Internet of Things: Security and Privacy Issues

FARZAD KAMRANI, MIKAEL WEDLIN,
IOANA RODHE

FOI-R--4362--SE
ISSN-1650-1942 December 2016

Defence and Security, Systems and Technology

Farzad Kamrani, Mikael Wedlin,
Ioana Rodhe

Internet of Things:
Security and Privacy Issues

Titel	Sakernas internet: säkerhet och integritetsfrågor
Title	Internet of Things: Security and Privacy Issues
Rapportnr / Report no	FOI-R--4362--SE
Månad / Month	December
Utgivningsår / Year	2016
Antal sidor / Pages	50p
ISSN	1650-1942
Kund / Customer	FM
FoT område	Avskanning av forskningsfronten
Projektnr / Project no	I621053
Godkänd av / Approved by	Matts Gustavsson
Ansvarig avdelning	Försvars- och säkerhetssystem

FOI Swedish Defence Research Agency
Defence and Security, Systems and Technology
SE-164 90 STOCKHOLM

Abstract

This report presents the results of a survey project at FOI with the goal of comprehending the security and privacy implications of the Internet of Things (IoT).

There are several characteristics such as uncontrolled environment, heterogeneity, requirements for scalability, and constrained resources, which make the security and privacy issues of the IoT more challenging. Moreover, one of the requirements of IoT systems is a high degree of availability, which sometimes may compromise confidentially and integrity of the data. In the light of these conditions, the report discusses some examples of IoT security problems in specific areas, such as medical technology and transportation. Stuxnet worm that targeted control systems of Iranian nuclear enrichment plants, along with a demonstration in which researchers used a vehicle's Internet connection to take over the control of the car remotely, indicate that the IoT security encompasses new dimensions previously not observed.

Other incidents such as the recent attack on the USA-based company Dyn shows that IoT devices like video cameras, digital video recorders, and home routers can be used as a platform to orchestrate distributed denial-of-service (DDoS) attacks.

We have already observed a gradual change in attitudes, both in public administrations and in business entities towards collecting, processing and storing as much personal data as possible. It is plausible to assume that the emergence of the IoT will reinforce this development. One of the challenges of privacy in the IoT is that the data collection process is more passive and more pervasive, which results in that the users are less aware of whether and when they are being tracked.

Existing approaches to anonymize personal data, such as *k*-anonymity, *l*-diversity and *t*-closeness add some safeguards to the data, however, none of these methods can guarantee anonymity, and researchers have demonstrated how these privacy measures can be breached in practice. Differential privacy, which is a randomization-based notion of privacy, provides a mathematical model of privacy and quantifies the individual privacy loss. Differential privacy systems based on this model, ensure that the privacy loss will be significantly the same regardless of joining or withdrawing private information from a database.

While confidentiality of the data has always been critical in the military domain, the advance of the IoT creates new challenges in a world where armed forces become more and more dependent on connected devices and services. Privacy issues that the explosion of the IoT might entail are not only a concern for military personnel as individuals. Exposure of personnel's private information to adversary is a potential severe threat that should be considered seriously.

Keywords

Internet of Things, Security, Privacy, Privacy-preserving methods

Sammanfattning

Denna rapport presenterar resultaten av ett avskannande forskningsprojekt vid FOI med målet att förstå säkerhets- och integritetskonsekvenserna av sakernas internet (IoT).

Det finns flera egenskaper såsom okontrollerad miljö, heterogenitet, krav på skalbarhet och begränsade resurser som gör att säkerhets- och integritetsfrågor i IoT blir mer utmanande. Dessutom är en hög grad av tillgänglighet ett av kraven på IoT system, vilket ibland kan äventyra konfidentialitet eller riktighet i informationen.

Mot bakgrund av dessa betraktelser, diskuterar rapporten några fall av säkerhetsproblem med IoT på vissa områden, såsom medicinteknik och transport. Stuxnet-masken som riktades mot kontrollsystemen för de iranska nukleära anrikningsanläggningarna, tillsammans med en demonstration där några forskare använde ett fordons internetanslutning för att ta över kontrollen över bilen från avstånd, visar att säkerhet för IoT omfattar nya dimensioner som tidigare inte observerats.

Andra händelser som den nyligen inträffade attacken på det USA-baserade företaget Dyn visar att IoT enheter så som videokameror, digitala videoinspelare och hemroutrar kan användas som en plattform för att iscensätta distribuerade belastningsattacker (DDoS).

Vi har redan sett en gradvis förändring av attityder, både inom den offentliga förvaltningen och i privata företag mot insamling, bearbetning och lagring av så mycket personuppgifter som möjligt. Det är rimligt att anta att framväxten av IoT kommer att förstärka denna utveckling. En av utmaningarna för den personliga integriteten inom IoT är att datainsamlingsprocessen är mer passiv och mer genomträngande, vilket resulterar i att användarna är mindre medvetna om och när de spåras.

Befintliga metoder för att anonymisera personuppgifter, såsom k-anonymity, l-diversity och t-closeness är användbara och ger något skydd, men ingen av dessa metoder kan garantera anonymitet, och forskare har visat hur dessa åtgärder kan brytas i praktiken. Differential privacy, vilket är en slumpbaserad metod, ger en matematisk modell av integritet och kvantifierar den enskildes integritetsförlust. Differential privacy försäkrar att den personliga integritetensförlusten för en individ kommer att vara (ungefär) densamma oavsett om dess data är med eller utesluts från en databas.

Även om informationssäkerhet har alltid varit kritisk för den militära domänen, skapar IoTs frammarsch nya utmaningar i en värld där väpnade styrkor blir mer och mer beroende av anslutna enheter och tjänster. Integritetsfrågor som explosionen av IoT kan innebära är inte bara ett problem för militär personal som individer. Exponering av personalens privata information till motståndaren är ett potentiellt allvarligt hot som bör beaktas.

Nyckelord

Sakernas internet, Säkerhet, Integritet, Integritetbevarande metoder

Contents

1 Introduction

This document is the final report of a survey project at FOI with the goal of covering some, for FOI and the Swedish Armed Forces, relevant aspects of the *Internet of Things* (IoT). The scope of the project was limited to literature study and synthesis and there was no room for conducting new research.

The IoT covers many different areas and overlaps with several research fields, ranging from enabling technologies (e.g. sensors, protocols, sensor networks), to software architecture (e.g. middleware, cloud solutions, data management, big data), services and applications (e.g. smart homes, smart cities, connected cars), social impacts of the IoT (e.g. acceptance of users, change in the societal organizations, change in control of the infrastructure), and security and privacy issues [1].

In an early stage of the project, the authors decided to delimit the scope of the study to security and privacy issues of the IoT as they found it more appealing. Although we cannot postulate that we provide a complete overview of the security and privacy issues of the IoT, we believe that the report can serve as a good starting point to obtain a grasp of the subject and become acquainted with the most current developments in the area.

1.1 What is IoT?

Loosely speaking, the IoT is the collection of physical devices that are connected to the Internet; however, despite the diversity of research on IoT, its definition remains fuzzy. A good start point to seek for the definition of the IoT is [1], where the *IEEE Internet of Things Initiative*[1] aims to give an all-inclusive definition of the IoT ranging from small localized to large, distributed and complex systems. The document provides an excellent example of how diverse the definition of IoT might be.

Even if IoT devices are often small appliances that encompass a wireless transmission channel, it is not a necessary part of the IoT. A more common feature is that at least one of the systems is embedded and in control of something in the environment.

The definition that best reflects the authors' view of the IoT is found in [2] where IoT is defined as:

> "The term IoT refers to the connection of systems and devices with primarily physical purposes (e.g. sensing, heating/cooling, lighting, motor actuation, transportation) to information network (including

[1] IEEE Internet of Things (IoT) Initiative is an IEEE (Institute of Electrical and Electronics Engineers) platform for the global technical community working on the IoT (http://iot.ieee.org/ [accessed Nov 11, 2016]).

> the Internet) via interoperable protocols, often built into embedded systems."

This definition is much broader than most of the others and includes embedded systems that we have previously seen in larger energy producers and other critical infrastructure.

1.2 Security and privacy challenges

Like every new technology, the IoT brings with its benefits new risks and challenges. As the IoT is closely related to communication and information technology, it is justified to consider security and privacy challenges already known in information security and examine how these concerns are transferred to the current and future state of the IoT. At the first glance, the similarities seem so many and the differences so subtle that one may be deceived that security and privacy concerns in the IoT are the same challenges as those known in the information security and the same measures are sufficient to face these challenges. However, some characteristics make the security and privacy challenges in the IoT so distinct that a more careful investigation of the subject is required.

- The number of connected devices to the Internet has already surpassed the number of humans on the planet. This number continues to increase dramatically and is predicted to be between 26 billion and 50 billion by 2020 (e.g. see [3]). Several factors facilitate this development, among others the introduction of Internet Protocol version 6 (IPv6), which allows that every device has a unique IP-address, leading to much easier communication between devices. However, the security and privacy issues for the IoT does not increase linearly with the number of Internet-connected devices, but grows in a much faster rate. This is due to the fact that, even assuming everything is the same, the number of communication channels in a network increases faster than the number of nodes.

- Computer networks are often heterogeneous in nature, which can induce security challenges. The IoT is expected to be far more heterogeneous than current computer networks, integrating a multitude of various devices from different manufacturers, software platforms and communication protocols.

- While servers and workstations are protected in server rooms and offices, and personal computers, notebooks and handheld devices are protected by the owner's presence, in an IoT setting, sensors and other devices are located everywhere, and exposed to theft, malicious damage and intrusion. Vicious attackers can use the increased physical accessibility of devices to find more vulnerabilities in IoT systems.

- IoT devices are usually battery-driven, fault-prone systems and generally have lower processing power and memory and used to not fulfil the requirements for implementing appropriate security and anonymization services, which require sufficient amount of processing and memory resources. Although in recent years, this has been less of an issue, many constructions still have a lack of security functions built in.
- The IoT is expected to be ubiquitous and pervasive. Connected devices are worn, carried or seamlessly embedded in the world around us. They may collect data, communicate and interact with other devices, without our permission or even our knowledge, simply because we do not own them (e.g. video surveillance cameras in a shopping mall, or connected vehicles that we travel in as a passenger).
- As the number of connected things increases, the amount of gathered and accumulated information about us in different databases increases continuously. Although sensitive data might be removed or protected by anonymization when the data is disseminated, an unpredictable combination of seemingly non-sensitive data from different sources can create a unique identifier resulting in privacy breaches.
- While until recent years, cyber-attacks have mainly threatened information systems, computer networks, and personal computers, the IoT will escalate security risks to a different level. In the IoT era, as actuators and control systems will be interconnected with other systems, attackers may be able to directly target connected devices and achieve physical destruction of the equipment and infrastructures, such as self-driving cars, smart houses, electric grids, oilfields, transportation systems, and nuclear plants. Stuxnet [4] was the first malicious code that attacked the control system of a nuclear facility; however, with the explosion of the IoT, it will not be the last one that leaves the cyber realm to cause physical destruction.
- The IoT inherently has a dynamic characteristic. Pervasive devices such as wearables can join and leave the IoT network (e.g. smart homes) anytime. This, in combination with multiprotocol communication characteristics, makes the traditional information security measures insufficient for the IoT [5].

1.3 The outline of the report

The outline of the report is as follows. In chapter 2, the security issues related to the IoT are discussed. While in chapter 2, the target of attacks are IoT devices, in chapter 3, security attacks are discussed where IoT devices are used as a platform for a large-scale attack on Internet infrastructure. In chapter 4, privacy in general, and in relation to the IoT is reviewed. Chapter 5 considers security and privacy in cloud-supported IoT and chapter 6 discusses some of the findings of the study and concludes the report.

2 IoT Security

> "The S in IoT stands for Security."[1]
>
> – Melvin Lammerts

IT security is a large and diverse field and it is not always obvious what the term really means. It is also closely related to the broader concept of information security. In the context of IoT, information security is intently related to protecting the privacy of the user.

Many researchers and industry partners, however, agree on that securing the IoT is very crucial and that this should happen from the beginning of IoT development [6]. At the same time, many researchers agree that securing the IoT is one of the most serious security challenges we are facing[2] today.

We already have many things connected to the Internet, mobile platforms, connected kitchen devices, cars and industrial control systems. Therefore, there are already many systems, both small and large, which collect and process data in our daily lives. Nevertheless, the greatness of the IoT comes from connecting all these systems together and allowing devices to communicate with each other across the systems. This demands new architectures for the IoT, and here the IoT is only at its initial stage of development. The area is not yet mature, there are no accepted standards and architectures, but there is much work in progress in various organizations [7]. In Europe alone, there are several research projects where architectures for IoT are developed, such as IoT Open Platforms[3], IoT European Research Cluster[4], IoT European Platforms Initiative[5], and IoT Architecture[6].

Georgia Institute of Technology highlights several threats related to the IoT in its report on emerging cyber threats[7] 2015:

- "Attackers target the trust relationship between users and machines" - trust is considered as a major challenge in the IoT where different devices will share data with each other on their users' behalf.
- "Technology enables surveillance, while policy lags behind."
- "Mobile devices fall under increasing attack, stressing the security of the ecosystem."

[1] https://twitter.com/showthread/status/776089345069555713 [accessed Sept 14, 2016].

[2] https://www.infosecurity-magazine.com/news/securing-the-iot-next-big-challenge/ [accessed Nov 11, 2016].

[3] http://open-platforms.eu/ [accessed Dec 12, 2016].

[4] http://www.internet-of-things-research.eu/ [accessed Dec 12, 2016].

[5] http://iot-epi.eu/ [accessed Dec 12, 2016].

[6] http://www.iot-a.eu/ [accessed Dec 12, 2016].

[7] http://www.cc.gatech.edu/sites/default/files/images/2015emergingcyberthreatsreport.pdf [accessed Dec 29, 2016].

Security and privacy challenges of the IoT originate from the specific characteristics of IoT networks, which make them unique [8]. These characteristics are:

- uncontrolled environment,
- heterogeneity,
- need for scalability, and
- constrained resources.

It can, however, be argued that the last item about resources is less valid than it is suggested. Even the smallest processor platforms today contain a decent crypto engine and enough program memory to implement security functions.

Vasilomanolakis et al. [8] further propose security requirements for IoT systems, based on their unique characteristics and divide the requirements into the following groups: network security, identity management, privacy, trust and resilience. The authors consider several architectures that have been proposed for the IoT in the research community and analyse whether different architectures meet the proposed security requirements. The analyses show that many of the security requirements are considered but none of the architectures covers all of them. Most uncovered are the privacy and trust requirements.

As long as there have been computers, there has been a widely accepted model for IT security based on three desired security features of the systems, often abbreviated as *CIA*, *confidentiality* (i.e., preventing unauthorized access to data), *integrity* (i.e., ensuring data is not altered), and *availability* (i.e., ensuring data is accessible when needed).

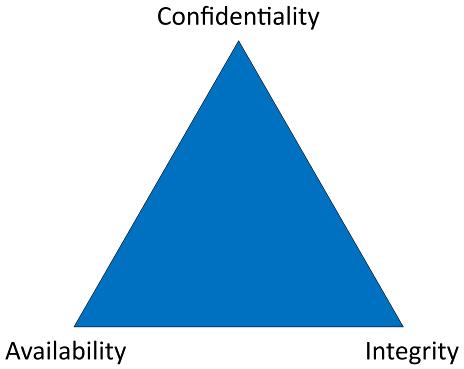


Figure 2.1: CIA (confidentiality, integrity, and availability) of IT systems are depicted as vertices of a triangle.

These properties have often been described in the form of a triangle in which the properties are placed in vertices (see Figure 2.1). Through the years, the model has been modified with a number of alternative key properties, but the core properties, CIA, has always remained. Something that has not been equally highlighted is that these three properties are never possible to be fully achieved simultaneously, as they are mutually exclusive. For instance, given the same resources, it is impossible to increase the availability, without compromising the confidentiality, accuracy, or both.

For the general information-processing computer systems, traditionally security almost entirely has focused on the confidentiality property, but for most of the embedded systems and the IoT it can be argued that the other two aspects are the most important ones, or at least much more important than it is in office information systems. One effect of this is that the security mechanism often is exclusively focused on the protection of the confidentiality of the data. Another observation is that the differences in approach in many cases seriously hampers cooperation between the administrators of the control system and standard IT systems. If you have confidentiality as the most important parameter, the implication will be to advocate more and more intrusive safeguards than what you can accept if it is availability that is the most prioritized feature. The delicate trade-off between confidentiality and availability introduces challenges that are the source of many IoT security problems. In this study, we have seen several examples of this. The knowledge is not new, but it has not previously been discussed in this context.

By the recent explosion of computerization of everything, from the microwave ovens to weapons systems, however, we cannot avoid the challenge any longer. Many of IoT systems will be linked together in larger networks and they will not work at all if you consistently apply a security strategy that primarily consists of confidentiality. However, we cannot ignore the other aspects of security, either. These systems will be subjected to antagonistic threats that we somehow have to deal with.

Even if IoT domain is not mature yet, we already have a multitude of devices connected to the Internet, which are part of systems of various sizes. To understand the threats and vulnerabilities of the IoT, one can start by investigating some of these devices.

2.1 Sloppy implementation of protocols and standards

Many weaknesses come from the implementations of various protocols and standards. One example is the ZigBee[8] standard, which includes security, but when some implementations of the standard have been analysed, several vulnerabilities could be identified. Zillner and Strobl [9] show that often only the minimum security requirements are implemented according to the ZigBee standard, and that user-friendliness is prioritized over security. They also show how the ini-

[8] https://en.wikipedia.org/wiki/ZigBee [accessed Aug 12, 2016].

tialization key is sometimes sent in plaintext in the initialization phase, which makes it possible for several encryption keys to be retrieved later on. Jun and Qing [10] also show, by looking at the firmware or by using a network protocol analyser, some encryption keys can easily be retrieved.

2.2 Medical technology

One area where embedded systems completely has exploded in recent years is in the medical-centres and hospitals. An example that shows how it can go wrong is from Panama where “The National Oncology Institute” right at the turn of the century had a problem with a radiation gun for the treatment of cancer patients [11]. A software program that was used to calculate the exposure dose occasionally computed inaccurately, so that a large number of patients were given a too high dose, some as high that they did not survive. This is not an example of security breach in the IoT per se, but more of an indication that the medical tools are complex and prone to error even when they are used by qualified users.

A significant factor in all these failed systems is that almost all regulation is against protecting the secrecy and privacy of the patients’ data. This is rather obvious in [12] where a few of the incidents in Swedish health care system are analysed.

2.3 Transportation

Another major area where the software begins to play an increasing role is in transportation. Two researchers made a general review of vulnerabilities in a number of automobile brands [13] already in 2014. In the autumn of 2015, the same researchers could also demonstrate that the hack that they had previously performed sitting in the car, also worked remotely and exemplified this in a Jeep Cherokee [14].

2.4 Stuxnet

When the first reports of the worm Stuxnet began to appear during the summer of 2010, it created some interest in the IT security community. It was the first time anyone had seen a malware that was specially created to manipulate control systems.[9]

[9] It is not the first time physical systems are destructed by using manipulated control software. During the cold war, the US Central Intelligence Agency (CIA) orchestrated a counter-intelligence operation that delivered modified computer chips and software to the Soviets military equipment and infrastructure (gas turbines, chemical plants, etc.) [15]. The Siberian gas pipeline sabotage in 1982 is attributed to this operation, where CIA allegedly manipulated some PLCs that the Soviet Union smuggled out from Canada. The manipulation eventually had the effect that a large part of the Siberian gas pipelines exploded [16]; however, this claim has been challenged.

To begin with, it just seemed to be an attempt to gather information from project databases in a Siemens WinCC system using a default password that was common to all installed systems of this type. For instance, the authors' very first impression was that the worm was probably just the result of someone's experiment who found a new default password and tested how widespread it was. It soon became evident that it was much more serious than that.

This description of Stuxnet is primarily based on three sources: (i) first and foremost, the white paper [17] published by the antivirus company Symantec[10], which conducted a study of the worm, (ii) FOI projects that studied a sample of the Stuxnet code (e.g., see [18]), and (iii) *Confront and Conceal* [19], a book by New York Times reporter David E. Sanger[11].

Although there is no solid proof, and no state has officially admitted responsibility for designing Stuxnet, much evidence points to Israel and United States as the origin of the malware. The attack was designed to manipulate the *programmable logic controller* (PLC), which controlled operating speed of centrifuges in the nuclear facility in Natanz and by destroying them, disrupt the enrichment process. These centrifuges were a central part of the nuclear enrichment process and additionally sensitive to variations in rotation speed. The attack required exact information about the structure of the system in all its details, and Israelis, which allegedly had infiltrated Iran's nuclear program, acquired this intelligence [19, pp 195].

Control systems were standard systems from Siemens and were easy to obtain. The centrifuges were more difficult to purchase. However, the United States had managed to take over a few that was left over when Libya had suspended its nuclear program, and these could be used for "destructive testing". Libya's centrifuges were of exactly the same design as Iran.

The oldest documented infection was in June 2009 [17, pp 9] and the large public outbreak started a year later, in summer 2010. There are slightly different data about when the operation began to succeed with wrecked centrifuges, ranging from that they started to fail almost immediately to November 2010 [19, pp 188].

Although, the Stuxnet attack is not a typical IoT security issue (the facility's network was an air-gapped network), however, it demonstrates the potential and impact of security problems in the IoT, where devices are seamlessly connected to the Internet and attacks might be designed to manipulate the physical systems directly.

[10] https://www.symantec.com/ [accessed Oct 23, 2016].

[11] https://en.wikipedia.org/wiki/David_E._Sanger [accessed Oct 23, 2016].

3 The IoT as a Platform for Attacks

> “In a relatively short time we’ve taken a system built to resist destruction by nuclear weapons and made it vulnerable to toasters.”[1]
>
> – Jeff Jarmoc

One effect of the discrepancy between protecting the secrecy and maintaining the availability is that many small gadgets that we connect to Internet have an insufficient security level. A recent example of this occurred on 21 October 2016 when the USA-based company Dyn was hit by a *distributed denial-of-service* (DDoS)[2] attack[3].

Two characteristics make the attack more significant from a security perspective. First, the attack made several large services on Internet unavailable by targeting the Domain Name System (DNS) supplier of these services. Services such as Spotify, Amazon, HBO, Netflix, Twitter and Reddit among others were affected. Notably for Sweden was that *the national website for emergency information* (www.krisinformation.se) was unavailable during the attack[4].

More interesting is the second characteristic of the attack. The botnet[5] Mirai[6], used for attacking Dyn was built of web cameras, digital video recorders, home routers and other IoT appliances.

The Mirai bot takes over a node by testing the default login names and passwords set by the manufacturer. Even with this naive operation, these botnets can grow fairly large. In the Dyn case they counted to more than 100k bots in the botnet. When an unprotected device is connected to the Internet, it will be reinfected within minutes[7].

According to Scott Hilton[8], the vice president of Dyn, the significance of the attack is not only due to its severity, but also because that it has highlighted vulnerabilities in the security of IoT devices that need to be addressed. The attack has started conversations about Internet security and volatility in information security community, as well as debate in the infrastructure community about the future of the Internet.

[1] https://twitter.com/jjarmoc/status/789637654711267328 [accessed Oct 21, 2016].

[2] https://en.wikipedia.org/wiki/Denial-of-service_attack [accessed Nov 27, 2016].

[3] http://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/ [accessed Oct 26, 2016].

[4] http://www.dn.se/nyheter/sverige/sakerhetsbristerna-fick-vara-kvar-hos-msb-i-fyra-ar/ [accessed Oct 23, 2016].

[5] https://en.wikipedia.org/wiki/Botnet [accessed Nov 18, 2016].

[6] https://en.wikipedia.org/wiki/Mirai_(malware) [accessed Nov 18, 2016].

[7] https://www.deepdotweb.com/2016/11/06/analysis-record-ddos-attacks-mirai-iot-botnet/ [accessed Nov 6, 2016].

[8] Ibid, 3.

4 IoT Privacy

4.1 Privacy definition

Privacy is a multifaceted concept and literature provides a broad array of definitions and nuances of the concept of privacy [20, 21, 22]. One of the most influential approaches to privacy associates the concept of information privacy to the control of personal information. Alan Westin is widely credited with the well-known definition of information privacy as:

> "The claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to others." [23]

While control is highlighted as one of the key factors in explaining privacy, this definition is not undisputed, and some researchers in philosophy have noted that the concept of privacy should be distinguished from the notion of control[24]. Further, some literature indicates, "increasing perceived control over the release of private information will decrease individuals' concern about privacy"[25]. The authors argue that this paradoxical behaviour has its roots in the same mechanisms that make people perceive driving safer than flying, in part, based on the misleading feeling that they have more control when driving [25].

Regardless of the definition, privacy threat (the potential risk of loosing control over personal information), is generally one of the major concerns of users and has a significant influence on the adoption level of a new technology. This will probably also be true for the IoT. For instance, an empirical investigation shows that the security and privacy has a significant correlation with the willingness of users to provide personal information to IoT services [26]. In a study about the acceptance of the IoT in the home, respondents consistently demonstrate reluctance to share their data with commercial organizations (notably they are more willing to share information publicly than with commercial organizations) [27].

4.2 Status of privacy

The IoT is the "natural" next step in the evolution of the Internet. Hence, it is plausible to analyse the effect of recent trends in information technology (e.g. social network media, smartphone and big data) on individuals' privacy to foresee the potential influence of the IoT on privacy of the users.

One of the consequences of the current rapid technological developments and globalization is that the scale of the collection and sharing of personal data has increased significantly across all sectors of the society. Furthermore, the number of actors, application areas, storage time of the data, distribution and exchange of the information between actors all show enormous increase. Individuals increasingly make personal information available publicly and globally. Data mining and multiple and further use of the data by actors is reality, huge amounts of data are processed in real-time, cross-border flow of personal data has increased substantially[1]. Some large corporates, as a result of the development in general and their own business strategies, have access to an increasing amount of personal data and thus are able to depict a more complete picture of an individual.

This development is due to digitalization and a gradual change of attitudes towards information processing, both in public administration and in business. The following simplified comparisons, by the *Swedish Privacy Committee*[2], provides a graphic description of the status of privacy:

- Previously organizations had a specific goal in building a personal database. They now have a variety of purposes.
- In the past, they collected information because there was a clear need. Now they gather data in order that it "may be useful"[3].
- In the past, it was important, not least for cost considerations, to keep storage times short. Now it is considered to be of great advantages to retain the data.
- Previously, searching and analyses had a specific objective. Now, big data and data mining is a reality.
- In the past, personal data was gathered through a specific registration. Now, it arises more or less automatically when an individual acts and uses an online service.
- Personal data has become a commodity, collected, traded and sold.

4.3 Privacy by design

To address the privacy concerns of the customer, a joint team of the *Information and Privacy Commissioner of Ontario*, Canada, the *Dutch Data Protection Authority* and the *Netherlands Organisation for Applied Scientific Research*, introduced the concept of *Privacy by Design*[28], in the mid-1990's. According

[1] The General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679): http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf [accessed Dec 27, 2016].

[2] http://www.regeringen.se/rattsdokument/statens-offentliga-utredningar/2016/06/sou-201641/ [accessed Nov 29, 2016].

[3] For instance, Jason Hoffman, head of technology for cloud systems at Ericsson, in an interview with NyTeknik on 2nd November stated: "We should measure everything in the real world. . . . take all data and hopefully make our lives better." (http://www.techsite.io/p/476128 [accessed Nov 03, 2016]).

to this approach, privacy may be accomplished by practicing the following seven foundational principles [29].

1. *Proactive not reactive, preventative not remedial:* It aims to prevent privacy violation events from occurring.

2. *Privacy as the default setting:* It should be the default rule of any IT system. The user is not required to take any action to protect her privacy.

3. *Privacy embedded into design:* It is an integral component of the core functionality of IT systems and is not an add-on part that is bolted to the system afterwards.

4. *Full functionality – positive-sum, not zero-sum:* It avoids trade-offs between different objectives and seeks to achieve all desired (seemingly conflicting) goals (e.g. privacy and functionality) by a "win-win" approach.

5. *End-to-end security – full lifecycle protection:* It is embedded into the IT system prior to collecting the first data record and to the end of the lifecycle of the data, ensuring that information is securely retained, processed and destroyed at the end of process.

6. *Visibility and transparency – keep it open:* It assures all stakeholders that information is managed according to the stated promises and objectives and its components are visible and transparent.

7. *Respect for user privacy – keep it user-centric:* The interests of the users are its first priority. It provides strong privacy by default and chooses appropriate notice measures when required.

The concept of privacy by design is recognized in different recommendations for data protection, by different policy makers and actors, among others by European Union[4]. However, these principles are considered to be vague and to leave many open questions about how they should be implemented when designing a system [30]. Consequently, there have been several attempts to concretize the term privacy by design and exemplify how it should be applied in practice. For instance, the parliamentary committee appointed by the Swedish Government to survey and analyse the risks of privacy (so called *Privacy Committee*), recommends the following principles[5]:

1. *Data minimization:* Prevent privacy risks in a proactive manner by systematically minimizing the amount of collected and processed data.

[4] http://ec.europa.eu/justice/news/consulting_public/0006/com_2010_609_en.pdf [accessed Nov 29, 2016].

[5] http://www.regeringen.se/rattsdokument/statens-offentliga-utredningar/2016/06/sou-201641/ [accessed Nov 29, 2016].

2. *Informed consent:* Terms are presented in an understandable, relevant and transparent way, which gives the user the ability to choose not to share certain information.
3. *Transparency:* Provide users with insight into how their data is treated and used.
4. *Verifiable preventive protection:* Prevent threats by security measures whose effectiveness are verifiable.
5. *Possibility to withdraw consent:* Offer the users the possibility to conveniently withdraw their consent and remove the shared information.

4.4 Privacy threats in the IoT

Potential threats in the IoT are hard to quantify due to the uncertainties of whether and how the IoT will influence society. On the other hand, we can already today observe the effect of collection of large data quantities from e.g., social media, smartphone sensors, and mobile network operators. The implication of the breakthrough of the IoT on privacy could be best studied using the experiences of how these technologies have affected the privacy.

Using this parallel, we can infer that even if the data transmitted by an endpoint device might not cause any privacy issues per se, the accumulated data from multiple devices can still create privacy problems. Furthermore, some characteristics make privacy threats of the IoT more challenging. The data collection process is more passive, more pervasive and less intrusive, which results in that users are less aware that they are being tracked [20].

In the sequel of this section, we render different types of privacy threats and discuss (mostly based on [20]), how the evolution of the IoT may affect these threats.

4.4.1 Identification

The IoT by definition is pervasive, where different devices sense and collect data about the users and their environment to provide some kind of service. The collected data is typically processed at service providers, which are located outside of the users' control. Identification is the threat of associating an identifier (e.g. name, address), with private data about an individual.

Data anonymization, that is, replacing personal information by randomly generated unique IDs, is not sufficient to guarantee the anonymity of the users and it has been shown that identity of the users can be inferred from the anonymized data sets. Renowned examples of such privacy breaches despite anonymization [31, 32] are:

- In Massachusetts, a government agency, Group Insurance Commission (GIC), which purchased health insurance for state employees, released records of every state employee's hospital visits at no cost to any researcher who requested them. To protect patient privacy, the data was anonymized by removing fields such as name, address, and social security number; however, ZIP code, birthdate and gender were not removed.

Sweeney [33] showed that despite removing explicit identifiers from medical data, in most cases, the remaining data (such as gender, date of birth, and ZIP code) was sufficient to re-identify individuals by linking it with public voter database. More spectacularly, Sweeney could reveal and send the health records of the governor William Weld (including diagnoses and prescriptions) to his office. William Weld, then Governor of Massachusetts, had assured the public that the release of the GIC data did not compromise the patient privacy [34, 35].

- A serious privacy breach of AOL search data occurred, when AOL released 20 million (seemingly anonymized) search inquiries online, to engage academic researchers [36]. AOL had assigned each of the users a unique number; however, the information was so detailed and personal that it was possible to reveal some users' identity and compromise their privacy[6].

- In 2006, the online DVD rental company Netflix announced an open competition to develop an algorithm that could improve its movie recommendation system by 10%. Along with the contest, Netflix released a massive training data set to the competitors, consisting of more than 100 million movie ratings, given by around 480000 unique users to 17770 movies. The name of the users and movies were replaced by numerical IDs to anonymize the data set[7]. However, Narayanan and Shmatikov [37] could identify several anonymized Netflix users by comparing the data set with reviews posted on the Internet Movie Database (http://www.imdb.com). The experiment showed that it was possible to identify users' political leanings and sexual references. The scandal led to a lawsuit against Netflix, which ultimately resulted in termination of the second round of the contest in 2010 due to privacy concerns.

In the IoT, new technologies and interconnection of these features and techniques further enlarge the threat of identification [20].

The use of surveillance camera technology, in non-security contexts is an example of such techniques, where customers' behaviour is studied for analysis and marketing (e.g. see [38]). More recently, in 2013, one of the Russia's largest cosmetics chain stores (Ulybka Radugi) started using emotion recognition software as a pilot test to sense the customers' facial expression at the

[6] http://www.nytimes.com/2006/08/08/business/media/08aol.html [accessed May 10, 2016].

[7] https://www.wired.com/2009/12/netflix-privacy-lawsuit/ [accessed May 10, 2016].

checkout counter[8]. The goal of this technology is for Ulybka Radugi to offer customized discounts in real-time for its customers based on how she is feeling and reviewing her purchase history and preferences.

Automated identification of individuals from a digital image or a video frame (*facial recognition systems*) is already in use by law enforcement agencies i many countries.

Speech recognition is widely used in mobile applications and huge databases of speech samples are already being built, which can potentially be used to recognize and identify individuals [20]. The increasing interconnection and vertical communication of everyday things, opens up possibilities for identification of devices through *fingerprinting* [20]. For example, it is shown that it is possible to use the *Radio-frequency identification* (RFID) profile of a person to trace him [39].

To address the problem, attribute-based authentication is suggested to minimize the data a device communicate in the IoT, and maintain control over the disclosure of data and improve user privacy in the IoT [40].

4.4.2 Localization and tracking

Localization and tracking are the threats of determining and recording a person's location through time and space by different means, e.g. cell phone location, Internet traffic, or GPS data [20].

The availability of vast and detailed spatial and spatiotemporal data, which has become possible due to data collection techniques, such as global positioning systems (GPS), high-resolution remote sensing, location-aware services, and Internet-based volunteered geographic information has led to an increasing interest in using geographic data and incorporating spatial information and analysis.

However, this data has become more diverse, complex, dynamic, and much larger than before and therefore is more difficult to analyse and understand. The emergence of the field of spatial data mining and knowledge discovery tries to address these difficulties by developing theory, methods and practice for the extraction of useful information and knowledge from massive and complex spatial databases [41].

One rich source of geo-location information is the mobile network operators' *call detail records* (CDRs). CDRs are data records containing detail information about telephone calls and other communication services (e.g. text message), which are automatically collected by mobile network operators, primarily for billing, accounting purposes and network traffic monitoring. CDRs consist of metadata (i.e. data about data) and do not include the content of the communications. Typical attributes included in a CDR are, phone numbers of the source and destination of a call, starting time (and date) of the

[8] http://www.humintell.com/2013/08/emotion-recognition-software-that-helps-you-shop [accessed May 10, 2016].

call, its duration, type of the call (voice, message, etc.). The information in CDRs is considered highly personal and people usually expect that it will not be disclosed to any third party by the network operators.

However, during the past decade, researchers have been provided access to a quite large amount of data, among others CDR data sets for different research and development projects [32]. In order to maintain the users' privacy, the CDR data sets are anonymized, that is, each phone number is replaced by a randomly generated unique ID, before they are transferred to a third party. Nevertheless, as discussed in section 4.4.1, such methods cannot guarantee anonymization and protect sensitive personal data.

Web 2.0 (i.e. social networking sites, social media, blogs, wikis, RSS feed, apps, etc.)[9] has changed the role of the user on the web, from consumer to producer of information. Smartphones, equipped with GPS sensors have allowed users to geo-locate themselves. Integration of Geographic Information Systems (GISs) and social networks has resulted in so-called Location Based Social Networks (LBSN), that is, social networks that include location information into their contents.

This capability has made real-time urban sensing possible. Urban sensing uses citizens as active and passive sensors with the goal to gain insights in human behaviour in the city. Possible use scenarios are imaginable, for example, community healthcare, public safety, city resource management and transportation management [42]. The data sources that can be inferred for urban sensing are heterogeneous and originates from three data sources: (a) mobile sensor data, (b) infrastructure sensor data, and (c) social data from social network and other Internet services [43]. Data sources can be used independently but the combination of data from different sources provide a comprehensive understanding of individual and group behaviour, social interactions, and community dynamics [43, 42].

Localization in the adjacent surrounding usually is not perceived as a privacy threat as we are used to observing others and being observed by other people when we are in their field of view. Localization is experienced as a threat mainly when this information is recorded, processed and stored without the permission and control of the subject. As with other privacy concepts, the lack of control is central to the concept of *location privacy*, which is defined as "the ability to prevent other parties from learning one's current or past location" [44].

By emergence of the IoT, several factors would presumably exaggerate the privacy threat of the localization: (i) expansion of location-aware applications and improvement of their accuracy, (ii) the ubiquity of data collection technology and process, and (iii) interaction with IoT devices that register the identity, location and activity of the user.

Current research on location privacy, such as *trusted third party*, *peer-to-peer*, etc. mainly deals with location-aware applications in smart phones and

[9] https://en.wikipedia.org/wiki/Web_2.0 [accessed Aug 12, 2016].

does not encompass location privacy threats in the IoT. Ziegeldorf et al. identify three main challenges: (i) addressing threats of pervasive data collection, (ii) how to control shared location data, and (iii) privacy-preserving protocols for interaction with IoT systems [20].

4.4.3 Profiling

Profiling is the practice of collecting and processing data about individuals' activities (e.g. sites visited, products purchased, product pages viewed, and emails sent) over long periods in order to categorize them according to some key feature. The information usually is collected with little or no consent of users, and combined with other personal data to create a more comprehensive profile. Profiling is used currently in a large spectrum of domains, for example, e-commerce, targeted advertising and credit scoring [20, 45].

Profiling poses several potential privacy risks. Users demonstrate concerns[10] associated with unsolicited marketing, invisibility of data collection process, and the risk that undesired third party access the data. In recent years, data marketplaces have emerged, which trade data that have been collected from a variety of sources, aggregated, enriched and processed.

Another risk associated with profiling is that personal information may be revealed to other users, as other users who share the same computer and browser may view one's targeted advertisement (due to cookies saved on the computer and depending on the settings). Moreover, many users are disturbed by the mere awareness of being watched and tracked [45].

Several cases of privacy violating profiling efforts have been reported, for instance Facebook's racially discriminatory[11] and Google's gender-discriminatory[12] advertisings.

By evolving the IoT, data collection increases quantitatively by orders of magnitude, due to the explosion of data sources and connected devices. Moreover, data will change also qualitatively as data is collected from previously inaccessible parts of people's private lives [20], for example, data collected by wearables and different devices at home.

4.4.4 Lifecycle transitions

This type of privacy threat refers to disclosure of private information where the owner of a consumer product is changed during its lifecycle (e.g. photos and other private information on a second-hand smartphone or computer) [20].

[10] Among consumers who show a medium degree of privacy concern, people can be clustered into two different groups: (i) those who are more "identity aware" (i.e. users who worry more about sharing e-mail address, physical address or phone number), and (ii) those who are more "profile aware" (i.e. users who do not wish to share their hobbies, age, interests, or preferences) [46].

[11] http://fortune.com/2016/10/28/facebook-ad-propublica-race [accessed Oct 28, 2016].

[12] http://www.theverge.com/2015/7/7/8905037/google-ad-discrimination-adfisher [accessed Oct 28, 2016].

Since currently, consumer products that hold private information (e.g. smartphones, cameras and laptops) are mostly under the control of the same owner during their entire lifecycle, this problem is not observed very often. However, as more and more everyday things will be connected and will contain private data (e.g. smart homes, connected cars, etc.), the risk for privacy disclosure due to change of owner will increase.

4.4.5 Inventory attack

Inventory attacks are related to illegitimate gathering of information about the existence and characteristics of things in a specific place (e.g. household, office, or factory) [20]. Inventory attacks can usually be performed by using the fingerprint[13] of IoT devices, for instance, their communication speed, reaction time, and so on. Assuming that the promise of the IoT will be fulfilled, all smart things will be addressable over the Internet, opening the opportunity for unauthorized entities to exploit this and create an inventory list of things belonging to a target.

An inventory attack could be used for profiling individuals, since owning special items disclose private information about the owner. For instance, books, movies or music reveals personal interests and medicine or medical devices expose one's health state [20].

4.4.6 Linkage

Linkage threat refers to uncontrolled disclosure of information due to combination of separated data sources and linking different systems [20]. The combined information about an individual provides a much more detailed portrait of her. This occurs because aggregating information creates synergies. Aggregated information can reveal new facts, which the owner did not expect would be known about her, when the original isolated data was collected [48]. The revealed combined information does not need to be truthful to be conceived as a privacy breach. On the contrary, many users fear poor judgement and loss of context when data that have been gathered from different parties under different contexts and permissions are combined [48, 20].

As Solove [48] suggests, aggregating information is not a new activity and it has always been possible to combine separate pieces of personal information, to infer something new about an individual. What makes it different, nowadays, is that the aggregation's power and scope are different now; the volume of collected data about people is significantly more extensive, and the process to combine and analyse it is much more powerful.

The threat of linkage will deepen by the IoT development, for two main reasons: (i) the integration process will link system from different companies

[13] Fingerprinting is defined as the measuring of an identifying characteristic of an individual, or a physical or digital item [47].

and organizations to build a heterogeneous distributed system-of-systems delivering new services that no single system could provide on its own, and (ii) the linkage of systems will make data collection in the IoT even less transparent than what it already is expected to be [20].

4.5 Privacy-preserving data mining

Privacy-preserving approaches are efforts to prevent information disclosure generally due to legitimate access to the data and differs from data security measures (e.g. access control and encryption). While the aim of conventional data security is to prevent information disclosure against illegitimate actions such as hacking, access control violation, query-injection and theft, privacy-preserving methods attempt to prevent identification, profiling and linkage as a result of legitimate operations on the data.

Collection and analysis of the data to provide service to the user is one of the central promises of the IoT, and the main challenge in the IoT privacy is to balance data gathering and analysis with the users' privacy requirements [20].

In recent years, to address the privacy requirements of the users in the context of the big data and social online networks, several privacy-related fields have emerged.

Privacy-preserving data mining is a common name for different approaches with the aim of retrieving valid data mining results without learning the underlying data values. Although, the field has been receiving much attention in privacy research community and beyond, its meaning is not entirely clear [49].

As the name suggests, privacy-preserving data mining is primarily focused on privacy issues in databases and is related to data mining. However, since it is assumed that many applications based on data mining will provide intelligent services to the IoT, the field is highly relevant to the privacy in the IoT. Depending on the application area, settings and the goal, different privacy-preserving methods have been constructed.

4.5.1 Group-based anonymization

It is obvious that prior to publishing a data set, all unique identifiers such as social security and driving license numbers have to be removed from records to mask the identity of the individuals in the database. However, as examples discussed in section 4.4.1 show, combination of non-unique information such as gender, age and ZIP-code (so called quasi-identifiers) can be used to identify an individual uniquely. For instance, the combination of the date-of-birth and place-of-birth might be enough to re-identify a certain individual if the place-of-birth refers to a small village [50].

Therefore, different types of anonymization methods are introduced to construct groups of anonymous records, which are transformed in a group-specific way [51]. In the following, some of these methods are briefly discussed.

4.5.1.1 k-anonymity

Assume the owner of a database containing patient information would like to share parts of the database with researchers and aims to make it difficult that the published records can be tracked back to specific individuals. After removing all (explicit) identifiers, there is still a considerable risk to associate a record with a specific individual by using quasi-identifiers. A method to mitigate this risk is k-anonymization, which means reducing the granularity of the data in such a way that for each record, there are at least $k - 1$ other records that have the same values for their quasi-identifiers [52]. Attacks using combinations of quasi-identifiers are prevented by k-anonymization, since it assures that no individual can be identified with a certainty exceeding $1/k$.

While k-anonymization reduces the risk of sensitive data disclosure, it is still possible for attackers to make inferences about an individual that is known to be in the database. An example is *homogeneity attacks*, in which a lack of diversity among sensitive attribute values is exploited by the attacker (e.g. consider a case where an equivalence class of k records all have the same value for the sensitive data).

Moreover, k-anonymization is not a suitable method for high-dimensional data sets and in many cases, the level of information loss required in order to preserve k-anonymity (even for very small values of k) is so high that the data set is not useful for any data mining purposes [53].

4.5.1.2 l-diversity

The l-diversity model has been suggested as an extension to k-anonymity to handle vulnerability of the k-anonymization against homogeneity attacks. Homogeneity attacks, as above remarked, target cases where some sensitive values for a set of k records in a group are identical (or show little variation). To address this type of attacks, the l-diversity method is proposed which not only maintains the minimum group size of k, but also tries to provide a diversity among the sensitive attributes. In l-diversity, each group of k records with the same set of non-sensitive values must contain at least l “well represented” values for each sensitive attribute [51]. A linear-time algorithm for creating tables that obey the l-diversity privacy requirements is presented in [54].

Although l-diversity protects data against homogeneity attacks, it is not immune against attacks based on the distribution of sensitive values. Li et al. [55] show that l-diversity is insufficient to prevent attribute disclosure by two types of attacks: (i) *skewness attack* where the overall distribution of the sensitive data is strongly skewed (and known), and (ii) *similarity attack* when the sensitive attribute values in a group are distinct but semantically similar.

4.5.1.3 t-closeness

The t-closeness method [55] is an enhancement of the l-diversity approach with the objective of protecting against skewness attack. The main idea in t-closeness method is that the distance between the distributions of the sensitive attribute within each anonymized group should not differ from the global distribution of the same sensitive attribute, by more than a threshold t.

4.5.2 Distributed privacy-preserving data mining

Distributed methods for privacy-preserving data mining concerns scenarios, where several participants (database owners) wish to collaborate with each other and use data mining algorithms to compute aggregated statistics, however, they do not fully trust each other. Therefore, they are not inclined to share the data and compromise the privacy of the individual data records.

The data sets may be partitioned either horizontally or vertically. In horizontally partitioned data sets, each participant has data records with the same attributes, however, the data is spread between them and each owner has some partition of the data (e.g. two cities each having a database over both electricity and water consumption of their own population). In vertical partitioning, each participant may have different attributes of the same set of records (e.g. a water supplier having a database over water consumption and an electricity provider having a database over electricity consumption of the same individuals).

Distributed privacy-preserving data mining is closely related to *secure multiparty computation (SMC)*, which is a subfield of cryptography focusing on constructing algorithms such that several participants can jointly compute a function over their inputs while keeping those inputs private[51].

4.5.2.1 Algorithms for horizontally partitioned data sets

In order to illustrate, how these types of algorithms work and how they can possibly be used in an IoT context, consider a hypothetical scenario, where four IoT devices, a, b, c and d (e.g. belonging to four different individuals), cooperatively consume some critical resource. They are all interested to compile the (aggregated) instantaneous consumption of the resource in order to optimize their behaviour. None of the participants is, however, willing to reveal its own consumption of the resource, since it can provide the others with sensitive information about the owner's lifestyle. The problem can be solved with *homomorphic encryption techniques.*

Homomorphic encryption is a form of encryption, which allows specific types of operations to be performed on encrypted data and has the desired property that decrypted result matches the result of operations performed on the original data. For example, an additive homomorphic encryption scheme with the addition operator $\oplus$ is a scheme that satisfies

$$De(En(M_1) \oplus En(M_2)) = M_1 + M_2, \tag{4.1}$$

for all values of M_1 and M_2.

Given the above prerequisites, participant a generates a private and a public encryption key and sends the public key to participants b, c and d. Participant b uses the key to encrypt the value of its own consumption of the resource, $En(R_b)$, and sends the result to participant c. Participant c encrypts the value of its own consumption, $En(R_c)$, and computes $En(R_b) \oplus En(R_c)$. This value is sent to participant d, which in turn computes $En(R_b) \oplus En(R_c) + \oplus En(R_d)$ and sends it to participant a. Participant a uses its private key to decrypt the received encrypted value and computes $R_b + R_c + R_d$ using 4.1. Participant a adds its own consumption R_a to this value and broadcasts the total value of the consumption of the resource to participants b, c and d. Using this algorithm, no participant obtains more information than the total consumption of the resource. The sequence of messages between entities is shown in Figure 4.1.

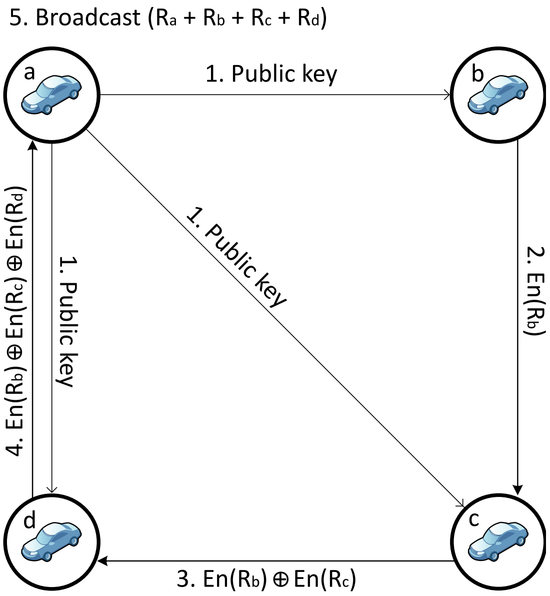


Figure 4.1: A schematic diagram over how encrypted messages flow between entities. No information other than the sum of consumed resource is disclosed to any entity.

This approach can be generalized across k participants. Moreover, since many data mining algorithms can be expressed as repetitive computations of different primitive functions such as addition, scalar product, etc., one can generalize the above method to design homomorphic encryption schemes for privacy-preserving data mining for horizontally partitioned databases and exchanging aggregated information without compromising privacy[51].

4.5.2.2 Algorithms for vertically partitioned data sets

A data set is considered vertically partitioned when different attributes associated to the same individuals are distributed between different parties that do not wish (or are not allowed) to share their data with other participants. However, they are permitted and interested in discovering interesting relations between the attributes. In general, IoT devices are not directly involved in such operations; nevertheless, service providers will be likely inclined to perform such operations on data gathered by different parties.

For instance, consider a hypothetical scenario where a municipal water provider wishes to investigate whether there is a correlation between the amount of consumed water and the type of dishwasher owned by households, assuming that the latter information is collected by the electricity provider, using a service that targets connected dishwashers. Privacy-preserving data mining algorithms for distributed databases with vertical data partitioning can be used to compute the average water consumption of households owning different types of dishwashers without disclosing the type of dishwashers or water consumption of each household. We outline an algorithm, which uses a *commutative encryption scheme.* A commutative encryption is an order-independent encryption, that is, it satisfies

$$En_{k1}(En_{k2}(M)) = En_{k2}(En_{k1}(M)), \tag{4.2}$$

for all values of encryption keys $k1$ and $k2$ and all messages M. Commutative encryptions further fulfils

$$De_{k2}(En_{k1}(En_{k2}(M))) = De_{k2}(En_{k2}(En_{k1}(M)) = En_{k1}(M), \tag{4.3}$$

and

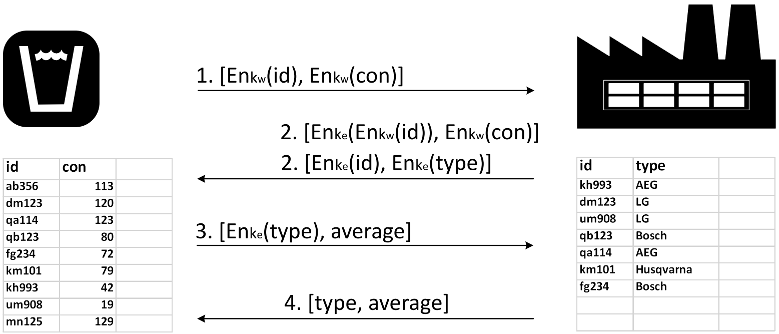
$$M_1 \neq M_2 \implies En_{k1}(En_{k2}(M_1)) \neq En_{k1}(En_{k2}(M_2)). \tag{4.4}$$

We assume that the water provider holds a set of tuples, $[(id, con)]$, where id is a unique global identifier for households and con is the water consumption of the corresponding household. The water supplier encrypts records in the database with its secret key k_w and sends the encrypted data to the electricity provider (i.e. $[(En_{k_w}(id), En_{k_w}(con))]$).

The electricity supplier encrypts only the ids in the received data using its own secret key, k_e, and returns the double encrypted ids and single encrypted water consumption values back to the municipal water provider (i.e. $[(En_{k_e}(En_{k_w}(id)), En_{k_w}(con))]$). Alongside with this data, the electricity provider, also encrypts and sends its own database to the water provider. The latter is the set of the tuples $[(En_{k_e}(id), En_{k_e}(type))]$, where id is the same unique identifier for households and $type$ is the type of the dishwasher the household possesses.

The water provider encrypts the encrypted ids in $[(En_{k_e}(id), En_{k_e}(type))]$ by its own key and composes $[(En_{k_w}(En_{k_e}(id)), En_{k_e}(type))]$. By comparing the double-encrypted ids, the water provider can now join the two tables $[(En_{k_e}(En_{k_w}(id)), En_{k_w}(con))]$ and $[(En_{k_w}(En_{k_e}(id)), En_{k_e}(type))]$, and create the table $[(En_{k_e}(En_{k_w}(id)), En_{k_e}(type), En_{k_w}(con))]$.

The water provider is now able to create the table $[En_{k_e}(type), con)]$ by decrypting $En_{k_w}(con)$ using its own key and calculate the mean value of water consumption for each type of the dishwashers, albeit the types are encrypted by k_e. The water provider sends the table $[(En_{k_e}(type), average)]$ to the electricity supplier, which decrypts the values for dishwasher types and sends back the table $[(type, average)]$ to the water provider. Both the water provider and electricity supplier have now access to a table consisting of the mean of water consumption for each dishwasher type, while no further information about the households has been disclosed. Figure 4.2 illustrates the sequence of the messages between the two entities.



id	con	
ab356	113	
dm123	120	
qa114	123	
qb123	80	
fg234	72	
km101	79	
kh993	42	
um908	19	
mn125	129	

id	type	
kh993	AEG	
dm123	LG	
um908	LG	
qb123	Bosch	
qa114	AEG	
km101	Husqvarna	
fg234	Bosch	

Figure 4.2: The flow of encrypted messages between a water provider, at the left side and an electricity supplier, at the right side of the figure. No information other than the average of water consumption of households having the same type of dishwasher machine is disclosed to any entity.

This approach can be generalized to other data mining rules (see e.g. [51] for a detailed discussion on privacy-preserving data mining algorithms and techniques).

4.6 Differential privacy

Data anonymization approaches such as k-anonymity, l-diversity and t-closeness, discussed earlier in section 4.5.1 are considered *non-interactive* publishing approaches, in which the data owner releases an anonymized data set (or a subset of that) to the public at once. However, as it has been shown, anonymization is not a trivial task and all these methods are, in some degree, prone to privacy breaches.

Another privacy-preserving approach that in recent year has received much attention is *differential privacy*, which is related to the *interactive* query model. In the interactive approach, the data owner does not release the anonymized data but provides a statistical database, where the analyst can run a sequence of queries and retrieve information without being able to compromise individual private data.

Differential privacy is a mathematical model of privacy invented by Cynthia Dwork [56] that quantifies the individual privacy loss in a statistical database while aggregate information about the data is released. In this framework, the goal is that by adding noise to the database ensure that the amount of any privacy loss and the ability of an adversary to cause harm remains the same for any individual independent of whether she opts in to, or opts out of, the database[14].

> Definition: A randomized function K gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one row, and all $S \subseteq Range(K)$,
>
> $$Pr[K(D_1) \in S] \leq e^{\epsilon} \times Pr[K(D_2) \in S]. \tag{4.5}$$

Using a mechanism K that fulfils this definition to provide answers to queries, will theoretically ensure any participants that the leakage of her personal information regardless of presence or absence of her data in the database will be significantly the same [56].

Note, that this definition does not describe how to achieve ϵ-differential privacy. However, departing from the definition, Dwork designs [56, 58] several differentially private mechanisms, which are in essence, calculating how to add just as required random noise to the database to hide the presence or absence of a single individual, while it is still possible to retrieve accurate aggregate information from the database. For example, the Laplace mechanism adds Laplace noise (i.e. random numbers generated from a Laplace distribution[15]) to the function.

Two desirable properties that are both satisfied by differential privacy are: (i) composability, meaning that if we query an ϵ-differential privacy database t times, then the result would be $t\epsilon$-differentially private, and (ii) robustness

[14] The underlying idea can be rooted back to the notion introduced by *Tore Dalenius*, in 1977 articulating a desideratum for statistical databases: "access to a statistical database should not enable one to learn anything about an individual that could not be learned without access" [57]. Dwork shows that although this characteristic is desirable, it is not achievable and suggests a new measure, differential privacy, which quantifies the increased risk to one's privacy as a result of participating in a database.

[15] https://en.wikipedia.org/wiki/Laplace_distribution [accessed Dec 12, 2016].

to auxiliary information, that is, differential privacy is independent of what auxiliary information is available to the adversary.

4.6.1 Attacks on differential privacy

Differential privacy is one of the most robust privacy models known so far; however, several attacks on differential privacy have been discussed in literature.

Haeberlen et al. [59] point out side-channel attacks on differential privacy systems considering information leakage from the system through characteristics of how it operates, e.g. long or rejected response. They describe several different kinds of covert channel attacks[16] (on differential privacy systems, showing these systems can be exploited by adversarial queries [59]. Although, these vulnerabilities are potentially serious, they are related to specific differential privacy systems, that is how they are implemented, and cannot be considered as a critique of the model per se. However, quite recently, some researchers have raised concerns about the differential privacy model itself. Liu et al. argue that the privacy guarantee of differential privacy relies on the assumption of independence of tuples in the database. This assumption generally is not fulfilled in real-world databases where various natural dependences between users can lead to degradation in expected privacy levels [60]. They demonstrate an inference attack, where an adversary uses the probabilistic dependence between tuples to extract users' sensitive information from differentially private query results; violating the differential privacy guarantees. Liu et al. introduce the notion of *dependent differential privacy (DDP)*, and propose a mechanism that takes into the account the dependence between tuples and achieve privacy guarantees in DDP [60].

It should also be emphasized, that in practice, every data publishing scenario has its own assumptions and requirements of the data publisher, the data recipients, and the data publishing purpose. For instance, there are two models of data publishers: (i) *untrusted* model, in which the data publisher is not trusted and may identify sensitive information about the record owners, and (ii) *trusted* model, in which the record owners trust the data publisher and are willing to provide their personal information to [61]. The differential privacy model is based on the latter assumptions. However, trustworthiness is not an immutable perceived property and may change of different reasons during time. Moreover, even if the data collector/publisher is well-intentioned and trustworthy, intrusions and thus privacy breach can occur for other reasons, including hacking (e.g. see [59]), subpoena, or *mission creep* (i.e. expansion of a project or mission beyond its original goals) [62]. These concerns, has prompted researchers to seek methods for learning from data without saving the data, something that is briefly discussed in the next section.

[16] https://en.wikipedia.org/wiki/Covert_channel [accessed Nov 08, 2016].

4.7 Pan-private streaming model

The main idea behind *pan-privacy* model is that there should be no tempting target and the data should never be saved. It is assumed that the data collector is trusted, however, the raw data should not be saved because the entity holding information may be subject to compulsory non-private data release (e.g. due to a subpoena), be purchased by another possibly less trustworthy entity, or be hacked. Note that in the two first cases (i.e. subpoena and purchase), intrusion is known, but in the latter case (hacking) the intrusion is unknown. Thus, the goal of a pan-private mechanism is to hold the internal state of the algorithm differentially private even against an adversary that can observe the algorithm's internal state on rare occasions [62].

Note that an streaming algorithm that after updating the internal state discards the input, in general, does not provide privacy against intrusions. For instance, consider the *counter problem*, in which the input is a stream of zeros or ones (for instance 10010011000100001 . . .) and the goal is to output a publicly observable counter that approximates, with reasonable accuracy, the total number of 1's, while protecting *individual* increments. The administrator, reads each x_i, updates its internal state to $\sum_i x_i$ and outputs the value observable to the adversary. Evidently, as can be seen in Figure 4.3, output 1 is sufficient to expose the value of all increments in the input. Nevertheless, by adding an independent Laplace noise (i.e. random numbers generated from a Laplace distribution), one can protect each increment and achieve differential privacy of the internal state and the output. For instance, consider output 2 and 3 in Figure 4.3, where with each input, a Laplace noise (with $mean = 0$, and $scale = 1$ and 2 respectively) is added to the accumulated sum.

Note that the administrator, accumulates statistical information, and never stores data about individual inputs. The accuracy of the information is obtained (to some extent), since the noise cancels out[17]. Research is ongoing to develop new methods that increase the accuracy of the statistics without compromising the privacy.

4.8 Visual privacy protection

Advances in computer vision technologies and development of indoor monitoring systems, which can be used for assisting a rapidly aging world population, will probably boost the use of IoT devices at home. These systems are able to automatically interpret visual data from the environment and provide home help services for the elderly. The use of indoor monitoring systems, poses a new threat to individual's privacy, implying the significance of visual privacy protection techniques. Literature on *privacy-aware monitoring system* and

[17] The example is taken from Moni Naor, Institute for Advanced Study video lecture, November 23, 2009, available at: https://video.ias.edu/csdm/dynamicdata [accessed 14 Dec, 2016].

Input

1	0	0	1	0	0	1	1	0	0	0	1	0	0	0	0	1

Output 1: counter without noise

1	1	1	2	2	2	3	4	4	4	4	5	5	5	5	5	6

Output 2: counter with Laplacian noise, scale = 1.0

0.45	0.38	-2.30	0.21	1.20	2.31	2.86	3.32	3.34	2.22	5.55	9.47	9.80	11.70	8.88	8.87	11.14

Output 3: counter with Laplacian noise, scale = 0.5

0.72	0.69	-0.65	1.10	1.60	2.15	2.93	3.66	3.67	3.11	4.77	7.23	7.39	8.34	6.94	6.93	8.56

Figure 4.3: Three different continual outputs of a counter: output 1 reveals the input completely. Outputs 2 and 3 preserve the secrecy of individual input values, while providing some degree of accuracy.

privacy-preserving photo sharing serves an opening to the privacy in the IoT realm.

Padilla-López et al. [63] provide a literature survey on several protection methods for visual privacy and existing privacy-aware monitoring systems. The authors identify five categories of privacy protection methods.

- *Intervention* methods that physically prevent the acquisition of an image by interfering with the camera's optical lenses (e.g. by directing pulsing light at detected cameras).

- *Blind vision* method deals with image or video processing in an anonymous way using *secure multi-party computation* (SMC) techniques applied to vision algorithms. SMC is a subfield of cryptography with the goal of creating methods for parties to jointly compute a function in such a way that their inputs and are not revealed (see section 4.5.2). Using blind vision techniques, vision algorithms among several parties can be performed in an anonymous manner.

- *Secure processing* refers to other methods that process the visual data in a privacy-preserving manner, but are not based on SMC. This category contains several approaches, for example *private content based image retrieval*, which makes it possible to query an image database by a sample image, without revealing the content of the query image to the database [64], and matching encrypted images using *phase-only correlation* [65].

- *Redaction* is modifying the sensitive parts of an image such as faces, number plates, etc. to protect the private information of the subjects. This can be achieved by using common image filtering (e.g. *blurring* or *pixelating*), concealing region of interest by *image encryption*, *face de-identification*, *object removal*, or *visual abstraction.*

- *Data hiding based methods* are approaches in which like redaction methods region of interest are modified, however the hidden information is also embedded inside the modified version so that the original images can be retrieved if required. Data hiding methods are similar to *digital steganography*[18] technique.

4.9 Data protection legislation

We conclude this chapter by a brief remark on privacy legislation, which presumably has a significant effect on how IoT privacy will be shaped.

Most countries do not have any regulation targeting IoT systems specifically, so general privacy laws for data collection, processing, and dissemination apply to the IoT. These privacy regulations vary between different countries, even within the EU. However, EU's General Data Protection Regulation (GDPR)[19], effective in May 2018, aims to unify the regulation for data protection within the Union.

Even if the GDPR is a compromise between various interests and does not go as far as some privacy advocates might wish, it is considered as a milestone in the data protection regulation and a step in the right direction. According to the GDPR fact sheet[20], by the new regulation, users can expect greater control over their personal data and corporates have to comply with more stringent requirements and build data protection into their systems from the very beginning of the design process.

A special Eurobarometer on data protection[21], which was carried out in 2015, shows that an overwhelming majority of Europeans are concerned about

[18] Steganography is an old practice used to hide (possibly encrypted) messages inside a cover message or image. The advantage of steganography over cryptography alone is that the intended secret message does not attract attention to itself and only the recipient is aware of its existence. The embedded message is recovered using a secret key (https://en.wikipedia.org/wiki/Steganography [accessed Dec 27, 2016]).

[19] The EU Regulation 2016/697, GDPR, is a regulation passed on 27 April 2016, which will enter into force on 25 May 2018 after a two-year transition period. It will replace, the Data protection Directive, and unlike the directive, does not require any enabling legislation to be passed by national governments (http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf [accessed Dec 27, 2016]).

[20] http://europa.eu/rapid/press-release_MEMO-15-6385_en.htm [accessed Dec 27, 2016].

[21] http://ec.europa.eu/public_opinion/archives/ebs/ebs_431_en.pdf [accessed Dec 27, 2016].

how their data is collected and processed by companies. For instance, more than eight out of ten respondents feel that they do not have complete control over their personal data and two-thirds of respondents are concerned about not having complete control over the information they provide online. In GDPR, these concerns are addressed through:

- Right to be forgotten, i.e., when an individual no longer wants her/his data to be processed, the data will be deleted.
- Easier access to one's data, i.e., individuals will have more information on how their data is processed.
- The right to know when one's data has been hacked, i.e., companies and organizations must notify the national supervisory authority and (if required) the data subject of data breaches.
- Data protection by design and by default, i.e., data protection safeguards will be built into products and services from the earliest stage of development.
- Stronger enforcement of the rules, i.e., data protection authorities will be able to fine companies who do not comply with EU rules up to 4% of their global annual turnover.

5 Security and Privacy in Cloud-Supported IoT

Cloud-based services are often considered as the natural infrastructure of the IoT that provide support for data storage, data processing and data sharing [66]. Therefore, it is justified to study security, privacy and personal safety risks in the context of cloud-based services.

Singh et al. [66] distinguish twenty security and privacy concerns that are beyond particular application domains, and are from the perspectives of cloud providers and end-users across the range of IoT technologies. In the following a brief discussion on some of these concerns are given. The reader interested in security and privacy issues of IoT-cloud is referred to [66], for a more detailed discussion.

5.1 Accessing the cloud

The flow of data from IoT devices to the cloud (either for storage or processing), and the flow of data (including actuating commands) to the devices are sensitive. Therefore, secure communication that prevents unauthorized access to data is required to both counter wiretapping and protecting data from tampering. Current technologies for secure communication of data and encrypting data by applications and modifications of these techniques are used to achieve confidentiality and integrity of data. However, this approach can be a complex endeavour, considering the number of participants in IoT systems and the dynamic nature of such systems, and does not scale well.

Another related concern is the access control for IoT-cloud. Access controls serve as a means to manage the privileges of users, that is only those that have the right and are authorized have access to appropriate data and services. For example, this can be implemented by imposing *access control lists*. A security challenge is that the IoT often involves interactions between IoT devices that have not been connected earlier. Moreover, the IoT may require more flexible and context-dependent access control policies, for instance, it may be desirable that in acute medical circumstances personal devices reveal information about the owner's health state, something that is not appropriate otherwise [66].

5.2 Data management within the cloud

Identifying potentially sensitive data is a crucial part of IoT security. It is important to note that it is not sufficient to identify devices that produce the data, to determine whether the data is sensitive. Hence, it is important to design security mechanisms in a way that resolve the sensitivity of the data in the current context.

However, the benefits from the IoT vision depends on information sharing between devices, and if the data gathered by each device is isolated, the wider vision of the IoT will be lost. While the current cloud design is focused on strong protection without sharing, the requirements of the IoT are both data protection and sharing. The main concern here is to reassure that personal data is protected and shared as the user specifies, even when it has been uploaded to the cloud.

A way to ensure protection against leakage of data by the cloud provider due to bugs, malicious insiders, or misconfiguration is to enforce encryption by IoT device. One problem here is that then IoT devices have to deal with all aspects of security, data and key management, which can be complex and hinder scalability. Another problem is that encryption by IoT devices, limits the cloud provider's capability to supply data processing and analytics services. A way to address this problem is using homomorphic encryption techniques (discussed in section 4.5.2.1), which enable computation on encrypted data, without access to the intelligible data. However, these methods are not currently mature enough to be used in practice [66].

Data generated or produced by IoT devices may have different level of sensitivity. While some data such as location of a personal device are intrinsically sensitive, others may be harmless per se. However, even if an individual data stream is not harmful, the combination of data with other information may compromise security and privacy of clients. The general problem is that it is impossible to foresee all imaginable security risks and privacy concerns that may appear from different combination of data produced by all devices. Anonymizing the data is often not sufficient to prevent re-identification of attributes (see section 4.4.1). This is increasingly meaningless as the amount and variety of available information about individuals grows exponentially [67].

5.3 Identity management

The IoT adds more complexity to identity management. One problem is that some devices could be shared between different individuals. For example, consider a home monitoring and control system, which should identify residents of a house to apply user-specific policies. However, different family members may have different conflicting preferences (e.g. about appropriate room temperature), thus some conflict detection and resolution mechanisms are required [66].

5.4 Managing scale for the IoT-cloud

The IoT produces a huge amount of data, enlarge communication requirements, and increase the number of clients with which an IoT-cloud should interact. This may lead to availability problems, and consequently failure in coordination of the devices. Another issue is how to manage the log information for devices and where to save them. Logs are essential for controlling the compliance with regulations and contracts [66].

5.5 Malicious devices

The IoT implies a drastic increase of the number of connected devices, which vastly increase opportunity of denial-of-service (DoS) attacks (see also chapter 3).

An essential problem is how to determine which devices have been compromised. Among other methods, analysing the data outputs and patterns of behaviour or reputation of an IoT device (i.e., trust value assigned by other nodes) are suggested methods to address the problem [66].

6 Discussion

6.1 IoT security

The Stuxnet example is a military, targeted operation that we do not see many of. However, it is extremely difficult to protect oneself against an opponent with large resources of this magnitude.

What one can at least say for sure is that this sort of operation has a time span of several years in which a long time is used for planning. One can also conclude from the FOI's own investigations that the code quality seems to have been degraded in more recent versions, probably due to a short development time of perhaps six months. The most of the sloppy code was added after this period, resulting in an uncontrolled spread of the worm, which presumably was not the intention.

Although, these centrifuges cannot be classified as IoT, it indicates that the antagonistic threat to embedded systems is not new. If it is possible to conduct such sophisticated attacks on air-gapped systems, we can argue a fortiori that the IoT will experience similar threats. Our tools tend to be more and more complex and that makes them more vulnerable to antagonistic threats unless protection against attacks is built in, which it seldom is.

Are we heading for disaster in all areas? Adequately, a regular practice of the developers of control systems is that they assume that systems can fail in different ways and therefore they provide redundant safeguards. Although it might not be possible to protect a system against all deliberate tampering, it must be generally difficult to manipulate a well-designed system in a "dangerous" manner. Hacking is just one more way that a control system can be broken. A relatively clear example of this approach is described in a report on traffic light controller [68]. The system that the authors study has several fully open lines of communication by which an attacker can easily manipulate the system. With one exception. Right in front of the traffic lights there is a separate unit called the MMU, Malfunction Management Unit, whose task is to not allow dangerous light combinations. It is relatively easy to obtain the advantage of green wave, but not to manipulate the system to display green in all directions simultaneously.

At least in larger systems that can fail in harmful ways, the protection level is usually at a higher level than in household appliances and they are often rather difficult to break. This is probably the main reason that we so seldom see any hacking terrorism. It is usually much easier to make a statement in more traditional ways.

6.2 IoT privacy

Smartphones and social network media, which are ranked as the first and second top 11 technologies of the first decade of the 21st century [69], have already changed our lives in a significant way[1]. In our pockets, we carry various embedded sensors (GPS, accelerometer, gyroscope, microphone, camera, and Bluetooth), as well as a traditional telephone. The wealth of (mostly free) available applications that creatively use these sensors and users' information has widespread social implications [70].

The IoT evolution continues to add billions of new sensors and devices to the Internet, generating an enormous amount of information about people, including their locations, connections, shopping records, financial transactions, pictures, voices, conversations, health state, etc., with or without explicit consent. This data will be further processed, stored, analysed and will be treated as a commodity (data is already proclaimed to be the new oil[2]).

What are the social consequences of these changes on our lifestyle and the way we perceive privacy? Geoff Webb, a security expert and IoT blogger argues that the IoT will affect everything and everyone in a way that is unprecedented from the invention of agriculture: "*The IoT will be* ***everywhere*** *which means that when the changes occur (and they will) those changes will impact everything, and everyone - there's no 'offline' no 'standby' for the IoT. No one will be able to escape its impact, because you won't* ***use*** *the IoT, you'll live inside it all day, every day*"[3]. He concludes that we have to *say goodbye to privacy* since we are not only going to lose our privacy, but we will have to watch that the very concept of privacy be redefined.[4] This is not a new insight. As early as in 1991, Mark Weiser (who also coined the term ubiquitous computing), already identified privacy as one of the main challenges of ubiquitous computing [71]: "Perhaps key among [the social issues that embodied virtuality will engender] is privacy: hundreds of computers in every room, all capable of sensing people near them and linked by high-speed networks, have the potential to make totalitarianism up to now seem like sheerest anarchy." [72].

Apparently, the value placed on privacy is culturally and historically relative and this value might change depending on technological growth and other social and cultural factors. However, it is implausible that the right to privacy is voluntarily abandoned.

On the other extreme, is it possible that human's insistence on privacy, reflected in various regulations will restrain the development of the IoT, to

[1] The number of global smartphone users as 2016 is around 2 billion and there is a noticeable rise in the percentage of people in the emerging and developing nations in adoption of the smartphones and Internet usage (http://www.pewglobal.org/2016/02/22/ [accessed Dec 15, 2016]).

[2] https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/ [accessed Dec 29, 2016].

[3] http://www.wired.com/insights/2015/02/say-goodbye-to-privacy/ [accessed Dec 29, 2016].

[4] Ibid.

that extent that its benefits will be completely lost? There are no simple and pre-determined answers to these questions.

Privacy and data utility are often perceived to be conflicting. In recent years, various privacy-preserving techniques have targeted the balance between privacy and utility. As discussed in the report, the idea that removing identifiers is sufficient to provide anonymity has lost its scientific support. Methods that aim to mitigate risk of privacy breach after removing identifiers (e.g. *k*-anonymity, *l*-diversity and t-closeness) have also all shown to be more or less vulnerable.

More recent techniques such as homomorphic encryption, differential privacy, and pan-privacy are comparatively powerful privacy-preserving methods, however, they are either in research phase or not completely mature.

Binding privacy legislation in combination with increased level of data protection enforcement will likely both promote the research on privacy-preserving methods and overcome any reluctance of IoT companies to address privacy concerns of the consumers. We are in the early stages of developing IoT technologies. Providing clear legal privacy requirements allows for developing methods that could address these requirements.

Bibliography

[1] Roberto Minerva, Abyi Biru, and Domenico Rotondi. IEEE: Towards a definition of the Internet of Things (IoT). http://iot.ieee.org/, May 2015. Last accessed December 09, 2016.

[2] U.S. Department of Homeland Security. Strategic principles for securing the Internet of Things (IoT). https://www.dhs.gov/securingtheiot, November 2016. Last accessed December 08, 2016.

[3] Ovidiu Vermesan, Peter Friess, Patrick Guillemin, Harald Sundmaeker, Markus Eisenhauer, Klaus Moessner, Marilyn Arndt, Maurizio Spirito, Paolo Medagliani, Raffaele Giaffreda, Sergio Gusmeroli, Latif Ladid, Martin Serrano, Manfred Hauswirth, and Gianmarco Baldini. Internet of Things strategic research and innovation agenda. In *Internet of Things - From Research and Innovation to Market Deployment*, chapter 3, pages 7–143. River Publishers Series in Communication, 2014.

[4] Kim Zetter. *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon.* Crown Publishing Group, New York, NY, USA, 2014.

[5] Joseph Bugeja, Andreas Jacobsson, and Paul Davidsson. On privacy and security challenges in smart connected homes. In *European Intelligence and Security Informatics Conference (EISIC)*, pages 172–175. IEEE Computer Society, 2016.

[6] Chema Alonso, Antonio Guzmán, Andrey Nikishin, John Moor, Luis Muñoz Jaime Sanz, Belisario Contreras, and Bertrand Ramé. Scope, scale and risk like never before: Securing the Internet of Things. Technical report, Telefónica and ElevenPaths, 2016.

[7] Vangelis Gazis, Manuel Görtz, Marco Huber, Alessandro Leonardi, Kostas Mathioudakis, Alexander Wiesmaier, Florian Zeiger, and Emmanouil Vasilomanolakis. A survey of technologies for the Internet of Things. In *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1090–1095, August 2015.

[8] Emmanouil Vasilomanolakis, Jörg Daubert, Manisha Luthra, Vangelis Gazis, Alexander Wiesmaier, and Panagiotis Kikiras. On the security and privacy of Internet of Things architectures and systems. In *International Workshop on Secure Internet of Things (SIoT 2015)*, 2015.

[9] Tobias Zillner and Sebastian Strobl. ZigBee exploited: The good, the bad and the ugly. *Black Hat USA*, 2015.

[10] Li Jun and Yang Qing. I'm a newbie, yet I can hack ZigBee: Take unauthorized control over ZigBee devices. *DEF CON 23*, August 2015.

[11] Cari Borrás. Overexposure of radiation therapy patients in Panama: Problem recognition and follow-up measures. *Revista Panamericana de Salud Pública*, 20(2-3):173–187, 2006.

[12] Lars-Göran Angantyr, Johan Carlstedt, Björn-Erik Erlandsson, Susannah Sigurdsson, Åsa Molde, Helena Andersson, and Svante Nygren. IT-haverier i vården: Erfarenheter och förslag till åtgärder från aktuella fall, Kamedo-rapport 96. Socialstyrelsen. http://www.socialstyrelsen.se/publikationer2011/2011-12-24, 2011. Last accessed December 09, 2016.

[13] Charlie Miller and Chris Valasek. A survey of remote automotive attack surfaces. *Black Hat USA*, 2014.

[14] Charlie Miller and Chris Valasek. Remote exploitation of an unaltered passenger vehicle. *Black Hat USA*, 2015.

[15] Gus W. Weiss. The farewell dossier. *Studies in Intelligence (declassified or unclassified articles from the CIA's internal journal)*, 39(5):121–126, 1995.

[16] Thomas C. Reed. *At the Abyss: An Insider's History of the Cold War.* Presidio Press, March 2005.

[17] Nicolas Falliere, Liam O Murchu, and Eric Chien. W32.Stuxnet dossier. Technical report, Symantec, February 2011. Last accessed December 09, 2016.

[18] Arne Vidström. Möjligheter och problem vid analys av fientlig kod riktad mot Siemens S7-serie [Opportunities and problems in the analysis of malware directed against Siemens S7]. Technical Report FOI-R–3567–SE, Swedish Defence Research Agency, FOI SE-164 90 Stockholm, Sweden, November 2012.

[19] David E. Sanger. *Confront and Conceal: Obama's Secret Wars and Surprising Use of American Power.* Broadway Books, April 2013.

[20] Jan Henrik Ziegeldorf, Oscar García Morchon, and Klaus Wehrle. Privacy in the Internet of Things: Threats and challenges. *Security and Communication Networks*, 7(12):2728–2742, 2014.

[21] Dora Gálvez-Cruz. *An environment for protecting the privacy of e-shoppers.* PhD thesis, Department of Computing Science, University of Glasgow, 2009.

[22] Karen Renaud and Dora Gálvez-Cruz. Privacy: Aspects, definitions and a multi-faceted privacy preservation approach. In *Proceedings of the 2010 Information Security for South Africa Conference (ISSA)*, pages 1–8. IEEE Computer Society, 2010.

[23] Alan F. Westin. *Privacy and Freedom*, page 7. Atheneum, New York, 1967.

[24] Heng Xu. Reframing privacy 2.0 in online social network. *Journal of Constitutional Law*, 14(4):1077–1102, March 2012.

[25] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. Misplaced confidences: Privacy and the control paradox. In *Proceedings of the 9th Annual Workshop on the Economics of Information Security (WEIS 2010), Harvard University*, June 2010.

[26] Tobias Kowatsch and Wolfgang Maass. Critical privacy factors of Internet of Things services: An empirical investigation with domain experts. In Hakikur Rahman, Anabela Mesquita, Isabel Ramos, and Barbara Pernici, editors, *Proceedings of the 7th Mediterranean Conference on Information Systems, MCIS '12, Guimarães, Portugal*, volume 129 of *Lecture Notes in Business Information Processing*, pages 200–211. Springer, 2012.

[27] Tim Coughlan, Michael Brown, Richard Mortier, Robert J. Houghton, Murray Goulden, and Glyn Lawson. Exploring acceptance and consequences of the Internet of Things in the home. In *Proceedings of the 2012 IEEE International Conference on Green Computing and Communications*, pages 148–155, Los Alamitos, CA, USA, 2012. IEEE Computer Society.

[28] Peter Hustinx. Privacy by design: delivering the promises. *Identity in the Information Society*, 3(2):253–255, August 2010.

[29] Ann Cavoukian. Privacy by design: The 7 foundational principles. White paper, Information and Privacy Commissioner of Ontario, Canada, 2009.

[30] Seda Gürses, Carmela Gonzalez Troncoso, and Claudia Diaz. Engineering privacy by design. *Computers, Privacy & Data Protection*, 2011.

[31] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, SP '08, pages 111–125, Washington, DC, USA, 2008. IEEE Computer Society.

[32] Vincent D. Blondel, Adeline Decuyper, and Gautier Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4, 2015.

[33] Latanya Sweeney. Weaving technology and policy together to maintain confidentiality. *The Journal of Law, Medicine & Ethics*, 25(2-3):98–110, June 1997.

[34] Daniel C. Barth-Jones. The 're-identification' of Governor William Weld's medical information: A critical re-examination of health data identification risks and privacy protections, then and now. Available at SSRN: http://ssrn.com/abstract=2076397, July 2012. Last accessed December 14, 2016.

[35] Paul Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57:1701–1771, 2010.

[36] Saul Hansell. AOL removes search data on group of web users. The New York Times, August 8, 2006.

[37] Arvind Narayanan and Vitaly Shmatikov. How to break anonymity of the Netflix prize dataset, 2006. arXiv:cs/0610105 [cs.CR].

[38] Xiaoming Liu, Nils Krahnstoever, Ting Yu, and Peter Tu. What are customers looking at? In *Proceedings of the IEEE International Conference on Advanced Video and Signal-Based Surveillance*, London, UK, September 2007.

[39] Ton van Deursen. 50 ways to break RFID privacy. In *Privacy and Identity 2010*, volume 352 of *IFIP AICT*, pages 192–205. Springer, 2011.

[40] Gergely Alpár, Lejla Batina, Lynn Batten, Veelasha Moonsamy, Anna Krasnova, Antoine Guellier, and Iynkaran Natgunanathan. New directions in IoT privacy using attribute-based authentication. In *Proceedings of the ACM International Conference on Computing Frontiers*, CF '16, pages 461–466, New York, NY, USA, 2016. ACM.

[41] Diansheng Guo and Jeremy Mennis. Spatial data mining and geographic knowledge discovery - an introduction. *Computers, Environment and Urban Systems*, (33):403–408, 2009.

[42] Daniele Sacco, Gianmario Motta, Linlin You, Nicola Bertolazzo, and Chen Chen. Smart cities, urban sensing and big data: Mining geo-location in social networks. *AICA*, September 2013.

[43] Daqing Zhang, Bin Guo, and Zhiwen Yu. The emergence of social and community intelligence. *Computer*, 44(7):21–28, July 2011.

[44] Alastair R. Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, January 2003.

[45] Eran Toch, Yang Wang, and Lorrie Faith Cranor. Personalization and privacy: A survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction*, 22(1-2):203–220, April 2012.

[46] Sarah Spiekermann and Lorrie Faith Cranor. Engineering privacy. *IEEE Transaction on Software Engineering*, 35(1):67–82, January 2009.

[47] Saša Radomirović. Towards a model for security and privacy in the Internet of Things. In *Proceedings of the 1st International Workshop on the Security of the Internet of Things*, SecIoT'10, Tokyo, Japan, December 2010.

[48] Daniel J. Solove. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3):477–560, 2006.

[49] Chris Clifton, Murat Kantarcioglu, and Jaideep Vaidya. Defining privacy for data mining. In *Proceedings of the National Science Foundation Workshop on Next Generation Data Mining*, pages 126–133, November 2002.

[50] Joel Brynielsson, Fredrik Johansson, and Magnus Jändel. Privacy preserving data mining: a literature review. Technical Report FOI-R-3633-SE, Swedish Defence Research Agency, FOI SE-164 90 Stockholm, Sweden, 2013.

[51] Charu C. Aggarwal and Philip S. Yu. A general survey of privacy-preserving data mining models and algorithms. In Charu C. Aggarwal and Philip S. Yu, editors, *Privacy-Preserving Data Mining - Models and Algorithms*, volume 34 of *Advances in Database Systems*, pages 11–52. Springer, 2008.

[52] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based System*, 10(5):557–570, October 2002.

[53] Charu C. Aggarwal. On k-anonymity and the curse of dimensionality. In *Proceedings of the 31st International Conference on Very Large Data Bases, Trondheim, Norway, August 30 - September 2, 2005*, pages 901–909. ACM, 2005.

[54] Xiaokui Xiao and Yufei Tao. Anatomy: Simple and effective privacy preservation. In *Proceedings of the 32nd International Conference on Very Large Data Bases*, VLDB '06, pages 139–150. VLDB Endowment, 2006.

[55] Ninghui Li and Tiancheng Li. t-closeness: Privacy beyond k-anonymity and l-diversity. In *Proceedings of the 23rd IEEE International Conference on Data Engineering (ICDE '07)*, 2007.

[56] Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming: 33rd International Colloquium, ICALP '06, Venice, Italy, Proceedings, Part II*, pages 1–12, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[57] Tore Dalenius. Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 15:429–444, 1977.

[58] Cynthia Dwork and Adam Smith. Differential privacy for statistics: What we know and what we want to learn. *Journal of Privacy and Confidentiality*, 1(2):135–154, 2009.

[59] Andreas Haeberlen, Benjamin C. Pierce, and Arjun Narayan. Differential privacy under fire. In *Proceedings of the 20th USENIX Security Symposium*, San Francisco, CA, USA, August 2011. USENIX Association.

[60] Changchang Liu, Supriyo Chakraborty, and Prateek Mittal. Dependence makes you vulnerable: Differential privacy under dependent tuples. In Srdjan Capkun, editor, *Proceedings of the Network and Distributed System Security Symposium, NDSS '16*, Geneva, Switzerland, February 2016. Internet Society.

[61] Benjamin C. M. Fung, Ke Wang, Rui Chen, and Philip S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42(4), 2010.

[62] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, August 2014.

[63] José Ramón Padilla-López, Alexandros André Chaaraoui, and Francisco Flórez-Revuelta. Visual privacy protection methods: A survey. *Expert Systems with Applications*, 42(9):4177–4195, 2015.

[64] Jagarlamudi Shashank, Palivela Kowshik, Kannan Srinathan, and C.V. Jawahar. Private content based image retrieval. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pages 1–8. IEEE, 2008.

[65] Izumi Ito and Hitoshi Kiya. One-time key based phase scrambling for phase-only correlation between visually protected images. *EURASIP Journal on Information Security*, 2009:1–11, January 2009.

[66] Jatinder Singh, Thomas F. J.-M. Pasquier, Jean Bacon, Hajoon Ko, and David M. Eyers. Twenty security considerations for cloud-supported Internet of Things. *IEEE Internet of Things Journal*, 3(3):269–284, 2016.

[67] Arvind Narayanan and Vitaly Shmatikov. Myths and fallacies of "personally identifiable information". *Communication ACM*, 53(6):24–26, June 2010.

[68] Branden Ghena, William Beyer, Allen Hillaker, Jonathan Pevarnek, and J. Alex Halderman. Green lights forever: Analyzing the security of traffic infrastructure. In Sergey Bratus and Felix F. X. Lindner, editors, *Proceedings of the 8th USENIX Workshop on Offensive Technologies, WOOT '14*. USENIX Association, August 2014.

[69] Philip E Ross. Top 11 technologies of the decade. *IEEE Spectrum*, 48, 2011.

[70] Juha K. Laurila, Daniel Gatica-Perez, Imad Aad, Jan Blom, Olivier Bornet, Trinh Minh Tri Do, Olivier Dousse, Julien Eberle, and Markus Miettinen. From big smartphone data to worldwide research: The mobile data challenge. *Pervasive Mob. Comput.*, 9(6):752–771, 2013.

[71] Marc Langheinrich. Privacy in ubiquitous computing. In John Krumm, editor, *Ubiquitous Computing Fundamentals*, chapter 3, pages 95–159. Chapman & Hall, 1st edition, 2009.

[72] Mark Weiser. The computer for the 21st century. *Scientific American*, 265(3):94–104, September 1991.

FOI is an assignment-based authority under the Ministry of Defence. The core activities are research, method and technology development, as well as studies for the use of defence and security. The organization employs around 1000 people of whom around 800 are researchers. This makes FOI the largest research institute in Sweden. FOI provides its customers with leading expertise in a large number of fields such as security-policy studies and analyses in defence and security, assessment of different types of threats, systems for control and management of crises, protection against and management of hazardous substances, IT-security an the potential of new sensors.

FOI
Swedish Defence Research Agency
SE-164 90 STOCKHOLM

Tel: +46 8 5550 3000
Fax: +46 8 5550 3100

www.foi.se