chapter four

New challenges for user privacy in cyberspace

Adam Wójtowicz and Wojciech Cellary

Contents

4.1 Introduction 77
4.2 Privacy and the Internet of things 79
4.3 Augmented reality applications 83
4.4 Biometric access control 85
4.5 Cloud computing 87
4.6 Big data 89
4.7 Conclusions 93
References 94

4.1 Introduction

Each new technology, always, is related to opportunities and threats. To take advantage of opportunities but eliminate or at least reduce threats, legal solutions are applied, which forbid some practices that are technically possible but socially inacceptable. A technology that has provided people with endless opportunities and deeply changed human lives is information technology (IT). It, however, is not free of threats, which are particularly hard to deal with. Among them, one of the most significant is breach of privacy. The concept of *privacy* is broad. It may concern individuals, organizations, businesses, public institutions, and states. It is also multifaceted. Among the facets are, for example, the following:

- *Statutory aspects of privacy*—e.g., the United Nations Universal Declaration of Human Rights, the Data Protection Directive of the European Commission, the US Health Insurance Portability and Accountability Act, and the US Family Educational Rights and Privacy Act
- *Technical aspects of privacy*—e.g., information gathering, information flow control, and information leakage
- *Societal aspects of privacy*—e.g., what information is private and how it is handled
- *Political aspects of privacy*—e.g., the surveillance state and population control

In this chapter, we focus on the privacy of individuals immersed in a world saturated with IT. In such a world, the problem of people's privacy has become more important than ever due to the accessibility of digital data describing not only a person's possessions, actions, and relationships with other people, but even their wishes, intentions, and emotions. The problem of privacy breach is critical since it may lead to restrictions of individual liberty and erosion of our society's foundations of trust.

The concept of user privacy has no precise definition that is commonly accepted. Certainly, user privacy is related to the concepts of "personal data" [1] and personally identifiable information (PII) [2]. Privacy concerns the right of a person to not disclose specific information about himself or herself, or more precisely, to disclose that information only to selected entities, but not to others. As such, privacy is inherently related to data confidentiality. Data confidentiality is usually related to strict secrets, e.g., a bank account password. The notion of privacy is broader; it may concern, for example, a person's medical history. A breach of confidentiality may lead to a breach of privacy. User privacy is also related to concepts of a user's anonymity, unobservability, and unlinkability. Anonymity is defined as the state of not being identifiable; unobservability is defined as a state of being undistinguishable; and unlinkability is defined as the impossibility of the correlation of two or more actions/items/pieces of information related to a user [3].

In general, privacy concerns the will of persons to control the disclosure of information about them. The awareness of threats to privacy performed by an agency or entity via intrusion or eavesdropping is nowadays high and constantly raised by many organizations collecting private data, e.g., financial institutions, telecommunication operators, and e-services providers. For example, a bank's business depends much on the trust of customers. A hack over the Internet into a bank's system may heavily impact the bank's business. Banks know that the weakest point in their security systems are naive customers, so they constantly raise the awareness of customers against hackers. Unfortunately, this is only one side of the privacy problem. The other side is the risk related to violating people's privacy by persons and organizations to which customers confidently entrust their private data. Continuing the example of banks, privacy may be breached by banks as organizations and by their employees as individuals. The problem of privacy breach by trust abuse is different from common security issues and—unfortunately—is not fairly highlighted by organizations collecting private data.

To throw light on the privacy problem, we present the points of views of individuals, businesses, and states in the following. We start with explaining the reasons why an individual's privacy should be protected.

The first reason why private data should be kept secret is to reduce the possible distress caused by the change in social relations: a person who has lost some aspect of his/her privacy can consequently be subject to judgment by other people, hardly ever favorable. The problem is amplified by the fact that it is difficult to stop the mass spread of disclosed private information.

The second reason for privacy protection is to reduce vulnerability to business-related attacks, such as (1) aggressive marketing, (2) refusing to enter certain contracts, or (3) aggravating contractual provisions. It is possible to imagine a scenario where a suffering patient calls a doctor for help and the doctor first analyzes the patient's financial situation and then sets the price of medical care based on patient's savings. In other words, if privacy is not protected, the price of a good or service paid by the customer may depend on the customer's wealth, instead of on the value of the good or service equal for every customer. The disclosure of private medical data of an important person, e.g., a chief executive officer of a company listed on the stock exchange, may influence the valuation of that company. Increased vulnerability may also be used to initiate attacks for political or social reasons.

The third reason to protect privacy is to minimize the probability of criminal attacks. Private data may be used by criminals to target potential victims and to minimize risk when planning a crime.

Finally, the last, but not least, reason to protect privacy is to minimize vulnerability to identity theft. Identity theft has serious consequences for a victim. It is very hard to prove that decisions, such as bank transfers, were made by an identity thief, instead of a true

bank client, while the credentials used were true and correct. Banks are rightly afraid of a fraud—a dishonest client withdrawing money from his/her account and then claiming that it was not done by him/her but by an identity thief.

The reason why a business is interested in violations of the privacy of its clients is, unsurprisingly, to reduce the business risk and increase profits. A privacy breach is intended to detect person's needs and vulnerability to arguments and suggestions to purchase goods or services to meet those needs. A privacy breach is also used for price discrimination, i.e., charging varying prices when there are no cost justifications for the differences [3], as illustrated earlier by the example of a doctor having access to a patient's private financial data. Business often argues that permitting access to private data will enable it to better inform the client about the possibilities of meeting his/her needs. Also, a client is not forced to take advantage of the advertised offers. Although the latter may be formally true, business hides the risk of privacy abuse, because a reason of privacy breach by businesses is also the identification of vulnerabilities aiming at weakening a client's negotiating position—making them more susceptible to arguments for adopting a worse proposal or for refusal to conclude a contract [4]. Private data and other knowledge about a client may also be used by, or sold to, untrusted and unauthorized parties.

Most nations include the protection of privacy into law. Surveillance, and hence a reduction in privacy, is legally possible only with regard to particular citizens who are formally suspected of committing crimes, and only with the consent of the court guarding civil liberties and supervising law enforcement authorities. In practice, surveillance is used by different governments in a legal, semilegal, or illegal way to prevent activities deemed undesirable, from criminal acts to civil disobedience or political opposition. However, recently, an approach of governments to privacy is undergoing change driven by the phenomenon of terrorism, in particular suicide attacks. The current legal system is based on the assumption that punishment follows the committed crime. This assumption obviously does not work for suicide terrorists, because when they commit the crime of killing inadvertent innocent people, they inflict the highest punishment on themselves—death. Facing the danger of suicide terrorist attacks, the only way for the state to assure public safety is to preventively isolate suspects. This, however, implies the need to violate suspects' privacy to find out about their plans, in advance of the criminal act. Hence, there is a change in the attitude about the state's surveillance of its citizens. In the age of terrorism, the state tries to collect all possible data about all citizens—in other words, to keep the whole society under surveillance—and to analyze collected data when a suspect appears. With such approach, a person is treated as the sum of his/her social relationships, electronic interactions, and favorite content. A citizen becomes suspicious not because he/she has committed an illegal act, but just because his/her online activity patterns indicate that he/she is more prone to commit a crime than an average citizen [5].

The remainder of this chapter is organized as a set of sections devoted to privacy risks that are specific to various emerging IT and electronic business trends. Section 4.2 concerns privacy issues specific to the Internet of things; Section 4.3, to augmented reality (AR); Section 4.4, to biometrics; Section 4.5, to cloud computing; and Section 4.6, to big data. Section 4.7 summarizes these considerations by adopting a holistic view that considers the mutual dependencies of various technologies and trends.

4.2 Privacy and the Internet of things

The Internet of things (IoT) is defined as "a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on

existing and evolving interoperable information and communication technologies" [6]. As the IoT evolves from the early experiment phase to become a ubiquitous infrastructure processing sensitive data, the number of various privacy challenges to be met increases. One of the main obstacles, seen from the business point of view, is related to the fact that in the rush to provide the market with new IoT solutions before the competition does, the privacy issues are perceived only as a roadblock to productivity and cost-effectiveness. Many IoT investments initially focus on functional requirements only for solutions that can be rapidly marketed and are expected to produce the desired return on investment [7].

From the technical perspective, every sensor or device that collects, sends, stores, and/or processes sensitive data provides a potential privacy risk. As many as 70% of the most commonly used IoT devices contain security vulnerabilities [8]. Furthermore, the diversity of IoT devices makes achieving privacy protection challenging. This is related to the problems of the growing number of unstandardized devices and technological/business fragmentation [9]. Standardization allows developers to build on fewer software/hardware platforms and have more resources allocated for security protection. In turn, in standardized environments, privacy breaches affect a bigger number of devices and users [7]. Above all, the scale of the IoT networks alone is challenging: 11.2 billion connected devices will be in use in 2018, 20.4 billion in 2020 [10], usually containing several sensors and complex software-based logic.

IoT devices are often physically accessible to intruders. Since IoT interconnects physical "things," not only can the intruders perform usual digital privacy-targeted attacks, e.g., stealing data, but they can also take advantage of tampering with devices or attack networks [11] (e.g., healthcare devices, electrical grids, or traffic signals). For example, if a smart thermostat is not able to protect data from eavesdropping and unauthorized usage, specifically when transmitting energy usage data to the utility operator for dynamic billing or real-time power grid optimization, then the sensitive data leak could contain the information that the power usage level has decreased, which indicates that a person's home is left empty [12], which may provoke burglary. The network connections that the devices use may also give subsequent access to central applications and databases.

Another IoT privacy threat comes from the lack of adopting a privacy-focused approach to build systems. A strong focus on security from the beginning of the project is often missing, especially when dealing with emerging technologies and underdeveloped markets [7]. Trade-offs, e.g., a choice of solid security at the cost of compromising user experience, are very challenging. If a company plans to develop its own IoT infrastructure, or deploy an existing solution, it must do research and stay as informed as possible while putting much effort to the training for their personnel. For instance, software designers of IoT solutions, specifically "smart home" systems, who build connections between various devices, face new security engineering challenges specific to the new domain they are often not familiar with. In a recent work [13], four IoT smart home devices (a Sense sleep monitor, a Nest Cam Indoor security camera, a WeMo switch, and an Amazon Echo) have been analyzed. The results of this research prove that the network traffic rates of the devices can reveal a user's physical behavior even if the traffic is encrypted. Preserving user privacy would require special network traffic obfuscation to hide variations that reflect real-world interactions, which is not specified as a requirement by designers of such systems.

At the same time, users of connected devices practically do not realize that their security is in play, or at least, at risk. For an average user, a smart TV or smart watch is still just a TV set or watch. Users are not aware that it is a fully equipped network node, which can be used to collect data describing its owner and his/her environment, or that smart wrist wearables (particularly smartwatches and fitness trackers with embedded sensors such as

accelerometer, gyroscope, or magnetometer that are paired with a networked smartphone) might be exploited to steal user's automated teller machine password [14]. Consequently, the user behavior creates another group of privacy risks.

Solid IT security controls that have been developed over the past three decades should be adapted to the specific constraints of the embedded devices popular in the IoT. Applying existing security practices to these devices requires significant reengineering to address device constraints [7]. This is caused by the fact that they are designed for low power consumption, typically have only as much processing power and storage as needed for their purpose, and often have limited connectivity. More powerful processors needed for encryption, and other data security functionalities can be used in some smart products, but it is impractical for disposable devices with no displays and with limited power consumption. More powerful processors have bigger size, and they need additional appliances and space for heat dissipation. Moreover, they need more power, which requires bigger, heavier, and more expensive batteries. In order to reduce weight and size, higher costs of research and materials are required as well as longer time-to-market [7]. With higher prices and more complex builds, such devices could not be considered disposable. Creating access control methods that can be implemented in cheap and compact IoT devices without compromising the user experience, or without adding additional hardware, represents an engineering trade-off challenge. Outsourcing computationally intensive encryption to the cloud is not a privacy-preserving solution either (cf. Section 4.5).

The other IoT privacy-related problem is a result of the fact that in M2M usage scenarios (e.g., telemetry or traffic control), there is no human operating the IoT devices who can input authentication credentials or decide whether an application should be trusted or not. The devices must make their own decisions about whether to disclose their data or trust in some process or other device. In turn, in IoT usage scenarios with a potential presence of human operator (e.g., telemedicine or wearables), connected devices have little or no interfaces that clearly present choices and explain their privacy-sensitive consequences; or even if the choices can be effectively presented in the initial setup of the devices, they can shortly become too hard to understand and remember for average user, because of the highly dynamic nature of IoT networks. Often, even if it is technically possible, service providers do not provide users with clear messages and choices for unexpected collection or uses of their data and a choice to opt out of data collection is not given [15]. For instance, not only do users not know that a smart meter is collecting data about their air-conditioning habits and that a smart watch is collecting data about their physical habits, but also they do not know to what extent this information is shared with data brokers or marketing companies.

The next privacy issue is an effect of ubiquitous data collection, which is possible with IoT devices. Service providers that collect PII do not follow the principle of data minimization. This principle states that only the data needed for a specific purpose should be collected and then safely disposed [15]. Data that have not been collected or have already been deleted cannot be used for unintended purposes. Conversely, of course, collecting and storing large amounts of data increases potential privacy risks that could result from a data leak. Unfortunately, there is an increasingly popular trend that promotes unlimited collection and storing of data because of the high value expected from its potential, but yet unknown future uses—cf. Section 4.6. This leads to putting user privacy at real risk on the off chance an organization might discover a valuable use for the private data at some future point in time.

The aforementioned business hopes related to future data uses decrease the willingness of data operators to deidentify consumer data where possible. Many IoT data uses

could still be accomplished by using deidentified data [15]. However, even once anonymized, data can still be reidentified [3]. Therefore, a technical means for data anonymization should be coupled with administrative controls. Organizations should legally commit that they will not try to reidentify personal data. They should also require this commitment from those with whom they share data [15].

In the world of regular mobile devices, there are millions of unpatched and insecure devices in use. Even reputable vendors of costly devices, such us Apple and Google [16], do not update their software on devices that are only a few years old. In the case of inexpensive disposable IoT devices that can operate on the network for years, the lack of updates to past generations can be an even more significant issue, in terms of its scale as well as its consequences [7]. According to a recent report [17], 26% of IoT professionals, including developers, vendors, and enterprise users, find long-term support as the "biggest immediate challenge faced by IoT professionals." Further magnification of this challenge can be expected, when the IoT market forces the rapid release of new products based on emerging technologies and when the well-known phenomenon of "planned obsolescence" becomes common in this market.

The next challenge is related to the technical difficulty to apply a patch if a vulnerability is known or even just to provide users with a message about a new fix. The difficulty lies in receiving software updates or security patches in a timely manner without consuming the bandwidth, impairing functional security or causing significant recertification costs every time a patch is published [12]. Service providers need to publish patches, and devices need to authenticate them, in a seamless and secure way. This is the problem of thousands of devices processing sensitive data that are dependent on security patches to protect against attacks on the confidentiality of these data. Secure IoT devices must either be secure "by design" and protected from the beginning of operation or be able to receive updates throughout their life cycle. Neither option is realistic [7].

In the IoT, as in conventional networks, devices need a firewall or packet analysis to control traffic incoming to and outgoing from the device, to protect data confidentiality. A host-based firewall or intrusion prevention system is required even if network-based devices are installed. The problem results from the fact that often embedded devices use specific protocols, distinct from standard Internet protocols [12], and at the same time, there is lack of industry-specific protocol filtering tools, which could identify attack schemes or malicious payloads in nonstandard protocols at IoT devices. IoT devices do need to filter the incoming data in a way that makes optimal use of its limited computational resources.

As the described privacy risks result from the inherent specificity of IoT, no single solution can be ultimate for every deployment either currently or in the future. However, depending on the particular system, the protection strategies can be applied by combining solutions from different categories, to mitigate those risks. The solution categories include technical, organizational, and legal measures. Technical solutions include system design methods that minimize the attack surface and enforce data anonymization, such as the one presented by Wójtowicz and Wilusz [18], new access control models (e.g., supporting context awareness and addressing embedded device constrains), new encryption schemes, new methods for IoT patching, and reengineering network threat detection/prevention systems to be suitable for the IoT. Organizational solutions include IoT project management focused on user privacy from the requirement specification phase to the long-term support provisioning and end user training as well as continuous training for developers facing new threats. In turn, legal or administrative solutions include introducing obligatory information and opt-out options for users regarding the collection or usage of their data, forbidding violations of principle of data minimization, and data reidentification.

The privacy threats described in this section are common for many different applications of IoT. Moreover, each IoT application has its specific threats not covered here. However, one group of them, i.e., mobile AR systems, has particularly distinct characteristics and high impact potential for the personal data flow. Therefore, the next section is exclusively devoted to privacy threats resulting from the usage of AR systems.

4.3 Augmented reality applications

An AR system is defined as one that "combines real and virtual objects in a real environment; runs interactively, and in real time; and registers (aligns) real and virtual objects with each other" [19]. Modern AR systems operating on mobile devices require access to several data streams coming from a number of sensors. They include not only camera-captured images and audio streams, but also geolocalization data, accelerometer/gyroscope data, temperature, or data from peripheral devices. A significant risk related to AR applications results from the difficulty to set trade-off boundaries for the required levels of application access to those streams and the probability of data confidentiality violation, their unintended usage, and, therefore, violation of the users' privacy. Limited privacy controls in the AR domain are a good example of the classic usability vs. security dilemma. This risk can be illustrated with the following cases:

- AR application uploading the user video stream or geolocalization data to its server-side software components
- Shape detector reading credit card numbers or text on electronic displays or on a bottle of medicine that reveals a medical condition or identifying a person
- Object or gesture tracker recording user's activity; even anonymized skeleton stream allows the inference of potentially sensitive gestures, movements, proximity of faces, bodies, etc.
- Collecting geometrical three-dimensional (3D) data to create models of users' indoor spaces
- Face recognizer intended for device user identification, gathering data about other persons in the camera's field of view
- Quick response (QR) code scanner, apart from code scanning, records data about its environment

These examples imply that privacy risks are much higher in AR than in conventional systems because of the continuous mode in which AR systems operate. Complex AR applications require an always-recording feature, e.g., an AR application that automatically recognizes and decodes QR codes requires continuous access to a video stream. The always-on sensing of AR applications and wearables can disclose sensitive data such as personal images, health information, or enterprise intellectual property. This privacy risk is called data aggregation [20] and is related mostly to temporal and spatial accumulations of raw visual data. Apart from privacy issues enabled by data aggregation followed by applying reasoning and data mining techniques, the aggregation alone inherently introduces privacy breaches, since the human consciousness of the presence of always-on recording devices can alter one's "attitude, behavior, and physiological state" [21]. Also, the accumulation of spatial data in AR services raises the risks related to location disclosure (identity privacy, user's position privacy, and user's movement path privacy) in the context of user anonymity, unlinkability of the user's actions, and the strongest requirement of complete unobservability of user actions [3].

Today, AR applications perform data collection, rendering, and user input interpretation, aided by third-party software libraries or cloud-based recognition services. These applications provide some level of functional access control, but users do not have fine-grained control [22] over the confidentiality of particular pieces of the data against third-party applications. The main reason for that is related to the fact that today's operating systems (desktop and mobile) are built without AR applications in mind. Only coarse-grained controlling of access to data streams is offered, instead of AR-specific privileges [23]. For example, an application should only be provided with an access limited to specific objects that are recognized by the operating system with skeleton tracker, without access to the whole video stream.

Therefore, it is difficult to build an AR application that follows the "principle of least privileges," i.e., to ensure that every application and user is able to access only the data and resources that are necessary for their legitimate purpose. Policy-based mechanisms applied by the software distribution services tend to be ineffective against applications that collect users' PIIs at the back end [20]. It is not probable that these threats can be fully mitigated, except for specific classes of applications, e.g., requiring only numerical data aggregated from multimedia streams. Only in such cases could privacy-enhancing techniques that have been developed for years be utilized, such as differential privacy [24].

AR systems employ new input techniques such as voice, gaze-tracking technologies, or glove-based haptic sensors. The use of these input methods while running multiple applications simultaneously produces new privacy threats related to the inaccurate identification of the application that is seeking the input and should receive it [25]. Malicious applications can steal user input intended for another application, e.g., they could attempt to register a verbal command that sounds similar with that of another privacy-sensitive application. This threat is even more significant since multiple AR applications expose their application programming interfaces (APIs) to each other and users share multimedia content between these applications. Cross-application data sharing can also be implicit, e.g., in AR systems that automatically use video streams of nearby users to build a 3D model of the given user at the runtime [25].

An individuals' personal privacy (as well as information-gathering rights or device ownership rights) can be lessened by AR services that selectively disable sensing capabilities according to server-side rule-based logic, e.g., prohibiting the ability of AR devices to record during a music concert [26]. On the other hand, AR applications can provide users with correct information that cannot be legally used to make business decisions. This could be a case of an AR application collecting face images and performing face recognition during a job interview followed by mining in candidate social media profiles in jurisdictions where discrimination based on marital status, arrest history, etc. is illegal [27]. By providing a number of informational elements about a real person in real time, AR applications increase the risk of conscious or even unintended discrimination in various aspects of life. Data aggregation also creates privacy risks for bystanders, who are not able to opt out or be anonymized in AR streams. AR services are not able to consider anonymization requests from other users' devices or the environment.

Data processed with AR sensors can be used for human body detection and subsequently for user identification and authentication with biometric means. Biometrics, although potentially useful due to its convenience and usually robust security properties, brings additional privacy-related concerns, which are described in the following section.

4.4 Biometric access control

The use of biometric methods for user identification or authentication introduces various privacy concerns. In this section, they have been classified into six main groups of risks. The first group of risks is related to the fact that biometric attributes encode the biological properties of parts of the human body (physiological biometrics) or some human behavior (behavioral biometrics). It is relatively easy for access control systems designed for acquiring biometric samples and processing encoded templates to perform the additional analysis of templates or sample data and infer information describing users based on these data [28]. The inferred information can be deterministic or stochastic. Not only may the information refer to the body, or the medical condition of the user, but it can even be used to estimate cultural or social characteristics of the user. Examples of biometric attributes and their impact on privacy are listed in the following:

- Voice sequences—language spoken (nationality), accent (cultural/social characteristics), age, gender, and emotional state
- Face images or 3D head models—medical condition, age, gender, race, estimated cultural/social characteristics, and emotional state
- Fingerprints—medical condition (e.g., malformed fingers can be correlated with genetic disorders [28])
- Iris—medical condition
- Vein patterns and electrocardiogram patterns—medical condition
- Behavioral biometrics such as gait—medical condition
- Behavioral biometrics such as style of typing or style of touchscreen usage can reveal, directly or indirectly, privacy-sensitive input.

The second biometric-related risk of privacy breach comes from the side of service providers with regard to the unambiguity of biometric identifiers. The presence of biometric identifiers (even if they are not originally referring to any PII in the system) makes it possible to bind a person's virtual identity (anonymous or pseudonymous) used in cyberspace with his/her real-world identity, or to bind several persons' virtual identities with each other [29]. Moreover, biometric identifiers could be used not only to bind identities themselves, but also to bind data (and metadata) describing actions of a particular user biometrically authenticated in various distributed services if service providers collude. In emerging ubiquitous services which are naturally decentralized and untrusted on the one hand, and require new seamless and convenient access control methods (such as biometrics) on the other, this threat is of special significance. The derived information could become a basis for discrimination against a person if their characteristics are considered unwanted [28].

The third group of privacy risks is also related to the anonymity and pseudonymity of users of biometric systems. It follows from the fact that not all biometric attributes are as difficult to collect without one's knowledge or permission as vein patterns or electrocardiogram patterns. Generally, biometric systems cannot rely on the secrecy of biometric samples [30]. Samples such as face images, fingerprints (left frequently, e.g., on a glass), or various behavioral biometrics are relatively easy to collect without a person's knowledge. Subsequently, they can be used for two groups of purposes. The first purpose is to instantly infer additional information (e.g., emotional state from face images or voice samples) and take advantage of them (e.g., in dynamic marketing applications). The second group of purposes is related to identity theft (cf. Section 4.1): collecting one's biometric

samples can be followed by preparing fake authenticators imitating corresponding parts of the human body (artificial finger, face mask, high-resolution iris image, etc.), in order to conduct unauthorized authentication and ultimately to violate the confidentiality of user's private data within the system.

To protect the biometric access control against attacks by authentication imitations, some researchers and systems engineers propose so-called liveness detection functionality [31] (e.g., based on the presence of the pulse and eye blinking detection). Although liveness detection can indeed reduce the likelihood of success of the attacks, it introduces new privacy-related risk, since it increases the amount of sensitive data that are collected in a continuous manner. Similarly, multimodal biometric systems combine multiple biometrical recognizers. They have been developed to reduce the false acceptance rate (FAR) and false reject rate (FRR) and to collect more streams of sensitive data. Multimodal biometric systems therefore increase the possibility of data cross analysis accompanied by the associated increased risks of privacy breach. They constitute the fourth group of privacy-related risks that also includes privacy risks following from the unexpected (from the users' point of view) cross analysis of voluntary biometric databases created for user verification purposes with mandatory screening databases [32].

The fifth, similar but distinct, privacy concern is related to the fact that as opposed to conventional authenticators such as passwords, once the biometric sample or template is eavesdropped or disclosed by an attacker, the countermeasures are not straightforward. Compromised password, digital certificate, or credit card data can be effectively revoked and reissued. In the case of a biometric pattern reflecting an immutable attribute of a person's body, the act of eavesdropping on the pattern has permanent consequences. Revocation or cancellation is possible only in specific cases with a priori use of special techniques of cancellable biometrics and/or biometric cryptosystems. However, these techniques cannot assure both provable security and practical FAR/FRR at the same time, and they introduce new issues [33]. Thus, a person's sensitive biometric templates are at constant risk while employed for practical access control. Despite obvious advantages, the fact that biometric patterns are immutable over time can also introduce privacy-related risks beyond just compromising the system. Potentially, there are many circumstances in which a user might want to change his/her identifier, but its biological uniqueness persists even though the sample as well the template are recoded to different digital representations.

The sixth risk arises from the practical limitation of a great majority of biometric access control systems that assumes the existence of nonzero FAR. Biometric systems usually allow their managers to adjust the sensitivity level and find an optimal trade-off between FAR, FRR, and other recognition parameters for a given application. However, the adjustment rarely allows these rates to be reduced to zero, especially in large-scale systems [30]. Disclosing private data because of false acceptances and allowing authentications followed by unauthorized penetration of the system (intentional or accidental) are inherent risks that cannot be omitted in the consideration of privacy.

Finally, a design solution that allows the reduction of some of the aforementioned privacy risks is a shift toward "distributed architecture." In this solution, biometric templates are stored in an encrypted form within devices (e.g., smartcard or smartphone) over which a user has full control [34]. Each device has a biometric sensor built in. User identification, authentication, or transaction authorization is performed locally by comparing the acquired sample with the stored template (according to a more robust verification instead of identification scheme). In some applications, such an approach is possible to implement and effective from the privacy perspective. However, unfortunately, the current dominant trend is just the opposite—to store and process as much data as possible in the cloud-based,

centralized manner that is potentially privacy destroying. The cloud computing problem seen from the user privacy perspective is the subject of the following section.

4.5 Cloud computing

Cloud-based data processing requiring privacy assurance can be successfully deployed in private clouds [35]. However, it is the public cloud model that is the most popular architecture when cost reduction is concerned. Relying on a public cloud service provider to store and process user data raises serious privacy concerns since the user is forced to cede control to the cloud provider on many issues affecting data privacy.

The first group of privacy risks follows the cloud operator's difficulty in providing privacy controls required to protect the users' data. These risks result from technical, organizational, and legal limitations of the public cloud model. The loss of control over the physical as well as logical aspects of the system and data reduces the user's ability to keep actual knowledge about the processes and to make accurate and aware decisions regarding the privacy protection of his/her data or of the data of his/her organization. Also, verifying the functional requirements of the service and the effectiveness of privacy controls is not feasible to the same extent as with an internal organizational system [36]. The knowledge of a cloud provider's privacy protection measures and controls is needed if the user is to perform continuous privacy risk assessments. However, cloud providers are not eager to provide users with descriptions of their privacy measures and controls for several reasons. One of them is the fact that such descriptions are considered proprietary and could be used to develop an efficient attack scheme on cloud infrastructure. Providing detailed system-level monitoring by a cloud user is not part of most service-level agreements (SLAs), which limits the user's ability to conduct audits [36].

The result of migrating to a public cloud infrastructure is usually losing a direct point of contact with the entities responsible for data management and losing an influence over decisions made about the data environment. This makes the user dependent on the cooperation of the cloud provider to perform the responsibilities of both parties, such as the passive and active protection of data confidentiality. Also, compliance with data protection laws is an area of joint responsibility that requires cooperation and coordination with the cloud provider [36]. Consequently, there may be data security breaches of which the controller is not notified by the cloud provider and possibly unrecognized conflicts between cloud customer data security procedures and the cloud environment.

Redundant data storage in multiple physical locations is a common feature of cloud computing services. This can lead to the data proliferation phenomenon. Detailed information about the location of a user's data is unavailable or not disclosed to the user. Therefore, often, it is unclear which party is responsible for ensuring legal requirements and data handling standards for PII processing or whether it is possible to audit them for compliance with these requirements and standards. Moreover, it is not clear to what extent cloud subcontractors involved in processing can be identified and verified as trustworthy, particularly in a dynamic environment [35]. Trust is not transitive, which requires disclosing such third-party contracts in advance of reaching an agreement with the cloud provider, and maintaining the terms of these contracts throughout the agreement. In practice, it is rarely fulfilled, so privacy guarantees can become an issue with composite cloud services [36]. If cloud computing providers outsource certain tasks to third parties, the level of privacy protection of the cloud provider depends on the level of privacy protection of each "supply chain" link and the level of dependency of the cloud provider on the third party.

Any corruption in this chain or a lack of coordination of responsibilities between any parties involved can lead to loss of data privacy [37].

Moreover, business events such as an acquisition of the cloud provider could increase the probability of business strategy modification and introduce data privacy risks [37]. In turn, in the event of the confiscation of physical hardware because of a subpoena by law enforcement agencies or civil suits, the centralization of storage as well as shared tenancy of physical hardware results in a higher number of users at risk of the disclosure of their data to third parties [37]. If data centers are located in high-risk countries, e.g., lacking the rule of law and having an unpredictable legal framework and enforcement and states that do not respect international agreements, sites could be raided by local authorities, and private data could be subject to enforced disclosure [37]. Thus, a cloud computing service, which combines outsourcing and offshoring may raise very complex issues; hence, it can be difficult to ascertain privacy compliance requirements [35]. Moreover, a sealed search warrant served at the cloud provider may allow law enforcement to search the tenant's systems while forbidding the cloud provider from notifying the tenant that a search took place [38].

The threat of a "malicious insider" is considered especially important in the case of the cloud computing model, since cloud architectures employ user roles, which are particularly high risk. As cloud services use increases, employees of cloud providers increasingly become targets for criminal groups [37]. Insider threats also include business partners, contractors, and other parties that have any access to a cloud provider's systems. Incidents may involve various types of fraud, sabotage of information resources, and theft of sensitive information. Incidents may also be caused unintentionally. From a user's point of view, moving data and applications to the cloud computing environment operated by a cloud provider expands the circle of insiders not only to the cloud provider's staff and subcontractors, but also potentially to other customers using the service, thereby increasing risk [36].

Multitenancy and shared resources are two of the main attributes of the cloud computing model. Since computing power, storage capacity, and network resources are shared between multiple users, an attacker can exploit vulnerabilities from within the cloud environment, overcome the separation mechanisms, and gain unauthorized access to private data. This class of risks includes the failure of mechanisms separating storage, memory, routing, and reputation between tenants of the shared resources. An attacker can compromise the service engine by hacking it from inside a virtual machine, the runtime environment, the application pool, or through its APIs. The probability of this incident scenario depends on the cloud model considered; it is likely to be low for private clouds and higher in the case of public clouds [37].

An infrastructure of public cloud computing is complex compared with that of a conventional data center. Many software components (for both general computing purposes and management purposes) comprise a public cloud service, which results in a large attack surface. Components evolve in time as new features are deployed and existing ones are upgraded. Data security depends not only on the correctness and effectiveness of particular components, but also on interactions among them. Challenges exist in understanding and securing APIs that are often proprietary to a cloud provider. The complexity also results from the fact that the number of possible interactions between components increases proportionally to the square of the number of software components. The increasing complexity is followed by an increasing number and probability of privacy risks related to the loss or unauthorized access, deletion, use, modification, or disclosure of sensitive data [36].

The privacy risk related to ineffective data deletion can occur in several ways, e.g., when a provider is changed, resources are scaled down, or physical hardware is reallocated.

Also, fundamental cloud-related features impact this risk: data may be available beyond the lifetime specified in the security policy since in-depth data removal requires destroying its physical carrier, which frequently stores data from other users at the same time. When in-depth data removal from the cloud is requested, standard procedures that were developed before cloud emergence (e.g., certification requirements) are inefficient if only the software API is applied to data removal [37]. Also, the risk is impacted by the lack of knowledge of who controls retention of data or what the regulatory requirements are in that respect [35].

Probably the biggest privacy risk related to cloud services is related to the information that the cloud provider accumulates or calculates about user-related activity in the cloud. This would include data collected to measure and charge for resource consumption, logs and audit trails, and application-specific data. Such data, if sold or leaked, or in case of their release in the form of user-scoring service or organization-rating service, are a huge threat to user privacy. For example, in the case of organizations, the data could be used to infer the status and outlook of an organization's initiative [36]. At present, there are no technological barriers to such secondary uses [35]. Encrypting stored data is straightforward, but despite advances in homomorphic encryption, there is no prospect of commercial systems being able to maintain this encryption during real-time processing of large datasets [39]. This means that nowadays, and probably also in the foreseeable future, cloud customers doing anything other than storing encrypted data in the cloud must trust the cloud provider [37] or put their trust in ex post law enforcement. While the focus is mainly on protecting application data, cloud providers also store and process metadata. Regardless of whether the metadata is stored within or outside the cloud resources, metadata includes details about the accounts of cloud users that could be used by the cloud provider for unauthorized purposes or compromised by a third party and used in subsequent attacks [36].

Threats to the user's ownership rights over the data constitute the next group of privacy risks. Rarely does the service contract state clearly that the user or organization retains exclusive ownership over all its data and that the cloud provider acquires no rights or licenses through the agreement. Specifically, the service contract should exclude intellectual property rights and licenses to use the user's data for the cloud provider's purposes. It should also exclude any interests in the data even for security purposes and exclude any cloud provider's unilateral amendment to these data ownership rights [36]. Furthermore, SLAs are expressed in natural languages, as opposed to machine-readable formal languages, making automatic assessment whether data usage rules are respected by cloud service provider impossible. Also, it is hard to prevent data rights transferability to other third parties upon bankruptcy, acquisition, or merger, and it is hard to ensure that a data subject can get access to all his/her PII [35]. To summarize, the usage of public cloud infrastructures makes it difficult to assure effective controls of privacy compliance verification in an automated way, so the end user has no means to verify that his/her privacy requirements are fulfilled [35].

4.6 Big data

The big data phenomenon is a consequence of cheap data storage and transmission, and the explosion of digital data sources. There are two categories of digital private data, related to their source: data collected explicitly with the awareness of the person affected and data collected implicitly, without personal awareness that data are being collected. In the latter case, he/she just has the general knowledge that private data collection might

happen, but has no specific knowledge if it really has happened, which data have been collected, how long they have been stored, who had or has access to them, and for which purpose they are processed.

Within the category of explicitly collected data, three cases are distinguished. In the first case, the user is fully aware of his/her disclosure of private data to a service provider in order to use the service's core functionality. In the second case, the service provider generates or collects the private data of a stakeholder/customer in a situation where the customer is, or should be, aware of and a participant to the data being used, e.g., a doctor generating the medical history of a patient or a bank keeping track of customer financial transactions. In those cases, providing private data is an unquestionable requirement for being served. The doctor cannot help the patient without being able to collect samples and analyze or process results. Without providing a courier with the private address, a parcel cannot be delivered, etc. The problem of privacy arises when data collected for the sake of a particular service are used for other purposes without the consent of the concerned person.

The third case of the disclosure of private data by a person includes the data collected by a wide range of digital services: social media, media sharing, games, education, training, coworking, etc. People voluntarily disclose their private data, in general, to be in contact with other people or to get a higher social position which is considered a benefit. The problem of privacy arises when such data go beyond people for whom they were intended. In practice, the spread and processing of such data are impossible to stop. It is also important to notice that the definition of benefits resulting from private data publishing evolves during one's lifetime. A video or collection of photos leveraging the popularity of a registered high school student may be an obstacle for his/her professional, social, or political career 20 or 30 years later.

In the category of implicit private data collection, automated data collection and big data analysis are distinguished. Data are collected automatically when digital services are provided. Examples are tracking credit card payments, mobile phones locations, websites visited, web user's queries inserted to search engines, recognizing objects in camera images, and data coming from sensors. Storing these historical private data is not a necessary condition for providing digital services. These data are stored for marketing or safety reasons, but they may be legally or illegally abused for violating people privacy. It is worth to stress that data retention and processing that is illegal in one country may be legal in another one.

Big data opens new possibilities for data analysis. Due to massive computational powers of modern computers, the availability of raw datasets from multiple sources, and the development of new data processing methods, instead of analyzing a random data sample, as it is done with classic statistical methods, it is possible to analyze all available data. Since all data are analyzed, it is possible to accept a higher level of their disorder and lower level of exactitude. It turns out that such approach often yields more objective and accurate knowledge [32]. Finally, one of the most important characteristics of big data analysis is the possibility of a paradigm shift. Instead of discovering knowledge by searching for causality, one can discover it by searching for correlation. Knowledge obtained in this way can form the ground for effective actions; however, it does not provide understanding. In other words, by analyzing big data, it is possible to learn with high probability what is happening, and even what will happen (predictive big data analysis), but not why it happens or why it will happen [5]. Correlation is a statistical relationship between two data values. If one of the data values changes, then the other data value (the correlate) is also likely to change. The correlate's change can be preceding, which permits predicting (with some probability, not with certainty). However, correlations may be meaningless and spurious.

A collection of such correlations is presented by Tyler [40]. It is also worth emphasizing that big data analysis can be personalized, and as such, it is different from profiling. Profiling, as this term is used in IT, is based on the classification and on assumptions that people belonging to the same class will behave in the same way so that they can be treated in the same way. For example, every man over 50 is at risk of heart attack, so every man over 50 should visit a cardiologist. Predictive big data analysis is based on calculating the probability of an event that will happen to a particular person. So, some men over 50 will be at high risk of heart attack, while others will not, depending on case-specific variables. Big data analysis is used to predict what decisions an individual will make in the future, e.g., what product he/she will buy as the next one, what holiday destination he/she will choose, or what next word he/she will type when texting. Big data analysis permits us to go beyond profiling due to the personalization of prediction. However, it must be stressed that both big data analysis and profiling are based on machine-learning techniques [41].

Big data analysis not only increases the risk of privacy violations, but it also changes the character of the risk. The value of big data analysis is in the data reuse for purposes different from the primary use. Some types of big data analysis may undermine the current, broadly used, legal principle of notice and consent of individuals for using their personal data for a specific purpose and a prohibition of using these data for any other purpose. One cannot consent in advance to processing his/her data in a way that does not exist yet. Due to the massive volume of data, the number of data owners is often counted in the millions. Due to their dispersion, an individual who is the owner of his/her private data cannot be asked again for consenting to processing those data when a secondary purpose arises. However, the lack of consent does not necessarily protect privacy. People protested showing their houses in Google Street View for fear of burglaries; however, blurring the image of a particular house could in itself provide a clue for the burglars [5].

One of the fundamental means of protecting user anonymity in the datasets that are to be published (e.g., medical or census data) is a process called data deidentification, which is composed of removing explicit or implicit identifiers, such as name, Social Security number, or driving license number. However, the efficiency of this process is brought into question by big data analysis which—to a large extent—permits to reidentify previously deidentified data. Reidentification tends to be persistent: once data are linked to an identified person, they become difficult to separate from his/her identity. Reidentification applied on a mass scale will gradually erode an individuals' privacy [42]. A significant challenge arises from the fact that it is difficult to develop formal constraints for deidentification that would prove to be robust enough to protect data from the threats of both present and future techniques for reidentification.

To bypass the current legal regulations forbidding personal data processing without explicit consent of a person concerned, service providers make such consent a condition for beginning service. This strategy for gaining consent is particularly significant in digital service markets operating according to the "winner takes all" rule, which leads to their monopolization (observed single social network, single search engine, single online auction service, etc.). Therefore, a person is faced with the hard dilemma: either surveillance acceptance or digital/social exclusion. What is even worse, the personal data collected are subject to trade. To mitigate laws limiting personal data interchange, these data are not traded directly, but shared in a form of recommendation services.

Big data analysis permits systems to go beyond the reidentification of raw data, namely, to create derivative datasets describing sensitive attributes of an individual [43]. This is done through the analysis of relationships of an individual with other persons, products, services, themes, opinions, etc. based on publicly available information and cross-referencing

of different datasets. As such information is not directly collected from the individual, companies analyzing big data have no legal obligation to gather his/her consent or give notice in the way required by the laws regulating conventional PII collection [43].

Data derivation permits systems to generate a detailed picture of different aspects of an individual's life, including information that he/she has never explicitly disclosed [43], which is a real threat to his/her privacy. Taking advantage of the individual's sensitive data, a service provider gets to know the preferences of the transacting person. The service provider can therefore takeover the entire "added value" of the transaction by dynamic service pricing in an optimal (from the service provider perspective) distance from the user's reservation price [42]. These information asymmetries and "price discriminations" have been present in the online markets for years, but now, they are further escalated by big data techniques.

Narrowing the users' access and choice (called the "filter bubble" phenomenon) is also a consequence of the big data analysis of sensitive data. When searching the Internet, a person will be always limited to the same fragment of information and knowledge resources considered the most appropriate to him/her. So he/she will never get information from outside the "glass walls." In the long term, it may influence person's cultural capital and even impact free information interchange and the freedom of speech. Clustering Internet users leverages the division of the society into groups of similar-thinking clones [42].

An important problem of big data analysis, which belongs to probabilistic approaches, is that predictive algorithms are often themselves unpredictable. Techniques of machine learning including neural networks, which are the basis of predictive big data analysis, run in two phases. The first phase is devoted to training from examples; the second phase is devoted to prediction. The quality of prediction is highly dependent on examples used for training the network. If a real case does not conform to training examples, the prediction will be false; thus, decisions based on that prediction will be wrong. The consequences of such wrong predictions may be different, from negligible or severe. If a person is not properly prompted when texting, consequences are negligible, but if a person is wrongly qualified as a potential terrorist, the consequences may be very severe. In practice, it is impossible to prove that a prediction is right or wrong, because, as we mentioned earlier, big data analysis is not based on a cause-and-effect relationship but on correlations among different datasets and the analysis of a big number of training examples. Moreover, it is impossible to know in advance when a learning algorithm will predict a user's PII. Therefore, it cannot be planned where and when to assemble privacy protections related to these data [43].

The legal status of different datasets is ambiguous. Some of them are publicly or semipublicly accessible, some of them are owned by communication and digital service providers. It is unclear whether an individual's data can be simply used (alone or as a part of a larger aggregate) without requesting permission whether the data can be taken out of the context and analyzed in a way likely opposite the subject's will, who can benefit from the access to big data; who is responsible for ensuring that individuals are not harmed by the big data analysis, how informed consent should be defined and how it can be executed, and what constitutes the set of best ethical practices for data analytics [44].

Even with the use of nonsensitive data, predictive big data analysis may have a discriminatory effect on individuals. Those who have a privileged digital history since their childhood, e.g., having well-situated family in their social network, having digital interests (likes and clicks) more attractive from the commercial point of view, or having more promising financial prospects will automatically receive even more measurable benefits in the digital society; for those with questionable or ambiguous digital records, even the social status that they have had thus far will be hard to maintain. "Predictive analysis becomes a self-fulfilling prophecy that accentuates social stratification" [42].

People attempting nonstandard behaviors or brave enough to take challenges, but who failed, will risk immediate decline in the digital society based largely on big data analysis. What is even worse, "as the ramifications of big data analytics sink in, people will likely become much more conscious of the ways they're being tracked, and the chilling effects on all sorts of behaviors could become considerable" [45]. What will result is a gradually emerging "surveillance society, a psychologically oppressive world in which individuals are cowed to conforming behavior by the state's potential panoptic gaze" [42]. The worst kind of censorship is autocensorship.

4.7 Conclusions

As follows from the preceding sections, emerging information technologies including the IoT, AR, biometrics, cloud computing, and big data analysis increase the risk of privacy breaches and, in many cases, make current approaches to protecting privacy inefficient and insufficient. Moreover, within the e-society, all these technologies are used simultaneously, so their cumulative and reinforcing effects apply to each person. Thus, it is required to adopt a more holistic view of privacy protection.

The IoT, AR, and biometrics may be seen as data providers. Those data are stored in the cloud. Data aggregated in the cloud, coming from different sources, are perfect objects for big data analysis.

The IoT provides new challenges for privacy because it follows the principle of ubiquitous computing. Sensors and actuators deployed in a particular environment (smart home, smart car, smart road, etc.) adapt that environment to individual or group needs automatically, without explicit human interaction. Therefore, there is no space for explicit consent. Adaptation requires knowledge of preferences, i.e., private data. If a smart home or a smart car is adapted to the needs of its owner, the owner's private data are not disclosed to third parties. If a road, an office, or a public building is adapted to the needs of an individual, his/her private data have to be entrusted to companies managing them, so the risk of trust abuse is much higher. It is also worth noting that IoT extends the risk of private data abuse from the digital world to the real world, i.e., real installations deployed in buildings, cars, roads, etc. The malfunctioning of these installations may cause physical damages.

AR provides private data coming from information-rich raw multimedia data streams that are temporally and spatially interrelated. These data concern not only the owner of an AR device, but all the people who surround him/her in a particular place and moment, who are captured in photos and videos.

Biometric access control provides unique personal identification for life. As such, it eliminates a possibility of privacy protection by several virtual identities of the same person devoted to different services. As mentioned in Section 4.4, data collected for biometric access control may be used to infer information describing persons, not only his/her body or medical condition, but even cultural or social characteristics.

Cloud computing is currently the most economical option of providing computing power and storage capacity. It is particularly useful in the case of small electronics devices of limited capabilities including power supply such as sensors, actuators, and mobile devices. The application of cloud computing requires entrusting private data to cloud computing providers, i.e., pass control over them. A client of cloud computing services can only trust that his/her data are not processed for purposes that he/she never agreed to. The risk of trust abuse is increased by the fact that data stored in the cloud are often replicated and spread among different locations in different countries governed by different law regulations. To reduce the risk associated with could computing, private or

community clouds are used which restrict clients to one company or several companies of similar characteristics, e.g., only banks. In a contract with cloud a computing service provider, a client may consider restricting the storage of his/her data to one data center or centers located in one country governed by one system of privacy protection regulations.

As mentioned earlier, massive data aggregated in the cloud, coming from different sources, partially identified, are perfect objects for big data analysis based on correlations instead of on a cause-and-effect relationship. Big data technologies have the potential to bring together all the specific risks following from the technologies described earlier and amplify them. Big data analysis permits us to not only reidentify data deidentified prior to release, but also generate with high probability a detailed picture about different aspects of a person's life, including information the person has never disclosed to any service. Moreover, big data analysis permits—again with high probability, not certainty—the prediction of future behavior and future actions of a person. In cases when certainty is not required and a person retains the right to free choice, predictive big data analysis may provide a person with advantages, otherwise—not. If an individual is an object of a decision made by somebody else, e.g., he/she may get a loan or not, may get a job or not, or may be invited to an event or not, big data analysis may lead to discrimination.

As explained by Mayer-Schonberger and Cukier [5], predictive big data analysis challenges the current justice system. Thinking about committing a crime is not illegal, only progressing from the thought to the criminal act is. Individual responsibility is linked to the individual choice of an action. Finding someone guilty of an anticipated crime he/she has not yet committed is a mistake made by using predictive big data analysis based on the correlation with making decisions about one's individual responsibility that requires a proof of a cause-and-effect relationship. The abuse of big data analysis leads to a society in which there is no individual choice of action based on free will, but the individual moral compass is replaced by predictive algorithms, and individuals are exposed to the unlimited coercion of collective decisions made in the past used to calculate the probability of their actions to be made in the future. This poses the risk of enslaving of the society.

As follows from this chapter, in the era of emerging technologies, in particular predictive big data analysis, a new approach to protect individuals' privacy has to be developed. The risk of wrong predictions may be mitigated by providing access to data and algorithms for their verification and certification by trusted third parties. The same big data should be analyzed by a regulatory authority, to limit the monopolization of benefits coming from predictive analysis. Also, it is necessary to create legal procedures for rebutting a prediction about a specific person. Organizations using predictive big data analysis should be legally responsibility for its effects. Also, if provided open access to big data, governments or third-sector organizations could use the same big data techniques to discover who is being discriminated against and by whom the discrimination is being activated.

References

1. European Parliament and the Council of the European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), *Official Journal of the European Union*, L 119, 2016.
2. Stevens, Gina Marie. Data security breach notification laws. Congressional Research Service, Washington, DC (2012).
3. Cremonini Marco, Chiara Braghin, and Claudio Agostino Ardagna. *Privacy on the Internet Computer and Information Security Handbook* (Second Edition), Morgan Kaufmann, Burlington, MA (2013).

4. Cellary, Wojciech, and Jarogniew Rykowski. Challenges of smart industries—Privacy and payment in visible versus unseen Internet. *Government Information Quarterly* (2015).
5. Mayer-Schonberger, Viktor, and Kenneth Cukier. *Big data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, Boston, MA (2013).
6. International Telecommunication Union, Recommendation ITU-T Y.2060 (06/2012), 2012. http://handle.itu.int/11.1002/1000/11559.
7. Hajdarbegovic, Nermin. Are We Creating An Insecure Internet of Things (IoT)? Security Challenges and Concerns, 2015, Accessed April 5, 2017. https://www.toptal.com/it/are-we-creating-an-insecure-internet-of-things.
8. Hewlett-Packard Development Company. HP Study Reveals 70 Percent of Internet of Things Devices Vulnerable to Attack, 2014, Accessed April 5, 2017. http://www8.hp.com/us/en/hp-news/press-release.html?id=1744676.
9. Light Reading. Poll: Standardization Biggest Challenge in IoT, 2015, Accessed April 5, 2017. http://www.lightreading.com/iot/iot-strategies/poll-standardization-biggest-challenge-in-iot/a/d-id/714062.
10. Gartner. Gartner Says 8.4 Billion Connected Things Will Be in Use in 2017, up 31 Percent from 2016, Gartner Press Release, 2017, Accessed September 11, 2017. http://www.gartner.com/newsroom/id/3598917.
11. Verizon. State of the Market: The Internet of Things 2015, 2015, Accessed April 5, 2017. http://www.verizonenterprise.com/resources/reports/rp_state-of-market-the-market-the-internet-of-things-2015_en_xg.pdf.
12. Wind River System. Security in the Internet of Things. Lessons from the Past for the Connected Future, 2015, Accessed April 5, 2017. https://www.windriver.com/whitepapers/security-in-the-internet-of-things/wr_security-in-the-internet-of-things.pdf.
13. Apthorpe, Noah, Dillon Reisman, and Nick Feamster. A smart home is no castle: Privacy vulnerabilities of encrypted IoT traffic, *Workshop on Data and Algorithmic Transparency (DAT'16)*, 2016.
14. Wang, Chen, Xiaonan Guo, Yan Wang, Yingying Chen, and Bo Liu. Friend or foe? Your wearable devices reveal your personal pin. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, pp. 189–200. Association for Computing Machinery, New York (2016).
15. Ramirez, Edith. *Privacy and the IoT: Navigating Policy Issues*. US Federal Trade Commission, Washington, DC (2015).
16. Lobao, Martim. Software Updates: A Visual Comparison of Support Lifetimes for iOS vs. Nexus Devices., Android Police, 2015, Accessed September 11, 2017. http://www.androidpolice.com/2015/09/17/software-updates-a-visual-comparison-of-support-lifetimes-for-ios-vs-nexus-devices/.
17. Canonical, Defining IoT Business Models. Monetising IoT investments, maximising IoT skills and addressing IoT security, 2017, Accessed September 11, 2017. https://pages.ubuntu.com/IOT_IoTReport2017.html.
18. Wójtowicz, Adam, and Daniel Wilusz. Architecture for adaptable smart spaces oriented on user privacy, *Logic Journal of the IGPL*, vol. 25, issue 1, pp. 3–17 (2017).
19. Azuma, Ronald, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications* vol. 21, issue 6, pp. 34–47 (2001).
20. Jana, Suman, Arvind Narayanan, and Vitaly Shmatikov. A Scanner Darkly: Protecting user privacy from perceptual applications. In *2013 IEEE Symposium on Security and Privacy (SP)*, pp. 349–363. Institute of Electrical and Electronics Engineers, Piscataway, NJ (2013).
21. Calo, Ryan. People can be so fake: A new dimension to privacy and technology scholarship. *Penn State Law Review* vol. 114, p. 809 (2009).
22. Raval, Nisarg, Animesh Srivastava, Kiron Lebeck, Landon Cox, and Ashwin Machanavajjhala. Markit: Privacy markers for protecting visual secrets. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1289–1295. Association for Computing Machinery, New York (2014).
23. Jana, Suman, David Molnar, Alexander Moshchuk, Alan M. Dunn, Benjamin Livshits, Helen J. Wang, and Eyal Ofek. Enabling fine-grained permissions for augmented reality applications with recognizers. In *USENIX Security*, pp. 415–430. USENIX Association, Washington, DC (2013).

24. Dwork, Cynthia. Differential privacy. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming*, pp. 1–12 (2006).
25. Roesner, Franziska, Tadayoshi Kohno, and David Molnar. Security and privacy for augmented reality systems. *Communications of the ACM* vol. 57, issue 4, pp. 88–96 (2014).
26. Bell, Michael, and Vitali Lovich. Apparatus and methods for enforcement of policies upon a wireless device. US Patent 8,254,902, issued August 28, 2012.
27. Roesner, Franziska, Tamara Denning, Bryce Clayton Newell, Tadayoshi Kohno, and Ryan Calo. Augmented reality: Hard problems of law and policy. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1283–1288. Association for Computing Machinery, New York (2014).
28. Prabhakar, Salil, Sharath Pankanti, and Anil K. Jain. Biometric recognition: Security and privacy concerns. *IEEE Security & Privacy* vol. 99, issue 2, pp. 33–42 (2003).
29. Grijpink, Jan. Privacy law: Biometrics and privacy. *Computer Law & Security Review* vol. 17, issue 3, pp. 154–160 (2001).
30. Crompton, Malcolm. Biometrics and privacy the end of the world as we know it or the white knight of privacy? *Australian Journal of Forensic Sciences* vol. 36, issue 2, pp. 49–58 (2004).
31. Derakhshani, Reza, Stephanie AC Schuckers, Larry A. Hornak, and Lawrence O'Gorman. Determination of vitality from a non-invasive biomedical measurement for use in fingerprint scanners. *Pattern Recognition* vol. 36, issue 2, pp. 383–396 (2003).
32. Bolle, Ruud M., Jonathan Connell, Sharath Pankanti, Nalini K. Ratha, and Andrew W. Senior. *Guide to Biometrics*. Springer, Berlin (2013).
33. Rathgeb, Christian, and Andreas Uhl. A survey on biometric cryptosystems and cancelable biometrics. *EURASIP Journal on Information Security* vol. 2011, issue 1, p. 3 (2011).
34. Rejman-Greene, Marek. Privacy issues in the application of biometrics: A European perspective. In Wayman, James, Anil Jain, Davide Maltoni, and Dario Maio (eds), *Biometric Systems*, Springer, London, pp. 335–359 (2005).
35. Pearson, Siani. Privacy, security and trust in cloud computing. In *Privacy and Security for Cloud Computing*, pp. 3–42. Springer, London (2013).
36. Jansen, Wayne, and Timothy Grance. *Guidelines on Security and Privacy in Public Cloud Computing*. NIST Special Publication 800-144. National Institute of Standards and Technology, Gaithersburg, MD (2011).
37. European Union Agency for Network and Information Security. *Cloud Computing: Benefits, Risks and Recommendations for Information Security*. European Network and Information Security, Heraklion (2009).
38. Molnar, David, and Stuart E. Schechter. Self hosting vs. cloud hosting: Accounting for the security impact of hosting in the cloud. In *Proceedings of the Ninth Workshop on the Economics of Information Security (WEIS 2010)*. Microsoft Research, Cambridge, MA (2010).
39. Zang, Wanyu, Meng Yu, and Peng Liu. Privacy protection in cloud computing through architectural design. In Vacca, John R. (ed), *Security in the Private Cloud*, pp. 319–343. CRC Press, Boca Raton, FL (2016).
40. Vigen, Tyler. *Spurious Correlations*. Hachette Books, New York (2015).
41. Alpaydin, Ethem. *Introduction to Machine Learning*. MIT Press, Cambridge, MA (2014).
42. Tene, Omer, and Jules Polonetsky. Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property* vol. 11, p. xxvii (2012).
43. Crawford, Kate, and Jason Schultz. Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review* vol. 55, p. 93 (2014).
44. Boyd, Danah, and Kate Crawford. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* vol. 15, issue 5, pp. 662–679 (2012).
45. Stanley, Jay. *The Potential Chilling Effects of Big Data*. American Civil Liberties Union, New York (2012).