STUDIES IN COMPUTER SCIENCE NO 7, LICENTIATE THESIS

JOSEPH BUGEJA

SMART CONNECTED HOMES: CONCEPTS, RISKS, AND CHALLENGES

MALMÖ UNIVERSITY

SMART CONNECTED HOMES: CONCEPTS, RISKS, AND CHALLENGES

Malmö University
Studies in Computer Science No 7,
Licentiate Thesis

ISBN 978-91-7104-929-2 (print)
ISBN 978-91-7104-930-8 (pdf)
Holmbergs, Malmö 2018

JOSEPH BUGEJA

SMART CONNECTED HOMES: CONCEPTS, RISKS, AND CHALLENGES

Malmö University, 2018
Department of Computer Science
Faculty of Technology and Society

Studies in Computer Science

Faculty of Technology and Society
Malmö University

1. Jevinger, Åse. *Toward intelligent goods: characteristics, architectures and applications,* 2014, Doctoral dissertation.
2. Dahlskog, Steve. *Patterns and procedural content generation in digital games: automatic level generation for digital games using game design patterns,* 2016, Doctoral dissertation.
3. Fabijan, Aleksander. *Developing the right features: the role and impact of customer and product data in software product development,* 2016, Licentiate thesis
4. Paraschakis, Dimitris. *Algorithmic and ethical aspects of recommender systems in e-commerce*, 2018, Licentiate thesis
5. Hajinasab, Banafsheh. *A Dynamic Approach to Multi Agent Based Simulation in Urban Transportation Planning*, 2018, Doctoral dissertation
6. Fabijan, Aleksander. *Data-Driven Software Development at Large Scale*, 2018, Doctoral dissertation
7. Bugeja, Joseph. *Smart Connected Homes: Concepts, Risks, and Challenges*, 2018, Licentiate thesis

Electronically available at:
https://muep.mau.se/handle/2043/25061

I dedicate this thesis to my parents, who instilled in me the virtues of perseverance and patience, and relentlessly encouraged me to strive for excellence. Your unconditional love and prayers carry me.

ABSTRACT

The growth and presence of heterogeneous connected devices inside the home have the potential to provide increased efficiency and quality of life to the residents. Simultaneously, these devices tend to be Internet-connected and continuously monitor, collect, and transmit data about the residents and their daily lifestyle activities. Such data can be of a sensitive nature, such as camera feeds, voice commands, physiological data, and more. This data allows for the implementation of services, personalization support, and benefits offered by smart home technologies. Alas, there has been a rift of security and privacy attacks on connected home devices that compromise the security, safety, and privacy of the occupants.

In this thesis, we provide a comprehensive description of the smart connected home ecosystem in terms of its assets, architecture, functionality, and capabilities. Especially, we focus on the data being collected by smart home devices. Such description and organization are necessary as a precursor to perform a rigorous security and privacy analysis of the smart home. Additionally, we seek to identify threat agents, risks, challenges, and propose some mitigation approaches suitable for home environments. Identifying these is core to characterize what is at stake, and to gain insights into what is required to build more robust, resilient, secure, and privacy-preserving smart home systems.

Overall, we propose new concepts, models, and methods serving as a foundation for conducting deeper research work in particular linked to smart connected homes. In particular, we propose a taxonomy of devices; classification of data collected by smart connected homes; threat agent model for the smart connected home; and identify challenges, risks, and propose some mitigation approaches.

Keywords: Smart Connected Homes, Internet of Things, Smart Home Devices, Data Lifecycle, Security Risks, Privacy Manage-

ment, Vulnerability Assessment, Security Mitigations, Threat Agents, Smart Home Services, System Architecture.

PUBLICATIONS

Included Papers

Paper I: Bugeja, J., Jacobsson, A., & Davidsson, P. (2018). Smart Connected Homes. In: *Internet of Things A to Z Technologies and Applications* (1st ed., pp. 359–384). IEEE John Wiley & Sons.

Paper II: Bugeja, J., Jacobsson, A., & Davidsson, P. (2016). On Privacy and Security Challenges in Smart Connected Homes (pp. 172–175). In: *Proceedings of the 2016 Intelligence and Security Informatics Conference (EISIC 2016)*. IEEE.

Paper III: Bugeja, J., Jacobsson, A., & Davidsson, P. (2017). An Analysis of Malicious Threat Agents for the Smart Connected Home (pp. 557–562). In: *Proceedings of the International Conference on Pervasive Computing and Communications Conference (The First International Workshop on Pervasive Smart Living Spaces 2017)*. IEEE.

Paper IV: Bugeja, J., Jönsson, D., & Jacobsson, A. (2018). An Investigation of Vulnerabilities in Smart Connected Cameras. In: *Proceedings of the International Conference on Pervasive Computing and Communications Conference (The Second International Workshop on Pervasive Smart Living Spaces 2018)*. IEEE.

Paper V: Bugeja, J., Davidsson, P., & Jacobsson, A. (2018). Functional Classification and Quantitative Analysis of Smart Connected Home Devices (pp. 144–149). In: *Proceedings of the Global IoT Summit (GIoTS 2018)*. IEEE.

Paper VI: Bugeja, J., Jacobsson, A., & Davidsson, P. (2018). An Empirical Analysis of Smart Connected Home Data (pp. 134–149). In: *Proceedings of the Internet of Things (ICIOT 2018)*. Lecture

Notes in Computer Science, vol 10972. Springer International Publishing.

Personal Contribution

For all publications above, the first author was the main contributor with regard to inception, planning, execution, and writing of the research.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to my supervisor Dr. Andreas Jacobsson for his unwavering support, advice, guidance, and for affording me the chance to make this work a reality. I also like to particularly express my heartfelt gratitude and sincere thanks to Prof. Paul Davidsson who has been my co-supervisor. Thank you for sharing your wealth of knowledge and expertise with me and applying it to this thesis. I am so grateful to have had you both as my study leaders. Thank you both for your words of encouragement, professionalism, and for the time and attention you put in me throughout my educational experience.

Before thanking any others, I would like to especially thank the research profile "Internet of Things and People" (IOTAP) funded by the Knowledge Foundation and Malmö University in collaboration with various industrial partners. I would also like to thank all the members of the research profile project "Intelligent Support for Privacy Management in Smart Homes" in particular Verisure for their support in my research.

Furthermore, I would like to also convey my thanks and appreciation to my PhD examiner Dr. Jan Persson, review group members – Prof. Bengt J. Nilsson and Assoc. Prof. Helena Holmström Olsson – for their help in reviewing this thesis, assessing my individual study plan, and also for bestowing invaluable advice necessary to ensure the relevance and quality of work.

Thank you also to Dr. Annabella Loconsole for stepping in to ensure a healthy work and research schedule on my part. I am also grateful to Dr. Åse Jevinger, Solveig-Karin Erdal, and Susanne Lundborg in particular for their help in coordinating the logistics for the licentiate seminar. Thanks also to Assoc. Prof. Christina Bjerkén especially for your invaluable support during the past years when you were the director of PhD studies.

Last, I would like to thank all my previous and present research scholars working in the IOTAP Research Center, and the Depart-

ment of Computer Science and Media Technology, Malmö University for their kindness and friendship.

Joseph Bugeja
Malmö, 2018

TABLE OF CONTENTS

PART I

PART II

PART I

THESIS

1. INTRODUCTION

"Wireless cameras within a device such as the fridge may record the movement of suspects and owners. Doorbells that connect directly to apps on a user's phone can show who has rung the door and the owner or others may then remotely, if they choose to, give controlled access to the premises while away from the property. All these leave a log and a trace of activity."

-- Mark Stokes

In 1991, Mark Weiser, introduced the term of ubiquitous, also known as pervasive, computing in his seminal paper "The Computer for the 21st Century" [1]. His vision was that computing should be integrated seamlessly in the background, allowing people to employ it when needed without shifting their attention from their main tasks. Eight years later, the idea of Internet of Things (IoT) was introduced by Kevin Ashton while working on the Auto-ID Center at the Massachusetts Institute of Technology. Ashton originally coined the term "Internet of Things" in a presentation he made at Proctor & Gamble (P&G), where he made the first association between the new idea of Radio Frequency Identification in P&G's supply chain and the emerging Internet [2].

The Internet of Things (IoT) can be thought of as a computing paradigm where physical objects (e.g., devices, vehicles, and buildings) are augmented with identifying sensing/actuation, storing, networking, and processing capabilities, allowing them to communicate with each other and with other devices and services over the Internet to accomplish some objective [3]. These objects are typically referred as smart objects, smart devices, or simply as connected things. Smart objects can interact with other smart devices and people, and can collect information from their surroundings and exchange data with each other including remote servers on the Internet. Because of their capability to make sense of and leverage their environment, these objects are often called "smart" and

can enable context-aware automation without human operation in nearly every field.

Smart homes is a domain of IoT, essentially an automated building, composed of a network of devices that provide "electronic, sensor, software, and network connectivity inside a home" [4]. This setup gives the residents the ability to get information, control, and automate different parts of the home and improve the quality of daily chores in a household possibly from anywhere and at anytime, typically over the Internet through a smartphone application [5]. As smart home technology has evolved, smart devices have been networked to form smart home ecosystems. These ecosystems have enabled smart devices to combine efforts and provide benefits beyond just convenience [6] including that of enhancing the residents security/safety, entertainment, health/fitness, and overall the quality and efficiency of occupants' lives. In our work, we refer to IoT-based smart homes as smart connected homes.

In recent years, the development of the IoT and smart connected homes, has been gaining increasing momentum due to a range of advancements in wireless protocols, sensors, processors, data analytics, cloud technologies, and widespread availability of smartphones. Some survey studies, in particular Gartner [7], estimate the amount of connected devices to increase from about 11 billion in 2018 to 20 billion by 2020 with consumer devices representing the largest group. In terms of amount of smart home devices there were about 33 million Wi-Fi enabled devices being shipped globally in 2016, and this figure is expected to increase to 320 million by 2020 [8]. These estimations are graphically depicted in Figure 1 (the figure is an adaptation of [8]). In reality, the total amount of shipped devices is more than the previously cited statistic when device types supporting other protocols, such as Bluetooth, are factored in. Indeed, some analysts estimate that an average household could contain over 500 smart objects by 2022 [9]. Noting the potential of the market, commercial Information and Communication Technology (ICT) organizations like Google, Apple, and Samsung have started to show interest in the technology launching their own products, e.g., Nest smart thermostat, platforms, e.g., Apple HomeKit, and communication protocols, e.g., Google Weave, to compete on the market for building the next smart home ecosystem. Today, the IoT is part of daily life, with

smart assistants like Siri and Alexa being added to toasters, thermostats, lights, and the list goes on.

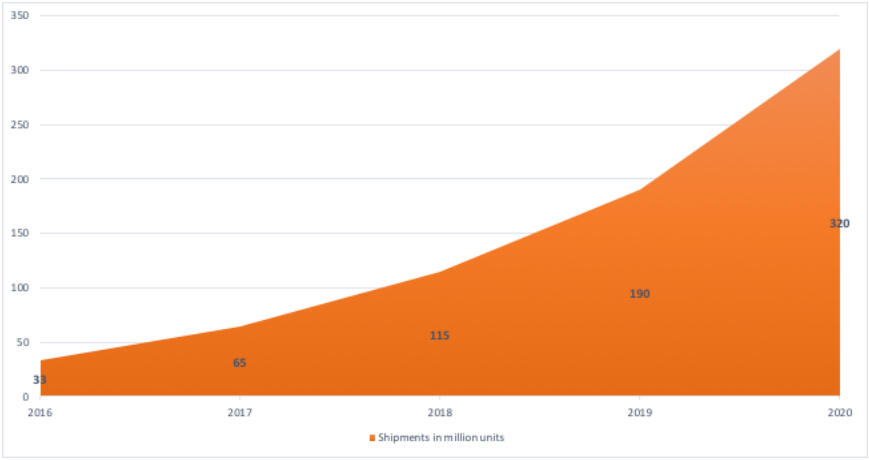

Figure 1. Shipments of units of Wi-Fi enabled smart home devices worldwide from 2016 to 2020.

1.1 Research Objectives

The home locks within a digital trove of sensitive personal data. This data are collected by smart devices that tend to lie in close proximity to the users but are often transmitted to the Internet and remote cloud services. In fact, smart devices have been shown to be able to collect a diverse and increasing range of user information, including sleeping patterns, exercise routines, medical information, and more [10] [11]. As the number and type of smart home ecosystems and the data being generated by them are increasing at a fast pace, so are the risks and challenges introduced by these devices. Simultaneously, it also becomes increasingly harder to gain a deeper understanding of the smart connected home ecosystem especially in terms of its technical composition, supported functionality, and the type of data it deals with.

Such comprehension is necessary to build a more robust, resilient, secure, and privacy-preserving smart connected home. Likewise, it is needed as a precursor to perform a comprehensive security and privacy analysis of the smart connected home. Complicating this is the fact that the smart home market is fragmented with a diverse selection of unstandardized devices and a broad spectrum of stakeholders that operate

without security and privacy expertise. Moreover, research work in the smart home field is being segmented by multiple academic disciplines such as networking, ubiquitous, and mobile computing each bringing their own concepts and assumptions. Together, these factors increase the difficulty of attaining a common understanding of the smart home, and also add to the creation of different vulnerabilities and risks.

In relation to this, we are interested in organizing commercial smart home devices in a systematic manner, surveying their technical capabilities, and identifying data being collected by them. These components are pivotal for developing a rigorous analysis of the smart connected home environment. Additionally, we identify threat agents, vulnerabilities, risks, and propose some mitigation approaches suitable for home environments. Identifying these is core to characterize what is at stake, and to gain insights into what is needed to bolster security and privacy in smart home systems. By recognizing what is being exploited by attacks, we also contribute to raising awareness and motivating discussions about security and privacy challenges that IoT technologies bring forth to the home environment and to our society in general.

1.2 Research Questions

In this thesis we want to answer the following main research questions:

RQ1: How can smart connected home devices and the data collected by them be categorized?

RQ2: What security and privacy risks does the introduction of IoT technologies inside the home bring to the residents?

RQ3: What are the characteristics and challenges in mitigating security and privacy risks in smart connected homes?

RQ1 lays out the technical composition of a smart connected home including its devices and data. Classifying and grouping the different devices and data types is key for reasoning about security and privacy, especially for conducting risk assessments. RQ2 deals with the investigation of security and privacy risks associated with the installation and use

of smart home products. Especially, we seek to explore both actual and probable attack scenarios, and likewise motivations and capabilities required to perform attacks on smart homes. This is especially important to better understand what assets are being targeted and likewise to observe the effort involved in building effective security strategies. RQ3 examines the characteristics of IoT environments, in particular the challenges that hinder or make the design of effective security safeguards particularly difficult to implement in smart home environments. At the same time, in RQ3 we aim to identify and discuss mitigations working at different architecture layers of the IoT-based home. Recognizing the current mitigations is core to assess what has been done therein and what remains further to be done to build more secure and privacy-preserving smart home systems.

1.3 Contributions

Overall, our main contributions to the research community with this thesis are summarized as follows:

i. A taxonomy and quantitative analysis of devices in smart connected homes;
ii. An analysis and classification of data collected by smart connected homes;
iii. A threat agent model for the smart connected home;
iv. Identification of state-of-the-art security challenges and their mitigations in smart connected homes.

Contribution i) and ii) are the answer to RQ1. Essentially, the taxonomy of devices created as part of contribution i) serves as input to the data categorization as is needed for contribution ii). Contribution iii) answers RQ2 by proposing a new threat agent model identifying different malicious intruders, risks, and typical compromise methods used by each threat agent. Contribution iv) is proposed as an answer to RQ3. In this regard, contribution iii) is also considered a pivotal component in answering RQ3.

The contributions above are elaborated on in Chapter 4 of this thesis.

1.4 Thesis Outline

We divide the thesis into two parts: Part I: Thesis and Part II: Publications. In Part I, we provide an extensive introduction to the thesis area and summarize answers to the posed research questions. In Part II, we include the six publications that form the actual research of this thesis. An outline of Part I is presented below:

Chapter 1: Introduction. The first chapter presents the theme of the thesis and introduces the research questions and motivation of the thesis.

Chapter 2: Central Concepts. The second chapter introduces the conceptual framework to understand the rest of this thesis. This includes a short history of smart homes, primer of smart home technologies, and fundamental notions connected to security and privacy.

Chapter 3: Research Methodology. The third chapter describes the methodology that has been applied during the research process of this thesis.

Chapter 4: Contributions. The fourth chapter presents the main contributions to the research community mapping them to the posed research questions.

Chapter 5: Discussion. The fifth chapter discusses the relevance of our findings and some implications of our contributions.

Chapter 6: Conclusions and Future Work. The sixth chapter concludes the thesis, summarizing it and identifies some opportunities for future work.

2. CENTRAL CONCEPTS

"If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle."

-- Sun Tzu

2.1 Smart Connected Homes

There is no generally standard definition or consensus of what a "smart home" is. The definition of the term varies according to the technology or the functionality the home implements. In fact, several alternative names have been used across the years to refer to the smart home, e.g., "intelligent living", "digital house", "smart environments", and more [12]. A common, simple, and well accepted definition has been developed by the UK Department of Trade and Industry (DTI). The DTI's Smart Home project defined a "smart home" as: "A dwelling [residence] incorporating a communication network that connects the key electrical appliances and services, and allows them to be remotely controlled, monitored or accessed." [13].

While DTI's definition works for most smart home scenarios, nowadays homes are evolving into smart living spaces or ecosystems incorporating diverse services such as optimized entertainment, security/safety, energy management, and more. Furthermore, in addition to the automation and control aspects, smart homes are also providing proactive services, e.g., providing timely physical support, to the residents through sensor technologies and sophisticated algorithms based on artificial intelligence and machine learning.

2.2 Smart Connected Home Evolution

The history of smart home technology goes back many years. In fact, the actual term "smart home" was originally coined by the American Association of House Builders in the year 1984 [14].

Although the concept of a smart home has been around for a while, the smart home has only taken momentum in recent years. Here an important milestone for making the development of smart home technology a reality was when electricity was brought to households in the beginning of the 20th century [15]. Electricity stimulated the introduction of new equipment in the home, e.g., electrical machines and domestic appliances.

Another important landmark, introduced in the last quarter of the 20th century, was the introduction of information technology in the homes. This created new possibilities for exchanging information sparking the evolution of smart home technology [15].

More recently, we observe another important milestone in the smart home evolution brought about by the IoT and the ensemble of technologies surrounding it, in particular innovations in sensors and microelectronic devices.

2.2.1 Pre-IoT Smart Connected Homes

The first smart home devices emerged in the late 1960s with the invention of the Electronic Computing Home Operator (ECHO IV) and Kitchen Computer [16]. The ECHO IV was used for family bookkeeping, inventory taking, and climate control [17]. A year later, the Kitchen Computer came out. This machine allowed people to store recipes [18].

In the 1970s, X10, was established and was used as a standard communication protocol for wiring houses for home automation. This is often touted as the ancestor of home automation.

When "personal computers" appeared in the consumer market in the late 1970s, controlling and automating home appliances was mainly conducted by hobbyists in Do-It-Yourself (DIY) projects [19]. Here, some form of remote control was possible by decoding Dual-Tone Multi-Frequency (DTMF) signals through telephone lines [20]. However,

the turning point in smart home development occurred when the domestic Internet, appeared on personal computers in the mid 1990s [21].

At the same time, in the 1990s, ubiquitous computing technologies arose. Using these technologies, researchers started developing smart home projects all across the globe [22]. In the majority of the cases these homes were real-life living space testbeds [22].

We refer to these types of systems as "smart homes". Such systems tend to use proprietary protocols, offer no or rather limited integration facilities, and allow few control options to end-users, typically limited to local (in-house) control and using specific controllers.

2.2.2 IoT-based Smart Connected Homes

In recent years, the IoT became a commercial reality allowing for home devices to be remotely observed and controlled through the Internet. Hereunder, is a chronological list of some of the most popular commercial smart home systems appearing in the consumer market in 2010 and onwards:

In 2010, the Nest Learning Thermostat [23] (nowadays owned by Google) enters the smart home scene. This device functions as a smart thermostat learning the residents' preferred house temperature and adjusting it automatically. Nest is sometimes identified as the flagship product that introduced the era of the (modern) smart home [24].

In 2013, Microsoft launched "Lab of Things" [25]. This is an open-source platform that eases the process of interconnecting smart home devices together and implementing application scenarios or workflows.

In 2014, SmartThings (later acquired by Samsung) issued a device that functioned as a residential gateway (sometimes called hub or home controller) linking together nearly every connected gadget at home [27].

In 2015, Apple released HomeKit [26]. This is a developer framework and an interoperability protocol that allows different devices to communicate with each other.

In 2016, Amazon launched a smart speaker system – Amazon Echo – that could be used to control the smart home by using the voice as an input channel and providing a full ecosystem of programmable skills (capabilities).

Later, as competition arose, Google released Google Home. Just like Amazon Echo, Google Home also acts as an intelligent (digital) personal assistant allowing for home automation and other things such as searching the web, get a personalized daily briefing, checking weather report, etc., by speaking a command to the device.

We refer to these types of systems as "smart connected homes". These systems tend to be Internet-connected, feature multimodal user interface channels, various networking protocols, and "intelligent" logic making it possible to make some autonomous decisions.

The focus of our work is on this category of smart homes.

2.3 Existing Systems

Several smart home projects have been conducted over the last several decades conveying different ideas, functions, and utilities. We divide these systems into two types: systems that are essentially laboratory systems and commercial systems. Laboratory systems are fundamentally used for research purposes and often involve dedicated housing facilities, whereas commercial systems involve platforms and off-the-shelf products retrofitted into actual finished homes.

2.3.1 Laboratory Systems

Over the last decade, a number of smart home live-in laboratory or experimental houses have been built [14]. Many of these projects were initially developed to study human behavior and in-home automation. Typically, this involved monitoring and recording of activities and interactions of residents in a purposely designed setup. Some prominent examples are: Aware Home project [28], MavHome project [29], GatorTech Smart House project [30], House_n project [31], and PlaceLab [32].

Most of the mentioned systems are linked to the pre-IoT smart connected homes. In general, these are essentially test-beds for technological components and an early attempt to bring the ubiquitous computing paradigm into the home.

2.3.2 Commercial Systems

Nowadays, there is a growing trend of developing ready-to-use off-the-shelf solutions. These are sometimes referred to as smart home gateway (hub) ecosystems. Here, the idea is to provide the residents with a central hub that is capable of connecting and interacting with various smart devices present in a home.

Various large manufacturing companies have launched similar products such as Samsung Smart Home, Google Home, Apple HomePod, and many more. Most of these systems leverage the cloud infrastructure to deploy their services. Another characteristic of these systems is that they support a number of different applications (beyond that of home automation), tend to be programmable, and allow end-users options to customize them according to their liking.

In comparison to the laboratory systems, commercial systems tend to be installed (or rather retrofitted) in actual residences. Here, the residents (or rather the homeowner) tend to have an active role to select and bring into their household the technology they want and oftentimes install it themselves without relying on a professional user [33]. In comparison to the laboratory systems, commercial systems bring forth added complexities (e.g., in relation to the sophistication of the underlying and evolving technologies), new dynamics (e.g., in relation to the ecosystem of stakeholders, assets, and services), and likewise challenges (e.g., given the plethora of unregulated and unstandardized devices and services). Thereby, this raises more research opportunities to the academic and industry communities.

Given these factors, in our work, we put our attention on commercial systems. These systems are associated with the IoT-based smart connected homes we explored earlier.

2.4 System Properties

In this section, we describe the applications, generic architecture, and technical specifications of smart connected home systems.

2.4.1 Applications

Application Area	Device Type	Collected Data Types
Energy and resource management	Plug, light bulb, shower head water meter	Location data, consumption data
Entertainment systems	Music player, TV, audio speaker	Voice commands, features accessed, search queries
Health and wellness	Blood pressure monitor, scale	Body metrics, social networking services related
Networking and utilities	Gateway/hub, wireless signal extender	Network/connectivity-related data, personal preferences
Human-machine interface	Remote control	Battery charge level
Household appliances and kitchen aids	Vacuum cleaner, oven, floor mopper	Location data, operating schedules
Security and safety	Cloud camera, door bell, smoke detector	Contact preferences, location data, interaction data
Sensors	CO2 sensors, rain sensor, air quality sensor	Sensor status

Table 1. Smart connected home application areas and examples of device and data types.

The smart connected home encloses multiple applications belonging to the different areas. Common application areas include: energy, entertainment, security, and healthcare [34]. In smart connected homes, the smart devices form the core of the concept, as they create the foundation of the user experience.

There is a remarkable number of smart devices available in the market. These devices, in particular through the use of sensors, collect data on which decisions are made. Smart devices deal with different types of data, some of which can be of privacy sensitive nature. The smart connected home application areas alongside the devices and type of data captured by each is summarized in Table 1.

In Part II of this thesis, we elaborate on the application areas, device types, and as well on the collected data types of devices.

2.4.2 Architecture

The technical composition of a smart connected home consists of various components that are controlled by different stakeholders each having different interests, incentives, and obligations that they need to ad-

here to. These components interact with each other exchanging data about the state of the home, the environment, and the activities and behavior of its residents. A generic smart connected home environment consists of the following assets:

- **Smart device.** These are hardware units, e.g., domestic appliances, lights, or sensors, that can sense, actuate, process data, and communicate. Three core devices are sensors, actuators, and end-user client devices. Sensors detect, monitor, and measure properties of objects such as room temperature. Actuators perform actions in the physical environment such as switching on or off lights. End-user client devices such as smartphones are commonly used by the residents to interact and manage the smart connected home.

- **Gateway.** The gateway (hub) is a specialized smart device that collects data from other smart devices and acts as the central point of connectivity for end-users to access and manage the home devices and to external networks. Gateways can also act as bridges translating between different communication protocols.

- **Cloud.** The main task of the cloud is to store data but it is also often used for computation power, e.g., as is needed for voice processing and as well data analytics. Depending on the architecture and communication model adopted, some devices, can send sensed data directly to the cloud, however this is often facilitated through the gateway.

- **Service.** Software applications that provide the facility to control, manage, and operate the smart home system. Services may be available in smart devices, gateways, and clouds. Cloud services often expose APIs (Application Programming Interfaces) for controlling devices over HTTP, are often utilized to implement "intelligent" logic, and frequently used for interconnectivity, e.g., through middleman cloud services like IFTTT (If This Then That).

- **User.** The stakeholder that uses and benefits from the services offered by the smart connected home. Typically, this represents the

residents that manage the different smart connected home devices and services.

Smart devices use different networking protocols to communicate with other smart devices, services, and users. Some of the most commonly used wireless standards in the home include: IEEE 802.11 (Wi-Fi), Bluetooth Low Energy (BLE), ZigBee, Z-Wave, and Thread [35]. It is common that standalone smart devices, e.g., smart thermostats such as Nest thermostat, connect to the Internet through existing Wi-Fi networks, while others, e.g., smart locks, use low energy protocols like Zigbee and BLE, and communicate to the Internet through a gateway or bridge [36] [37].

In centralized architectures smart devices tend to communicate with a central gateway and the gateway implements all the decision logic; whereas in distributed architectures smart devices communicate with each other and decisions are done locally by each node. In reality, it is also possible to have hybrid or decentralized architectures combining the characteristics of both.

More details about the composition and architecture of a smart connected home are found in Paper I.

2.4.3 Technical Specifications

Smart home devices differ in terms of their hardware and software capabilities. At one end, there are constrained devices, such as smart locks with limited CPU, memory, battery, etc. Then, we find resourceful or high-capacity devices, such as gateways, that are typically powered by the main supply [38], [39]. In Table 2, we show some of the capabilities of smart devices, in terms of their supported protocols, services, and as well their processing and storage capabilities.

As can be observed the actual specifications vary considerably between the different devices types. It can be also noted that both the storage capacity and processing power of such devices tends to be relatively low compared to that of a traditional computer system.

Device Type	Network Protocols	Services	Processing Power	Storage Capacity
Samsung SmartThings Hub	Wi-Fi, Zigbee, Z-wave, Bluetooth, Ethernet	Mobile apps, IFTTT	1 GHz	4 GB
Amazon Echo	Wi-Fi, Bluetooth	API, IFTTT, Web browser, mobile apps	1 GHz	4 GB
Nest Learning Thermostat	Wi-Fi, Bluetooth, Thread	API, IFTTT, mobile apps	800 MHz	-
August Smart Lock	Bluetooth	IFTTT, mobile apps	32 MHz	-

Table 2. Specifications of smart home devices.

2.5 Security and Privacy Concepts

This section is essentially a primer of computer security. Such background is needed to understand some of the included publications (in particular Paper III and Paper IV) where the focus is inclined towards security and privacy.

2.5.1 Information Security Terminology

In this section, we introduce some terminology that will be useful throughout this thesis. Here, we rely on RFC 4949 (Internet Security Glossary) [40].

- **Asset.** An asset is essentially anything within an environment that has value (to the organization or to its owner) and therefore requires protection. It can include both ICT resources, e.g., smart devices, and non-ICT resources, e.g., activity data.

- **Mitigation.** A mitigation is an action that reduces or removes a vulnerability or protects against a threat. An example of a mitigation against poor authentication requirements is that of enforcing two-factor authentication for users to gain successful access.

- **Threat.** Any potential occurrence that may result in an unwanted outcome for a specific asset. In other words, a threat is a possible danger that might exploit a vulnerability. An example of a threat is that of disclosing health related information of a user.

- **Threat Agents.** An individual or group of entities that can manifest a threat. An example of a threat agent is a hacker.

- **Vulnerability.** A weakness in an asset, e.g., in its design, implementation, or operation, that could be exploited to cause loss or damage. An example of a vulnerability is having no password set on a smart home device.

- **Risk.** The likelihood (possibility) that a threat will exploit a vulnerability to harm (or lose) an asset. Typically, this is written as a formula: *risk = threat · vulnerability*. The formula indicates that reducing either the threat agents or vulnerability directly results in a reduction in risks. An example of a risk in a smart home system is the chance that a threat agent captures the password of the smart home gateway to eavesdrop traffic in the home network.

Figure 2 is a conceptual map showing the relationship among the introduced terms. This is an extension of the diagram produced by Stallings et al. [41].

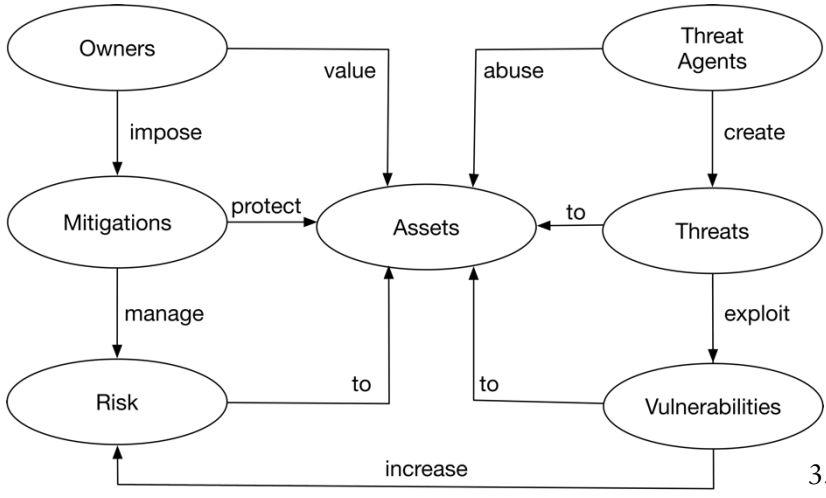

Figure 2. Security concept map.

2.5.2 Security and Privacy Goals

Almost from its inception, the key objectives of computer security have been threefold: confidentiality, integrity, and availability — the CIA triad of security [42]. These embody the fundamental security objectives for data and for tangible ICT resources. The purpose of confidentiality is to ensure that only authorized individuals can view a piece of (private and proprietary) information. Integrity ensures that only authorized individuals can generate, modify, or delete data. The goal of availability is to ensure that the data, or the system itself, is accessible when the authorized user wants it.

Privacy often overlaps with the field of security but implementing security does not assure privacy. The concept of privacy (often referred to as "data protection" in European policies [43]) is closely related to that of confidentiality. This is as it deals with protecting user's personal information from unauthorized entities, however the concept of privacy is broader than that. A popular definition of privacy is that of Alan Westin, defining 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" [44]. However, there is still no general consensus about the definition of the term privacy, it is evolving with time, and is influenced by societal and technological advances. For instance, with the IoT given the multitude and diversity of available devices asking users to explicitly control and manage all those to achieve privacy as implied in Westin's definition is impractical and may not be possible. This is especially as connected devices tend to be continuously and automatically collecting and transmitting data commonly without involving the users for decision making.

2.5.3 Smart Connected Home Assets

For an ICT or a socio-technical system, assets can be broadly categorized as: hardware, software, data, and communication facilities and networks [45]. In the case of smart connected homes, hardware can essentially be any smart home device, e.g., a domestic appliance, software represents services, e.g., smartphone application operated by the users, data can range from sensor data to human interaction data, and com-

munication facilities and networks can include dedicated network devices, e.g., routers, and infrastructures, e.g., a cloud data center.

In meeting the security and privacy goals, we are interested in protecting all the identified assets against risks.

2.5.4 Data, Metadata, and Information

RFC 4949 defines data as "information in a specific physical representation, usually a sequence of symbols that have meaning; especially a representation of information that can be processed or produced by a computer," and information as "facts and ideas, which can be represented (encoded) as various form of data". Thus, while the terms are related, in general information is often seen as data that has been processed into a meaningful form [46]. However, both terms are difficult to define in a useful way, and are oftentimes used interchangeably in legislation and regulations [46].

Legally, any data that can be linked to a person, directly or indirectly, is in general referred to as personal data [46]. Personal data have been intensely debated and given attention especially by the recent entry into force of the General Data Protection Regulation (GDPR) [47]. The GDPR is fundamentally an EU regulation that aims to protect and expand EU citizens' right to have their data processed safely and only when needed. Personal data in this context can include data that describes the person's economic, mental, or physical status. Sensitive personal data includes data on ethnicity, political opinion, religious beliefs, health, and generic and biometric data. Location data and online identifiers are also considered personal data. In a US context, personal data is oftentimes referred to as personal identifiable information.

Metadata is the term used by legislators for data about communications other than the actual content (text) [46]. In a way, this is data about data. Common examples of metadata include: time, date, activity duration, IP address, and more [48]. Metadata especially when combined with other data points, can be used to track or profile individuals, and when systematically collected and analyzed it can yield insights beyond what might reasonably be expected [49].

In this work, we use the terms data, metadata, and information interchangeably. Specifically, in Paper VI, we identify collected data from commercial systems.

2.5.5 Threats, Threat Agents, and Threat Modeling

Smart connected homes combine different types of technologies, devices, interfaces, and protocols. These factors add to the creation of numerous and different types of security and privacy threats [50]. Some categories of such threats include: device tampering, information disclosure, privacy breaches, denial-of-service (DoS), spoofing, elevation of privilege, signal injection, ransomware, and side-channel attacks [51], [52], [53], [38]. An example of a risk caused by a spoofing threat is that of having an intruder eavesdropping network traffic to gain sensitive information possibly allowing him to break-in the house at a time when the residents are away. Effectively, threats and risks impact the confidentiality, integrity, and availability of a system, and may be aggravating for slow processing, limited memory, and less power settings [54].

A threat is imposed or created on a specific asset by a threat agent. There are essentially, three different classes of threat agents: humans, technological, and environmental threat agents [55]. In terms of the human threat agent, instances of this can range from hackers to nation states, insiders to outsiders, and including those that can cause deliberate and accidental threats. Oftentimes, security literature refers to threat agents as anti-users [56].

To identify threats and threat scenarios for a system different models can be used. A threat model is a structured approach that allows a systematic identification and rating of security related threats that are likely to affect the system under consideration [57]. In general, threat modeling approaches can be categorized into three different groups: system-centric, asset-centric, and attacker-centric approaches [56] with main emphasis being on the system architecture, assets, and threat agents, respectively. Following the attacker-centric approach, to identify threat agents different methodologies are available. Examples of these are the Threat Agent Risk Assessment (TARA) developed by Intel [58], Verizon's Lists [59], and the threat assessment methodology developed by

Sandia National Laboratories at the Department of Homeland Security [60].

In our work, we were mainly interested in the attacker-centric method (cf. Paper III). This is especially since we noted a research gap in that aspect, and furthermore as anti-users arguably represent the most dynamic and versatile entity to the smart connected home posing the greatest risk to the residents.

2.5.6 Vulnerabilities, Vulnerability Analysis, and Attacks

Many types of threat agents can take advantage of several types of vulnerabilities, resulting in a variety of specific threats. Common examples of vulnerabilities of ICT resources are an asset becoming: leaky (i.e., some or all information of a resource becomes available to unauthorized users), corrupted (i.e., doing the wrong thing), and unavailable (i.e., the resource is impractical or impossible to reach) [45]. As an example in the case of a smart connected home, a threat agent, may attack a smart home device, e.g., a connected refrigerator, turning it into a zombie device (i.e., a compromised device that a remote attacker has hacked to forward data, including malware, to other Internet hosts), or as a proxy for future communications. One risk with this is that the infected appliance can then be used to conduct a DoS attack rendering the other networked home devices unavailable.

In order to discover vulnerabilities in a system against potential threats, a vulnerability analysis can be performed. For example, in the context of access control, vulnerability analysis attempts to identify the strengths and weaknesses of the different access control mechanisms and the potential of a threat to exploit a weakness (exploiting a weakness is oftentimes referred to as an attack). While threat modeling works at a higher abstraction level, vulnerability analysis works at a lower detail-oriented level. Nonetheless, both approaches are needed to understand and properly manage risk.

Different threat agents use different tools and methods for conducting vulnerability analysis and to conduct attacks. These can range from using specialized security distributions like Kali Linux [61] to online databases and search engines such as Shodan [62] and Censys [63]. Shodan – designed by the computer programmer John Matherly in 2009 – is a

vulnerability assessment tool [64] that crawls the Internet on a daily basis looking for Internet-connected devices (e.g., routers, printers, webcams) probing for their ports and indexing the retrieved banners and metadata. For web servers the banner would be the HTTP headers, and the metadata may include the operating system, hostname, geographic location, and more [65]. Censys has similar goals to that of Shodan but it uses different tools and methods to retrieve and document IoT devices.

In our case, given the flexibility, extensive documentation, and intuitive interfaces, we rely on Shodan to conduct a vulnerability assessment related to smart connected cameras (cf. Paper IV).

3. RESEARCH METHODOLOGY

"This approach [research methodology] is based on bringing together a worldview or assumptions about research, a specific design, and research methods. Decisions about choice of an approach are further influenced by the research problem or issue being studied, the personal experiences of the researcher, and the audience for whom the researcher writes."

-- John W. Creswell

3.1 Research Approach

There are in general two distinct research approaches: quantitative and qualitative. Quantitative approaches assume a positivist (empiricism) philosophy; whereas qualitative approaches follow an interpretive (constructivism) paradigm. Mixed methods is an alternative research approach commonly linked to the pragmatic worldview and featuring combinations of both qualitative and quantitative strategies [66].

Commonly, for answering problems that are typically not yet fully understood, not well researched, or still emerging, an exploratory research approach is deemed well-suited [67]. This makes such an approach applicable for our research objectives, i.e., to help develop a thorough understanding of the smart connected home ecosystem including its encompassing concepts, risks, and challenges. It is also ideal since limited existing theory has been identified possibly due to the rapid development of smart home technology interest. Nevertheless, we also adopt research strategies that tend to have a confirmatory nature and that tend to be associated with the positivist paradigm.

3.2 Research Strategy

A research strategy is the method used to answer the posed research questions. Typical research strategies used in information systems and

computing research include: survey, design and creation, experiment, case study, and action research [68].

Surveys try to systematically identify patterns in data so as to generalize to a larger population than the group being targeted. Design and creation (commonly called design science) focuses on the development of new information technology artifacts. Experiments prioritize testing hypothesis and investigating cause and effect relationships. Case study aims to obtain a detailed insight of one of the instances of a problem. Action research prioritizes conducting research in a real-world setting.

In our thesis, we use survey, design and creation, and case study as our research strategies. Specifically, for answering RQ1 multiple surveys, and design and creation are employed. Surveys are done to understand the overall distribution of devices, their technical capabilities, and data being collected by them. Design and creation is employed to build a taxonomy of devices, and a data model. For RQ2, we rely on literature survey as a strategy for identifying risks, and a case study focusing on smart connected cameras as a popular IoT device type present in smart homes. We also employ design and creation for RQ2 to develop a threat agent model. For RQ3, we rely on a literature survey to identify challenges and key characteristics of mitigations suitable for IoT-based homes. Experiment and action research, are not applicable as research strategies for answering the posed research questions. However, for future work both can be considered as alternatives or as complementary methods to support our research.

3.3 Data Generation Methods

A data generation method is the means by which empirical data or evidence is produced. Four examples of data generation methods are interviews, observation, questionnaire, and documents [68].

In our case, we rely on documents as our data generation method. However, when possible, we applied triangulation of data by referring to multiple source of evidence to increase reliability and validity of our findings. In future, we may also involve other sources in particular interviews for generating primary data. For instance, specific categories of users, e.g., smart home developers, householders, and security experts,

may be interviewed or surveyed to have an alternative perspective that can help further substantiate our findings and enhance validation.

Multiple document types were investigated in this thesis including: books, reports, journals, conferences and workshop proceedings, newspapers, policies and manuals, and online databases. Books were consulted to gain an initial understanding of smart homes and as a generic reference connected to the security and privacy aspects of this thesis. Reports, in particular penetration testing reports were used to identify real-life vulnerabilities and attacks on IoT devices, and for pinpointing statistics and trends pertaining to smart home technologies. Journals, conference, and workshop proceedings were used to attain updated theories, emerging concepts, and methods used by researchers working on similar domains and research problems. Newspapers were used to find updated information about latest smart home products, services, and trends. Policies, in particular privacy policies, were used to identify smart home data and data collection practices of commercial organizations. Product manuals were investigated to understand the technical composition of actual devices, e.g., in terms of its sensors. Online databases, in particular SmartHomeDB [69], was used as the main repository for collecting information about commercial smart home devices. SmartHomeDB is a comprehensive and community-supported database covering the technical specifications of commercial smart home devices.

To identify relevant documents for the literature review (specifically for Paper I and Paper II that are connected to the state-of-the-art of smart homes) various search terms have been used. In particular, the terms: 'smart home', 'connected home', 'smart home environment', 'intelligent home', 'home automation', 'internet of things', 'smart living', 'pervasive computing', 'ubiquitous computing', 'security', 'privacy', 'risk', 'threat' were rearranged and combined with Boolean operators (in particular 'AND' and 'OR') to retrieve documents from scientific databases. These databases included: IEEE Xplore, JSTOR, ScienceDirect, ACM Digital Library, GoogleScholar, and SpringerLink. Retrieved documents were stored, and later reviewed, and analyzed for relevant information in relation to the research questions. For the rest, i.e., for industry-related and other non-academic literature, Google was used as the main search engine to retrieve those using similar search terms. All

the utilized sources including the type of documents cited are identified in the actual publications (cf. Part II).

3.4 Data Analysis

There are two main paradigms used for analyzing data: quantitative data analysis and qualitative data analysis. Quantitative data analysis uses mathematical techniques such as statistics to examine and interpret data. Qualitative data analysis looks for themes and categories typically within the words or images people use or create.

In answering RQ1, we adopt primarily a quantitative data analysis approach. For the quantitative data analysis, we utilized two main software (statistical) packages: SPSS and R. SPSS was used to compute statistics about the occurrence of IoT devices in each of the identified functional group, and for calculating the distribution of technical capabilities across each smart home application area. The programming language R was used to analyze the privacy policies in terms of their collected data types. For RQ2 and RQ3, we mainly relied on qualitative analysis in particular to enumerate the security and privacy threats and risk scenarios, severity ranking, and capability levels of different threat agents, including state-of-the-art challenges and mitigations. In the future, a quantitative analysis approach may be considered especially as a method to validate our findings. Alas, at the moment, we observe the lack of open IoT databases that can be used for security and privacy research.

The adopted research strategies, data generation methods, and data analysis type for each research question are summarized in Table 3.

Research Question	Research Strategy	Data Generation	Data Analysis
RQ1	Design and Creation Literature Survey	Documents	Quantitative
RQ2	Case Study Design and Creation Literature Survey	Documents Documents	Qualitative
RQ3	Literature Survey	Documents	Qualitative

Table 3. Research methodology summary.

Two main techniques were used to analyze data – coding and content analysis. Coding refers to assigning tags or labels to annotate units of meaning to chunks of collected data, e.g., to words. Content analysis is concerned with the semantic analysis of a body of text to uncover dominant concepts.

In, our thesis, we used coding to detect key smart home functional areas. Specifically, here open coding was used to uncover and name concepts from within data, and then to group them into higher-level categories as are used for the taxonomy. This was implemented through a combination of hand coding and software. Furthermore, content analysis, specifically conceptual analysis was used in different survey studies to examine the presence, frequency, and centrality of concepts, often represented as words, e.g., as unigrams or bigrams in the case of privacy policies.

3.5 Survey

In our thesis, we conducted different types of surveys for each research question. Surveys research can serve different purposes – exploratory, description, or explanation [67]. Exploratory surveys are useful for attaining familiarity with a certain topic of interest. Description surveys focus on finding about the situation, events, attitudes, etc., that are occurring in a population. Explanatory surveys question the relations between variables. In our work, surveys were used for both exploratory purposes, e.g., to uncover challenges and risks in the form of a traditional literature survey, but also for description purposes, e.g., to describe the technical capabilities of devices.

The method for conducting the surveys varied between RQ2 and RQ3 versus that of RQ1. Whereas, in RQ2 and RQ3, we conducted a traditional literature review by examining documents manually; for RQ1, we performed three different technical surveys related to: i) device functionality, ii) device capabilities, and iii) device data; primarily leveraging web and data mining techniques.

The purpose of i) was to identify the total number of distinct smart home functional areas, including the number of devices (and their percentile distribution) for each identified category. Here, a sample size (i.e., the number of entities included in the survey) consisting of 1,193

smart home devices was used. This number represented the entire dataset of smart home devices that were available in the consumer market (as of May 2017) as per the utilized data source (SmartHomeDB).

For ii) the scope was that of identifying the total number of distinct technical capabilities (properties) including their overall distribution with respect to each of the identified functional area. In this survey, as sampling technique cluster sampling was used with the entire dataset used in survey i). Cluster sampling in our case refers to the selection on the basis of device types which might naturally occur together in groups. This was used since we were interested in calculating the capabilities of each functional area (one cluster at a time). The functional clusters were identified in survey i). To extract the actual capabilities of devices an algorithm was implemented in Python to download, preprocess the document, and to transform it into a binary vector with 1 representing a device supporting a capability, and 0 otherwise. In essence, the cumulative output was akin to that a term-document matrix, with rows representing devices, and columns representing capabilities.

In terms of survey iii), the purpose was to identify the type of data being collected by smart home devices, alongside the total number of device types that are associated with each. This survey involved a sample size of 87, and was chosen using purposive sampling. Purposive sampling in a non-probabilistic sample technique where the cases are selected as they possess properties of interest. In our case, this reflected devices of distinct types and that feature the most reviews (negative or positive) by the SmartHomeDB user community.

3.6 Design and Creation

The design and creation strategy aims at developing new artifacts. Types of artifacts commonly include constructs, models, methods, and instantiations [68]. Constructs are concepts or vocabulary used in a particular IT-related domain. Models combine constructs to abstract or represent a situation in such a way that it aids in problem understanding and solution development. Methods provide guidance on the models to be produced and process stages to be followed to solve problems using IT. Instantiations are essentially prototypes or working systems that demonstrate that constructs, models, and methods can be implemented

in a computer-based system. In our thesis, we propose two new constructs, two models, and a method.

A new construct in the form of a taxonomy for classifying home devices was put together in relation to RQ1. We believe that this taxonomy is useful especially for researchers coming from different academic disciplines as a common vocabulary representing the functionality and capabilities offered by smart connected home devices. Related to this, we also introduced the concept of 'smart connected home' in essence as a term to identify homes that leverage IoT technologies.

Two new models were proposed for answering RQ1 and RQ2. First, for RQ1, a model that categorizes the data collected by a smart connected home was laid out. This also identifies the data sources, and whether the user is given a choice in their collection. Secondly, in relation to RQ2, a model that identifies the motivations and capabilities of different malicious threat agents was devised. The model groups the capability levels of different threat agents at an increasing scale using three qualitative labels: "Low", "Moderate", and "High".

In terms of methods, a new method was proposed in relation to RQ1. Specifically, this method combines data mining with manual analysis to build a classification of home devices using actual technical specifications of devices.

3.7 Case Study

A case study is fundamentally an empirical investigation of a contemporary fact or situation within its real-life context [70]. Similar to survey research, there are three types of case studies: exploratory, descriptive, and explanatory [68].

In our thesis, we performed a descriptive short-term case study in relation to RQ2. This was done to attain a rich understanding of the vulnerabilities posed by real-life instances of smart connected cameras. Such devices are popular in smart living spaces, e.g., smart connected homes, and are found all across the world. Cameras were especially interesting to study as images/video feeds are often perceived as the most privacy invasive technologies [71]. In the case study, risks brought about by the introduction of such technologies in homes were discussed, including the identification of different mitigations. As a vulnerability

identification and assessment tool, Shodan was primarily used together with a comprehensive database of security vulnerabilities – Common Vulnerabilities and Exposures (CVE) system [72]. Here, we developed a proof-of-concept application in Python programming language that interfaced with Shodan API. This was built to efficiently identify the total number of smart connected cameras, including metadata being transmitted from them.

To identify the actual severity (risk) levels of each identified vulnerability the National Vulnerability Database (NVD) was utilized. The NVD is a widely used database containing millions of records about software vulnerabilities. Furthermore, it ranks vulnerabilities using qualitative labels, e.g., "Low", "Medium", and "High". In our case, we used the NVD to grade the identified vulnerabilities pertaining to smart connected cameras.

4. CONTRIBUTIONS

4.1 Research Question 1

In order to answer RQ1, we developed two main contributions related to the classification of devices and data of smart connected homes.

First, we proposed a taxonomy of smart home devices that categorizes devices according to their functionality as primary qualifier. This taxonomy is a first attempt at homogenizing the fragmented and multidisciplinary smart home space by utilizing the actual specifications of 1,193 commercial connected devices. In relation to this, we also derived from the data a set of twelve capabilities that can be used to describe and compare smart home devices in a generic way. Details about this contribution can be found in Paper V.

Second, we presented a model that groups together smart connected home data. This model uses as input our proposed taxonomy of devices, categorizing collected data from each functional category in terms of its collection mode, collection method, and collection phase. This model is a first effort at identifying and categorizing the different data types of an entire smart home system using privacy policies as a medium for doing so. Establishing the data collected by smart devices raises more awareness about what is at stake if a device is compromised, and serves as a foundation for building more theoretical research especially in connection to privacy management and as a first step for quantifying privacy risks. Details about this contribution can be found in Paper VI.

4.2 Research Question 2

In order to answer RQ2, we proposed two main contributions related to the identification and assessment of vulnerabilities and risks posed by smart connected home devices.

First, we developed a new threat model for the smart connected home. This model identifies different malicious intruders, risks, and typ-

ical compromise methods used by each threat agent. Identifying the malicious threat agents, including their motivations and capabilities, gives the smart home researchers and developers an alternative approach to reason about risk exposure, discoursing about threats, and as a stepping stone for building effective protection strategies, e.g., in terms of monetary costs, for smart connected homes. Details about this contribution can be found in Paper III.

Second, in order to better assess actual risks with smart connected home devices, we investigated the existing global vulnerability state of Internet-connected smart cameras around the world using free tools. Smart connected cameras are increasingly being targeted by different threat agents. Specifically, we discovered first-hand the kind of data (metadata) that is publicly accessible from networked cameras, and discussed whether the discovered data can cause security and privacy risks to the residents but also to other connected entities. In this investigative study, in line with the existing work, we realized how aggravating the current situation is, especially given the course-grained data, including personal data, that were obtained remotely and with minimal technical skills. Details about this contribution can be found in Paper IV.

4.3 Research Question 3

In order to answer RQ3, we identified the smart connected home characteristics and state-of-the-art challenges and some mitigations related to meeting security and privacy requirements.

In particular, we described the attributes of different type of smart home setups, including their architecture, operation, stakeholders, and data being captured by the different application areas. Furthermore, we identified challenges that adversely affect the design of more generic smart home solutions. In particular, we identified interoperability, security and privacy, reliability, and usability, as the overarching factors that need priority attention from the smart home developers and researchers especially in order to create more resilient security components. More details about this contribution can be found in Paper I.

With regards to the security and privacy challenges we identified characteristics at each architecture layer of the smart connected home that render the design of effective security mechanisms – including tradi-

tional ICT mitigations – particularly challenging to implement or port to smart homes without considerable modification. Additionally, we identified different state-of-the-art mitigations from both academia and industry. However, such approaches all leave behind residual risk(s) raising the need for better end-to-end solutions. More details about this contribution can be found in Paper II and as well in Paper IV which proposes mitigations for both the residents and security vendors related to smart connected cameras.

4.4 Paper Overview

The dissertation is based on six publications in peer-reviewed conferences, workshops, and a book chapter. In Figure 3, we depict the relationship between the different papers.

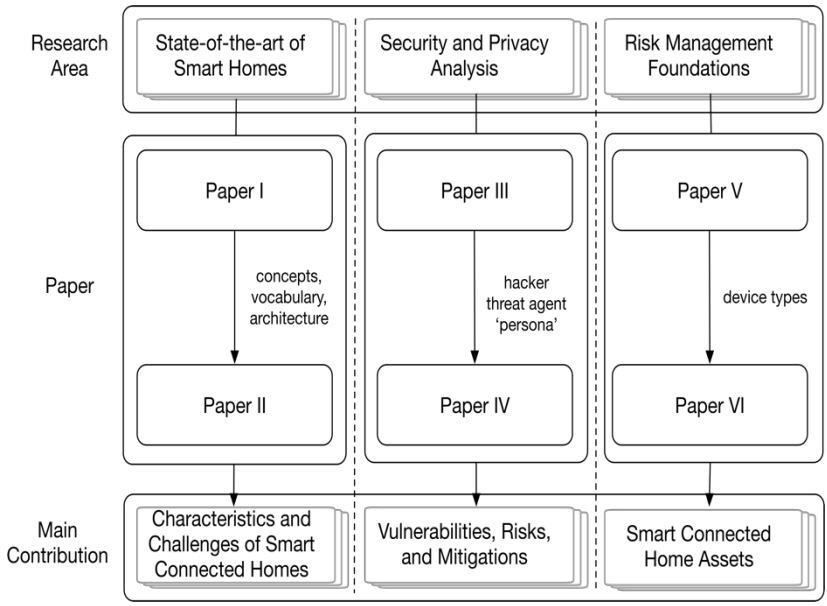

Figure 3. Relationship between the different papers, including their underlying research area and contributions. Paper II, Paper IV, and Paper VI use as input the items (represented as text located to the right of each corresponding arrow) produced by their respective related paper.

Paper I and Paper II contribute towards the identification of key characteristics and challenges brought about by the introduction of

smart connected home technologies. Paper III and Paper IV contribute towards the identification and assessment of vulnerabilities, risks, and mitigations. Paper V and Paper VI are quantitative studies that describe the assets of a smart connected home, with focus on the devices and data.

Overall, our main contributions to the research community are discussed in more detail in Part II of this thesis.

5. DISCUSSION

Homes are a vital structure in our everyday lives. They offer us physical comfort and safety, and a sense of stability and belonging, and are considered a building block for smart cities. In recent years, the traditional brick-and-mortar home has been transformed to a full-fledged IoT-based home – essentially an Internet-connected system that is capable of supporting advanced human-to-machine and machine-to-machine interactions – offering different services allowing for home automation and beyond that. Home owners may now choose from a range of IoT devices that make many aspects of a home "smart" – ranging from smart plugs, smart thermostats, and smart locks. Concurrently, as we discussed in this thesis, these devices are highly vulnerable to security and privacy risks including being subject to surveillance attacks by companies and sometimes governmental agencies, e.g., to assist in court cases and forensic investigations [73] [74] [75].

From our findings in Paper IV we discovered first-hand a plethora (more than half a million) of smart connected cameras broadcasting rich data ranging from port numbers, default passwords, pictures/videos, and so on. The risk of exploiting some of the discovered vulnerabilities could allow complete compromise of security and privacy. While intuitively, one may assume that the risk is only local affecting solely the target household, compromising a multitude of these or other IoT devices may cause havoc and inflict damage possibly on a global level. Indeed, Mirai botnet [76] incident in October 2016 used IoT devices, such as IP webcams, to cause the largest Distributed Denial-of-Service attack ever reported with an unprecedented strength of 1.2 terabits per second. This is especially worrying as most of the discovered devices have been noted to be unhardened, e.g., not updated with the latest security updates. This makes them an easy target for different malicious threat agents from those with the lowest skills to the most capable ones (cf. Paper III).

As a step to better understand the largely fragmented smart connect-

ed home domain we inspected the specification of 1,193 devices in an attempt to organize them in a systematic manner (cf. Paper V). Here, in line with the existing literature, we observed the huge heterogeneity of devices present in the home domain. Devices ranged from those that are intended for the provisioning of 'security and safety', e.g., cloud cameras that are typically stationary and serve a single-purpose, to devices such as those that are part of the 'human-machine interface', e.g., voice command devices, that tend to be mobile and may serve a generic purpose. These different factors complicate risk analysis posing different challenges especially when it comes to estimating the actual risk a device carries and in identifying appropriate, possibly more generic controls. In assisting with this, we derived from the data a set of generic and distinct capabilities that can be used to describe and compare the different devices. Understanding the capabilities is key to recognize distinct channels over which data can be collected and disseminated to various stakeholders, and thereby is critical for discoursing about privacy implications. Here, across the board, we noted first-hand that the majority of the devices tend to use the cloud architecture (66.2%) and support remote access (65.5%) and management typically through a smartphone application (69.6%). This finding is indicative that data has major potential to be transferred out of the home to external entities (and potentially stored indefinitely for future use), and thus with priority attention effort is needed especially to secure outgoing information flows since they may contain instances of personal data of the residents.

Smart home devices tend to be continuously monitoring and collecting data about the occupants and their surrounding environment. Some of the data (cf. Paper VI) are voluntarily provided by the residents, while other data are captured automatically by the devices typically through embedded sensor technology. In some instances, end-users cannot (easily) opt-out from having their data collected. In fact, consent over the collection or use of user data is often buried inside privacy policies, and is commonly accepted silently and unknowingly by users. But even if a user decides not to provide the requested data this sometimes hinders the services provided by the underlying smart home device. Indeed, after GDPR has been put into effect it has been reported repeatedly that some smart devices have suddenly stopped working apparent-

ly as personal data were ceased from being collected [77]. Our study indicates, that all surveyed manufacturers collect instances of personal data. Sometimes, this may be as sensitive as physiological data. While intuitively one may think that this data is only common in the 'health and wellness' domain of the smart home, it is also being used in other domains example to provide enhanced entertainment, e.g., in the case of smart TVs. This makes us reflect first about data ownership especially in the case of sensor data where the data might not be personally identifiable, as are for instance account data that are directly attributable to users. Secondly, it makes us think about the amount of data an organization may have obtained about the occupants especially when combining and comparing the fine-grained and continuous data captured by a smart device versus that gathered through traditional means, e.g., by a website. On the positive side, the acquired data, can contribute to the "smartness" of devices and likewise can help the companies improve their business operations e.g., in the case of smart meters for real-time dynamic pricing, demand forecasting, and power grid operations [78]. Nonetheless, the gathered data, can become a goldmine for hackers or companies, including Internet Service Providers, that want to target specific users or their social environment, e.g., with customized advertisements [79]. At the moment, the residents have no effective controls to identify who has access to the data generated, how their personal data are being used, and are not given the tools to delete collected data.

Traditional security measures that are used to safeguard security and privacy threats in a typical ICT environment are difficult to port to the smart connected home without considerable modifications. In fact, we have discussed several impediments present at different architecture layers of the smart home, e.g., resource constraints, that make the design of truly effective solutions particularly challenging to implement (cf. Paper I and Paper II). There are also other concerns, in particular, multi-tenancy, that was not identified in the surveyed literature but also complicates the design of generic smart home security solutions. Furthermore, beyond the technical challenges, there are many startups and companies operating in the smart home domain that lack the security expertise and resources needed to instill security and privacy at the early stages of their software design lifecycle. This situation inevitably re-

sults in insecure products being released to the market and bares the risk of vendor lock-in for consumers. In the long-term, we argue that regulation, legislation, and product liability are essential ingredients for improving the current state of security of the largely messy and unstandardized smart home. At the same time, we believe that adding security and privacy adds to increased costs and thus making the smart home inaccessible to some users. Thereby, we believe that finding the right balance between extra costs and improved security and privacy is an exercise that must be carefully thought through and justified well to consumers.

6. CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

The demand for smart home technology has increased substantially during recent years. This has attracted different commercial organizations to enter the smart home market offering a multitude of heterogeneous IoT devices. Alongside, the various benefits and convenience factors smart home technologies offer, the domain is advancing rapidly making it increasingly challenging to gain a deeper understanding of the smart home. Such understanding is core to build more robust, resilient, secure, and privacy-preserving smart home systems.

Nonetheless, in connection to this we identified multiple research gaps around two areas. First, when it comes to describing the technical composition of a generic IoT-based smart connected home setup especially in terms of its actual devices and data. Second, in identifying threat agents, security and privacy risks, and mitigations that can bolster the security and privacy of the smart home residents.

For the first area, we noted that most of the available scholarly work assumes a specific type of smart home, e.g., energy-based smart home, focuses on certain categories of devices, e.g., constrained devices, and does not look into the commercial reality of smart home manufacturers, e.g., when it comes to data collection practices. To different extents, this limits the applicability and reusability of existing work when it comes to modelling real-life scenarios. In filling this gap, we provided a generic description of the smart connected home ecosystem in terms of its underlying technologies, architectures, and offered services; presented a taxonomy and quantitative analysis of smart connected home devices based on the actual specification of commercial devices; and conducted an analysis and classification of data collected by smart connected homes using actual policies issued from smart home manufacturers.

For the second area, we observed especially the lack of threat agent methodologies that cater for IoT environments such as the smart con-

nected home, the shortage of studies that elaborate on the technical and social challenges, and as well mitigations intended for the smart home environment. In filling this gap, we developed a new threat agent model for the smart connected home; identified through surveying the literature state-of-the-art security and privacy challenges; investigated actual vulnerabilities by conducting a case study using a proof-of-concept software application; and presented mitigations at different architecture levels in smart connected homes.

Overall, our contributions, in terms of new concepts, models, and methods, together with the attained insights of the smart connected home, lay the groundwork for conducting deeper studies into security and privacy of the smart connected home.

6.2 Future Work

Looking towards the future, there are several avenues that are interesting to pursue. In particular, we seek to develop contributions related to some of the areas below:

User controllable privacy artifact. An interesting idea that can be explored, perhaps through action research, is the design of a privacy "knob" artifact (akin to a light dimmer) that allows the residents the option to tune their desired level of privacy accordingly. Example by turning the control clockwise the smart home system could start disabling or limiting the transmission of certain categories of data, and vice versa. In practice, this can be implemented as a software control on an smartphone device, or otherwise as a dedicated hardware device. Through this control privacy can then be adjusted to tradeoff, possibly automatically, the loss of privacy that comes with sharing data to third parties with the benefit offered by the service that comes from sharing the data.

Proactive network security approach. The idea is to devise an approach that automatically minimizes the risk of security attacks occurring within smart home systems. An effective way of doing so is to develop a network layer solution, working similar to an intrusion detection system used in a traditional ICT environment, that complements existing secu-

rity mechanisms already deployed at the device or service side. This solution can be designed to probe the local network in real-time looking for suspicious activity and simultaneously blocking attacks targeting smart home devices. At a pragmatic level, concepts borrowed from Software-Defined Networking can be explored as a means to filter out network traffic in a flexible manner. Furthermore, machine learning can be leveraged as a method to classify traffic and to dynamically detect unknown attacks.

Smart connected home formal model. As a natural progression of this thesis, we plan to capture the description and behavior of a generic smart connected home including the data it deals with in a formal model. While there is considerable research activity on IoT, such efforts fall short of providing a fundamental formal model that captures the computational and communication properties of a smart connected home system including any information exchanges. Such a model can serve as the basis for the development of robust and provably correct privacy implementations.

Privacy risk assessment method. Risk assessment in the smart connected home is especially complicated because of the heterogeneity of devices, and due to varying perception of risk by different users. Given this, it would be beneficial to develop a method that can be used to augment the description of devices with information about data being collected by them, showing the purpose of such collection, and as well whether that data are being divulged to third parties. Ideally, this method, should also make it possible to enlist common threat scenarios and calculate risk levels posed by different devices, including when they are connected to cloud services over the Internet. In a sense, the development of this method may help the users make conscious choices about using or interconnecting certain smart devices to meet their goals.

BILBIOGRAPHY

[1] M. Weiser, "The computer for the 21st century," *Scientific American*, vol. 265, no. 3, pp. 94–104, 1991.

[2] K. Ashton, "That 'internet of things' thing," *RFID journal*, vol. 22, no. 7, pp. 97–114, 2009.

[3] A. Whitmore, A. Agarwal, and L. Da Xu, "The Internet of Things—A survey of topics and trends," *Information Systems Frontiers*, vol. 17, no. 2, pp. 261–274, 2015.

[4] M. Alaa, A. A. Zaidan, B. B. Zaidan, M. Talal, and M. L. M. Kiah, "A review of smart home applications based on Internet of Things," *Journal of Network and Computer Applications*, vol. 97, pp. 48–65, 2017.

[5] V. Ricquebourg, D. Menga, D. Durand, B. Marhic, L. Delahoche, Christophe, "The Smart Home Concept : our immediate future," presented at the *1st IEEE International Conference on E-Learning in Industrial Electronics*, pp. 23–28, 2006.

[6] E. Fernandes, J. Jung, and A. Prakash, "Security Analysis of Emerging Smart Home Applications," *IEEE Symposium on Security and Privacy*, pp. 636–654, 2016.

[7] Gartner Inc., "Gartner Says 8.4 Billion Connected 'Things' Will Be in Use in 2017, Up 31 Percent From 2016," 2017. [Online]. Available: https://www.gartner.com/newsroom/id/3598917. [Accessed: 08-Aug-2018].

[8] Statista, "Unit shipments of Wi-Fi enabled smart home devices worldwide from 2016 to 2020 (in millions)," 2018. [Online]. Available: https://www.statista.com/statistics/671838/global-wi-fi-enabled-smart-home-device-shipments/. [Accessed: 08-Aug-2018].

[9] Gartner Inc., "Gartner Says a Typical Family Home Could Contain More Than 500 Smart Devices by 2022," 2014. [Online]. Available: https://www.gartner.com/newsroom/id/2839717. [Accessed: 08-Aug-2018].

[10] C. Gao, V. Chandrasekaran, K. Fawaz, and S. Banerjee, "Traversing the Quagmire that is Privacy in your Smart Home," presented at the *IoT S&P'18: ACM SIGCOMM 2018 Workshop on IoT Security and Privacy*, pp. 22–28, 2018.

[11] R. Garg, C. Schmitt, and B. Stiller, "Information Policy Dimension of Emerging Technologies," *SSRN Journal*, 2017.

[12] M. Chan, E. Campo, D. Estève, and J.-Y. Fourniols, "Smart homes — Current features and future perspectives," *Maturitas*, vol. 64, no. 2, pp. 90–97, 2009.

[13] N. King, "Smart home – a definition," 2003.

[14] S. Solaimani, W. Keijzer-Broers, and H. Bouwman, "What we do - and don't - know about the Smart Home: An analysis of the Smart Home literature," *Indoor and Built Environment*, vol. 24, no. 3, pp. 370–383, 2015.

[15] R. Harper, "Inside the smart home: Ideas, possibilities and methods," in *Inside the smart home*, Springer, pp. 1–13, 2003.
[16] D. Hendricks, "The History of Smart Homes," 2014. [Online]. Available: http://www.iotevolutionworld.com/m2m/articles/376816-history-smart-homes.htm. [Accessed: 08-Aug-2018].
[17] K. Gotkin, "When Computers Were Amateur.," *IEEE Annals of the History of Computing*, vol. 36, no. 2, pp. 4–14, 2014.
[18] J. Bugeja, A. Jacobsson, and P. Davidsson, "On Privacy and Security Challenges in Smart Connected Homes," presented at the *IEEE Intelligence and Security Informatics Conference (EISIC)*, pp. 172–175, 2016.
[19] J. Abbate, "Getting small: a short history of the personal computer," presented at the *Proceedings of the IEEE*, vol. 87, no. 9, pp. 1695–1698, 1999.
[20] T. K. L. Hui, R. S. Sherratt, and D. D. Sánchez, "Major requirements for building Smart Homes in Smart Cities based on Internet of Things technologies," *Future Generation Computer Systems*, vol. 76, pp. 358–369, 2017.
[21] S. R. Katre and D. V. Rojatkar, "Home Automation: Past, Present and Future," *International Research Journal of Engineering and Technology IRJET*, vol. 4, no. 10, pp. 343–346, 2017.
[22] T. Yamazaki, "Beyond the Smart Home," presented at the *IEEE International Conference on Hybrid Information Technology (ICHIT)*, vol. 2, pp. 350–355, 2006.
[23] Google Inc., "Nest Thermostat," 2016. [Online]. Available: https://nest.com/thermostat/meet-nest-thermostat/. [Accessed: 08-Aug-2018].
[24] K. Xu, X. Wang, W. Wei, H. Song, and B. Mao, "Toward Software Defined Smart Home," *IEEE Communications Magazine*, vol. 54, pp. 116–122, 2016.
[25] A. J. B. Brush, E. Filippov, D. Huang, J. Jung, R. Mahajan, F. Martinez, K. Mazhar, A. Phanishayee, A. Samuel, J. Scott, and R. P. Singh, "Lab of things: a platform for conducting studies with connected devices in multiple homes," presented at the *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pp. 35–38, 2013.
[26] Apple Inc., "HomeKit," 2018. [Online]. Available: https://developer.apple.com/homekit/. [Accessed: 08-Aug-2018].
[27] A. Herceg, "Defusing the hype in the smart home space," *Reinforced Plastics*, vol. 17, no. 3, pp. 102–104, 2016.
[28] C. D. Kidd, R. Orr, G. D. Abowd, C. G. Atkeson, I. A. Essa, B. MacIntyre, E. Mynatt, T. E. Starner, and W. Newstetter, "The Aware Home: A Living Laboratory for Ubiquitous Computing Research," presented at the *International Workshop on Cooperative Buildings*, pp. 191–198, 1999.
[29] D. J. Cook, M. Youngblood, E. O. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja, "MavHome: an agent-based smart home," presented at the *First IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 521–524, 2003.
[30] S. Helal, W. Mann, H. El-Zabadani, J. King, Y. Kaddoura, and E. Jansen, "The Gator Tech Smart House: A Programmable Pervasive Space," *Computer*, vol. 38, no. 3, pp. 50–60, 2005.
[31] S. S. Intille, "Designing a home of the future," *IEEE Pervasive Computing*, vol. 1, no. 2, pp. 76–82, 2002.

[32] S. S. Intille, K. Larson, E. M. Tapia, J. Beaudin, P. Kaushik, J. Nawyn, and R. Rockinson, "Using a Live-In Laboratory for Ubiquitous Computing Research.," presented at the *4th International Conference on Pervasive Computing*, vol. 3968, pp. 349–365, 2006.
[33] G. Mone, "Intelligent living," *Communications of the ACM*, vol. 57, no. 12, pp. 15–16, 2014.
[34] T. Mendes, R. Godina, E. Rodrigues, J. Matias, and J. Catalão, "Smart Home Communication Technologies and Applications: Wireless Protocol Assessment for Home Area Network Resources," *Energies*, vol. 8, no. 7, pp. 7279–7311, 2015.
[35] S. Pradeep, T. Kousalya, K. M. A. Suresh, and J. Edwin, "IoT and Its Connectivity Challenges in Smart Home," *International Research Journal of Engineering and Technology IRJET*, vol. 3, no. 12, pp. 1040–1043, 2016.
[36] E. Zeng, S. Mare, and F. Roesner, "End User Security & Privacy Concerns with Smart Homes," presented at the *Symposium on Usable Privacy and Security (SOUPS)*, pp. 1–16, 2017.
[37] T. Yumura, K. Akashi, and T. Inoue, "BluMoon: Bluetooth Low Energy Emulation System with Software-Implemented Controller," presented at the *International Workshop on Pervasive Flow of Things (PerFoT 2018)*, pp. 896–901, 2018.
[38] L.-B. Cédric, E. Darra, G. Tétu, G. Dufay, and M. Alattar, "*Security and resilience of smart home environments*," 2015.
[39] C. Bormann, M. Ersue, and A. Karanen, "RFC 7228 - Terminology for Constrained-Node Networks," 2014.
[40] R. Shirey, "RFC 4949: Internet Security Glossary, Version 2," 2007.
[41] Stallings, W., & Brown, L. *Computer Security: Principles and Practice*. Pearson Education. 2015.
[42] A. Mohsen Nia and N. K. Jha, "A Comprehensive Study of Security of Internet-of-Things," *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 4, pp. 586–602, 2016.
[43] G. Rosner, "Privacy and the Internet of Things," pp. 1–62, Oct. 2016.
[44] A. F. Westin, "Privacy and freedom," *Washington and Lee Law Review*, vol. 25, no. 1, p. 166, 1968.
[45] W. Stallings, L. Brown, M. D. Bauer, and A. K. Bhattacharjee, *Computer security: principles and practice*. Pearson Education, 2012.
[46] J. M. M. Rumbold and B. K. Pierscionek, "What Are Data? A Categorization of the Data Sensitivity Spectrum," *Big Data Research*, pp. 1–11, 2017.
[47] European Commission, "2018 reform of EU data protection rules," 2018. [Online]. Available: https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en. [Accessed: 08-Aug-2018].
[48] D. Haynes, "Politics and ethics of metadata," in *Politics and ethics of metadata*, 2nd ed., no. 14, 2018.
[49] S. Creese, M. Goldsmith, J. R. C. Nurse, and E. Phillips, "A Data-Reachability Model for Elucidating Privacy and Security Risks Related to the Use of Online Social Networks," presented at the *IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pp. 1124–1131, 2012.

[50] I. Vujacic, I. Ognajanovic, and R. Sendelj, "Sm@rt Home Personal Security and Digital Forensic Issues," presented at *The Eight International Conference on Business Information Security (BISEC)*, pp. 1–6, 2016.
[51] A. W. Atamli and A. Martin, "Threat-Based Security Analysis for the Internet of Things," presented at the *2014 International Workshop on Secure Internet of Things (SIoT)*, pp. 35–43, 2014.
[52] B. Violino, "IoT pushes IT security to the brink," 2016. [Online]. Available: http://www.csoonline.com/article/3081228/internet-of-things/iot-pushes-it-security-to-the-brink.html. [Accessed: 08-Aug-2018].
[53] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the Internet of Things: perspectives and challenges," *Wireless Networks*, vol. 20, no. 8, pp. 2481–2501, 2014.
[54] M. Dorodchi, M. Abedi, and B. Cukic, "Revisiting Trust of Integrating IoT in Enterprise," presented at *the PerIoT - Second International Workshop on Mobile and Pervasive Internet of Things*, pp. 540–545, 2018.
[55] A. Amini, N. Jamil, A. R. Ahmad, and M. R. Z. aba, "Threat Modeling Approaches for Securing Cloud Computing," *Journal of Applied Sciences*, vol. 15, no. 7, pp. 953–967, 2015.
[56] A. Shostack, *Threat Modeling: Designing for Security*. John Wiley & Sons, Inc., 2014.
[57] G. Martins, S. Bhatia, X. Koutsoukos, K. Stouffer, C. Tang, and R. Candell, "Towards a systematic threat modeling approach for cyber-physical systems," presented at the *IEEE 2015 Resilience Week (RWS)*, 2015, pp. 1–6.
[58] M. Rosenquist, "Prioritizing Information Security Risks with Threat Agent Risk Assessment," *Intel Corporation White Paper*, pp. 1–8, 2009.
[59] Verizon, "2018 Data Breach Investigations Report." 2018. [Online]. Available: https://www.verizonenterprise.com/verizon-insights-lab/dbir/. [Accessed: 08-Aug-2018].
[60] M. Mateski, C. M. Trevino, C. K. Veitch, J. Michalski, J. M. Harris, S. Maruoka, and J. Frye, "Cyber Threat Metrics," *Sandia National Laboratories*, 2012.
[61] R. W. Beggs, *Mastering Kali Linux for Advanced Penetration Testing*. Packt Publishing Ltd., 2014.
[62] "Shodan." [Online]. Available: https://www.shodan.io/. [Accessed: 08-Aug-2018].
[63] "Censys," *https://censys.io/*. [Online]. Available: https://censys.io/. [Accessed: 08-Aug-2018].
[64] B. Genge and C. Enăchescu, "ShoVAT: Shodan-based vulnerability assessment tool for Internet-facing services," *Security and Communication Networks*, vol. 9, no. 15, pp. 2696–2714, 2016.
[65] J. Matherly, "The Complete Guide to Shodan: Collect. Analyze. Visualize. Make Internet Intelligence Work for You.," pp. 1–70, 2016.
[66] J. W. Creswell, "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches," pp. 1–342, 2015.
[67] D. Haynes, "Metadata for Information Management and Retrieval,". Facet Publishing. 2012.
[68] B. J. Oates, *Researching Information Systems and Computing*. SAGE Publications, pp. 1–326, 2006.

[69] "Smart Home DB - The smart home database," [Online]. Available: http://www.smarthomedb.com. [Accessed: 08-Aug-2018].
[70] R. K. Yin, *Case study Research: Design and methods*, 3rd ed., vol. 5. Sage Publications, 2003.
[71] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring Activities of Daily Living in Smart Homes: Understanding human behavior," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 81–94, 2016.
[72] D. Papp, Z. Ma, and L. Buttyan, "Embedded systems security: Threats, vulnerabilities, and attack taxonomy," presented at the *2015 13th Annual Conference on Privacy, Security and Trust (PST)*, pp. 145–152, 2015.
[73] K. Hill, "Police have asked Dropcam for video from people's home cameras," 2015. [Online]. Available: https://splinternews.com/police-have-asked-dropcam-for-video-from-peoples-home-c-1793845407. [Accessed: 08-Aug-2018].
[74] K. Hill, "Fitbit data just undermined a woman's rape claim," 2015. [Online]. Available: https://splinternews.com/fitbit-data-just-undermined-a-womans-rape-claim-1793848735. [Accessed: 08-Aug-2018].
[75] H. Chung, J. Park, and S. Lee, "Digital forensic approaches for Amazon Alexa ecosystem," *Digital Investigation*, vol. 22, pp. S15–S25, 2017.
[76] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and Other Botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017.
[77] J. Cook, "'Smart' home devices stop working after European data law comes into force," 2018. [Online]. Available: https://www.telegraph.co.uk/technology/2018/05/25/smart-home-devices-stop-working-european-data-law-comes-force/amp/. [Accessed: 08-Aug-2018].
[78] N. Saputro, A. Yurekli, K. Akkaya, and A. Uluagac, "Privacy Preservation for IoT Used in Smart Buildings," in *Security and Privacy in Internet of Things (IoTs)*, CRC Press, pp. 135–166, 2016.
[79] B. Habegger, O. Hasan, L. Brunie, N. Bennani, H. Kosch, and E. Damiani, "Personalization vs. Privacy in Big Data Analysis," *International Journal of Big Data*, vol. 1, no. 1, pp. 25–35, 2014.

References

1. Bugeja, J., Jacobsson, A., Davidsson, P.: On privacy and security challenges in smart connected homes. In: Proceedings of the IEEE Intelligence and Security Informatics Conference (EISIC), pp. 172–175 (2016)
2. Ahlam, A., Laila, B., Slimane, B.: An overview of privacy preserving techniques in smart home wireless sensor networks. In: Proceedings of the IEEE 10th International Conference on Intelligent Systems Theories and Applications (SITA), pp. 1–4 (2015)
3. Apthorpe,N.,Reisman,D.,Sundaresan,S.,Narayanan,A.,Feamster,N.:Spyingonth eSmart Home: Privacy Attacks and Defenses on Encrypted IoT Traffic (2017). arXiv preprint arXiv: 1702.03681
4. Seralathan,Y.,Oh,T.T.,Jadhav,S.,Myers,J.,Jeong,J.P.,Kim,Y.H.,Kim,J.N.:IoTse curity vulnerability: a case study of a web camera. In: Proceedings of the IEEE 20th International Conference on Advanced Communications Technology (ICACT), pp. 172–177 (2018)
5. Boztas,A.,Riethoven,A.R.J.,Roeloffs,M.:SmartTVforensics:digitaltracesontelevi sions. Digital Invest. 12, S72–S80 (2015)
6. Anscombe, T.: IoT and Privacy By Design in the Smart Home. https:// www.welivesecurity.com/wp-content/uploads/2018/02/ ESET_MWC2018_IoT_SmartHome.pdf. Accessed 06 May 2017
7. Ziegeldorf, J.H., Morchon, O.G., Wehrle, K.: Privacy in the Internet of Things: threats and challenges. Secur. Commun. Netw. 7(12), 2728–2742 (2014)
8. Massey, A.K., Eisenstein, J., Anton, A.I., Swire, P.P.: Automated text mining for requirements analysis of policy documents. In: Proceedings of the Requirements Engineering Conference (RE) (2013)
9. Schaub, F., Balebako, R., Cranor, L.F.: Designing effective privacy notices and controls. IEEE Internet Comput. 21(3), 70–77 (2017)
10. Breaux,T.D.,Hibshi,H.,Rao,A.:Eddy,aformallanguageforspecifyingandanalyzin gdata flow specifications for conflicting privacy requirements. Requirements Eng. 19, 281–307 (2013)
11. Zimmeck, S., Wang, Z., Zou, L., Iyengar, R., Liu, B., Schaub, F., Wilson, S., Sadeh, N., Bellovin, S.M., Reidenberg, J.: Automated analysis of privacy requirements for mobile apps. In: Proceedings of the Network and Distributed System Security (NDSS) Symposium (2017)
12. Alohaly,M.,Takabi,H.:Better privacy indicators: a new approach to quantification of privacy policies. In: Proceedings of the WPI SOUPS (2016)
13. Bhatia, J., Breaux, T.D.: Towards an information type lexicon for privacy policies. In: Proceedings of the IEEE Eighth International Workshop on Requirements Engineering and Law (RELAW) (2015)
14. Bhatia, J., Evans, M.C., Wadkar, S., Breaux, T.D.: Automated extraction of regulated information types using hyponymy relations. In: Proceedings of the IEEE 8th International Requirements Engineering Conference Workshops (REW), pp. 19–25 (2016)
15. Costante, E., Hartog, den, J., Petkovic, M.: What websites know about you⋆ privacy policy analysis using information extraction. In: Data Privacy Management and Autonomous Spontaneous Security, pp. 146–159 (2013)

16. Costante, E., Sun, Y., Petkovic, M., Hartog, den, J.: A machine learning solution to assess privacy policy completeness. In: Proceedings of the ACM Workshop on Privacy in the Electronic Society (2012)
17. Liu, F., Ramanath, R., Sadeh, N., Smith, N.A.: A step towards usable privacy policy: automatic alignment of privacy statements. In: Proceedings of the 25th International Conference on Computational Linguistics (COLING) (2014)
18. Zimmeck,S.,Bellovin,S.M.:Privee: an architecture for automatically analyzing web privacy policies. In: Proceedings of the USENIX Security Symposium (2014)
19. Sadeh, N., Acquisti, A., Breaux, T.D., Cranor, L.F., McDonald, A.M., Reidenberg, J.R., Smith, N.A., Liu, F., Russell, N.C., Schaub, F., Wilson, S.: The Usable Privacy Policy Project: Combining Crowdsourcing, Machine Learning and Natural Language Processing to Semi- Automatically Answer Those Privacy Questions Users Care About. Carnegie Mellon University (2013)
20. Zhang,L.-J.,Li,C.:Internet of Things Solutions. Services Transactions on Internet of Things (STIOT) 1, 1–22 (2017)
21. Habegger, B., Hasan, O., Brunie, L., Bennani, N., Kosch, H., Damiani, E.: Personalization vs. privacy in big data analysis. Int. J. Big Data (IJBD) 1, 25–35 (2014)

ISBN 978-91-7104-929-2 (print)
ISBN 978-91-7104-930-8 (pdf)

MALMÖ UNIVERSITY
205 06 MALMÖ, SWEDEN
WWW.MAU.SE