oversight of countermeasures that are known to be effective when combatting malware and botnets as well as against DDoS attacks.

3.1 Botnet Tracking

National CERTs and security organizations write trackers to monitor botnets. It is a detective and reactive measure that helps keeping the situation under control. Tracking botnets means analysing the malware and trying to mimic the behaviour to get new configurations and information about the current C&C infrastructure. This information can be used to block malicious behaviour. This measure is effective against the phenomenon itself, not against single actors or victims.

3.2 Botnet Takedowns

Coordination actions by law enforcement, CERTs and security organizations may take down all C&C servers at once and thus destroy the botnet. Doing so is effective against single botnets and raises the price of operating a botnet. Botnet tracking is a prerequisite of a successful takedown. It should be clearly stated that a takedown makes only sense when at the same time the operators can be arrested, as otherwise botnets are quickly being rebuilt, often in a much more resilient way.

3.3 Lawful Interception

Law enforcement organizations may intercept traffic by the malware and their operators. However, this endeavour gets difficult when the botnet infrastructure is composed of systems being used by innocent citizens who have been hacked by the attackers (i.e., zombie computers).

Captured network traffic can be used to track the criminals and may lead to penal prosecution. It also helps informing potential victims about their infection. It is effective against criminal backend systems as well as for preserving evidence of the actors behind the scene.

3.4 Active Countermeasures

“Hacking back” is an often-proposed measure. C&C servers have as many vulnerabilities as every other software, and so can be hacked, too. When having access to the C&C infrastructure, it is possible to inform new victims or to get information about the criminals which leads to better penal prosecution. There may be huge collateral damage and there are many ethical questions open, especially if this is not done in a regulated and transparent way. One high-level question is obviously what distinguishes us from attackers if we employ the same methods to compromise their systems. Active countermeasures

[13] UK: National Cyber Security Strategy 2016-2021, https://www.gov.uk/government/publications/national-cyber-security-strategy-2016-to-2021

[14] NL: National Cyber Security Strategy: https://www.ncsc.nl/english/current-topics/national-cyber-security-strategy.html

also pose problems if innocent persons are involved when their system is being abused by criminals or in cases of misattribution, where the wrong individual or institution is targeted in "hacking back".

3.5 Regulatory Measures

If states increase regulations for the access to the Internet and the devices connected to it, they can make attacks much harder. This is especially true for DDoS attacks where a regulatory introduction of certain techniques such as BCP 38[15] would make amplification and reflection attacks much more difficult. Another important area of regulation with great efficiency would be minimal standards for devices connected to the Internet. Regulatory measures are therefore effective for the resilience of critical infrastructures as well as to reduce the attack surface.

3.6 Defence against DDoS Attacks

The most effective countermeasures a state can currently make against DDoS attacks are at the regulatory level. The actual reactive defence against DDoS attacks is mostly left to commercial organizations such as CloudFlare or Akamai.[16,17] Google funds an interesting project called "Project Shield" that helps news websites and freedom-of-speech activists that are often silenced by DDoS attacks.[18]

3.7 Incident Handling Capabilities

As security incidents cannot be avoided in all cases, every nation should have a national CERT that helps dealing with critical situations and that coordinates security measures to improve the overall security situation.

3.8 Countermeasures at Organisational or Individual Level

Apart from the already well-documented "classical" antivirus protection, which we are only shortly describing, we would like to enumerate some interesting techniques that help reduce the malware risk.

3.8.1 Traditional AV Protection

Current AV protection measures include:

- Antivirus programs (be they signature, behaviour or heuristic based).
- Sandboxing techniques that quickly run a program and try to detect malicious behaviour.
- System hardening.
- Prevention of program execution (using tools like AppLocker).
- Intrusion Detection / Intrusion Prevention systems, both at network and host level.
- Usage of 2FA (2 Factor Authentication) wherever possible.

3.8.2 Segmentation at Various Levels

It is important to have a segmentation on various levels in an organization:

[15] BCP38, or Ingress filtering, is a technique used to ensure that the IP addresses of Internet packets are legitimate and have not been spoofed. Packets with spoofed IPs are dropped. https://tools.ietf.org/pdf/bcp38.pdf
[16] https://www.cloudflare.com
[17] https://www.akamai.com/us/en/resources/protect-against-ddos-attacks.jsp
[18] https://projectshield.withgoogle.com/public/

- By task: a device being used for financial transactions should not be used for surfing the Internet. The same is true for the management of systems.
- By data: Data should be compartmentalized to protect the most important assets.
- By network zone: A good zoning of the network helps reducing the exposure and enables a better detection of malware flows.

3.8.3 Micro Virtualization

An interesting approach is the usage of virtualization techniques not to prevent the infection but to reduce its impact by isolating every process into its own virtual compartment.

3.8.4 Sinkholing

Sinkholing[19] is very effective to reduce the impact of malware. Using sinkholing techniques, a fraudulent domain is taken from the criminal and given to a security organization. This has two main benefits:

- The attackers lose control over at least a part of their botnet.
- The organization collecting the sinkholed data can distribute information about infected devices to national CERTs, ISPs and other organizations which in turn can inform affected customers.

This technique can also be done internally in an organization by using a technique called Response Policy Zone (RPZ).

3.8.5 DDoS Protection and Mitigation

DDoS protection and mitigation are something that cannot be solved at an individual level. In most cases, the help of the upstream provider is necessary. However, every organization should have emergency plans ready and should try to separate vital systems from publicly facing systems by using separate upstreams. In order not to be part of a DDoS, every organization should protect and harden their systems.

As mentioned before, a noteworthy project is Project Shield by Google which helps defending free speech from DDoS attacks without any cost.[18]

3.8.6 Incident Handling

Every organization should be prepared to handle security incidents. Depending on the size and strategy of an organization this can be done using an internal CERT / CSIRT or with external resources. However, some basic capabilities should always be available internally.

3.9 Ethical Concerns Raised by Cybersecurity Activities

3.9.1 Metadata Collection

During the past decade, end-point protection services have turned to collecting metadata from customer environments in order to provide adequate protection against high-volume commodity malware threats. Metadata from customer systems is utilized when determining verdicts on new samples, as a form of crowdsourcing to keep a large number of systems protected, and to look for trends and patterns in the threat landscape.

Collected metadata are normalized in order to strip away any private or sensitive data. Examples of normalization include the removal of sensitive information from file paths and URLs, and discarding of customer-specific data such as source IP addresses. Unsurprisingly, some customers of security software still object for privacy reasons to metadata collection, and even to cloud queries that security solutions send to their back-ends, and turn those features off.

[19] https://www.hitachi-systems-security.com/blog/sinkholing-a-critical-defensive-tool/

3.9.2 Malware Handling

Working groups that formed around the Anti-Virus industry follow an agreed-upon set of ethics principles regarding the sharing and handling of malware samples. (Examples of such groups include CARO, AMTSO, AVAR, and VirusTotal.) For instance, in order to prevent malware from leaking into the wild, or falling into wrong hands, the sharing of samples is limited to verified members of the group.

Incoming samples are all treated equally, regardless of source. Metadata are normalized, and samples deemed "confidential" (e.g., document file types) are flagged as such and stored in a way that prevents human access. Ethics extend to the storage and handling of malicious samples (files, URLs, etc.). For instance, if malware needs to be executed for analysis purposes, it must happen in controlled, isolated environments in order to prevent further spread of an infection.

At the same time, a grey area exists regarding the creation of exploits and malware. Some malware has the ability to spawn a uniquely identifiable copy of itself based on identifiers in a system for communication with its C&C infrastructure. Does this constitute as a new piece of malware? Many companies that were formerly only in the business of creating anti-malware software have since moved into other cybersecurity areas, such as threat assessment, penetration testing, and red teaming. These disciplines require a different set of tools and processes. Hence, stances are changing in one of the industry's traditional values.

3.9.3 Customer Support

Due to their complexity and low-level nature, anti-virus solutions can have many unforeseen interactions with other processes on a system. Hence, in some customer support cases, detailed diagnostics are required to debug a problem.

When a complex support case arises, detailed system information can be collected via a vendor-supplied diagnostic tool. Given the privacy-invasive nature of such a tool, informed consent is required from the owner of the system in question. Once the tool has collected diagnostics, it is up to the owner of the system to send the collected data to support staff, who will proceed to debug the issue.

In extremely infrequent cases, a support representative or engineer assigned to a customer case may find data suggestive of criminal activity. This situation may pose several ethical dilemmas, depending on the nature of the relationship between the customer and the vendor. Should the engineer notify authorities directly? If the support work was done on behalf of a reseller or partner, should they instead be notified? What steps should be taken in the case where a support engineer is dealing with an admin, and not the owner of the machine?

3.9.4 Malware Investigations

Malware analysis and threat hunting often present data that can be shared with other entities. Examples of such data may include IP addresses of command and control servers, data pointing to other systems or organizations that have been compromised, malicious URLs, file hashes, and so on. Some of this data are referred to as indicators of compromise (IoCs).

Processes established by the cyber security industry facilitate open sharing of some of this information. For the rest, the process of sharing information is dependent upon the nature of the data. For instance, certain types of IoCs can be provided to third parties (such as other vendors, CERT organizations, and other businesses.) For the most confidential pieces of data, decisions are not so easy. For example, imagine if a vendor, in the process of performing an investigation, finds information revealing that machines in other (non-customer) organizations have been compromised. If the vendor contacts those organizations directly, it may be perceived as a sales call. Sharing that information with local authorities requires high trust and is not always an obvious way of action.

3.9.5 Open Source Intelligence

Open source intelligence (OSINT) is data that can be gathered from public sources (available to non-government, non-military, non-law enforcement individuals). Many cyber security organizations gather and utilize open source intelligence for a variety of applications. OSINT can be obtained from areas of the Internet that are not indexed by search engines, and therefore not easy for the public to find. This type of information, whilst publicly obtained, is often used cautiously.

3.9.6 Penetration Testing

During the process of penetration testing, security researchers may find vulnerabilities in publicly available software or devices. In these cases, a process of responsible disclosure is followed whereby the manufacturer of the device or the maintainer of the software are given details of the vulnerability, along with a time-frame within which a fix is expected to be made.

If the maintainer fails to adhere to their end of the bargain, it falls upon the party who discovered a vulnerability to decide how to proceed. Publishing the vulnerability can create public awareness that particular systems or devices are insecure, but also may give good exploitation chances to attackers. Not publishing the information can lead to a false sense of security. Cybersecurity researchers' and vendors' desire to strengthen their image and recognition in the community and in the eyes of existing and potential customers, their intention to help users protect against unpatched vulnerabilities and limit the attack surface, and their needs related to cybersecurity consultancy services, including red teaming, protection design, and technology choices for customers, via releasing information about vulnerabilities and Proof-of-Concept code utilizing those can unfortunately lead to such side-effects as weaponizing of criminals and other parties that may cause harm to organizations and individuals, violating of 3rd party IPR, and even breaking of laws.

3.9.7 Detection Capability vs. Precision

Vendors of anti-malware and attack detection solutions normally face a trade-off between the detection rate and the false positive (FP) rate. For instance, they may opt for more aggressive blocking to achieve near 100% detection, but that usually results in more FP's and potentially significant problems for the customers. Given the importance of good performance in tests, such as comparative tests and reviews for antivirus software, the decision sometimes becomes an ethical issue.

3.9.8 Investigation of Nation-State Operations

Ethical concerns can be subdivided into passive concerns related to the awareness of an operation and active concerns arising from contemplating or enacting measures to identify or disrupt those operations[20]. In particular, legitimacy of espionage operations may not be immediately questionable. A layer of difficulty is added by the common practice of mixed use, where campaign operators deploy the same malware to infect radical targets along with more questionable targets (like research institutions in adversarial countries or opposition politicians within their own borders). The question of whether to turn a blind eye becomes more complicated as victims in clear need of protection are mixed in with extremists.

[20] https://media.kaspersky.com/pdf/Guerrero-Saade-VB2015.pdf

4. Cryptographic Techniques

This section introduces some of the most commonly employed cryptographic techniques. Since the invention of public key cryptography, many advances have been made which grant cryptosystems and digital signature schemes additional properties that go beyond those of data confidentiality, integrity and authenticity, for example by enabling the computation of functions on encrypted data. These techniques, therefore, can be used not only to protect data from attackers, but to protect individuals from the service providers themselves, which in some cases may have secondary goals when dealing with personal data. We argue that some of these techniques could be a very interesting addition to current cybersecurity countermeasures to protect the values of individuals while still being protected from cyber threats.

4.1 Secret-key Encryption

Secret key encryption, also known as symmetric encryption is the cryptographic paradigm in which two users, the sender and the receiver, share a secret key. In this setting, every user holds a secret shared key for every other user she wishes to communicate with. While existing symmetric encryption schemes are thought to be resistant to quantum computers, they have an intrinsic problem, which is the number of keys each user has to store. In modern day applications, a combination of public key cryptography and secret key cryptography is used to protect communications, using public key cryptography to generate and share random temporal (session) symmetric keys, with which they later encrypt communications, see for example TLS/SSL and digital envelopes.[21,22]

The current recommended symmetric encryption scheme is the Advanced Encryption Standard (AES),[23] known also by its original name Rijndael and selected by NIST in 2001. AES is a block cipher with 128 bit blocks and keys of length 128, 192 or 256 bits.

4.2 Public-key Encryption

Public-key encryption designates the cryptographic paradigm in which two players can communicate in a private way without sharing a common key.[24,25] Public key is also called asymmetric encryption, in contrast to symmetric encryption, where the users share a common key that is private. In a public-key encryption scheme, it is not assumed that users have shared private information before the communication starts. In the public-key encryption setting, each party holds two keys: a secret key, that is kept private by the party, and a public key, that is known by everybody. In a typical public-key encryption scheme, if a party Alice wants to send a message to Bob, Alice picks the public key of Bob and encrypts her message with this key. Then, Bob can recover the message from the ciphertext by using his secret key. In this setting, Alice does not need to share any secret information with Bob before sending him a message.

[21] Transport Layer Security v1.2 - https://tools.ietf.org/html/rfc5246
[22] RSA Laboratories: What is a digital envelope? - https://www.emc.com/emc-plus/rsa-labs/standards-initiatives/what-is-a-digital-envelope.htm
[23] FIPS 197: Advanced Encryption Standard - http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.197.pdf
[24] H. Delfs and H. Knebl, *Introduction to Cryptography, Principles and Applications, Second Edition*. Springer-Verlag Berlin Heidelberg 2007.
[25] J. Katz and Y. Lindell. *Introduction to Modern Cryptography, 2nd ed*. Chapman & Hall/CRC, 2014.

Indeed, the public keys can be published in a public board along with the user's name, and anybody willing to send a message to Bob can do it. From the key management point of view, this option is simpler, because each user is associated with a pair of keys, and for symmetric encryption we need a different key for every different pair of users. However, in general, symmetric encryption schemes are more efficient. Therefore hybrid approaches combining public-key encryption and symmetric-key encryption are very common: public-key encryption is used to distribute secret keys among users, and later these keys are used in symmetric encryption schemes.

The idea of public-key cryptography was introduced in 1976 by Diffie and Hellman.[26,27] The most famous public-key encryption scheme is the RSA cryptosystem, presented by Rivest, Shamir and Adleman in 1978.[28] Currently, there are many public-key cryptosystems, and they are broadly used in many different settings.

4.2.1 Homomorphic Encryption

Some encryption schemes are homomorphic in nature. Given two ciphertexts encrypting two plaintexts, certain operations can be performed on the ciphertexts such that the result can be decrypted to produce the outcome of applying an operation (not necessarily the same) on the plaintexts themselves. Thus, some computations can be performed on encrypted data. Schemes that exhibit homomorphic properties for a specific operation are known as partially homomorphic encryption schemes. On the other hand, if the set of permissible operations enable arbitrary computations to be performed then the schemes are referred to as fully homomorphic.[29,30,31] Such schemes are very powerful as they allow arbitrary computation on encrypted data. Unfortunately, current schemes can sometimes be limited in the number of operations that can be applied before decryption will no longer succeed (such schemes are referred to as somewhat homomorphic), or are inefficient in terms of speed or parameter or ciphertext size.

4.2.2 End-to-end Encryption

End-to-end encryption refers to the encryption of messages exchanged by two or more parties without the intervention of a centralized server. The centralized server may exist and support the transportation of the messages, but all this server sees is encrypted content. This behaviour is the opposite of the traditional message exchange protocols, in which the messages are only encrypted while in transit from the parties to the central server or from the central server to the parties.

End-to-end encryption is typically supported by having all participants have a key pair from a public-key encryption scheme. The centralized server, in addition to supporting the exchange of messages, works as a public key repository, so that users can find the public keys of the users to which they want to send messages. Once a user has the public key of another user, she can use this public key to encrypt the messages, which will only be decryptable by the owner of the associated private key. A more technically efficient variant is for users to exchange random session keys for symmetric encryption schemes using their key pairs, and then encrypt the messages with a symmetric encryption scheme using these random temporal session keys.

[26] W. Diffie and M. E. Hellman. Multiuser cryptographic techniques, In AFIPS National Computer Conference, 1976, pp. 109-112.
[27] W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Trans. Inform. Theory*, 22:644-654, 1976.
[28] R. L. Rivest, A. Shamir and L. M. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120-126, 1978.
[29] Z. Brakerski, C. Gentry and V. Vaikuntanathan, (Leveled) fully homomorphic encryption without bootstrapping" in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 2012, pp. 309–325.
[30] C. Gentry, S. Halevi and N. P. Smart, Fully homomorphic encryption with polylog overhead, in *Advances in Cryptology--EUROCRYPT 2012*,, pp. 465–482. Springer, 2012
[31] C. Gentry, *A Fully Homomorphic Encryption Scheme*, PhD Thesis, Stanford University, 2009.

A popular implementation of an end-to-end encrypted instant messaging application is Signal by Open Whisper Systems, which is also incorporated into WhatsApp and Google Hangouts.[32]

4.3 Hash Functions

Hash functions are fundamental in modern cryptography. They are used for many different purposes, such as checking the integrity of a message, generating pseudorandom numbers, and checking passwords. In general, a hash function takes as input an arbitrarily long string of bits, and outputs a bit string of fixed length. Hence, it can be seen as a function that compresses information. Hash functions are easy to compute and hard to invert. [24,25]

In certain situations, in order to guarantee additional properties, the hash functions are chosen at random from a given family of functions. We then consider an algorithm that takes a key and an arbitrary-length input string, and outputs a fixed-length string. These functions are called keyed hash functions. A very important application of hash functions is message authentication, which is about authenticating the origin of the message and, at the same time, checking its integrity.

The most common hash functions are in the family of SHA-2 functions. SHA-2 includes the algorithms SHA-224, SHA-256, SHA-384 and SHA-512. Currently, SHA-224 and SHA-256 are the most popular, and are recommended for a broad use. The previous standardized family of hash functions was SHA-1, but nowadays SHA-1 is not recommended. Only the SHA-1 hash function of 160 bits is recommended for some restricted applications. In 2012, NIST organized a competition in order to select a new family of hash functions. The new family of hash functions is called SHA3, and there are four main hash functions: SHA3-224, SHA3-256, SHA3-384 and SHA3-512.

4.4 Digital Signatures

In handwritten letters on paper, signatures guaranteed the authenticity of the document, and the author could not repudiate the document. Moreover, the paper support gave certain guarantees of non-manipulation of the message. Digital signatures were created in order to guarantee the authenticity and integrity of the messages, and to avoid the repudiation of the messages by their authors. Digital signatures appeared with the deployment of public-key encryption in order to provide a similar guarantee in the context of digital communications.

The Digital Signature Algorithm (DSA) was one of the most common signature schemes, and was adopted as a Federal Information Processing Standard (FIPS) in the early 90's. In the last twenty-five years, the recommended key lengths have been increased, but DSA is not recommendable anymore. A variant of DSA that is defined over elliptic curves, ECDSA, was adopted in recent years because of efficiency reasons. For the same level of security, ECDSA requires shorter keys and provides shorter signatures. RSA-Probabilistic Signature Scheme (RSA-PSS) was standardized by the International Standards Organization (ISO). In ISO 9796-2, they defined three digital signature schemes based on RSA, named Digital Signature 1, Digital Signature 2 and Digital Signature 3. Right now, ISO-9796-2 RSA Digital Signature 2, which is a variant of RSA-PSS, is one of the recommended signature schemes. [33,34]

If both the sender and the receiver share some information, an alternative to digital signatures is message authentication codes (MAC). MACs, which are based on keyed cryptographic hash functions, can

[32] https://whispersystems.org/

[33] ISO/IEC 9796-2. Information technology. Security techniques: Digital signatures giving message recovery. Part 2: Integer factorization based schemes. International Organization for Standardization, 2010.

[34] Algorithms, key size and parameters report, 2014. European Union Agency for Network and Information Security (ENISA).

be used in order to guarantee the integrity of the message and are commonly used in the context of symmetric encryption communications, where sender and receiver share a secret key.

4.4.1 Threshold Signatures

In a *(t, n)*-threshold signature scheme, the generation of a signature requires the cooperation of *t* out of *n* users who have shares of a private key. In the verification process, a verifier can check that the signature is correct, which means that a verifier can check that at least *t* authorized users participated in the creation of the signature.

4.4.2 Group Signatures

In a group signature scheme, a set of users, called members of the group, can issue signatures of arbitrary messages on behalf of the group. A verifier can check the validity of the signature using the group public key. The main interest of this kind of signatures is that it ensures privacy of signers against potential verifiers because a potential verifier cannot distinguish two signers from the same group. Group signatures were introduced by Chaum and van Heyst.[35]

One of the issues of group signatures is that if a member does not behave properly, the other members of the group will want to revoke the membership of the disrupting user without having to set up a new key. In order to facilitate the revocation of users, some members are distinguished with the capability to revoke the membership. These players are called group managers.

4.5 Secure Multiparty Computation

Secure multiparty computation protocols allow a set of parties to compute a joint function of their inputs in a secure way without requiring a trusted third party. During the execution of the protocol the parties do not learn anything about each other's input except what is implied by the output itself.

A general solution for the secure computation of functions among two players was introduced by Yao in 1986.[36] The main idea of these protocols was to describe the function as a circuit, and to compute every gate of the circuit in a secure way. This idea was extended to the multi-partite setting by Goldreich, Micali and Wigderson.[37] They showed how to create a secure multiparty computation protocol that allows playing any game and does not leak any information if the majority of the participants are honest. These protocols are computationally secure. The first unconditionally secure multi-party computation protocols were presented by Ben-Or, Goldwasser, Wigderson[38] and Chaum, Crépeau and Damgård.[39] They showed a protocol to compute any arithmetic function in a secure way that at least two thirds of the parties are honest.

Some of the main open problems in secure multiparty computations are to reduce the restrictions on the behaviours of the players, and to reduce the computational and communication costs of the protocols for interesting families of functions. Observe that in the general solutions described above, the computational costs of the protocol depend on the size of the circuit defining the function.

---

[35] D. Chaum and E. van Heyst. Group signatures. In *Advances in Cryptology-EUROCRYPT 1991*, pp. 257-265. Springer, 1991.
[36] A.C.-C. Yao, How to generate and exchange secrets. In *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 1986, pp. 162-167.
[37] O. Goldreich, M. Micali and A. Wigderson, How to play any mental game or a completeness theorem for protocols with honest majority. In *ACM Symposium on the Theory of Computing STOC*, 1987, 218-229.
[38] M. Ben-Or, S. Goldwasser and A. Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *ACM Symposium on the Theory of Computing (STOC),* 1988, pp. 1-10.
[39] D. Chaum and C. Crépeau, I. Damgård. Multi-party unconditionally secure protocols. In *ACM Symposium on the Theory of Computing* (STOC), 1988, pp. 11-19.

The main properties for secure multiparty computation protocols are privacy and correctness. Another important property of secure multiparty computation is fairness. A protocol is fair if there are no differences between the players with respect the obtention of the output. That is, a protocol is fair if either everybody receives output, or no one does.

4.6 Functional Cryptographic Schemes

Functional cryptographic schemes extend public-key cryptographic schemes to endow them with some additional properties. For example, identity-based cryptographic schemes use arbitrary length strings (such as an email address) as public keys, thus eliminating the need for a public key infrastructure. Attribute-based cryptography, on the other hand, allows the decryption of a ciphertext only by those who possess certain attributes (for example, a specific role in a company) instead of a specific private key, in the case of attribute-based encryption; or to sign messages for some set or function of the users' attributes in the case of attribute-based signature schemes, thus achieving some level of privacy while still ensuring the integrity and authenticity of the message.

4.7 Ethical Considerations Raised by Cryptography

One of the most discussed topics in the last years is the role of encrypted communications in the fight against criminal (and especially terrorist) organisations. Cases such as the FBI-Apple dispute after the San Bernardino terror attack and the proposal by former UK Prime Minister David Cameron on banning end-to-end encryption after the Charlie Hebdo attack, among others, have brought this debate to the public opinion.[40,41] The underlying debate in both cases is whether law enforcement agencies (LEAs) should be given unrestricted access to private communications and data from citizens. This debate is fuelled by the false dichotomy between security and privacy. Imposing restrictions on encryption technologies or including backdoors to provide LEAs with more data about the citizens directly leads to possible security flaws, and therefore a less secure Internet. On the other hand, the benefits to the fight against criminal organizations may not be as clear, taking into account that metadata from communications is still available to LEAs even if they do not have access to the contents of such communications.

Limitations on encryption technologies, however, have existed for a long time, although many of those have relaxed in the last decade.[42] The US government and those countries within the Wassenaar arrangement limit imports and exports of cryptography technologies and key sizes (thus, their security levels).[43] Other limitations exist within local legislations.[42] These controls limit the autonomy of common citizens to protect themselves (either from others or from their governments). Proponents and defendants of such controls, however, seem to fail to understand that criminal groups can develop their own encrypted messaging services and distribute them among their members, leaving as a consequence a less protected general population and equally protected powerful and/or criminal groups.

Cybersecurity experts and software systems designers should have a deeper understanding of the advances in cryptographic technologies since the 80's. Several advanced schemes, such as group signatures, homomorphic and functional encryption, among others, could highly improve cybersecurity systems with regards to the protection of the consumer's rights and values.

[40] https://www.washingtonpost.com/world/national-security/us-wants-apple-to-help-unlock-iphone-used-by-san-bernardino-shooter/2016/02/16/69b903ee-d4d9-11e5-9823-02b905009f99_story.html
[41] https://www.theverge.com/2015/1/12/7533065/whatsapp-imessage-ban-uk-government-encryption
[42] Survey on existing and proposed laws and regulations on cryptography. http://www.cryptolaw.org/
[43] The Wassenaar arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies - http://www.wassenaar.org/

5. Data Anonymization and Processing

The impact on privacy and fairness by cybersecurity services that collect and process personal data can be mitigated somehow by applying anonymization techniques. This section describes the current state of the art in such techniques.

5.1 Database Anonymization

Traditionally, institutes and governmental statistical agencies have systematically gathered information about individual respondents, either people or companies, with the aim of distributing this information to the research community. Commonly, statistical agencies make this information public by releasing a *microdata set*, essentially a database table whose records carry data referring to the respondents. While these databases may be extremely useful for researchers, it is fundamental that their publication does not compromise the respondents' privacy in the sense of revealing information about specific individuals. Statistical disclosure control (SDC) is the discipline that deals with the inherent trade-off between protecting the privacy of the respondents and ensuring that the protected data are still useful to researchers.

Usually, a microdata set contains a set of attributes that may be classified as *identifiers*, *key attributes* (a.k.a. *quasi-identifiers*), or *confidential attributes*. First, identifiers allow to unequivocally identify individuals. It would be the case of social security numbers or full names, which would be removed before the publication of the microdata set. Secondly, key attributes are those attributes that, in combination, may be linked with external information to re-identify the respondents to whom the records in the microdata set refer. Examples include job, address, age, gender, height and weight. Last but not least, the microdata set contains confidential attributes with sensitive information on the respondent, such as salary, religion, political affiliation or health condition.

Several methods have been proposed in the literature to protect such microdata sets.[44] Some of them aim to prevent identity disclosure, whereas others try to avoid the disclosure of the confidential attribute. Next, we briefly explain each of them.

5.1.1 Non-Perturbative Masking

In SDC, masking refers to the process of producing a modified safe data set X' from the original X. It can be *perturbative* masking or *non-perturbative* masking. In the former approach, the data are modified in such a way that some properties are retained. In the latter type of masking, X' is obtained by removing some data cells and/or by making certain specific data values more general, yet the data in X' is still true; as an example, a data value might be replaced by a range of values containing the original data.

Common non-perturbative methods include:

- *Sampling*. Instead of publishing the whole data set, only a sample of it is released.
- *Generalization*. The values of the different attributes are recoded in new, more general categories such that the information remains the same, albeit less specific.
- *Top/bottom coding*. Similarly to the previous approach, values above (resp. below) a certain threshold are grouped together into a single category.
- *Local suppression*. If a combination of quasi-identifiers is shared by too few records, this may lead to re-identification. This method relies on replacing certain individual attribute values with

[44] A. Hundepool, J. Domingo-Ferrer, L. Franconi, S. Giessing, E. S. Nordholt, K. Spicer and P.-P. de Wolf, *Statistical Disclosure Control*. Chichester, UK: Wiley, 2012.

missing values, so that the number of records sharing a particular combination of quasi-identifiers becomes larger.

5.1.2 Perturbative Masking

Perturbative masking generates a modified version of the microdata set such that the privacy of the respondents is protected to a certain extent and, at the same time, certain statistical properties of the data are preserved. Well-known perturbative masking methods include:

- *Noise addition*. This is the most popular method, which consists in adding a Gaussian noise vector to each record in the data set.
- *Data swapping*. This technique exchanges the values of the attributes randomly among individual records. Clearly, univariate distributions are preserved with this technique.
- *Microaggregation*. It aims to group similar records together and release the average record of each group.[45] The more similar the records in a group are, the more data utility is.

5.1.3 Synthetic Microdata Generation

An important anonymization method consists in generating a synthetic data set. That is, instead of modifying the original data set, a whole synthetic data set is generated such that some properties of the original one are preserved. The main advantage of synthetic data is that no respondent re-identification seems possible since the data are artificial. However, if, by chance, a synthetic record is very close to an original one, the respondent will not feel safe when supplying their data. Besides, the utility of synthetic data sets is limited to the properties initially selected by the method.

Some examples of synthetic generation include methods based on multiple imputation[46] and methods that preserve means and co-variances.[47] A good alternative to the drawbacks of synthetic generator methods is hybrid data, which mix original and synthetic data and are, therefore, more flexible.[48]

5.1.4 Privacy Models

For an anonymized data set X' to be safe/private enough, it needs to be sufficiently anonymized. The level of anonymization can be assessed after the generation of X' or prior to it.

Ex post methods rely on the analysis of the output data set and, therefore, it is possible to generate a data set that is not safe enough according to a certain criteria; several iterations with increasingly strict privacy parameters and decreasing utility may be needed. The most commonly used *ex post* approach is masking followed by record linkage, where the latter evaluates the proportion of masked records that can be linked to the respective original records they come from.

On the other hand, the *ex-ante* approach relies on *privacy models* that allow selecting the desired privacy level before producing X'. In this way, the output data set is always as private a specified by the model, although it may fail to provide enough utility if the model parameters are too strict.

K-Anonymity and Extensions: A well-known prior privacy model is *k-anonymity*, which is the requirement that each tuple of key-attribute values be shared by at least *k* records in the database. This condition may be achieved through generalization and suppression mechanisms, and also through more complex

[45] J. Domingo-Ferrer and J. M. Mateo-Sanz, Practical data-oriented microaggregation for statistical disclosure control, *IEEE Trans. Knowl. Data Eng.*, 14(1):189–201, 2002.
[46] D. B. Rubin, Discussion: statistical disclosure limitation," *J. Off. Stat*., 9(2):461–468, 1993.
[47] J. Burridge, Information preserving statistical obfuscation, *Stat. Comput*., vol. 13, pp. 321–327, 2003.
[48] J. Domingo-Ferrer and Ú. González-Nicolás, Hybrid microdata using microaggregation, Information Sciences, 180(15):2834–2844, 2010.

procedures such as microaggregation.[49,50] Unfortunately, while this privacy model prevents identity disclosure, it may fail to protect against the disclosure of the confidential attributes. The definition of this privacy model establishes that complete re-identification is unfeasible within a group of records sharing the same tuple of perturbed key-attribute values. However, if the records in the group also share a common value of a confidential attribute, the association between an individual linkable to the group of perturbed key attributes and the corresponding confidential attribute value remains disclosed. In order to prevent this problem, some extensions of *k-anonymity* have been proposed, the most popular being *l-diversity*[51] and *t-closeness*[52]. The property of *l-diversity* is satisfied when there are at least *l* "well-represented" values for each confidential attribute in all groups sharing the values of the quasi-identifiers. The property of *t-closeness* is satisfied when the distance between the distribution of each confidential attribute within the groups and the whole data set is no more than a threshold *t*.

Differential Privacy: Another example of prior privacy model is *differential privacy*. This model was originally defined for queryable databases and consists in perturbing the original query result of a database before outputting it. However, this is equivalent to perturbing the original data and then computing the queries over the modified data and, thus, *ε-differential privacy* can also be seen as a privacy model for microdata sets. A differentially private algorithm ensures that given two datasets that differ by only a single record, the algorithm will perform almost equally on both datasets. That is, the existence or not of a single record will not alter significantly the output of the algorithm. Typically, ε-differential privacy is attained by adding Laplace noise with zero mean and parameter $\Delta(f)/\varepsilon$, where $\Delta(f)$ is the sensitivity of the algorithm $t$and ε is a privacy parameter; the larger ε, the less privacy.

5.1.5 Permutation Model for Anonymization

A permutation model addresses the verifiability of anonymization by subjects.[53] Generally, *ex post* privacy models (as well as some *ex ante* ones such as differential privacy) do not allow subject verifiability and, thus, respondents need to trust the data holder. This can lead to biased answers. The novelty of the permutation model lies in that each respondent can evaluate the anonymization level of their own answer and verify that the data holder is trustworthy.

The permutation model views all anonymization methods as being functionally equivalent to a two-step procedure consisting of a permutation step (mapping the original data set to the output of the reverse mapping procedure[54]) plus a noise addition step (adding the difference between the reverse-mapped output and the anonymized data set). Since the ranks in the reverse-mapped version and in the anonymized version are the same by construction, the noise added in the second step needs to be small, since, otherwise, ranks would change. This shows that any anonymization method basically amounts to permutation.

The most interesting feature, however, is that each subject can check whether a privacy model called (d,**v**)-permuted privacy with respect to her original record is satisfied by the anonymized data set for some d and **v** of her choice; in plain words, each subject can check whether her response has been

[49] P. Samarati and L. Sweeney, Protecting privacy when disclosing information: k -anonymity and Its enforcement through generalization and suppression,, Technical Report, SRI International, 1998.
[50] J. Domingo-Ferrer and V. Torra, Ordinal, continuous and heterogeneous k-anonymity through microaggregation, *Data Min. Knowl. Discov*., 11(2):195–212, 2005.
[51] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-Diversity: Privacy beyond *k*-anonymity" *ACM Trans. Knowl. Discov. Data*, 1(1,) art. no. 3, 2007.
[52] L. Ninghui, L. Tiancheng, and S. Venkatasubramanian, t-Closeness: Privacy beyond k-anonymity and ℓ-diversity, in *Proceedings - International Conference on Data Engineering*, 2007, pp. 106–115.
[53] J. Domingo-ferrer and K. Muralidhar, New directions in anonymization: permutation paradigm, verifiability by subjects and intruders, transparency to users, Information Sciences, 337-338:11-24, 2016.
[54] K. Muralidhar, R. Sarathy and J. Domingo-Ferrer, Reverse mapping to preserve the marginal distributions of attributes in masked microdata, in *Priv. Stat. Databases-PSD 2014,* LNCS, 8744, pp. 105–116. Springer, 2014.

permuted enough in the anonymized data set. The subject only needs to know her original record and the anonymized data set.

5.2 Redaction and Sanitization of Documents

In this subsection, we focus on data protection methods for documents rather than databases. *Document redaction* consists of removing or blacking out sensitive terms in plain textual documents. Alternatively, when sensitive terms are replaced (instead of removed) by generalizations (e.g., AIDS → disease), the process is more generically referred to as *document sanitization*[55]. Document sanitization is more desirable than pure redaction, since the former better preserves the utility of the protected output. Moreover, in document redaction, the existence of blacked-out parts in the released document can raise awareness of the document's sensitivity to potential attackers[55], whereas sanitization gives no such clues.

In both cases, two tasks should be performed: i) detection of textual terms that may cause disclosure of sensitive information, and ii) removal or obfuscation of those entities. Traditionally, the detection of sensitive terms has been tackled in a manual way. This requires a human expert who applies certain standard guidelines that detail the correct procedures to sanitize sensitive entities[56]. Manual redaction has proven to be quite time-consuming and does not scale to currently required levels of information outsourcing[55,57].

In recent years, numerous *automatic redaction* methods have been proposed. Some approaches rely on specific or tailored patterns to detect certain types of information based on their linguistic or structural regularities (e.g., names, addresses and social security numbers).[58,59,60] Some schemes capitalize on specific patterns to remove sensitive terms from medical records.[59,60] These patterns are designed according to the HIPAA "Safe Harbor" rules that specify eighteen data elements which must be eliminated from clinical data in order to anonymize a clinical text[61]. As an alternative to manually-specified patterns, several authors have proposed using trained classifiers that recognize sensitive entities. Others proposed a tool that focuses on the sanitization of documents directly linked to certain companies.[62] The data to be detected include words and phrases that reveal the company the document belongs to.

With regard to automatic sanitization, Abril and Navarro-Arribas[63] propose a general scheme that uses a trained classifier for named entity recognition (NER) (*i.e.*, the Stanford NER[64]) to automatically recognize entities belonging to general categories such as person, organization and location names. This

[55] E. Bier, R. Chow, P. Golle, T. H. King and J. Staddon, The Rules of redaction: identify, protect, review (and repeat), *IEEE Secur. Priv. Mag.*, vol. 7, no. 6, pp. 46–53, 2009.
[56] National Security Agency, Redacting with confidence: How to safely publish sanitized reports converted from word to pdf, 2005.
[57] V. T. Chakaravarthy, H. Gupta, P. Roy and M. K. Mohania, Efficient techniques for document sanitization., in *17th ACM Conference on Information and Knowledge Management (CIKM'08)*, 2008, pp. 843–852.
[58] L. Sweeney, Replacing personally-identifying information in medical records, the scrub system, in *1996 American Medical Informatics Association Annual Fall Symposium*, 1996, pp. 333–337.
[59] M. Douglass, G. Cliffford, A. Reisner, W. Long, G. Moody and R. Mark, De-identification algorithm for free-text nursing notes, in *Computers in Cardiology*, 2005, pp. 331–334.
[60] A. Tveit, O. Edsberg, T. B. Rost, A. Faxvaag, O. Nytro, T. Nordgard, M. T. Ranang and A. Grimsmo, Anonymization of general practitioner medical records, in *Second HelsIT Conference*, 2004.
[61] Department of Health and Human Services, The Health insurance Portability and Accountability Act of 1996, *Public Law*, 104-191 1996.
[62] C. Cumby and R. Ghani, A machine learning based system for semi-automatically redacting documents, in *Twenty-Third Conference on Innovative Applications of Artificial Intelligence*, 2011, pp. 1628–1635.
[63] D. Abril and G. Navarro-Arribas, On the declassification of confidential documents, in *Modelling Decisions for Artificial Intelligence*-MDAI 2011, pp. 235–246. Springer 2011.
[64] J. Finkel, T. Grenager and C. Manning, Incorporating non-local information into information extraction systems by gibbs sampling, in *43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, 2005, pp. 363–370.

mechanism suggests generalizing sensitive entities instead of removing them from the sanitized document. The goal is to achieve a certain degree of privacy while preserving some of the semantics. Along the same line, Finkel et al.[65] provide a theoretical measure ("t-plausibility") that guides the sanitization process to balance the trade-off between privacy protection and utility preservation. To preserve the utility of the sanitized documents, the proposed solution suggests using a general-purpose ontology/taxonomy to generalize specific terms. Finally, Sánchez et al.[66] present a system that relies on information theory to quantify the amount of information provided by each term of the document. The work is built on the basis presented by Sánchez et al.[67], although this later work successfully addresses the generalization of the sensitive terms.

5.3 Data Stream Anonymization

A data stream is a sequence of data items that become available over time. This type of data is common to some environments, such as sensor networks, web logs, etc. Stream data are quite different from static data sets. In particular, stream data are potentially infinite, may be fast flowing, and may require a fast response. Because of these particularities, anonymization methods that target data streams have been specifically designed.

Data protection regulations require taking into account the privacy of the individuals when analysing the collected data. This is the case for both static data and data streams. However, because of their properties, data streams are particularly difficult to anonymize. Whereas there is a large body of disclosure control methods for static data, the disclosure risk control literature on data streams is limited. Disclosure risk control methods for data streams follow three main approaches: perturbative masking, non-perturbative masking and counterfeiting.

In the perturbative masking approach, some noise is added to conceal the real value of the records. In a work by Li et al.[68] the correlation and the autocorrelation of multivariate data streams is tracked in an attempt to find a good trade-off between privacy and utility. Differential privacy has also been used to anonymize data streams in some constrained scenarios. Dwork et al.[69] propose a differentially private counter of the number of 1's in a data stream is released at each step. Bolot et al.[70] generalize the previous method to compute differentially private sums over restricted windows.

In the non-perturbative masking approach, one seeks to hide each record in the stream within a group of records. In the static data setting, k-anonymity and its extensions are well-known privacy models that follow this approach. Cao et al.[71] adapt these privacy models to stream data. As to make groups we need to accumulate records, this approach necessarily introduces some delay in the release of the anonymized stream.

---

[65] W. Jiang, M. Murugesan, C. Clifton and L. Si, t-plausibility: Semantic preserving text sanitization, in *International Conference on Computational Science and Engineering (CSE'09)*, 2009, vol. 3, pp. 68–75.

[66] D. Sánchez, M. Batet and A. Viejo, Automatic general-purpose sanitization of textual documents, *IEEE Trans. Inf. Forensics Secur.*, 8(6):853–862, 2013.

[67] D. Sánchez, M. Batet and A. Viejo, Detecting sensitive information from textual documents: an information-theoretic approach, in *Modeling Decisions for Artificial Intelligence- MDAI 2012*, pp. 173–184. Springer, 2012.

[68] F. Li, J. Sun, S. Papadimitriou, G. A. Mihaila and I. Stanoi. Hiding in the crowd: Privacy preservation on evolving streams through correlation tracking. In *IEEE 23rd International Conference on Data Engineering-ICDE 2007*, pp. 686-695, 2007

[69] C. Dwork, M. Naor, T. Pitassi and G. N. Rothblum. 2010. Differential privacy under continual observation. In *Proceedings of the forty-second ACM symposium on Theory of computing (STOC '10).* ACM, New York, NY, USA, 715-724.

[70] J. Bolot, N. Fawaz, S. Muthukrishnan, A. Nikolov and N. Taft. 2013. Private decayed predicate sums on streams. In *Proceedings of the 16th International Conference on Database Theory (ICDT '13),* W.-C. Tan, G. Guerrini, B. Catania, and A. Gounaris (Eds.). ACM, New York, NY, USA, 284-295.

[71] J. Cao, B. Carminati, E. Ferrari, and K.-L. Tan. *Castle: Continuously anonymizing data streams*. *IEEE Trans. on Dependable and Secure Computing,* 8(3):337–352, 2011.

In the counterfeiting approach, we also seek to hide a record within a group of records. By hiding each record within a group of fake records, we avoid the delay inherent to the previous approach[72]. The main drawback is the overhead introduced by the addition of fake records.

5.4 Discrimination Prevention in Data Mining

Other than privacy implications, automated data collection and processing may have a secondary negative impact, which is discrimination. Automated data mining is used in several services to derive association and classification rules, which are then applied to a variety of decisions, such as loan granting, personnel selection, insurance premium computation, etc. While an automated classifier may be seen as a fair decision-making tool, if the training data are inherently biased, the generated rules will result in possibly discriminatory decisions.

Some works tackle this issue by pre-processing the training data using techniques akin to those from statistical disclosure control, but aiming at reducing the inherent bias in the data. Others act directly on the automatically mined rules, either by eliminating some of them or by generalizing some of the conditions of these rules.[73,74,75]

5.5 Ethical Considerations on Anonymization applied to Cybersecurity Countermeasures

The tools described in this section help to ensure the privacy of users which participate (willing or not) in data collection processes, additionally, anti-discrimination technologies ensure that the automatic processing of the collected data does not produce discriminating results or classification mechanisms that are discriminating.

Cybersecurity providers face the dilemma between more protection, which entails having a tighter control on the observed networks and systems and collecting more data, and the protection of the privacy of their clients. The current discussion, however, misses the benefits that anonymization technologies have to offer. Anonymization technologies are a powerful tool which is often ignored in the cybersecurity community. Stream anonymization, differential privacy, and sanitization could be used at data collection endpoints (points at which surveillance occurs) to limit the personal information that security providers collect about their clients. Additionally, database anonymization could be used on intelligence data in order to publish or share it among other organisms.

Using such tools, the compromise between control and privacy of the customers can be relaxed, since more data can be collected, and thus a higher protection capability, while protecting especially sensitive information.

[72] S. Kim, M. K. Sung, and Y. D. Chung. A framework to preserve the privacy of electronic health data streams. *Journal of biomedical informatics*, 50:95-106, 2014.

[73] S. Hajian and J. Domingo-Ferrer. A methodology for direct and indirect discrimination prevention in data mining. *IEEE transactions on Knowledge and Data Engineering*, 25(7):1445-1459, 2013.

[74] S. Hajian, J. Domingo-Ferrer, A. Monreale, D. Pedreschi and F. Giannotti, Discrimination- and privacy-aware patterns, *Data Mining and Knowledge Discovery*, 29(6):1733-1782, 2015.

[75] S. Hajian, J. Domingo-Ferrer and O. Farràs, Generalization-based privacy preservation and discrimination prevention in data publishing and mining, *Data Mining and Knowledge Discovery*, 28(5):1158-1188, 2014.

6. Conclusions

The authors would like to emphasize the fact that there are many effective countermeasures available. It is possible to find solutions and mitigation to most of the current threats and to improve the situation step by step. While all mitigation measures have side effects on the values, doing nothing would harm these values even more.

It is important to understand that most situations need balancing the various security goals, the methods used, their efficiency, whether they are effective for prevention, detection or reacting and their possible side effects. There is no magic silver bullet that can solve all these problems at once. Rather, one should take a multi-layered approach that consists of many different pieces in different domains that must be followed with great endurance over a long-time period. Working with Deming Cycles[76] is helpful as this enforces an evolutionary approach towards more secure systems.

When it comes to actions at a national level, transparency is crucial and a political process should balance the various opinions and avoid neglecting views of minority groups. One should try to avoid isolated views only defending the values deemed important, but try to have a more holistic view on the problem and be willing to make trade-offs for a greater benefit.

For many countermeasures, it is possible to provide guidance on how negative impacts can be reduced. One major countermeasure is that every person involved in this area is ready to question his/her actions and projects for their impact on security, safety and ethical core values. If such values can be made part of the engineering education, technological development is going to be better aligned with core values. Technologies overviewed in this document are powerful tools to achieve these goals.

For a cybersecurity technology and services provider, a key ethical dilemma today is how intrusive they should be, that is, a choice between (potentially) better protecting their customers and avoiding deeper knowledge and control of the customer environments. This is a multi-faceted issue, involving ethical, legal, business and competition, technological, and cultural aspects. The choice is further complicated by the fact that the boundaries between attacker types and primary threats for a specific customer are becoming blurred and our assumptions about those are a moving target. As an example, most of us are likely to agree that a security service provider should be more intrusive and controlling when protecting a critical infrastructure from state-sponsored attackers compared with protecting an SME from cybercrime. However, we have been observing more and more cases where criminals, once arrested, were co-opted by certain governments and evolved towards mercenary services. So, protecting against cybercrime today may easily turn into protecting against state-supported attacks tomorrow. At the same time, high-profile targets are often attacked by compromising first some of their business partners, which may not be of any direct interest for attackers and may feel quite safe. Such blurred lines may require security providers to adopt "security paranoid" protection strategies and configurations for most of their customers, which is difficult to reconcile with respecting the customers' privacy and autonomy. Differences in business culture, legislation, and other aspects between European and other cybersecurity players just make the matters more challenging. As a result, in practice, the cybersecurity domain today appears to be mostly self-regulated with respect to dealing with ethical issues. In particular, choices of sharing important and sensitive information are often based on personal experience, relationships, and trust.

Clearly, the application of cybersecurity measures is necessary for privacy; however, not all practices should be acceptable. For example, full monitoring of the IT infrastructure will very surely help protecting it, but seems an excessive measure. Therefore, we consider transparency of every security measure to be one of the most important factors to take into account regarding the application of cybersecurity measures, along with the discussion about its effectiveness as well as potential collateral damages. As

[76] PDCA – Plan-Do-Check-Act is an iterative management method.

privacy is only achieved when confidentiality and integrity are ensured, which are security goals, it is clear that privacy cannot be achieved without safeguards. Even though some security measures clearly endanger privacy, the question is not whether we need security but rather how the security measures are applied. The objective of building privacy-aware solutions, which is of paramount importance in sectors such as eHealth, drives research into security mechanisms, mostly in the cryptographic and data privacy research communities.

Appendix

A.1 Glossary

APT Advanced persistent threat.

AV Antivirus software.

Botnet A botnet is a number of networked devices that are coordinated by a command and control server (C&C).

CERT Computer Emergency Response Team.

CMS Content management system.

CSIRT Computer Security Incident Response Team.

DDoS Distributed Denial-of-Service attack. A DoS attack in which the attacker (or attackers) has more than one unique IP address (i.e. machines), often thousands.

DNS Domain name system. Hierarchical, distributed naming system for computers and networks. It translates easy to remember names (URLs) to IP addresses.

DoS Denial-of-Service attack. A cyber-attack which aims to make a certain target machine or network resource unavailable, typically by flooding the target with spurious requests.

HTTP Hypertext transfer protocol. Core communication protocol for the transmission of web traffic.

ICT Information and communication technologies.

IRC Internet Relay Chat. Real-time text-based communication protocol, where users communicate through channels. It is the precursor of internet forums, and of nowadays social networks.

ISO International Standards Organization.

NIST National Institute for Standards and Technology (USA).

NSA National Security Agency. Intelligence agency of the USA.

SDC Statistical Disclosure Control.

SMB Server Message Block protocol. A network file sharing protocol implemented in Microsoft Windows.

SMTP Simple mail transfer protocol. Core protocol for the transmission of electronic mail.

SSDP Simple Service Discovery Protocol. A network protocol based on the Internet Protocol Suite for advertisement and discovery of network services and presence information.

TOR The Onion Router. Traffic anonymization service. TOR directs network traffic through a series of volunteer relays.

UPnP Universal Plug and Play. A set of networking protocols that permits networked devices to seamlessly discover each other's presence on the network and establish functional network services.

A.2 Further Reading

The authors recommend the following literature in order to gain deeper insights into the topics discussed in the White Paper:

ENISA Threat Taxonomy - A tool for structuring threat information. *This report by ENISA contains a taxonomy of threats to information and communication systems, both in the cyber- and physical space. The proposed taxonomy is a compilation of works carried out from 2012 to 2015 and has been used internally by ENISA as a common reference in other cybersecurity related reports.* Available at: https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends/enisa-threat-landscape/etl2015/enisa-threat-taxonomy-a-tool-for-structuring-threat-information

ENISA Privacy and Data Protection by design - from policy to engineering. *This document collects previous works on Privacy-by-design and the strategies to develop software that preserves the privacy of their users. These strategies inspire part of the GDPR and should be followed by compliant software. Next, it describes specific tools and technologies to develop said strategies.* Available at: https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design/at_download/fullReport

Scientific advice mechanism scoping paper: Cybersecurity. *This work provides an introduction on the cyber-security topic and the state of the policies of the EU.* Available at: https://ec.europa.eu/research/sam/pdf/meetings/hlg_sam_012016_scoping_paper_cybersecurity.pdf

Delfs, H., Knebl, H. (2015): Introduction to cryptography (Vol. 2). Springer, ISBN: 978-3-662-47973-5 (Print) 978-3-662-47974-2 (Online). *This book covers key concepts in modern cryptography, from encryption and digital signatures to more advanced cryptographic protocols. Topics on algebra, number theory, probability theory and information theory are included, so no previous background is required.*

Lecture notes on cryptography. Summer course "Cryptography and computer security" at MIT. *Collection of lecture notes from the summer course on cryptography at MIT by Shafi Goldwasser and Mihir Bellare. The notes cover most of the basic topics in cryptography, including an introduction to information theory and number theory, which are the basis of modern cryptography.* Available at: https://cseweb.ucsd.edu/~mihir/papers/gb.pdf

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E.S., Spicer, K., de Wolf P.-P. (2012): Statistical disclosure control. Wiley and Sons, ISBN: 978-1-119-97815-2. *This book introduces the topic of statistical disclosure control and provides several strategies for protecting the privacy of respondents in microdata databases. The work most prominently contains the definition of k-anonymity and other privacy models, along with mechanisms to achieve k-anonymity.*

A.2 References

Abril, D., Navarro-Arribas, G., & Torra, V. (2011). On the declassification of confidential documents. *Modeling decision for artificial intelligence*, 235-246.

Ben-Or, M., Goldwasser, S., & Wigderson, A. (1988, January). Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *Proceedings of the twentieth annual ACM symposium on Theory of computing* (pp. 1-10). ACM.

Bier, E., Chow, R., Gollé, P., King, T. H., & Staddon, J. (2009). The rules of redaction: Identify, protect, review (and repeat). *IEEE Security & Privacy*, *7*(6).

Brakerski, Z., Gentry, C., & Vaikuntanathan, V. (2014). (Leveled) fully homomorphic encryption without bootstrapping. *ACM Transactions on Computation Theory (TOCT)*, *6*(3), 13.

Burridge, J. (2003). Information preserving statistical obfuscation. *Statistics and Computing*, *13*(4), 321-327.

Chakaravarthy, V. T., Gupta, H., Roy, P., & Mohania, M. K. (2008, October). Efficient techniques for document sanitization. In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 843-852). ACM.

Chaum, D., Crépeau, C., & Damgard, I. (1988, January). Multiparty unconditionally secure protocols. In *Proceedings of the twentieth annual ACM symposium on Theory of computing* (pp. 11-19). ACM.

Chaum, D., & Van Heyst, E. (1991). Group signatures. In *Advances in Cryptology—EUROCRYPT'91* (pp. 257-265). Springer Berlin/Heidelberg.

Cumby, C. M., & Ghani, R. (2011, August). A Machine Learning Based System for Semi-Automatically Redacting Documents. In *IAAI*.

Diffie, W., & Hellman, M. E. (1976, June). Multiuser cryptographic techniques. In *Proceedings of the June 7-10, 1976, national computer conference and exposition* (pp. 109-112). ACM.

Diffie, W., & Hellman, M. (1976). New directions in cryptography. *IEEE transactions on Information Theory*, *22*(6), 644-654.

Domingo-Ferrer, J., & Mateo-Sanz, J. M. (2002). Practical data-oriented microaggregation for statistical disclosure control. *IEEE Transactions on Knowledge and data Engineering*, *14*(1), 189-201.

Domingo-Ferrer, J., & González-Nicolás, Ú. (2010). Hybrid microdata using microaggregation. *Information Sciences*, *180*(15), 2834-2844.

Domingo-Ferrer, J., & Torra, V. (2005). Ordinal, continuous and heterogeneous k-anonymity through microaggregation. *Data Mining and Knowledge Discovery*, *11*(2), 195-212.

Domingo-Ferrer, J., & Muralidhar, K. (2016). New directions in anonymization: Permutation paradigm, verifiability by subjects and intruders, transparency to users. *Information Sciences*, *337*, 11-24.

Douglass, M. M., Cliffford, G. D., Reisner, A., Long, W. J., Moody, G. B., & Mark, R. G. (2005, September). De-identification algorithm for free-text nursing notes. In *Computers in Cardiology, 2005* (pp. 331-334). IEEE.

Finkel, J. R., Grenager, T., & Manning, C. (2005, June). Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd annual meeting on association for computational linguistics* (pp. 363-370). Association for Computational Linguistics.

Gentry, C. (2009). *A fully homomorphic encryption scheme*. Stanford University.

Gentry, C., Halevi, S., & Smart, N. P. (2012, April). Fully homomorphic encryption with polylog overhead. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques* (pp. 465-482). Springer, Berlin, Heidelberg.

Goldwasser, S., Micali, S., & Wigderson, A. (1987). How to play any mental game, or a completeness theorem for protocols with an honest majority. In *Proc. of the Nienteenth Annual ACM STOC* (Vol. 87, pp. 218-229).

Hajian, S., & Domingo-Ferrer, J. (2013). A methodology for direct and indirect discrimination prevention in data mining. *IEEE transactions on knowledge and data engineering*, *25*(7), 1445-1459.

Hajian, S., Domingo-Ferrer, J., Monreale, A., Pedreschi, D., & Giannotti, F. (2015). Discrimination-and privacy-aware patterns. *Data Mining and Knowledge Discovery*, *29*(6), 1733-1782.

Hajian, S., Domingo-Ferrer, J., & Farràs, O. (2014). Generalization-based privacy preservation and discrimination prevention in data publishing and mining. *Data Mining and Knowledge Discovery*, *28*(5-6), 1158-1188.

Jiang, W., Murugesan, M., Clifton, C., & Si, L. (2009, August). t-plausibility: Semantic preserving text sanitization. In *Computational Science and Engineering, 2009. CSE'09. International Conference on* (Vol. 3, pp. 68-75). IEEE.

Kim, S., Sung, M. K., & Chung, Y. D. (2014). A framework to preserve the privacy of electronic health data streams. *Journal of biomedical informatics*, *50*, 95-106.

Machanavajjhala, A., Gehrke, J., Kifer, D., & Venkitasubramaniam, M. (2006, April). l-diversity: Privacy beyond k-anonymity. In *Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on* (pp. 24-24). IEEE.

Muralidhar, K., Sarathy, R., & Domingo-Ferrer, J. (2014, September). Reverse mapping to preserve the marginal distributions of attributes in masked microdata. In *International Conference on Privacy in Statistical Databases* (pp. 105-116). Springer, Cham.

Li, N., Li, T., & Venkatasubramanian, S. (2007, April). t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on* (pp. 106-115). IEEE.

Rivest, R. L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, *21*(2), 120-126.

Rubin, D. B. (1993). Discussion statistical disclosure limitation. *Journal of official Statistics*, *9*(2), 461.

Samarati, P., & Sweeney, L. (1998). *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression*. Technical report, SRI International.

Sanchez, D., Batet, M., & Viejo, A. (2013). Automatic general-purpose sanitization of textual documents. *IEEE Transactions on Information Forensics and Security*, *8*(6), 853-862.

Sweeney, L. (1996). Replacing personally-identifying information in medical records, the Scrub system. In *Proceedings of the AMIA annual fall symposium* (p. 333). American Medical Informatics Association.

Yao, A. C. C. (1986, October). How to generate and exchange secrets. In *Foundations of Computer Science, 1986., 27th Annual Symposium on* (pp. 162-167). IEEE.