The Handbook of Privacy Studies

An Interdisciplinary Introduction

Edited by
Bart van der Sloot & Aviva de Groot

Amsterdam University Press

Cover design: Moker Ontwerp
Lay-out: Crius Group, Hulshout

ISBN 978 94 6298 809 5
e-ISBN 978 90 4854 013 6
DOI 10.5117/9789462988095
NUR 740

Contents

5. Privacy from an Informatics Perspective

Matthijs Koot & Cees de Laat

5.1 Introduction

Both 'privacy' and 'informatics' are semantically overloaded concepts; no broad consensus exists on a single definition of either. This chapter has the following objectives:

- to provide an intuition of 'privacy' and of 'informatics';
- to provide an understanding of relations between privacy and informatics;
- to provide references to academic and other authoritative sources for further research.

Elaboration is provided on selected topics in this theme. For topics that are already described and discussed in existing sources, references are provided.

5.1.1 An intuition of privacy

At the risk of minor overlap with other chapters, a short characterization of privacy follows to keep this chapter self-contained. It is adapted from earlier work.[1]

Privacy entails some desire to hide one's characteristics, choices, behaviour, and communication from scrutiny by others. A corollary is that privacy entails some desire to exercise control over the use of personal information, for example to prevent future misuse. Phrases commonly associated with privacy include[2] 'the right to be let alone', meaning freedom of interference by others; 'the selective control of access to the self or to one's group', meaning the ability to seek or avoid interaction in accordance with the privacy level desired at a particular time; and 'informational self-determination', meaning the ability to exercise control over disclosure of information about oneself.

Contrary to what some believe, the rise of social media and ubiquitous computing does not imply the 'end' or 'death' of privacy. Rather, as Evgeny

1 Koot 2012.

2 Warren 1890; Altman 1975.

Morozov paraphrased from Helen Nissenbaum's book[3] on contextual integrity in *The Times Literary Supplement* of 12 March 2010: 'the information revolution has been so disruptive and happened so fast (…) that the minuscule and mostly imperceptible changes that digital technology has brought to our lives may not have properly registered on the social radar'. In her two and a half-year ethnographic study of American teens' engagement with social network sites, danah boyd observed[4] that teens 'developed potent strategies for managing the complexities of and social awkwardness incurred by these sites'. So, rather than privacy being irrelevant to them, the teens found a way to *work around* the lack of built-in privacy. In conclusion: privacy is not dead. At worst, it is in intensive care, beaten up by overzealous and careless use of technology. It can return to good health as policymakers, technologists, and consumers learn why, what, where, when, and how to define privacy objectives.

Privacy can also be conceived of as a means of personal security: by controlling disclosure of one's own personal information, one can self-protect against known and unknown threats stemming from potential (future) uses of that information, such as identity fraud or yet-unforeseen uses of profiling.

Now that a broad intuition of privacy has been given, an intuition of informatics follows. Further on, the relation between privacy and informatics will be defined in terms of the importance of information security to privacy.

5.1.2 An intuition of informatics

In this chapter, 'informatics' is meant in the sense of 'Information and Communication Technology' (ICT): the hardware and software that spawn from science, technology, engineering, and mathematics (STEM) and enable storage, processing, and communication of data. Relevant academic disciplines include, inter alia, computer science, electrical engineering, information science, and logic.

For two reasons, this chapter does not focus on a single STEM discipline, but on applications of their, often joint, outcomes. First, legibility must be maintained for readers that have no background in STEM disciplines. Second, privacy issues are often not yet sufficiently clear in the course of practising any single discipline without considering specific applications. For instance, design of computer networking and wireless communication

3 Nissenbaum 2010.

4 boyd, 2008. (Note: boyd spells her Christian name and surname in lowercase, as explained here: http://www.danah.org/name.html.)

protocols may focus firstly on achieving robust and efficient means of communication, and not always take security and privacy requirements into account that emerge in their use in certain application domains. Similarly, the fundamentals of artificial intelligence are purely mathematical, and not until the mathematics are applied to specific domains (healthcare, public security, insurance, and so on), specific security and privacy risks start to become clear.

The design and use of ICT for the processing of personal data by definition relates to privacy. The use of technology results in increased frequency and size of collection, retention and use of personal data, and generates forms of personal data that did not exist before: for instance, sensors inside personal devices that make measurements about the user and/or the user's environment, such as the pedometer, gyroscope, location-related sensors based on the Global Positioning System[5] (GPS), and data trails due to Wi-Fi, Bluetooth, ZigBee, and so on. These measurements are not a privacy problem per se, but the relation between the measurements, the user's identity, and other data results in new potential privacy hotspots, depending on who can access the data. This is especially relevant when devices are tethered to a service provider or corporate environment where the user's real, verified identity is already known, such as in the case of personal devices tethered to Apple or Google, or enrolled in a corporate Mobile Device Management (MDM) environment.

ICT functions can be grouped into three areas:

- **storage**: solid-state disks, hard disks, etc.;
- **networking:** network equipment, communication protocols, etc.;
- **computation**: Central Processing Units (CPUs), Field Programmable Gate Arrays (FPGAs), Systems-on-Chip (SoCs), algorithms, etc.

These functions respectively map to three main states of data:[6]

- **data at rest**: data while stored;
- **data in transit**: data while transferred over computer networks;
- **data in use**: data while calculations are performed on it.

Software applications run on devices and communicate via network infrastructures to provide functionality to end-users. The distinction between

5 Or based non-US alternatives to GPS such as Galileo (EU), BeiDou (China), and Glonass (Russia).

6 The three-states model is useful to provide an understanding of ICT, but is not formally defined.

the three states of data is not apparent to the end-user, but does matter for those who want to understand data protection from a technological perspective. There is no single mechanism that protects data in all states: the mechanisms to protect data in transit are different from mechanisms to protect data at rest, and so on; although basic building blocks can serve purposes in more than one data state, such as cryptographic algorithms.

When a smartphone user takes a photo and shares it via Facebook's mobile app, for instance, what happens can be approximated in simplified terms as follows:

– First, the image sensor ('camera') of the phone generates data, which is then processed by the CPU (data in use) and finally stored on the phone (data at rest).
– Second, the Facebook app reads the photo from disk (the photo then becomes data in use) to send it to Facebook's data centre (data in transit).
– Third, in Facebook's data centre, the photo is processed while being received (data in use) and then stored on disk in Facebook's data centres (data at rest).

Being aware of these three states helps grasp data and communications privacy from an informatics perspective, including potential threats to privacy and countermeasures to protect against such threats. A selection of available protective measures in each state will be discussed shortly, after first introducing basic security and privacy controls which can be a part of those protections.

Digital privacy requires digital security. Security is a systems property: all components must be secure in order for the system as a whole to be secure and by extension to protect user privacy: hardware, operating systems, and applications. If the security of one component fails, other components can fail, undermining security and as a result potentially undermining privacy; for instance when the vulnerabilities result in data breach. Vulnerabilities in software and hardware are still a fact of life. For that reason, an elaboration on digital security follows in the next section.

Whereas the concept of 'privacy' is not well defined in informatics, a proposal for common definitions of 'anonymity' and related concepts exists in the area of anonymity research due to Pfitzmann and Hansen.[7] A simplified explanation of 'anonymity':

– a **subject** can be said to be sufficiently anonymous;
– from the perspective of an **observer**;

7 Pfitzmann and Hansen 2010.

– with regard to an **item of interest**;
– if the **observer** cannot link the **item of interest** to the **subject** with sufficiently high probability to be useful to the observer's objective.

The subject is a person, the item of interest is an activity or data (e.g. an online transaction, a database record, or knowledge of the subject's social network), and the observer is an entity from which the subject seeks to hide its link to the item of interest ('unlinkability'). Depending on context, potential observers may include untrusted peers on a shared system or network, Internet providers, or a so-called 'global passive observer' who is attributed the ability to eavesdrop on large parts of global Internet traffic (e.g. multinational cooperation between intelligence agencies, CloudFlare, and so on).

Informatics affects privacy of personal information, privacy of personal behaviour, and privacy of personal communications; and with the emergence of wearables, millimetre wave body scanners, and e-health devices, also privacy of the person ('bodily privacy'). The use of technology such as mobile apps generates a continuous stream of 'items of interest' that are, from the perspective of its creators, linkable to an identified or identifiable subject. The latter certainly applies to mobile apps that require the user to register via a social media account ('social login').

5.1.2.1 Security and privacy controls

A characterization of information security that gained popularity since its conception at NASA in the 1970-1980s, is the so-called 'CIA triad':

1. **confidentiality**: protecting data against unauthorized read access. Example measures: logical access control (make sure only user X or group Y can read a file or a certain record in a database), physical access control (access to server rooms), encryption (make sure only users who have the right cryptographic key can access data);
2. **integrity**: protecting data against unauthorized write access. Example measures: cryptographic signatures, logical access controls;
3. **availability**: making sure data is available to authorized users. Example measures: redundant data storage and connectivity, making backups of data.

Privacy can be a motivating factor for deciding on these controls. While the CIA triad, in its simplicity, is still widely present in expert publications, it has been argued that these three controls alone are insufficient for the

proper understanding of reality and advancing security.[8] For instance, authentication, authorization, and non-repudiation have been suggested to be included as separate controls, rather than implied to be part of the three traditional controls.

A popular approach to threat modelling named STRIDE[9] captures this. Threat modelling can help detect security threats (or privacy threats[10]) that may exist despite security controls, or due to a lack of security controls.[11] STRIDE was created by Microsoft in 1999, and is an acronym for six types of threats, each of which has an associated security control to counter it:

- **Spoofing**: possibility to impersonate a user
  Security control: **authentication**
- **Tampering**: possibility to perform unauthorized changes
  Security control: **integrity**
- **Repudiation**: possibility to deny that an action was performed
  Security control: **non-repudiation**
- **Information disclosure** (data breach): possibility to access/obtain data
  Security control: **confidentiality**
- **Denial of service**: possibility to render a service unavailable to legitimate users
  Security control: **availability**
- **Elevation of privilege**: possibility to obtain more or higher privileges
  Security control: **authorization**

Distinguishing six security controls and types of threats, rather than three, provides a more fine-grained way to identify potential threats and decide on countermeasures.

The STRIDE threat modelling process is informal and, at a minimum, consists of drawing a high-level diagram about a system or infrastructure, and subsequently identifying 'trust boundaries'. For an internet-facing web application, for instance, a trust boundary exists at least between the web application and its end-users: systems should never trust user input to conform to what the application (implicitly) expects. Failing to do so

8 Ross 2016.

9 Shostack 2014.

10 Threat modelling can also be applied to privacy. For instance, see Adam Shostack, 19 February 2018: 'Threat Modeling the Privacy of Seattle Residents'. Available at https://seattleprivacy.org/threat-modeling-the-privacy-of-seattle-residents/

11 Threat modelling can also be applied to privacy. For instance, see Adam Shostack, 19 February 2018: 'Threat Modeling the Privacy of Seattle Residents'. Available at https://seattleprivacy.org/threat-modeling-the-privacy-of-seattle-residents/

may result in vulnerabilities that can be exploited to gain access to the system, the data and/or or underlying infrastructure. Everywhere a data flow crosses a trust boundary, the STRIDE elements can be considered to determine which threats are relevant and necessitate protective controls.

Which protective controls should be implemented is a context-specific matter and depends on risk management and the economics of information security and privacy. It is important to note that technologies that provide confidentiality, integrity, availability, authentication, authorization, and non-repudiation can serve security objectives and privacy objectives simultaneously.

It is important to validate whether security controls are implemented adequately. This is usually done through mandatory compliance requirements. Ideally, these are not merely approached as a 'checkbox exercise' that should be passed with the least possible effort, but embraced by upper management as critical to values. Requirements can include operational security testing such as subjecting ICT infrastructure or applications to (authorized) penetration tests, social engineering, and so on. This provides insight into the vulnerabilities in technology, procedures, and human behaviour. Security testing is already mandatory for certain categories of ICT: for instance, systems that offer their users a login via the Dutch national authentication scheme DigiD must be subjected to such testing every year. This is in accordance with a norm[12] issued by the Dutch government. Similar requirements exist or may emerge in other domains.

As long as vulnerabilities in software and hardware exist, there is a potential risk to security and privacy. The 'legacy problem' exacerbates this: organizations that keep business-critical systems that contain known vulnerabilities operational because no patches, upgrades, or less vulnerable alternatives are readily available. The legacy problem can also exist at the level of individuals: not all vendors of personal devices provide patches for the entire expected device lifetime, not all users know how to install the patches, and not all can afford to buy newer, less vulnerable models; so individuals, too, can keep vulnerable devices in use.

Data protection regulation requires data controllers to ensure that personal data has 'appropriate security'. It does not make a distinction between states of data. To assess what 'appropriate' means, threats must be

12 Specifically, norm elements C.03 and C.04 of the 'Norm ICT-beveiligingsassessments DigiD' versio 2.0, issued by Logius, a body of the Ministry of the Interior. Available at https://www.logius.nl/fileadmin/logius/ns/diensten/digid/assessments/20161215_norm_V2_ict-beveiligingsassessments_digid.pdf

identified while taking into account available methods for digital security. The next sections provide an understanding of how data can be protected in its various states: at rest, in transit, and in use.

5.1.2.2 Protecting data at rest

Protection of data at rest can consist of physical, procedural, and logical measures. Logical measures include applying encryption, keeping encryption keys secure, and applying access controls (authentication and authorization) to disk storage (filesystem permissions) and to end-user applications (application access permissions) which can access data from storage. This holds for any computer: standalone computers at home, in-house corporate file servers, shared infrastructure in data centres,[13] and so on.

For data stored in a data centre, physical protection involves technical and procedural measures to prevent, detect, and (insofar possible) repress unauthorized physical access to the data centre and within the data centre itself (compartmentalization; customers should not be able to physically access equipment of other customers). Besides fences, security cameras, burglar alarms, and physical presence of security personnel, authorized persons should be trained to be mindful of attackers attempting to gain access through social engineering. For instance, attackers may attempt to impersonate an ICT vendor, cleaning company, elevator repair person, a customer, as well as leveraging tricks to distract or manipulate security personnel to gain access. Social engineering may also involve bribery or blackmail of authorized persons. Personnel at high-privilege positions, such as security personnel themselves, may need periodic screening for potential vulnerability to enticement by criminals or foreign states via for instance Money (bribery), Ideology (strong political or religious views), Coercion (blackmail), or Ego (e.g. self-importance or revenge) (MICE) or other angles.[14]

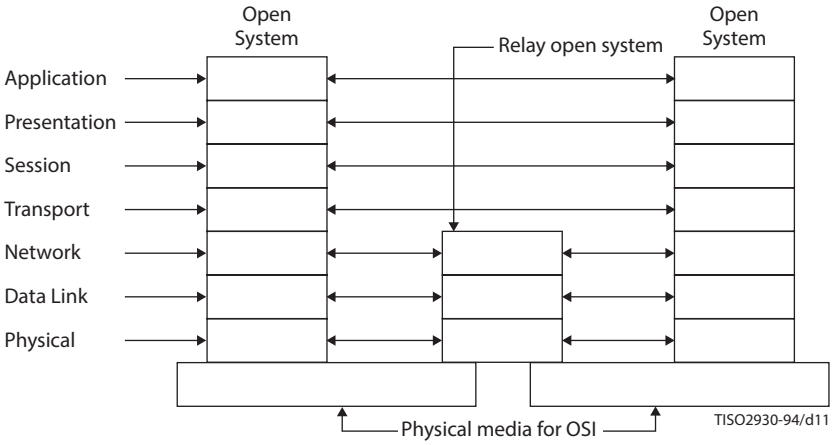
5.1.2.3 Protecting data in transit

Protection of data in transit, too, consists of physical, procedural, and logical measures. Internet exchanges are organizations that route network traffic between Internet providers and eventually, via Internet access providers, to end-users, including corporate consumers and individual consumers at home. The exchanges have to cope with risks that are similar to that of data

13 The term 'data centre' is used throughout this chapter. 'Cloud computing' is a marketing term that designates data centres: at all times, data is stored on real equipment, accessible by real operators, in a real jurisdiction.

14 Burkett 2013.

Fig 5.1: The seven-layer OSI model of data communication.



centres: the networking equipment should be protected against unauthorized physical or logical access. This is a responsibility of these exchanges.

The Open Systems Interconnection (OSI) model laid down in 1994 in ISO/IEC 7498-1 is a reference model used to characterize, design, and engineer protocols for communication between devices and applications running on those devices, including internet protocols (a topic that will be returned to later in this chapter). A basic understanding of the OSI model helps understand the protection of data in transit. The OSI model is a reference model for communication protocols. The picture below depicts a core part of the OSI model, namely the distinction of seven functional layers.

Here, 'Open System' refers to a device that participates as a sender or (final) receiver in the communication, for instance a smartphone, a laptop, or a web server in a data centre. 'Relay open system' refers to what is commonly referred to as a router. In a connection between two open systems on the global Internet, a packet travels across a series of intermediate routers, informally referred to as 'hops'. When browsing the web directly from a home computer, the home router is the first hop.

The OSI models specifies seven functional layers, seen at the left and right 'towers' in fig 5.1. Many common Internet protocols do not strictly fit in a single layer, but the model does serve a shared vocabulary in, firstly, engineering communities. The model can be (very) roughly simplified to four parts:

1. **application** + **presentation** + **session**: e.g. HTTP (web), SMTP (email), DNS ('the Internet's phonebook');
   intuition: a letter is typed by a user;

2. **transport**: e.g. TCP, UDP;
   intuition: the letter is put inside an envelope;
3. **network**: e.g. IPv4, IPv6;
   intuition: the recipient address is written on an envelope and the envelope is handed over to a postal service;
4. **physical** + **data link**: e.g. Ethernet over optic-fibre cables, Wi-Fi/Bluetooth over radio;
   intuition: the postal service hands over the envelope to an intermediate postal service, which hands it over to another intermediate postal service, and so on, until the envelope is delivered by the recipient's own postal service.

To make sure the letter in the envelope (example: an HTTP request sent by a browser, an email message, a DNS lookup) is delivered at the intended recipient, and that postal employees cannot read or change the letter or the envelope, measures can be taken at various layers. For instance, at the top layer, the sender and recipient may agree on a certain method and/or code for secret writing, so that the letter is only legible by them, unless an attacker has compromised the method or code. This is 'end-to-end encryption'.

In addition to that, the envelope can be sealed and tamper-evident. This can be done through SSL/TLS, as seen in e.g. HTTPS[15] and SMTPS.[16] Also, the postal vehicle can be armoured and protected against unauthorized road diversions: IP layer encryption may be used, DNSSEC (to protect attackers from tricking the phonebook into giving users a wrong number), and at the IP resource level through, i.a. Resource Public Key Infrastructure (RPKI).

The use of SSL/TLS, best known in relation to HTTP, where its use is referred to as HTTPS or informally 'the padlock in the browser', can provide confidentiality and integrity for communication. It provides confidentiality of communication against snooping by whoever is able to access a communication link between two communicating devices. For instance between a smartphone that runs a web browser and the web server that it connects to, or between servers in two data centres. It provides integrity through the use of cryptographic signatures over the contents of the communication: the sender cryptographically signs the communication content, and the receiver verifies this signature. If the verification fails, the data may have been tampered with, and the receiving system will reject the data. Both integrity and confidentiality are provided through cryptography and public

15 HTTP + SSL/TLS = HTTPS.
16 SMTP + SSL/TLS = SMTPS.

key infrastructure (PKI). For critical perspective on the latter topics, readers are referred to Asghari (2012) and Durumeric (2013).

SSL/TLS can be said to provide privacy in that the confidentiality it brings protects users against behavioural profiling by ISPs. A well-known example of (planned) snooping by ISPs is found in the UK around 2008: three ISPs considered deploying the Phorm Webwise system,[17] which would allow the ISPs to monetize on subscriber's web traffic through targeted advertising based on profiling built using keyword searches in individual users' web traffic. The plans led to public outcry, and were subsequently withdrawn. The Webwise system involved technology that is referred to as Deep Packet Inspection (DPI). HTTPS can help protect against such techniques.

5.1.2.4 Protecting data in use

Protection of data in use is a relatively state-of-the-art topic, and involves the use of novel cryptography to perform operations on encrypted data. That means that data never has to exist in unencrypted form on the system that performs calculations on it. Specifically, this involves 'fully homomorphic encryption'.[18]

At all times, the fundamental underlying question is: where is the data, and how does it need to be protected? One way to examine this properly is through the use of threat modelling:[19] an informal but structured approach to model threats and defences to data flows at any level of abstraction. This method can be applied to discover threats and decide on defences for data flows across the world, inside a single organization, inside a single device, or inside single application. The latter, for instance, is relevant when data is processed by a mobile app, and the mobile app must be robust against other apps running on the same mobile device.

Recognizing the three basic states of data is important to understand what data protection entails from a technical perspective. Data may be protected while transferred over the network using SSL/TLS, but be stored unencrypted on servers. Proper protection takes into account the entire lifecycle of data, from the moment it enters the system (from a sensor or from user input) until it is definitively removed.

Now that an intuition of both privacy and informatics is provided, the next section constructs two perspectives on the relation between both.

17 Clayton 2008.
18 Gentry 2009; Dulek 2017.
19 Shostack 2014.

5.2 Meaning and function of privacy

In simplified terms, the relation between privacy and ICT can be understood from two perspectives:
- ICT poses privacy challenges;
- privacy poses ICT challenges.

The first perspective gives examples of how the adoption of Internet technology – and vulnerabilities that come with it – gives rise to security needs at businesses and governments, and how the fulfilment of those needs can affect privacy. The second perspective focuses on how the need for privacy, whether expressed in policy and laws or expressed by individuals and groups, gives rise to requirements that technologists generally were not used to take (sufficiently) into account. Both perspectives are discussed next.

5.2.1 Perspective: 'ICT as a privacy challenge'

The first perspective, 'ICT as privacy challenge', pertains to the ever-increasing scale of computation, storage, and network power and use of that power in the private and public sector exacerbates existing privacy challenges. Examples include:
- private companies performing checks on social media. Besides legitimate uses, such as identifying insurance and welfare fraud, arguably less legitimate uses exist, such as screening and retaining employees' opinions expressed on social media that are not related to their job;
- privacy companies 'taking in' social media for commercial objectives, including marketing;
- the use of big data for safety and security, crowd control, behavioural analytics and prediction;
- automated facial recognition against public security camera footage;
- Automatic Number Plate Recognition (ANPR) on highways, but also in urban areas;
- Internet of Things (IoT): an increasing number of devices at home, at work, and/or worn by users are connected to the Internet. These may be built to provide convenience and functionality, not to protect their owner's privacy.

New means of ICT can generate personal data that did not exist before, or at least was not systematically stored and used. Humans are at all times connected to a time and place, and that connection is increasingly captured

by sensors and transactions (e.g. payments that require physical presence of a phone or credit card). The automotive industry introduces odometry sensors, ultrasonic sensors, front and back cameras, and light detection and ranging ('lidar') sensors generate data, as does car navigation equipment. The data may be non-personal data when considered in isolation, but longitudinal measurements that can be associated with a car owner become personal data. The data generated by sensors might be stored on the car for maintenance or insurance purposes; and may be 'phoned home' to the car manufacturer or measured by devices placed above or around highways. Additionally, individual movements may be tracked through electronic emissions from personal devices, which often emit information that is intended or can be repurposed as (partially or uniquely) identifying information. Physical characteristics of emissions themselves, both wired and radio, can be used for fingerprinting[20] with varying degrees of accuracy, reliability, and practicality. Physical characteristics may also be used to identify[21] rogue devices, for instance to detect cloned devices or illegal transmitters.

Whereas electrical appliances are subject to a mandatory (self-)certification scheme regarding safety, health, and environmental protection (the 'CE' marking for appliances traded within the EU), no such scheme exists in general for software or hardware with regard to security or privacy requirements. This is left up to the vendors. For specific domains, such as point-of-sale systems and payment cards, rigorous compliance tests are imposed, for instance by Mastercard. Whether mandatory certification can apply to software and hardware vendors in other domains, what tests should be part of such certification, and whether such certification should be carried out by the vendors themselves (self-certification; as is the case with CE markings) or by independent certification bodies, remain open questions.

5.2.1.1 Protection against digital threats can affect privacy

New categories of technologies come with new categories of threats and vulnerabilities, and countermeasures against those can affect privacy. A logical consequence of how Internet technology is designed and the rapid growth in global coverage and adoption is the emergence of botnets and phishing attacks. To protect against new phenomena that pose a risk to national security, such as the use of Internet by terrorists, organized crime,

20 Gerdes 2012; Shi 2011.

21 Hou 2014; Wang 2016.

and hostile nations, new methods and technologies are continuously being developed. These can involve big-data systems storing data that, at least in raw form, constitutes personal data. For instance DNS requests, that by definition describe an IP address performing an 'Internet phonebook lookup' for an Internet domain name when a user accesses a website. An example of a big-data system that collects DNS request data for the purpose of protecting against certain categories of new threats is SIDN's ENTRADA system. ENTRADA[22] is an experimental system that stores DNS requests received by the two authoritative name servers for the .nl top domain. Some 15,000 DNS requests per second are observed, and if stored with full IP and Ethernet headers, some 60GB[23] of data is added per day. The processing of such data can help detect botnet activity, and website spoofing; there have been court rulings[24] in the Netherlands on scammers setting up fake webshops that mimic real webshops for well-known brands. The data processed is obviously privacy-sensitive; SIDN itself took the initiative to establish an enforceable privacy framework that addresses privacy concerns associated with this data processing. This supports public trust in SIDN as maintainer of the .nl domain.

5.2.1.2 Digital espionage

Software that can be used for digitally spying on others is commercially available to individuals, or can be crafted by tech-savvy individuals. One recent example in the US is the case of Phillip Durachinsky, an American citizen who used malware dubbed 'Fruitfly' to spy on Americans. On 10 January 2018, Reuters reported[25] that the indictment states that Durachinsky collected data from thousands of computers belonging to individuals, companies, schools, a police department, and the US Department of Energy, from 2003 through early 2017. That would constitute no less than some thirteen years of computer hacking and spying without getting caught. The sensitive nature of digital espionage software becomes clear when realizing that such software, when only available to governments (as opposed to being available for the general public,

22 Wullink 2016.

23 Jansen 2016.

24 In the 'Meiberg' case, for instance, the Public Prosecution Office demanded up to three years imprisonment for large-scale scams involving falsified webshops. In Dutch: https://www.om.nl/@101212/eisen-3-jaar-cel/ (29 November 2017). The court ruling shows the defendants received between 48 and 146 weeks imprisonment: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/Gevangenisstraffen-voor-internetoplichting.aspx (22 December 2017).

25 Reuters, 10 January 2018: 'Ohio man indicted for using "Fruitfly" malware to spy on Americans'. Available at https://www.reuters.com/article/us-usa-justice-malware/ohio-man-indicted-for-using-fruitfly-malware-to-spy-on-americans-idUSKBN1EZ2KO

whether for free or paid), is subject to export controls under the Wassenaar Arrangement; 'intrusion software' was added to the List of Dual-Use Goods and Technologies in December 2013.[26] The purpose of the Wassenaar Arrangement is to support international peace by preventing military and 'dual-use' equipment, including hardware and software, from ending up in the hands of, for instance, governments that do not subscribe to nuclear non-proliferation treaties or that are known to abuse human rights. The addition of 'intrusion software' to this list was an initiative of Dutch MEP Marietje Schaake.

5.2.1.3 Cryptography vs. cryptanalysis and 'breaking' cryptography

With regard to cryptography, it is important to note that cryptographic algorithms tend to be broken over time. The typical lifetime of many cryptographic methods in the early days of the Internet was just about ten years. Advances in mathematics and cryptanalysis, and increases in computational resources made breaking encryption feasible. For certain classes of cryptographic algorithms, quantum computing may be able to break encryption using, for instance, Shor's algorithm[27] or Bernstein et al.'s GEECM.[28] Data that is encrypted and captured *today* may thus become decryptable in the near future. In some cases, existing methods may have longer lifetimes by imposing extended key-length requirements and/or key renewal schemes. In short, 'hygiene' with regard to the use of cryptographic methods and keys, such as timely re-encrypting data at rest with new algorithms or longer keys when necessary, is an important technical and procedural challenge to privacy.

5.2.2 Perspective: 'privacy as an ICT challenge'

A second perspective on the relationship between privacy and ICT is: 'privacy poses ICT challenges'. That is, ICT can mitigate or redress privacy challenges brought forth by ICT, or provide privacy where no privacy was possible before.

A well-known aphorism in Internet law is 'code is law'[29], attributed to Lawrence Lessig. This refers to the observation that the way hardware and software are designed and programmed ('coded') form a de facto regulatory

26 Matthijs R. Koot's Notebook, 12 December 2013, '"Intrusion software" now export-controlled as "dual-use" under Wassenaar Arrangement'. Available at https://blog.cyberwar.nl/2013/12/intrusion-software-now-export-controlled-as-dual-use-under-wassenaar-arrangement/.

27 Shor 1997.

28 Bernstein 2017.

29 Lessig 1999.

framework for cyberspace. John Borking contends[30] that this development is undesirable and undemocratic. Borking suggests that 'privacy law is code' is preferable, with privacy requirements laid down in legislation as (mandatory) guidelines to be followed by those who dream up and implement ICT. This relates to 'privacy by design'.[31] As stated earlier, privacy requires security. Besides privacy by design, there is the older notion of 'security by design'. The latter does not necessarily support privacy objectives. Rather, privacy by design and security by design are paradigms that can both be practised to pursue systems that are both reasonably secure and reasonably privacy-friendly.

Furthermore, the emergence of the General Data Protection Regulation (GDPR) in the EU motivates the organization of new academic events, in addition to existing recurring events, to advance privacy in ICT; one example being the IEEE International Workshop on Privacy Engineering (IWPE) (http://iwpe.info/), which has been co-hosted at the long-standing IEEE Symposium on Security & Privacy.

In 1994, a report[32] commissioned by the European Council, informally referred to as the 'Bangemann report', already identified personal data protection as a critical factor for consumer trust in the information society:

> The Group believes that without the legal security of a Union-wide approach, lack of consumer confidence will certainly undermine the rapid development of the information society. Given the importance and sensitivity of the privacy issue, a fast decision from Member States is required on the Commission's proposed Directive setting out general principles of data protection.

In other words: user confidence in the information society may suffer if 'the privacy issue', in the sense of data protection, is not properly dealt with. Regulatory points of view are discussed in other chapters in this book, for instance the chapter by Bart van der Sloot.

5.3 Classic texts and authors

The Internet era started some three decades ago, and developments have been so rapid and diverse that work published in the early days has often

30 Borking 2010.
31 Cavoukian 2009.
32 Bangemann 1994.

been superseded by new insights. A full historiography of computers, cryptology,[33] and digital security is beyond the scope of this chapter. Some insights described in early work however still apply today, or demonstrate that privacy and security challenges discussed today have existed before. Three topics are discussed below: the Ware report (a seminal work in the history of information security), the advent of public-key cryptography (notably RSA), and the creation of Pretty Good Privacy (PGP).

5.3.1 1970: The Ware report

One seminal work in computer security is due to the US Defense Science Board's Task Force on Computer Security which in 1970 released its report 'Security Controls For Computer Systems', also known as the 'Ware report', after its writer, Willis H. Ware. Prior to the task force and its report, Ware organized the 1967 Spring Joint Computer Conference session that discussed challenges that led to the establishment of the Task Force. The report, which has been characterized[34] as 'the paper that started it all, first raising computer security as a problem', states:

> Thus, the security problem of specific computer systems must, at this point in time, be solved on a case-by-case basis, employing the best judgment of a team consisting of system programmers, technical hardware and communication specialists, and security experts.

The report was written prior to the emergence of Internet, during early conceptualizations and advancements in computing and networking that eventually led to the Internet.

Now, close to 50 years after this report, that statement still applies, as do its seven conclusions:

1. Providing satisfactory security controls in a computer system is in itself a system design problem. A combination of hardware, software, communication, physical, personnel, and administrative-procedural safeguards is required for comprehensive security. In particular, software safeguards alone are not sufficient.

33 Macrakis 2010; De Leeuw 2015; Budiansky 2016.

34 Cited from the 'Seminal Papers' page of U.C. Davis' security lab, maintained by computer security scholar Matt Bishop. Available at http://seclab.cs.ucdavis.edu/projects/history/seminal.html

2. Contemporary technology can provide a secure system acceptably resistant to external attack, accidental disclosures, internal subversion, and denial of use to legitimate users for a *closed environment* (cleared users working with classified information at physically protected consoles connected to the system by protected communication circuits).
3. Contemporary technology cannot provide a secure system in an open environment, which includes uncleared users working at physically unprotected consoles connected to the system by unprotected communications.
4. It is unwise to incorporate classified or sensitive information in a system functioning in an open environment unless a significant risk of accidental disclosure can be accepted.
5. Acceptable procedures and safeguards exist and can be implemented so that a system can function alternately in a closed environment and in an open environment.
6. Designers of secure systems are still on the steep part of the learning curve and much insight and operational experience with such systems is needed.
7. Substantial improvement (e.g., cost, performance) in security controlling systems can be expected if certain research areas can be successfully pursued.

These findings were made in the context of (government) systems processing classified or otherwise sensitive information, but it is easy to see that the findings also largely apply to contemporary computer systems; one only needs to interpret 'open environment' as 'internet-connected'. Readers interested in lessons that can be learned from the Ware report regarding security certification of technology are referred to Murdoch (2012).

5.3.2 1976, 1978: advent of public-key cryptography (RSA)

One of the challenges in cryptography is key distribution. Before the advent of public-key cryptography, parties that want to communicate securely need to share a secret key. This is referred to as 'symmetric encryption', where 'symmetric' refers to the fact that parties use a single, shared secret key. To communicate a secret key, you need to have a secure channel, or rely on out-of-band methods, such as physical exchange via couriers. This changed with the introduction of public-key cryptography, which is also referred to as 'asymmetric encryption'. In public-key cryptography, each communicating party has two keys: a public key and a private key, derived

at the same time via a mathematical algorithm. The public key can only be used to encrypt and to verify cryptographic signatures, and thus does not need to be kept secret (hence, 'public' key). The private key can only be used to decrypt and to generate cryptographic signatures, and must be kept secret by its owner. Under assumptions of certain 'hard problems' in mathematics, deriving a private key from its associated public key is intractable. To communicate securely, parties only need to exchange their public key, which can be done via open channels.

The first published work that introduces the idea of public-key crypto systems is due to Diffie and Hellman[35] in 1976, under influence of Merkle who subsequently published[36] a seminal work in 1978. In that same year, Rivest, Shamir, and Adleman introduced[37] a crypto system that has since been known as 'RSA', an acronym of the authors' last names. The RSA system builds on the assumption expressed by Euler's theorem, which dates back to the 1700s, which essentially boils down to the assumption that it is very hard to factorize large prime numbers. RSA remains in widespread use today, for instance in SSL/TLS, and in PGP, the next topic.

5.3.3 Zimmerman (1991): Pretty Good Privacy (PGP)

In 1991, the US Senate drafted an anti-crime bill[38] that included the following clause, that would essentially require providers of encrypted communication services and manufacturers of encrypted communications equipment to place backdoors in their systems to allow the government to access plain-text (i.e. unencrypted) communications:

> SEC. 2201. COOPERATION OF TELECOMMUNICATIONS PROVIDERS WITH LAW ENFORCEMENT.
> It is the sense of Congress that providers of electronic communications services and manufacturers of electronic communications service equipment shall ensure that communications systems permit the government to obtain the plain text contents of voice, data, and other communications when appropriately authorized by law.

35 Diffie 1976.
36 Merkle 1978.
37 Rivest 1978.
38 'S.266 – Comprehensive Counter-Terrorism Act of 1991', 102nd US Congress. Available at https://www.congress.gov/bill/102nd-congress/senate-bill/266/text

This led US-based software engineer Phil Zimmermann to create software he dubbed 'Pretty Good Privacy' (PGP) and make it available to the general public via an Internet-connected file exchange server in that same year. PGP was the first publicly available software that implemented a public-key cryptography system: RSA. At the time, strong cryptography was considered to be subject to US Arms Export Control Act, but the PGP software nonetheless ended up outside the US.

Current versions of PGP, notably the open-source software GnuPG, remain in use today in a variety of high-security contexts, including communication with CERTs about incidents and vulnerabilities, and communication between journalists and their sources.

5.4 Traditional debates and dominant schools

The development of ICT has mostly taken place in politics-agnostic environments, and many technologists' attitude was, and remains, one of 'technology is neutral'. This neutrality is suspect when the rationale and funding for R&D have roots in organizations with a political agenda, and cannot always be seen as politics-agnostic. ICT exists in a habitat that is not isolated from personal choices, market forces, and government decisions, all of which are to some extent political. The development of Internet standards by communities of engineers is an example; also recall Lessig's 'code is law' and Borking's 'privacy law is code'.

A brief reflection on the history of standardization of Internet protocols follows, to illustrate that the 'technology-is-neutral' point of view is, for better or worse, no longer upheld, or at least faces increases opposing voices within some Internet engineering communities. Simply put, the below shows how privacy (and security) by design, notions that are not inherently politics-agnostic, gain presence in these communities.

How computers 'talk' to each other on the Internet is largely laid down in technical Internet standards. An Internet standard starts with an idea for change or new functionality. Under the umbrella of the Internet Engineering Task Force (IETF) that idea is further developed into a 'Request for Comments' (RFC) document. These are presently published on the IETF Datatracker.[39] This process is completely open: anyone who has relevant knowledge and insights can join IETF discussions. When an idea reaches a draft status, and sometimes earlier than that, ICT vendors implement

39 IETF's Datatracker is available at https://datatracker.ietf.org/.

the idea. Possibly after minor changes or corrections, and with sufficient adoption by industry, the idea can reach maturity and is promoted to the status of 'Internet Standard'. Roughly put this is how Internet technology has developed from the 1980s into what it is today. Examples of Internet standards include the protocol used for communication between web browsers and web servers (HTTP and HTTP/2,) email (SMTP), a protocol intended to protect the confidentiality and integrity of such communications (TLS), and the 'Internet address book' that resolves domain names to IP addresses (DNS).

The predecessor of the Internet, ARPANET, and the early Internet, were networks that consisted solely of parties that had some trust relation. Because of that, Internet protocols designed during the early Internet (1980s and early 1990s) did not take security or privacy into account. Concerns about inadequate security arose when the Internet expanded further and commercialized, and it was decided in 1993 that new RFCs must contain a 'Security Considerations' paragraph. This is laid down in RFC 1543.[40] The paragraph must contain a discussion about possible threats and attacks on the protocol described in a new standard. After several years of (sometimes bad) experiences with writing such paragraphs, it was clarified in 2003 what exactly should be in that section; this is laid down in RFC 3552.[41] This section should describe which digital attacks are relevant to the protocol, which are not, and why. For relevant attacks, it must describe whether the protocol protects against them. Among other things, it is mandatory to pay attention to eavesdropping (confidentiality), to the injection, modification, or removal of data (integrity), and to denial-of-service attacks that may interfere with services that use the protocol (availability). Such a paragraph will never be perfect, but requiring protocol designers to think about security properties should lead to improvement of security on the Internet. In addition, RFCs are 'living documents', in that updates and errata can be published.

Snowden's revelations have shown that intelligence services, especially the NSA (US) and GCHQ (UK), are actively gathering intelligence on the Internet on a large scale, using a wide variety of methods and techniques. These revelations, in conjunction with cases of ethically doubtful behaviour by nongovernment entities, eventually led to a rough consensus within the IETF that 'pervasive monitoring' should be considered to be an 'attack' that

40 RFC 1543: Instructions to RFC Authors, October 1993. Available at https://tools.ietf.org/html/rfc1543

41 RFC 3552: Guidelines for Writing RFC Text on Security Considerations, July 2003. Available at https://tools.ietf.org/html/rfc3552

designers of new internet protocols should take into account. Pervasive monitoring is defined as follows:

> Pervasive Monitoring (PM) is widespread (and often covert) surveillance through intrusive gathering of protocol artefacts, including application content, or protocol metadata such as headers. Active or passive wiretaps and traffic analysis, (e.g., correlation, timing or measuring packet sizes), or subverting the cryptographic keys used to secure protocols can also be used as part of pervasive monitoring. PM is distinguished by being indiscriminate and very large scale, rather than by introducing new types of technical compromise.

Furthermore: 'The motivation for [pervasive monitoring] can range from non-targeted nation-state surveillance, to legal but privacy-unfriendly purposes by commercial enterprises, to illegal actions by criminals'.

The consensus that pervasive monitoring should be considered to be an 'attack' was laid down in 2014 in RFC 7258 by Stephen Farrell, research fellow at the school of Computer Science and Statistics at Trinity College Dublin, and Hannes Tschofenig, a senior engineer at microprocessor manufacturer ARM Limited. It has the status of 'Best Current Practice' (BCP), and promotes mitigation of pervasive monitoring in new protocols. It should be noted that the BCP does not mandate prevention of monitoring by motivated attackers, which may include law enforcement and intelligence services. Rather, the BCP states the following: '"Mitigation" is a technical term that does not imply an ability to completely prevent or thwart an attack. Protocols that mitigate PM will not prevent the attack but can significantly change the threat.'

Adherence to the BCP is expected to result in better privacy-by-default in new Internet protocols. Readers interested in matters of privacy and ethics in Internet protocol design are also referred to RFC 8280[42] and RFC 6973.[43] In short, the aphorism 'architecture is politics', attributed to Mitchell Kapor, applies to the digital realm as well. Interested readers are also referred to Milan (2017) which provides a Science and Technology Studies (STS) perspective on policy related to the Internet architecture and infrastructure.

As a final example: governments may seek to influence standardization bodies for Internet protocols to protect national security interests;

42 RFC 8280: Research into Human Rights Protocol Considerations, October 2017. Available at https://tools.ietf.org/html/rfc8280

43 RFC 6973: Privacy Considerations for Internet Protocols, July 2013. Available at https://tools.ietf.org/html/rfc6973

classified documents leaked via Edward Snowden indicate the existence of government programmes that pursue this: NSA's Bullrun programme and GCHQ's Edgehill programme. A famous example of the alleged weakening of cryptography by government actors related to the cryptographic algorithm 'Dual_EC_DRBG', which turned out to contain a vulnerability that has the characteristics of an intentional backdoor crafted by cryptologist-mathematicians. Between 2006 and 2014, the US NIST agency recommended 'Dual_EC_DRBG' for use; and it was widely in use due to RSA Security products using that algorithm by default. Interested readers are referred to https://projectbullrun.org.

5.5 New challenges and topical discussions

In addition to the challenges regarding nternet standards as laid out in the previous section, current and new challenges include:[44]

- ethics of big data and artificial intelligence;
- ubiquitous identification and surveillability;
- privacy-enhancing technologies (PETs);
- digital vulnerabilities in current and emerging technology.

These are discussed in the next subsections.

5.5.1 Ethics of big data and artificial intelligence

Big data holds the promise of filtering out human cognitive bias in data analysis, but it isstill humans who programme algorithms and Interpret their outcomes. As such, logical fallacies must still be taken into account. Skepticism toward overzealous and questionable uses of big data, while avoiding techno-panic[45] and threat inflation, remains relevant. For instance, a digital vulnerability hitting mainstream news may indicate that a vulnerability of that statute occurs infrequently; media attention exacerbates perception of risk, which on the hand can at times be qualified as spreading 'Fear, Uncertainty, and Doubt' (FUD), but on the other hand can reinforce public

44 This list is necessarily incomplete. A plethora of other privacy challenges and topics exist, notably in specific application domains, such as healthcare, personal finance, law enforcement, and intelligence. The topics discussed in this chapter were selected on the basis of having relevance beyond a single application domain.

45 Thierer 2013.

awareness of the reality of technological fallibility and promote adoption of privacy-by-design and security-by-design by makers and buyers of ICT goods and services.

One recommended resource about fallibilities in big data and artificial intelligence is a Spring 2017 course taught at the University of Washington named 'Calling Bullshit: Data Reasoning in a Digital World', created by mathematical biologist Carl T. Bergstrom and data scientist Jevin West. The course aims 'to teach you how to think critically about the data and models that constitute evidence in the social and natural sciences'. From the website:[46]

> *Bullshit* involves language, statistical figures, data graphics, and other forms of presentation intended to persuade by impressing and overwhelming a reader or listener, with a blatant disregard for truth and logical coherence;
> *Calling bullshit* is a performative utterance, a speech act in which one publicly repudiates something objectionable. The scope of targets is broader than bullshit alone. You can call bullshit on bullshit, but you can also call bullshit on lies, treachery, trickery, or injustice.

This calls for awareness of the possibility of false positives and flaws in profile-building, both of which may unjustly result in unjust harms to privacy of individuals and groups. A toy example to explain the phenomenon of false positives: suppose that the algorithms have an 99% accuracy level, and one out of 100,000 people is a true threat. With 99% accuracy, there is 1% inaccuracy, i.e. unjustly indicating a person as a threat. Hence, a false positive. For every 100,000 persons, this will yield 1000 false positives, yielding a 0.1% overall false positive rate. Safeguards may be needed to prevent and redress the impact that 'false flagging' can have on an individual.

It can be noted that the Dutch legislator already recognizes this issue in the context of the Dutch intelligence services: the Memorandum of Explanation of the new Dutch intelligence and security services law explicitly[47] forbids the services from promoting or taking measures towards a person based on outcomes of automated data analysis alone. Human decision-making must augment automated data analysis[48].

46 Available at http://callingbullshit.org/ (includes course materials).

47 'Wet op de inlichtingen- en veiligheidsdiensten 2017' (Wiv2017), Memor. of Explanation, pp. 175-176.

48 This of course begs the question how human analysts interpret the outcomes of automated data analysis.

One way forward in addressing ethical questions in big data and artificial intelligence is algorithmic transparency and accountability. In January 2017, the US Public Policy Council of the Association for Computing Machinery (USACM) released[49] a statement that included a list of principles that support algorithmic transparency[50] and accountability: awareness, access and redress, accountability, explanation, data provenance, auditability, and validation and testing. In March 2018, the same council released[51] a statement on the importance of preserving personal privacy, in the context of big data and the Internet of Things. Interested readers are also referred to a survey[52] exploring potential malicious uses of artificial intelligence, published in February 2018.

Also worth noting are initiatives for codes of ethics in informatics. For instance, a programmers' equivalent to the Hippocratic Oath was proposed[53] in early 2018 by software developer Nick Johnstone, in a joint effort with other developers:

As a programmer, I swear to fulfill these tenets:

- I will only undertake honest and moral work. I will stand firm against any requirement that exploits or harms people.
- I will respect the lessons learned by those who came before me, and will share what I learn with those to come.
- I will remember that programming is art as well as science, and that warmth, empathy and understanding may outweigh a clever algorithm or technical argument.
- I will not be ashamed to say 'I don't know', and I will ask for help when I am stuck.
- I will respect the privacy of my users, for their information is not disclosed to me that the world may know.
- I will tread most carefully in matters of life or death. I will be humble and recognize that I will make mistakes.

49 ACM US Public Policy Council (USACM), 'Statement on Algorithmic Transparency and Accountability', 12 January 2017. Available at https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

50 'Algorithmic transparency' does not entail public disclosure of source code. Although such disclosure would provide a strong safeguard, other interests may be prohibitive to such disclosure, for instance protection of intellectual and business interests.

51 ACM US Public Policy Council (USACM), 'Statement on the Importance of Preserving Personal Privacy', 1 March 2018. Available at https://www.acm.org/binaries/content/assets/public-policy/2018_usacm_statement_preservingpersonalprivacy.pdf

52 Brundage 2018.

53 See https://github.com/Widdershin/programmers-oath.

- I will remember that I do not write code for computers, but for people.
- I will consider the possible consequences of my code and actions. I will respect the difficulties of both social and technical problems.
- I will be diligent and take pride in my work.
- I will recognize that I can and will be wrong. I will keep an open mind, and listen to others carefully and with respect.

Not much is known about the effects and (in)effectiveness of such ethics codes in informatics, however. Similar proposals have been seen in the past in the realm of system administrators[54] and database administrators who, due to the nature of their job, often have highly privileged access to systems and data. Administrators can be confronted with requests related to investigations fraud of incidents.

5.5.2 Ubiquitous identifiability and surveillability

New technology and increased connectivity come with new possibilities to identify, and subsequently track, devices and users. This topic can be illustrated in terms of the OSI model.

At the Data Link layer (OSI layer 2), protocols such as Bluetooth and Wi-Fi render personal devices identifiable via the Media Access Control (MAC) address[55] associated with a network interface (a Bluetooth interface or a Wi-Fi interface). MAC addresses are a key part in the mechanism that enables communication between devices over some wired or wireless physical medium (radio, copper, fibre); which is the whole idea of protocols at this layer. Although MAC addresses are usually not *globally* unique, they are, by intent, *locally* unique within smaller scopes; and may be unique within, for instance, a single country. The risk of ubiquitous surveillability via MAC tracking is addressed through 'MAC randomization', variations of which are already implemented in recent versions of Android and iOS. Flaws[56] in design or code may thwart this protection and still allow tracking. And even if the design and code are flawless, surveillability remains: if Bluetooth beacons

54 For instance, see https://www.usenix.org/system-administrators-code-ethics.

55 In the OSI model, MAC addresses reside within the Data Link layer. Wi-Fi, Bluetooth, and Ethernet are examples of protocols that provide functions at that OSI layer and implement MAC addresses.

56 Matte 2016; Matte 2017; Martin 2017. Also see The Guardian: 'MAC randomisation: A massive failure that leaves iPhones, Android mobes open to tracking' (by Thomas Claburn), 10 March 2017. Available at https://www.theregister.co.uk/2017/03/10/mac_address_randomization/

become widespread, incentives emerge to nudge[57] users into installing an app that requires Bluetooth pairing with a beacon, or with some different Bluetooth device controlled by the same company. Mobile phones may allow apps that are granted the Bluetooth permission to communicate over Bluetooth also when the app is not in use; communication can take place without the user being aware of being tracked.

At the Network layer (OSI layer 3) all the way up to the Application layer (OSI layer 7), protocol behaviour and artefacts can be found that allow web tracking. At the Application layer, every web visit discloses some technical information to one or more websites. Not only to the website the user knowingly visits, but also to any third parties from which that website includes content, such as systems controlled by online advertising brokers. A user can be tracked[58] on the web by (combination of) their IP address (Network layer), cookies, browser/device fingerprinting,[59] and other recurring patterns in observable device or user behaviour. Websites that contain, for instance, a 'Like' button (Facebook) or 'Tweet' button (Twitter) cause web browsers to load content from third-party servers. If a user makes an online purchase and discloses their real identity, address, and other information to a web shop, that web shop knows which real identity is associated with a certain unique combination of technical information. Depending on jurisdiction and terms of service, the web shop may monetize that data, for instance by selling (access to) it to third parties, who can leverage the data to enhance behavioural targeting.

Furthermore, if a website includes code from a third-party system and the website does not include proper security instructions for browsers,[60] the user is exposed to the possibility of malicious code being loaded if the third

57 For instance by offering a service or discount only via an app that requires the user to grant Bluetooth permission and enable Bluetooth; in addition to making it less easy to fully disable Bluetooth communication, as observed in a change made between iOS 10 and iOS 11. Also see The Guardian: 'iOS 11: toggling wifi and Bluetooth in Control Centre doesn't actually turn them off' (by Samuel Gibbs), 21 September 2017. Available at https://www.theguardian.com/technology/2017/sep/21/ios-11-apple-toggling-wifi-bluetooth-control-centre-doesnt-turn-them-off. Furthermore, Apple removed the audio jack from new iPhone models, requiring users to either purchase a Lightning-to-audio adapter, or use a Bluetooth headphone. The latter may increase the number of users that have Bluetooth enabled by default.

58 EFF's Panopticlick website allows visitors to test how uniquely identifiable their browser is. It was launched in 2010, received a significant update in 2015. Available at https://panopticlick.eff.org/ (suggestion: do the test both from a normal browser and from Tor Browser and see the difference in uniqueness, expressed in bits of entropy). The methodology is explained in the About page at https://panopticlick.eff.org/about.

59 Eckersley 2010; Mowery 2012; Acar 2014.

60 For instance by using 'Content-Security-Policy' (CSP), 'Subresource Integrity' (SRI), or 'Confinement with Origin Web Labels' (COWL).

party is compromised.[61] This risk is especially relevant to web applications that allow authenticated users to access sensitive data (e.g. personal data) or functions (e.g. security management).

The risk of web tracking can, to some extent, be mitigated through Tor Browser, a web browser that provides users with some degree of privacy while browsing the web. Tor Browser is implemented such that its users, by default, 'blend into the crowd' with other users by suppressing or generalizing information it emits and that would otherwise allow observers to 'zoom in' on a certain part of the users to link a web request to its real source. In addition to the digital footprint of Tor Browser being less identifying, it hides the user's IP address by routing web traffic via the Tor network ('dark web'), where the last hop, a so-called 'exit node', submits the web request to the web server, thus acting as a proxy. The Tor network is a decentralized network consisting of nodes, often volunteer-operated, physically spread around the world (though the highest-bandwidth exit nodes tend to be located in Western countries).

Web tracking is also mentioned[62] as an issue affecting the protection of the covert identity of intelligence agents deployed abroad. Tor Browser has some use in such contexts by hiding the digital exhaust from at least local, low-resourced[63] eavesdroppers. Tor Browser itself, while 'hardened', should be expected to remain vulnerable to 0-days, i.e. vulnerabilities that are found but not disclosed to the vendor so that the vulnerabilities can be exploited. Tor Browser is based on Firefox ESR and security vulnerabilities in Firefox ESR may also apply to Tor Browser. (0-days are one of the means law enforcement and intelligence services can deploy in attempt to deanonymize users. This has for instance been done by the FBI in Operation Pacifier, which targeted users of an onion service used to exchange child sex abuse imagery).

Another example of surveillability is users' DNS lookups. Each time the user visits a website with a browser (Internet Explorer, Firefox, Chrome, etc.), or sends a message via an email application running on the user's system (Thunderbird, Outlook, etc.), the user's system emits a DNS request that looks up information about the website's domain or email recipient's domain. Due

61 The Register, 11 February 2018: 'UK ICO, USCourts.gov… Thousands of websites hijacked by hidden crypto-mining code after popular plugin pwned', The Register, 11 February 2018. Available at https://www.theregister.co.uk/2018/02/11/browsealoud_compromised_coinhive/.

62 Dujmovic 2018.

63 Tor is not designed to protect against the so-called 'global passive adversary': this type of adversary is explicitly excluded from Tor's original design. It is assumed that an attacker who is able to simultaneously intercept the first link and last link in the three-hop, thus five-link, connection that Tor builds can deanonymize Tor users. Such capabilities would likely require multinational signals intelligence efforts; that topic is beyond the scope of this chapter.

to the hierarchical structure of the DNS ecosystem and absence of encryption, this traffic is observable at various systems and networks on the Internet. This includes: 1) the user's ISP,[64] 2) the operator of authoritative name servers for the top-level domain (example: '.nl' is operated by SIDN), and 3) the operator of the authoritative name servers for the second-level domain (example: lookups for 'uva.nl' are sent to 'dns-prod1a.uva.nl', 'dns-prod2a.uva.nl', or 'dns-prod3a.uva.nl', operated by the University of Amsterdam itself). To protect end-user privacy, various methods have been proposed that provide varying protection against surveillance by eavesdropping on network links, including[65] DNSCrypt, DNSCurve, DNS-over-HTTPS (DOH), QNAME minimization,[66] and Oblivious DNS[67] (ODNS). These methods could be characterized as privacy-enhancing technologies, the topic of the next section.

Default privacy and security settings in technology standards (RFCs, ISO norms, and so on), operating systems, applications, and communication providers determine the privacy and security settings that apply to most users: most users do not know or care to change these settings. An example of possible consequences can be found in a 'heat map' data visualization published in 2018 by the company that made a fitness tracking app called Strava: the map inadvertently revealed locations of secret US military bases abroad. This situation can be attributed to Strava's default setting regarding user location data, which was by default not set to 'private'.[68] It turned out that personnel at US military bases in Syria, Afghanistan, and Antarctica used the app in its default settings. The media reports about this also resulted in a question[69] raised by a Dutch member of parliament.

64 Or if a public hotspot is used, the operator of that hotspot, as well as *its* upstream ISP.

65 DNSSEC is not listed because it does not encrypt DNS lookups. DNSSEC provides authentication and integrity, not confidentiality.

66 QNAME minimization is laid down in RFC 7816, 'DNS Query Name Minimisation to Improve Privacy' (March 2016, still a draft), available at https://datatracker.ietf.org/doc/rfc7816/. QNAME minimization does not provide encryption, but does reduce unnecessary leakage of DNS lookups that is due to DNS resolvers directly communicating full domain names (e.g. 'www.google.com') to the authoritative root servers. With QNAME minimization, systems can traverse the DNS hierarchy in a step-by-step approach where the full domain name is only communicated to the DNS server that is authoritative for that particular domain.

67 See https://odns.cs.princeton.edu/ and https://freedom-to-tinker.com/2018/04/02/a-privacy-preserving-approach-to-dns/.

68 *The Guardian*, 28 January 2018: 'Fitness tracking app Strava gives away location of secret US army bases'. Available at https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases.

69 Transcript of the 45th meeting of the House of Representatives 2017-2018 that took place on 30 January 2018. Available in Dutch at https://www.tweedekamer.nl/kamerstukken/plenaire_verslagen/detail?vj=2017-2018&nr=45&version=2.

Providers of social media have been known to change default privacy settings, so as to increase web visits via accessible user-generated content and thus generate more ad revenue. Privacy-enhancing settings are often opt-in rather than an opt-out. In 2010, Matt McKeon illustrated erosion in Facebook's default privacy settings between 2005 and 2010 in a series of pictures; fig. 5.3 depicts this series.[70] And more recently, a German court in 2018 ruled[71] against Facebook in a court case brought by Verbraucherzentrale Bundesverband, the federation of German consumer organizations, over Facebook's default privacy settings. Facebook profiles are by default indexed by search engines, and its default settings, the mobile Facebook app shares users' location data. Furthermore, prior to the revelations[72] surrounding Cambridge Analytica, the Facebook API allowed third-party apps to obtain not only information of Facebook-enabled app users, but also of the *friends* of those users. In short, default settings remain an essential topic in the discussion and assessment of privacy (and by extension, security).

Another emerging topic in surveillability is Mobile Device Management (MDM) software. MDM software is used by organizations to allow employees to use mobile devices to access confidential corporate data while providing the organization controls to cope with threats such as device loss and mobile malware. Various MDM vendors exist, and their products differ in terms of potential impact on privacy of employees. If the device is a personal device, owned by the employee, and the employee enrols in the MDM solution, the employee grants the organization a certain degree of control over their device and data on the device. Functionality available to MDM administrators can include the ability to track the physical location of devices (and hence track the person who is carrying it), enumerate mobile apps installed on a device (which may include dating apps, medical apps, and so on), access the mobile browser history, or inspect the user's live web traffic by routing web traffic through a corporate web proxy.

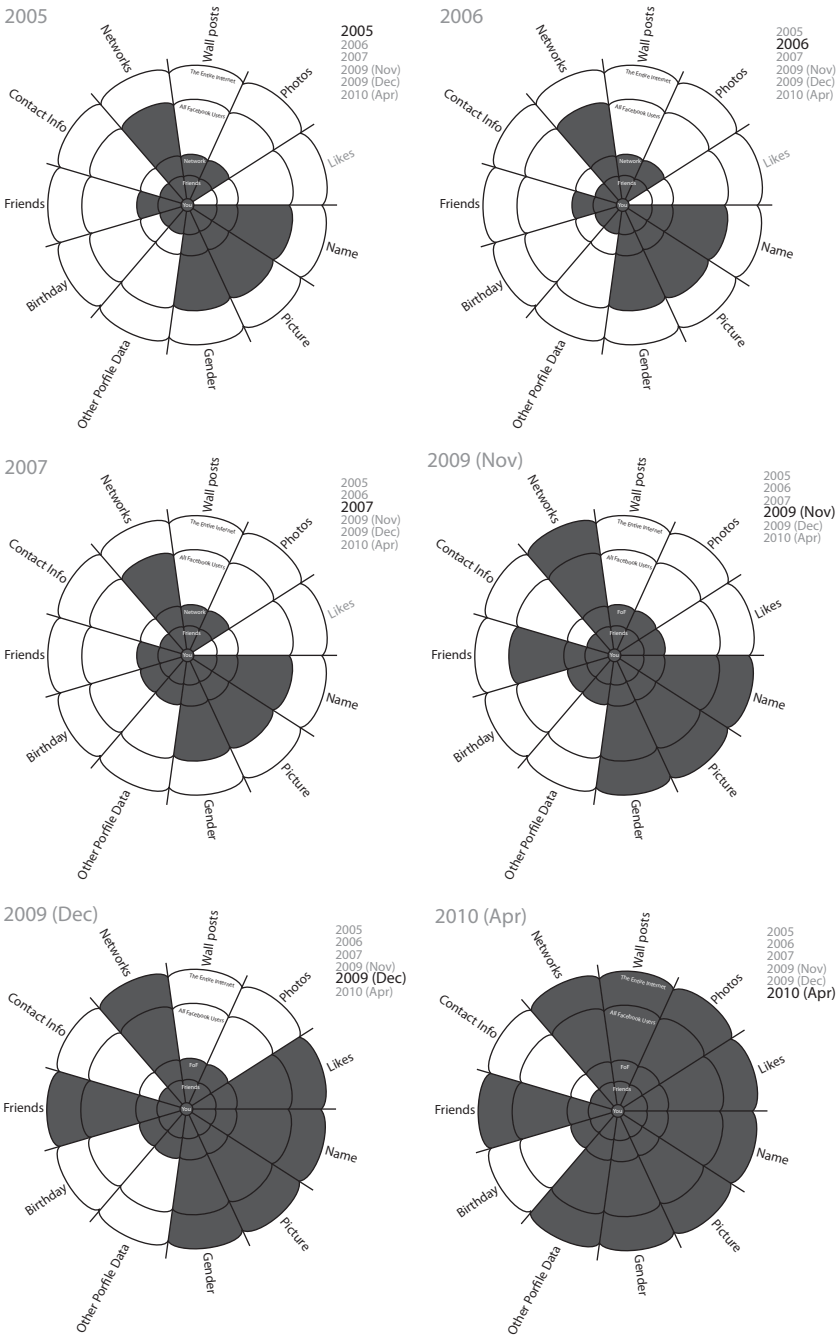
Readers interested in surveillance and privacy from an intelligence standpoint are referred to the chapter by Willemijn Aerdts and Giliam de

70 McKeon 2010. Figures used with permission of the author.

71 ZDNet, 13 February 2018: 'Facebook is breaking law in how it collects your personal data, court rules'. Available at http://www.zdnet.com/article/facebook-is-breaking-law-in-how-it-collects-your-personal-data-court-rules/.

72 *The Guardian*, 17 March 2018: 'Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach'. A copy of the old Facebook Developers API involved, taken offline by Facebook in 2015, can be found at the Internet Archive: https://web.archive.org/web/20131218130854/https://developers.facebook.com/docs/reference/login/extended-profile-properties/.

Fig. 5.2: Diminishing default privacy settings on Facebook from 2005 till 2010.



© Matt McKeon, http://mattmckeon.com/facebook-privacy/

Valk; Fidler (2015); and Petersen (2018). Readers interested in mass surveillance issues in Internet infrastructure are referred to a two-part study on mass surveillance published in 2015 by the Science and Technology Options Assessment (STOA) panel of the European Parliament:

– 'Mass Surveillance – Part 1: Risks and opportunities raised by the current generation of network services and applications',[73] 12 January 2015.
– 'Mass Surveillance – Part 2: Technology foresight, options for longer term security and privacy improvements',[74] 13 January 2015.

(In 2017, an article was subsequently published[75] in the *Computer Standards & Interfaces journal*.)

5.5.3 Privacy-enhancing technologies (PETs)

Whereas the worldwide web and most Internet protocols still used today were not designed with end-user privacy requirements in mind, as explained earlier in this chapter, Privacy-Enhancing Technologies[76] (PETs) such as Tor Browser provide privacy-enhanced alternatives. PETs can help overcome undesired effects of ubiquitous identifiability. Privacy while browsing websites is the strongest when using Tor Browser to access websites hosted as an onion service[77] within the Tor network, recognizable by the '.onion' top level domain. Also, a variety of peer-to-peer (P2P) based systems exist that provide closed-circuit networks designed to provide users anonymity with regard to various functions. One example is I2P (https://geti2p.net/), which provides a platform design from the bottom up using cryptography to achieve specific privacy and security properties. Users of I2P can host and browse .i2p sites (referred to as 'Eep sites'; not unlike the concept of onion services in Tor), send other I2P users email, and participate in anonymous instant messaging. Another example is GNUnet[78] (https://gnunet.org/). Similar functions are provided by RetroShare (http://retroshare.net/). Other platforms exist that seek to provide the user with anonymity for specific

73 Available at http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2015)527409.

74 Available at http://www.europarl.europa.eu/thinktank/nl/document.html?reference=EPRS_STU(2015)527410.

75 Schuster 2017.

76 The term 'Privacy-Enhancing Technology' (PET) was coined by John Borking.

77 'Onion service' refers to the concept that was formerly referred to as 'hidden service'. Maybe move this note to page 24, see comment 8.

78 Grothoff 2017.

functions, such as MUTE (http://mute-net.sourceforge.net/) for file sharing. Furthermore, research on mechanisms for anonymous authorization, that allow users to make use of a service without the service provider having to know the user's identity or even a pseudonym, remains relevant; as well as research on (vulnerabilities in) protocols and implementations of software that claims[79] to provide secure and private communication, such as Silent Circle, Signal, and Telegram.

Another relevant development is the emergence of self-hosted storage and communication platforms, such as the free and open source software ownCloud and Nextcloud. These platforms allow individuals and organizations to run their own 'cloud' and keep in full control over their data. Such platforms can incorporate PETs: for instance, end-to-end encryption for file sharing is scheduled to be part of Nextcloud 13. Self-hosted also means that the user or organization, rather than a provider, is responsible for security. And due to the amount of functionality and hence complexity, vulnerabilities are bound to be found in the platforms and in the underlying software; the discovery, disclosure, and timely patching of vulnerabilities will be a challenge, as is true for any complex system. It can be argued that cloud providers and self-hosted platforms protect against different threat models, have different user groups, and different usage scenarios; and hence complement rather than compete with each other.

PETs are developed within and outside academic contexts. Academic research on PETs is encouraged, for instance to improve their robustness, privacy, and security. A key dilemma is if, and how, PETs can and should be designed in a way that still caters to reasonable and legitimate interests of law enforcement agencies and intelligence and security services. Although criminals, too, make use of Tor (and other privacy-enhancing technologies and platforms), it is important to keep in mind that Tor is also widely in use to protect legitimate interests, such as whistleblower protection. The Dutch Publeaks[80] Foundation and Italian Anti-Corruption Authority[81] (ANAC) rely on onion services; and so do news media that use the SecureDrop[82]

79 Insofar the protocols and code of such systems are not openly published, a healthy level of skepticism is recommended. Academics can play an important role in discovering weaknesses, which may be caused by accidental bugs or be intentional backdoors.

80 The Publeaks Foundation is a joint initiative by various Dutch media. The Publeaks website was established in September 2013 and is available at https://publeaks.nl/.

81 'Italian Anti-Corruption Authority (ANAC) Adopts Onion Services', 13 February 2018, https://blog.torproject.org/italian-anti-corruption-authority-anac-adopts-onion-services. The ANAC website is available at http://www.anticorruzione.it/portal/public/classic/.

82 The SecureDrop project website is available at https://securedrop.org/.

software, which include The Intercept, The New York Times, The Guardian, The Washington Post, and Bloomberg News.

Research and development of PETs and of novel cryptologic building blocks for new, yet to be invented categories of PETs, is key to the future of privacy. The work on Privacy Patterns[83] is highly recommended for readers interested in privacy by design. Publications on applications of privacy by design can also be found in journals or at conferences in disciplines that are not focused on computer security and privacy. To give one example, work on privacy by design in the context of intelligent transportation systems has been published[84] in the domain-specific journal published in the *Journal of Transportation Planning and Technology*.

5.5.4 Digital vulnerabilities in current and emerging technology

Research into digital security and vulnerabilities has proven successful in improving the security of digital communication: the discovery of design flaws and bugs in implementations of older versions of SSL/TLS, novel cryptanalytic attacks against cryptographic methods supported by those older mechanisms, and so on, have led to TLSv1.2 (standardized in RFC 5246) and TLSv1.3 (still a draft at the time of writing). These newer protocols are significantly more robust in delivering security and privacy. Similarly, over the past decades, research into software vulnerabilities has led to the discovery – and subsequent patching – of a plethora of vulnerabilities in operating systems and end-user applications such as browsers. As mentioned earlier, security is a systems property, and failure of a component can mean failure of the system as a whole. Two recent examples of this are the Meltdown and Spectre[85] vulnerabilities that affect Intel processors in a way that essentially compromises the security of systems as a whole.

The digital 'threat landscape' is vast, and research into security and vulnerabilities will for the foreseeable future remain a crucial pillar in improving trustworthiness of digital systems. Readers interested in these matters are referred to the following publications by the European Network & Information Security Agency (ENISA), that lay out threat landscapes for big data, hardware, and the Internet:

83 Colesky 2015; Colesky 2016; Colesky 2018. Privacy Patterns website available at https://privacypatterns.eu/. For privacy by design, see Cavoukian (2009); Hoepman (2014).

84 Lederman 2016.

85 Kocher 2018. Also see https://meltdownattack.com/.

- 'Big Data Threat Landscape' (January 2016) https://www.enisa.europa.eu/publications/bigdata-threat-landscape
- 'Hardware Threat Landscape' (December 2016) https://www.enisa.europa.eu/publications/hardware-threat-landscape
- 'Cyber Threat Landscape' (November 2017) https://www.enisa.europa.eu/news/enisa-news/enisa-report-the-2017-cyber-threat-landscape

For security challenges related to the Internet of Things, readers are referred to the NIST Interagency Report 8200[86] and to the informational text[87] produced by the Internet Research Task Force (IRTF) Thing-to-Thing Research Group, both of which are still drafts at the time of this writing. Section 5.7 of latter document reinforces the need for testing IoT devices to discover (and patch) vulnerabilities:

> 5.7. Testing: bug hunting and vulnerabilities
>
> Given that IoT devices often have inadvertent vulnerabilities, both users and developers would want to perform extensive testing on their IoT devices, networks, and systems. Nonetheless, since the devices are resource-constrained and manufactured by multiple vendors, some of them very small, devices might be shipped with very limited testing, so that bugs can remain and can be exploited at a later stage. This leads to two main types of challenges:
>
> 1. It remains to be seen how the software testing and quality assurance mechanisms used from the desktop and mobile world will be applied to IoT devices to give end users the confidence that the purchased devices are robust.
>
> 2. It is also an open question how the combination of devices from multiple vendors might actually lead to dangerous network configurations, for example, if combination of specific devices can trigger unexpected behavior.

86 NIST Interagency Report (NISTIR) 8200 on the Status of International Cybersecurity Standardization for the Internet of Things (IoT), draft of February 2018. Available at https://csrc.nist.gov/publications/detail/nistir/8200.

87 IETF draft-irtf-t2trg-iot-seccons: 'State-of-the-Art and Challenges for IoT Security'. Available at https://datatracker.ietf.org/doc/draft-irtf-t2trg-iot-seccons/. At the time of writing, the current draft version is number ten, released in February 2018. This draft expires on 16 August 2018, after which a new draft or final release is expected.

Similar challenges exist in other current and emerging technology, such as virtualization,[88]software-defined networking,[89] speech recognition,[90] robotics, e-health technology, and so on.

Dilemmas can exist in computer vulnerability research where privacy interests collide with interests protected by law enforcement and intelligence. For instance, implemented and used correctly end-to-end encryption makes communication inaccessible to *anyone* but the sender and receiver, including to government agencies tasked with investigating crime and threats to national security. When no other viable means are available to carry out their legal tasks, these agencies resort to the exploitation of computer vulnerabilities to compromise devices (laptops, smartphones, etc.), for instance to locate and identify suspects or to eavesdrop on communication before it gets encrypted on devices ('pre-encryption'). This has led to the emergence of a market for 0-days, and knowledge about vulnerabilities is now often sold rather than publicly disclosed.[91] Many technology vendors and service providers have 'bug bounty' programmes, offering money to anyone who discovers a serious vulnerability and reports it to them in accordance with their guidelines, encouraging bug hunters to allow them to assess and patch the vulnerability before it is publicly disclosed. Bug bounties can be high, depending on the impact of a vulnerability and how difficult it is: for instance, Intel in February 2018 started a programme offering up to USD 250,000 for side-channel vulnerabilities. The 0-day market can be lucrative as well: in 2015, 0-day acquisition firm Zerodium, which was established by the French digital spyware company Vupen, offered[92] a million USD for a full iOS 9 jailbreak:

> Zerodium will pay out one million U.S. dollars ($1,000,000.00) to each individual or team who creates and submits to Zerodium an exclusive, browser-based, and untethered jailbreak for the latest Apple iOS 9 operating system and devices.

88 For instance: 'guest-to-host escapes', Rowhammer attacks, and other attacks that compromise the isolation of guests in shared virtualized environments.

89 For instance: unauthorized rerouting or mirroring of traffic.

90 Carlini 2018.

91 Allodi 2017. Also see Coriens Prins. (2014). 'Handel in geheime digitale lekken', *Nederlands Juristenblad* 89(17), 865-865.

92 See https://www.zerodium.com/ios9.html.

Zerodium reported[93] that 'only one team' received that bounty. An overview[94] is available of current bounties for 0-day exploits for various software, both desktop/server software and mobile software, ranging from USD 5,000 to USD 1,500,000 per submission. Pricing is based on the difficulty of finding exploitable vulnerabilities in a particular piece of software and market demand for a capability of exploiting that software. In a bug bounty programme seeking exploits against Tor Browser, an 0-day (or series of 0-days) that yield root/system access on Tor Browser users running Tails (based on GNU/Linux) or Windows 10 operating systems and have the Tor Browser security setting set to 'HIGH' (the default setting of Tor Browser, which blocks JavaScript) was awarded with bounties in the order of USD 200,000 to USD 250,000.

As long as systems remain vulnerable, hacking capabilities provide some redress for the challenges that strong end-to-end encryption pose to governments. Governments also seek alternative methods, for instance by imposing mandatory key escrow or pursuing 'kleptographic' methods, i.e. cryptographic methods that are designed to still allow access under certain conditions.[95] The 'crypto problem' remains an open problem to governments, policymakers and technologists. Interested readers are referred to two publications[96] released in February 2018.

5.6 Conclusion

An intuition of informatics and privacy has been provided, and it was argued that the relation between informatics and privacy can be viewed from two perspectives: ICT poses privacy challenges, and privacy poses ICT challenges. Selected topics relating to both perspectives have been discussed. From a

93 Ibid.

94 See https://www.zerodium.com/program.html.

95 For instance depending on some secret knowledge or cryptanalytic capabilities; that hopefully do not become available to criminals or hostile states.

96 1) 'The Risks of "Responsible Encryption"', February 2018. White paper by Stanford cryptologist Riana Pfefferkorn discussing risks of pursuing a requirement 'that vendors must retain the ability to decrypt for law enforcement the devices they manufacture or communications their services transmit'. Available at https://assets.documentcloud.org/documents/4374283/2018-02-05-Technical-Response-to-Rosenstein-Wray.pdf. 2) 'Decrypting the Encryption Debate – A Framework for Decision Makers', released 15 February 2018. Consensus Study Report of a study chaired by Fred Cate, with input from, among many others, noted Stanford cryptographer Dan Boneh. Available at https://www.nap.edu/catalog/25010/decrypting-the-encryption-debate-a-framework-for-decision-makers.

technical perspective, cryptography, PETs, and access controls are building blocks for privacy and data protection. Readers interested in privacy from a technological perspective are suggested to look at the resources listed below.

Finally, it can be noted that privacy-related publications exist in branches of informatics that directly deal with identified or identifiable personal data, for instance bioinformatics (e.g. processing genetic data), health informatics (e.g. processing electronic medication or health records), urban informatics (technologies for use in cities and urban environments), security informatics (e.g. identifying potential terrorists, spies, and criminals; and depending on regime, dissidents), and certain areas of robotics research. Due to length restrictions, these were not discussed here.

Further reading

Academic conferences

- PET Symposium (PETS) (https://petsymposium.org/). Recent proceedings:
  - Journal Proceedings on Privacy Enhancing Technologies (https://www.degruyter.com/view/j/popets)
- Computers, Privacy, and Data Protection (CPDP) (http://www.cpdpconferences.org/). Recent proceedings:
  - CPDP 2017: 'Data Protection and Privacy: The Age of Intelligent Machines', Ronald Leenes, Rosamunde van Brakel, Serge Gutwirth, and Paul De Hert (eds.), Oxford: Hart Publishing, 2017.
  - CPDP 2016: 'Computers, Privacy and Data Protection: Invisibilities & Infrastructures', Ronald Leenes, Rosamunde van Brakel, Serge Gutwirth, and Paul De Hert (eds.), Dordrecht: Springer, 2017.
  - CPDP 2015: 'Data Protection on the Move', Serge Gutwirth, Ronald Leenes, and Paul De Hert (eds.), Dordrecht: Springer, 2016.
  - CPDP 2014: 'Reforming European Data Protection Law', Serge Gutwirth, Ronald Leenes, and Paul De Hert (eds.), Dordrecht: Springer, 2015.
- IFIP International Information Security and Privacy Conference (IFIP SEC) (https://www.ifipsec.org/).
- Amsterdam Privacy Conference (APC), organized bi/tri-annually by the Institute of Information Law (IViR) of the University of Amsterdam. See e.g.: https://apc2018.com/
- ACM SIGSAC (https://www.sigsac.org/) conferences, including:
  - ACM Conference on Computer and Communications Security (CCS) (https://www.sigsac.org/ccs.html)
  - WiSec: ACM Conference on Security and Privacy in Wireless and Mobile Networks (https://www.sigsac.org/wisec/)
  - CODASPY: ACM Conference on Data and Application Security and Privacy (http://www.codaspy.org/)
- IEEE Symposium on Security & Privacy (S&P) (https://www.ieee-security.org/)
  - Co-hosted: IEEE International Workshop on Privacy Engineering (IWPE) (http://iwpe.info/)

- USENIX Security (https://www.usenix.org/) and co-hosted workshops, such as:
  - Workshop on Offensive Technologies (WOOT) (e.g. WOOT'18: https://www.usenix.org/conference/woot18)
  - Symposium on Usable Privacy and Security (SOUPS) (e.g. SOUPS'18: https://www.usenix.org/conference/soups2018)
- Network and Distributed System Security Symposium (NDSS) (https://www.ndss-symposium.org/)
- Events sponsored by the International Association for Cryptologic Research (IACR), a non-profit scientific organization. Including:
  - Crypto (https://www.iacr.org/meetings/crypto/)
  - Eurocrypt (https://www.iacr.org/meetings/eurocrypt/)
  - Asiacrypt (https://www.iacr.org/meetings/asiacrypt/),
  - Cryptographic Hardware and Embedded Systems (CHES) (https://ches.iacr.org/)
  - Real World Cryptography (RWC) (https://rwc.iacr.org/)
- Financial Cryptography and Data Security, organized by the International Financial Cryptography Association (IFCA) (https://ifca.ai)
- The International Symposium on Research in Attacks, Intrusions, and Defenses (RAID) (http://www.raid-symposium.org/)

Hacker conferences

Novel and high-quality work on privacy in relation to technology, both in defence (e.g. new PETs and security mechanisms) and offence (e.g. new vulnerabilities and attacks) is not only presented at academic conferences, but often also first, or even only, at hacker conferences. Large(r)-scale hacker conferences include:

- Chaos Communication Congress. See: https://ccc.de/en/
- DEF CON. See: https://www.defcon.org/
- Black Hat. See: https://www.blackhat.com/
- Hack in the Box. See: http://www.hitb.org/
- Four-yearly hacker conference organized in the Netherlands, new name for each event. Most recent event: Still Hacking Anyway (SHA) 2017. See https://sha2017.org/.

Small(er)-scale hacker conferences include, inter alia:

- INFILTRATE
- PHDays
- PH-Neutral (a speakers-only event)
- t2.fi

It is recommended to browse through conference materials (papers, slides, videos, code) of past conferences, which are usually publicly available and archived on the web.

Books

Academic works are included in the bibliography at the end of this chapter. These non-academic publications are further recommended:

- *Privacy in Technology*, J.C. Cannon (ed.), International Association of Privacy Professionals (IAPP), 2014.
- *Introduction to IT Privacy – A Handbook for Technologists*, Travis Breaux (ed.), International Association of Privacy Professionals (IAPP), 2014.

Miscellaneous resources

- *Anonymity Bibliography*, Freehaven. Selected papers and bibliography on anonymity, 1977-present. See https://www.freehaven.net/anonbib/
- *Dcypher.* Dutch platform for scientific research on information security. See https://www.dcypher.nl.

Bibliography

Acar, Gunes, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. (2014). 'The Web Never Forgets: Persistent Tracking Mechanisms in the Wild' in *Proceedings of the 2014 ACM Conference on Computer and Communications Security* (*CCS 2014*). New York: ACM.

Allodi, Luca. (2017). 'Economic Factors of Vulnerability Trade and Exploitation' in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (*CCS '17*). New York: ACM, 1483-1499. doi: 10.1145/3133956.3133960.

Altman, Irwin. (1975). *The Environment and Social Behaviour: Privacy, Personal Space, Territory, Crowding.* Monterey: Brooks/Cole Pub. Co.

Asghari, Hadi, Michel J.G. van Eeten, Axel M. Arnbak, and Nico A.N.M. van Eijk (2012). 'Security Economics in the HTTPS Value Chain', presented at *TPRC* 2012: the *R*esearch *C*onference on *C*ommunication, *I*nformation and *I*nternet *P*olicy.

Martin Bangemann et al. (1994). *Europe and the Global Information Society* ('Bangemann report'). Recommendations of the High-level Group on the Information Society to the Corfu European Council. *Bulletin of the European Union*, Supplement No. 2/94.

Bernstein, Daniel J., Nadia Heninger, Paul Lou, and Luke Valent. (year). *Post-quantum RSA*, Cryptology ePrint Archive: Report 2017/351.

Borking, John. (2010). *Privacyrecht is code. Over het gebruik van Privacy Enhancing Technologies*. Deventer: Kluwer.

Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff, Bobby Filar, Hyrum Anderson, Heather Roff, Gregory C. Allen, Jacob Steinhardt, Carrick Flynn, Seán Ó hÉigeartaigh, Simon Beard, Haydn Belfield, Sebastian Farquhar, Clare Lyle, Rebecca Crootof, Owain Evans, Michael Page, Joanna Bryson, Roman Yampolskiy, and Dario Amodei (year). 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation', arXiv:1802.07228 [cs.AI].

danah boyd. (2008). *Taken Out of Context: American Teen Sociality in Networked Publics*, PhD thesis University of Berkeley, California.

Stephen Budiansky. (2010). 'What's the Use of Cryptologic History?' in *Intelligence and National Security*, 25(6), 767-777. doi: 10.1080/02684527.2010.537875.

Randy Burkett. (2013). 'An Alternative Framework for Agent Recruitment: From MICE to RASCLS'. *Studies in Intelligence* 57(1).

Carlini, Nicholas and David Wagner(2018). 'Audio Adversarial Examples: Targeted Attacks on Speech-to-Text', arXiv:1801.01944 [cs.LG], submitted 5 January 2018.

Cavoukian, Ann (2009 [revised 2011]). *Privacy by Design: The 7 Foundational Principles*. Available at https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf.

Clayton, Richard. (2008). 'The Phorm Webwise System', technical analysis, University of Cambridge. Available at: http://www.cl.cam.ac.uk/~rnc1/080518-phorm.pdf.

Colesky, Michael, Jaap-Henk Hoepman, Christoph Boesch, Frank Kargl, Henning Kopp, Patrick Mosby, Daniel Daniel Le Métayer, Olha Drozd, José M. del Álamo, Yod-Samuel Martín, Mohit Gupta, and Nick Doty. (2015). *Privacy Patterns* (website), Contribution by EU FP7 project 'PRIPARE'. Available at https://privacypatterns.org.

Colesky, Michael, Jaap-Henk Hoepman, and Christiaan Hillen. (2016). 'A Critical Analysis of Privacy Design Strategies' in *Security and Privacy Workshops* (*SPW*), IEEE, 33-40.

Colesky, Michael, Julio C. Caiza, José M. del Álamo, Jaap-Henk Hoepman, and Yod-Samuel Martín. (2018). 'A System or Privacy Patterns for User Control' in *Proceedings of SAC 2018: Symposium on Applied Computing, Pau, France, April 9-13, 2018* (*SAC2018*).

Dingledine, Roger, Nick Mathewson, and Paul Syverson. (2004). 'Tor: the Second-generation Onion Router' in *Proceedings of the 13th conference on USENIX Security Symposium* (SSYM'04). Berkeley: USENIX,.

Dujmovic, Nicholas. (2018). 'Tech Stars on the Wall: The Human Cost of Intelligence Technology' in *International Journal of Intelligence and CounterIntelligence* 31(1), 126-138. doi: 10.1080/08850607.2017.1337447.

Dulek, Yfke, Christian Shaffner, and Florian Speelman. (2016). 'Quantum Homomorphic Encryption for Polynomial-sized Circuits' in *Advances in Cryptology – CRYPTO 2016*, Matthew Robshaw and Jonathan Katz (eds.), Lecture Notes in Computer Science 9816. Berlin/Heidelberg: Springer.

Durumeric, Zakir, James Kasten, Michael Bailey, and J. Alex Halderman. (2013). 'Analysis of the HTTPS Certificate Ecosystem' in *Proceedings of the 2013 Internet Measurement Conference* (IMC 2013). New York: ACM.

Peter Eckersley. (2010). 'How Unique Is Your Web Browser?' in M.J. Atallah and Nick J. Hopper (eds.). *Proceedings of the 10th International Conference on Privacy-enhancing Technologies*. Lecture Notes in Computer Science 6205. Berlin/Heidelberg: Springer. Lecture Notes in Computer Science 6205. doi: 10.1007/978-3-642-14527-8_1.

Fidler, David (ed.). (2015). *The Snowden Reader.* City: Indiana University Press.

Forer, Louis. (1989). *A Chilling Effect: The Mounting Threat of Libel and Invasion of Privacy Actions to the First Amendment.* City: Norton.

Gentry, Craig. (2009). *A Fully Homomorphic Encryption Scheme*, PhD thesis Stanford University. Available at https://crypto.stanford.edu/craig.

Ryan M. Gerdes, Ryan M., Mani Mina, Steve F. Russell, and Thomas E. Daniels. (2012). 'Physical-Layer Identification of Wired Ethernet Devices' in *IEEE Transactions on Information Forensics and Security* 7, 1339-1353.

Grothoff, Christian. (2017). *The GNUnet System*, PhD thesis University of Rennes 1. Available at https://grothoff.org/christian/habil.pdf.

Hellman, Martin and Whitfield Diffie. (1976). 'New Directions in Cryptography' in *IEEE Transactions on Information Theory* 22(6), 644–654. doi: 10.1109/TIT.1976.1055638.

Hoepman, Jaap-Henk. (2014). 'Privacy design strategies' in *ICT Systems Security and Privacy Protection*, 446-459.

Hou, Weikun, Xianbin Wang, Jean-Yves Chouinard, and Ahmed Refaey. (2014). 'Physical Layer Authentication for Mobile Systems with Time-Varying Carrier Frequency Offsets' in *IEEE Transactions on Communications* 62, 1658-1667.

Holvast, Jan. (1986). *Op weg naar een risicoloze maatschappij? De vrijheid van de mens in de informatiesamenleving.* Leiden: Publisher.

Horn, Gayle. (2005). 'Online Searches and Offline Challenges: the Chilling Effect, Anonymity and the New FBI Guidelines' in *New York University Annual Survey of American Law* 60, 735.

Jansen, J.R.P. and C.E.W. Hesselman. (2016). 'Ervaringen met privacybeheer voor DNS-"big data"-toepassingen' in *Privacy & Informatie*4.

Kocher, Paul, Daniel Genkin, Daniel Gruss, Werner Haas, Mike Hamburg, Moritz Lipp, Stefan Mangard, Thomas Prescher, Michael Schwarz, and Yuval Yarom. (2018). 'Spectre Attacks: Exploiting Speculative Execution', arXiv:1801.01203 [cs.CR], submitted 3 January 2018. Available at https://arxiv.org/abs/1801.01203.

Koot, Matthijs R. (2012). *Measuring and Predicting Anonymity*, PhD thesis University of Amsterdam.

Lederman, Jaimee, Brian D. Taylor, and Mark Garrett. (2016). 'A Private Matter: the Implications of Privacy Regulations for Intelligent Transportation Systems' in *Transportation Planning and Technology* 39(2). doi: 10.1080/03081060.2015.1127537.

Leeuw, Karl de. (2015). 'The Institution of Modern Cryptology in the Netherlands and in the Netherlands East Indies, 1914–1935' in *Intelligence and National Security* 30(1), 26-46. doi: 10.1080/02684527.2013.867223.

Lessig, Lawrence. (1999). *Code and Other Laws of Cyberspace*. New York: Basic Books.

Macrakis, Kristie. (2010). 'Confessing Secrets: Secret Communication and the Origins of Modern Science' in *Intelligence and National Security* 25(2), 183-197. doi: 10.1080/02684527.2010.489275.

McKeon, Matt. (year).'The Evolution of Privacy on Facebook', http://mattmckeon.com/facebook-privacy/. Graphics used with permission.

Martin, Jeremy, Travis Mayberry, Collin Donahue, Lucas Foppe, Lamont Brown, Chadwick Riggins, Erik C. Rye, and Dane Brown. (2017). 'A Study of MAC Address Randomization in Mobile Devices and When it Fails' in *Proceedings on Privacy Enhancing Technologies*, 4, 365-383. doi: 10.1515/popets-2017-0054.

Matte, Célestin, Mathieu Cunche, Franck Rousseau, and Mathy Vanhoef. (2016). 'Defeating MAC Address Randomization Through Timing Attacks' in *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks* (WiSec 2016). New York: ACM, 15-20. doi: 10.1145/2939918.2939930.

Matte, Célestin. (2017). *Wi-Fi Tracking: Fingerprinting Attacks and Counter-Measures*, PhD thesis INSA Lyon. Available at https://hal.archives-ouvertes.fr/tel-01659783.

Ralph C. Merkle, Ralph C. (1978). 'Secure Communication over an Insecure Channel' in *Communications of the ACM* 21(4), 294-299. doi: 10.1145/359460.359473.

Milan, Stefania and Niels ten Oever. (2017). 'Coding and Encoding Rights in Internet Infrastructure' in *Internet Policy Review* 6(1). doi: 10.14763/2017.1.442.

Mowery, Keaton and Hovav Shacham. (2012). 'Pixel Perfect: Fingerprinting Canvas in HTML5' in *Proceedings of W2SP 2012*, IEEE Computer Society.

Murdoch, Steven, Mike Bond, and Ross Anderson. (2012). 'How Certification Systems Fail: Lessons from the Ware Report' in *IEEE Security & Privacy* 10(6), 40–44. doi: 10.1109/MSP.2012.89.

Nissenbaum, Helen. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford: Stanford Law Books.

Pariser, Eli. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press.

Lund Petersen, Karen and Vibeke Schou Tjalve. (2018). 'Intelligence Expertise in the age of information sharing: public–private "collection" and its Challenges to Democratic Control and Accountability' in *Journal of Intelligence and National Security* 33(1) 21-35. doi: 10.1080/02684527.2017.1316956.

Pfitzmann, Andreas and Marit Hansen. (2010). 'A Terminology for Talking about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management', version 0.34, 10 August. Available at https://dud.inf.tu-dresden.de/Anon_Terminology.shtml.

Roosendaal, Arnold. (2013). *Digital Personae and Profiles in Law*.City: Wolf Legal Publishers.

Ross, Ron, Michael McEvilley, and Janet Oren. (2016). 'Systems Security Engineering – Considerations for a Multidisciplinary Approach in the Engineering of Trustworthy Secure Systems', NIST Special Publication 800-160 (updated January 2018). Available at https://csrc.nist.gov/publications/detail/sp/800-160/final.

Schuster, Stefan, Melle van den Berg, Xabier Larrucea, Ton Slewe, and Peter Ide-Kostic. (2017). 'Mass Surveillance and Technological Policy Options: Improving Security of Private Communications' in *Computer Standards & Interfaces* 50, 76-82. doi: 10.1016/j.csi.2016.09.011.

Shi, Yan and Micheal A. Jensen. (2011). 'Improved Radiometric Identification of Wireless Devices Using MIMO Transmission' in *IEEE Transactions on Information Forensics and Security* 6(4), 1346-1354. doi: 10.1109/TIFS.2011.2162949.

Shor, Peter. (1997). 'Polynomial-time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer' in *SIAM Journal of Computing* 26, 1484-1509.

Shostack, Adam. (2014). *Threat Modeling: Designing for Security*. Indianapolis: Wiley.

Thierer, Adam. (2013). 'Technopanics, Threat Inflation, and the Danger of an Information Technology Precautionary Principle' in *Minn. J.L. Sci. & Tech.* 14(1). Available at https://scholarship.law.umn.edu/mjlst/vol14/iss1/8.

Wang, Wenhao, Zhi Sun, Kui Ren, and Bocheng Zhu. (2016). 'Increasing User Capacity of Wireless Physical-Layer Identification in Internet of Things' in *Global Communications Conference* (*GLOBECOM*), 1-6. doi: 10.1109/GLOCOM.2016.7841894.

Samuel D. Warren, Samuel D. and Louis D. Brandeis. (1890).'The Right to Privacy' in *Harvard Law Review* 4(5).

Webb, Diana. (2007). *Privacy and Solitude in the Middle Ages.* London: Hambledon Continuum.

Wullink, Maarten, Giovane C.M. Moura, Moritz Müller, and Cristian Hesselman. (2016). 'ENTRADA: a High-Performance Network Traffic Data Streaming Warehouse' in *IEEE/IFIP Network Operations and Management Symposium* (*NOMS'16*). doi: 10.1109/NOMS.2016.7502925.