Automated Risk Assessment for Cyber Threats Identification in IoT Environments

Collen, Anastasija

Automated Risk Assessment for Cyber Threats Identification in IoT Environments

(Analyse Automatisée des Risques pour l'Identification des Cyber-menaces dans un Environnement IoT)

THÈSE

présentée à la Faculté d'Economie et de Management
de l'Université de Genève, Information Science Institute
par

Anastasija COLLEN

sous la codirection de
Dr. Niels A. NIJDAM
Prof. Dimitri KONSTANTAS

pour l'obtention du grade de
Docteur ès Economie et Management
mention *Systèmes d'Information*

Membres du jury de thèse:

Prof. Katarzyna WAC, Président du jury, Universitè de Genéve
Prof. Dimitri KONSTANTAS, Universitè de Genéve
Dr. Niels A. NIJDAM, Universitè de Genéve
Prof. Jean-Henry MORIN, Universitè de Genéve
Prof. Sokratis KATSIKAS, Norwegian University of Science and Technology (NTNU)

Thèse N° 107

Genève, le 13 mai 2022

La Faculté d'économie et de management, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par-là, émettre aucune opinion sur les propositions qui s'y trouvent énoncées et qui n'engagent que la responsabilité de leur auteur.

Genève, le 13 mai 2022

Dean
Markus MENZ

“Знание только тогда знание, когда оно приобретено усилиями своей мысли, а не памятью.”

Лев Толстой

“В жизни есть только один путь: знание. Знание раскрепощает человека.”

Конкордия Антарова

Acknowledgements

First of all I would like to thank for the endless support my thesis supervisors, Dr. Niels NIJDAM and Prof. Dimitri KONSTANTAS. I am very grateful for all the feedback and expertise you provided, which helped to shape the direction of this research and strengthened the final results of my work.

I also would like to sincerely thank the jury members, Prof. Katarzyna WAC, Prof. Sokratis KATSIKAS and Prof. Jean-Henry MORIN for their valuable time, guidance and discussions to advance and finalise my work.

I would like to express my gratitude to all my academic colleagues who directly and indirectly motivated and encouraged me in this long journey: Dr. George SPATHOULAS, Dr. Pankaj PANDEY, Prof. Bela GENGE, Javier AUGUSTO-GONZALEZ, Roland BOLBOACA, Prof. Melanie VOLKAMER, Maher BEN MOUSSA, Dr. Grigorios ANAGNOSTOPOULOS, Dr. Marios FANOURAKIS, Alexandre DE MASI, Dr. Allan BERROCAL and Meriem BENYAHYA.

I am indebted to my husband, Cyril COLLEN, who is always there for me, listening and advising, supporting and encouraging not to give up and always go forward. I could not have accomplished this academic step without him.

Nyon, May 2022
Anastasija COLLEN

Abstract

Internet of Things (IoT) enabled systems are steadily expanding their presence in all facets of industry and consumer lives. They enable regular citizens, consumers and manufacturers to easily interact with the digital world. A plethora of composing IoT objects is gradually employed in almost every domain. As their normal operation is becoming critical for society, abnormal behaviour of such system's composing elements poses significant implications – cyber risks – for their end-users, related to financial loss, privacy violation, critical services' outage or even human lives endangering. This is where a well established field of the Risk Assessment (RA) becomes indispensable. It studies various aspects of the identification of hazards and threats, analysis of their causes and consequences, and representation of the corresponding risks for further decision-making based on derived probabilities of encountered uncertainties. RA in Information Security shares a common notion of a future prediction, necessary to be equipped with, to understand the risk in a given situation. While traditionally performed in a static way, where analysis operates on historical and snapshot data of today, it is widely accepted that the future of RA relies on the Dynamic Risk Assessment (DRA) with conditions monitoring.

In this work, a complete framework on the DRA is applied to one of the most prominent examples of conjunction of physical and digital worlds – smarthomes. Stipulated by the studies on the challenges associated with the mobile properties of the IoT objects, we have conceptualised the representation of the generic object model – IoT Stack – and applied it to the DRA. On this side, our work began with the definition of the theoretical foundation for the establishment of the RA and its application in IoT environments through the evolution of the reference architecture from conceptualisation to deployment in real settings. Governed by the constantly evolving user and functional requirements, we have designed a complete workflow from data capture and network analysis to anomaly detection and operational DRA. It was further extended with the usability focused visualisation of the user interfaces for control and monitoring to support the decision-making process. Constant evolution of those requirements also shaped the input and output interfacing of the DRA, shifting the initial focus of behaviour comparison to anomaly processing integration into the RA process. We have integrated support of the real-time adjustment of the deployment infrastructure for a stronger system-level resilience. Finally, this work explored the possibilities to eliminate human interference in the RA process, aiming to develop a high level of automation for the decision-making to mitigate the confronted risks.

Bound by operating in the IoT environment, we faced the associated constraints and limitations on hardware and software level of the IoT objects. Automation, not always being possible or even desirable by lay users due to their risk perception, proved to be of crucial importance in the decision-making process. The DRA framework provides the tools for understanding, monitoring and addressing the risks encountered in the digital arena of our lives.

Résumé

Les systèmes de l'internet des objets (IdO) ne cessent d'étendre leur présence dans toutes les facettes de l'industrie et de la vie des consommateurs. Ils permettent aux utilisateurs réguliers, des consommateurs et des fabricants d'interagir facilement avec le monde numérique. Une pléthore des objets IdO est progressivement utilisée dans presque tous les domaines. Comme leur fonctionnement devient critique pour la société, un comportement anormal des éléments qui composent ces systèmes a des implications importantes - les cyberrisques - pour leurs utilisateurs finaux tels que des pertes financières, des violations de la vie privée, des interruptions de services critiques ou même des mises en danger de vies humaines. C'est là qu'un domaine bien établi, l'évaluation des risques (ER), devient indispensable. Elle étudie les différents aspects de l'identification des dangers et des menaces ainsi que l'analyse de leurs causes, de leurs conséquences et la représentation des risques correspondants pour une prise de décision ultérieure basée sur la probabilité des incertitudes rencontrées. L'ER en sécurité de l'information partage une notion commune de prédiction, dont il faut être équipé, pour comprendre les risques dans une situation donnée. Alors que l'analyse est traditionnellement effectuée de manière statique sur la base de données historiques et instantanées, il est largement admis que l'avenir de l'évaluation des risques repose sur l'évaluation dynamique des risques (EDR) avec une surveillance évoluée.

Durant ce travail, un cadre de gestion complet sur l'EDR est appliqué à l'un des exemples les plus proéminents des mondes physique et numérique : les maisons intelligentes. Stipulés par les études sur les défis associés aux propriétés mobiles des objets IdO, nous avons conceptualisé la représentation d'un modèle générique - IdO Stack - et nous l'avons appliqué à l'EDR. Notre travail a commencé par la définition des bases théoriques pour la mise en place de l'ER et de son application dans les environnements IdO avec l'évolution de l'architecture de référence depuis la conceptualisation jusqu'au déploiement dans des contextes réels. Orientés par l'évolution constante des besoins des utilisateurs et des exigences fonctionnelles, nous avons conçu un processus complet allant de la capture des données et de l'analyse du réseau à la détection des anomalies et à l'ER opérationnelle. Il a été étendu avec la visualisation axée sur des interfaces utilisateurs utilisable pour le contrôle et la surveillance afin d'encourager le processus de prise de décision. L'évolution constante de ces exigences a également façonné l'interface d'entrée et de sortie de l'EDR, en changeant l'objectif initial de la comparaison des comportements vers l'intégration du traitement des anomalies dans le processus de l'ER. Nous avons intégré le support de l'ajustement en temps réel de l'infrastructure de déploiement

pour une meilleure résilience au niveau du système. Enfin, ce travail a exploré les possibilités d'éliminer l'interférence humaine dans le processus de l'ER, visant à développer un haut niveau d'automatisation pour la prise de décision afin d'atténuer les risques rencontrés.

En opérant dans l'environnement IdO, nous avons été confrontés aux contraintes et limitations associées au niveau matériel et logiciel des objets IdO. L'automatisation, qui n'est pas toujours possible ou même souhaitable pour les utilisateurs réguliers en raison de leur perception des risques, s'est avérée d'une importance cruciale dans le processus décisionnel. Le cadre de l'EDR fournit les outils nécessaires pour comprendre, surveiller et traiter les risques rencontrés dans nos vies du monde numérique.

Резюме

Системы с поддержкой «интернета вещей» неуклонно расширяют свое присутствие во всех сферах промышленности и жизни потребителей. Они позволяют обычным пользователям и производителям легко взаимодействовать с цифровым миром. Всё больше и больше «умных» устройств используется в разных областях деятельности. Поскольку их нормальная работа становится критически важной для общества, аномальное поведение составных элементов таких систем создает значительные последствия для их конечных пользователей, кибер риски, связанные с финансовыми потерями, нарушением конфиденциальности, прекращением работы критически важных сервисов или даже угрозой для жизни людей.

Именно здесь становится незаменимой такая хорошо развитая сфера научных исследований, как оценка рисков (ОР). Она изучает различные аспекты идентификации опасностей и угроз, анализа их причин и последствий,выявляя соответствующие риски для дальнейшего принятия решений на основе рассчитанных вероятностей встречающихся неопределенностей. ОР в сфере информационной безопасности сопряжено с понятием прогнозирования будущего, которое помогает оценить риск в конкретной ситуации. Несмотря на то, что ОР традиционно выполняется статическим способом, производя анализ, оперируя статистическими и фактическими данными, широко признано, что будущее ОР зависит от динамической оценки рисков (ДОР) в совокупности с мониторингом состояний.

В данной докторской диссертации целостная система ДОР применяется к одному из наиболее ярких примеров соединения физического и цифрового миров - умному дому. Исходя из исследований проблем, связанных с мобильными свойствами умных объектов, мы разработали концепцию представления универсальной объектной модели - IoT Stack - и применили ее к ДОР. С этой точки зрения, наша работа началась с определения теоретической основы для создания ОР и ее применения в среде «интернета вещей» с помощью эволюции базовой архитектуры, от проектирования до внедрения в реальных условиях. Руководствуясь постоянно меняющимися пользовательскими и функциональными требованиями, мы разработали полный рабочий процесс от сбора данных и анализа трафика сети до обнаружения аномалий и функционирующей системы ДОР. Далее,этот процесс был расширен за счет визуализации пользовательских интерфейсов для управления и мониторинга, ориентированных на удобство использования и на поддержку процесса принятия решений. Постоянное развитие этих требований также сформировало интерфейс ввода и вывода ДОР, сместив первоначальный фокус сравнения поведения на интеграцию обработки аномалий в процесс ОР. Также, для повышения устойчивости к кибер атакам, на уровне системы была

внедрена поддержка корректировки инфраструктуры в режиме реального времени. В заключение, в данной работе исследовались возможности устранения человеческого вмешательства в процесс ОР, направленные на развитие высокого уровня автоматизации принятия решений для предотвращения возникающих рисков.

Работая в среде «интернета вещей», мы столкнулись с соответствующими ограничениями и сдерживающими факторами на уровне аппаратуры и программного обеспечения умных предметов. Автоматизация, которая не всегда возможна или даже желательна для рядовых пользователей из-за их восприятия рисков, оказалась крайне важной в процессе принятия решений. Система ДОР предоставляет средства для понимания, мониторинга и устранения рисков, возникающих на цифровой арене нашей жизни.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Overview

Chapter Contents

1.1 Context

"One of the main cyber risks is to think they don't exist. The other is to try to treat all potential risks." — Stephane Nappo

Modern world is essentially built around constant interchange of digital and physical artefacts, which are influencing our lives whether we want to acknowledge it or not. For some of us cybersecurity is just a distant notion, which somehow should be addressed in a mysterious and invisible way. For some others, we take it so seriously, that we are voluntarily recalling many digital activities and use of the modern technologies, unless we thoroughly study them and ensure appropriate safeguarding elements are put in place for its safe usage. Most of the time, however, we end up in some kind of compromised sacrifice, when we have to make a choice with a very limited knowledge, mostly deciding subconsciously whether the promised functionality is worth the risk taken or not. But what is this risk? What it is composed of? And how this risk can eventually cause any harm that can affect our life?

Starting with the Oxford dictionary definition, the risk is "the possibility of something bad happening at some time in the future; a situation that could be dangerous or have a bad result" [1]. This implies that to understand the risk, one needs to also have knowledge on those possibilities in the future and have some examples of those situations. But how can you gain such knowledge?

If we look at the legal definition, it states that risk is "the potential danger that threatens to harm or destroy an object, event, or person." [2]. Here we can see the same aspect of possible event to happen in the future cause harm. Once again, to identify associated risks to a specific action, one needs to have a deep understanding of the context, influencing factors and the scope of the danger to occur in the future.

Finally, computer security general definition notes a risk as "a measure of the extent to which an entity is threatened by a potential circumstance or event, and typically a function of: (i) the adverse impacts that would arise if the circumstance or event occurs; and (ii) the likelihood of occurrence. " [3]. This definition is more elaborated, including not only future potential event to occur, but also corresponding impacts and actual likelihood of this event.

What a careful reader can observe is a common notion of a future prediction, necessary to be equipped with, to understand the risk in a given situation. The digital world, similarly to a physical, is composed of many unknown factors, events and actions. In a physical world we often make decisions based on our intuition, experience and general common sense. Is it legit to apply the same logic in a digital world? And how do we obtain a necessary knowledge to transfer such skills in a digital arena?

This is where a well established field of Risk Assessment (RA) becomes indispensable. It studies various aspects of the identification of hazards and threats, analysis of their causes and consequences, and representation of the corresponding risks for further decision-making based on derived probabilities of encountered uncertainties. While traditionally performed in a static way, where analysis operates on historical and snapshot data of today, it is widely

accepted that the future of RA relies on the Dynamic Risk Assessment (DRA) with conditions monitoring [4, 5]. Furthermore, security and safety of the evolving digital systems of systems will pave the way to the development of novel RA methodologies and frameworks [4].

In this thesis we apply RA to the Internet of Things (IoT) environments, representing one of the most prominent examples of conjunction of physical and digital worlds [6]. This implicates studying challenges associated to the mobile properties of the IoT objects. The real-time readjustment of the deployment infrastructure brings us to the aspiration to eliminate the human interference on the RA process, aiming to develop a high level of the automation for the decision-making to mitigate the confronted risks. Besides, while operating in the IoT environment, we also face the associated constraints and limitations on hardware and software level of the objects. Automation, not always being possible or even desirable, rolls back to the actual endures, and, therefore, the aspects associated with risk perception in such environment are also of crucial importance, as they dictate the decision-making process. Finally, the RA process embraces the merging of two highly correlated notions on security and privacy, by addressing only the final results of one or other being violated. This is achieved by starting at the threat detection level, passing through RA and deriving the impact projection for informative decision-making.

1.2 Motivation

The ultimate goal of this work is to put expert knowledge on cybersecurity and privacy in the hands of a typical lay user, having limited technology understanding. We believe that doing so through the automation of the RA and visualisation of the decision-making is the most promising approach, as it creates awareness for the regular digital world inhabitants through the understanding of the consequences of the digital actions. Risk comprehension is an eventual fuse of the cybersecurity and privacy protection techniques but from the user-centric perspective.

RA is a process enabling the identification, estimation and prioritisation of risks associated in four different dimensions where the harm can be caused [5]:

- Activity: a specific action under evaluation;
- Operation: a process already in place;
- Subject: object or human that can be endangered;
- Environment: impact projection for the defined area.

Risks can be evaluated for each dimension in isolation or in mixed setup, where various combinations are possible. RA sets the foundations for the following risks evaluation stages [7]:

- Acceptance: acknowledgement of the possibility of the risk to occur in a specific setup, and taking the responsibility of dealing with the caused consequences.

- Mitigation: taking actions to limit the exposure of the risk and its consequences, by controlling and limiting its occurrence.
- Transfer: delegation or propagation of the risk occurrence to a third party, capable of taking responsibility and liability of the risk's consequences.
- Avoidance: ignorance of the risk occurrence likelihood and assumption of risk non-existence, as evidence of its occurring is too low or the associated cost of mitigation and transfer is too high.

When evaluating the risk stages and associated measurements costs, the RA is relying on five main variables [8]:

- Assets: any items of value (infrastructure or reputation);
- Vulnerabilities: how to exploit assets;
- Threats: action to exploit vulnerability (deliberate or accidental);
- Attack likelihood: probability of threat;
- Impact: estimation of the attack consequence.

The importance of the above mentioned variables are producing variations of the RA approaches and models. The asset-centric models, such as OCTAVE[1], are evaluating the impact of the risks occurrences. The threat-centric models, such as NIST SP800-30[2], are focused on the risk occurrence feasibility. Furthermore, the risks can be measured in two ways: qualitatively and quantitatively [9]. While the first method appears to be very simple, time and cost effective, it is also known to be not precise without impact measurement, as it uses non-numeric values as descriptive results. On the contrary, quantitative methods give a numeric probability, enabling easy measurement of the impact. However, its complex modelling relies on the historical data and, therefore, cannot provide values at loss at a particular time, especially for the risks that never occurred before. The hybrid RA methodologies are aiming to address those shortcomings by including user-centric concepts in traditional RA models, where the following properties are included [10]:

- Human System Integration: visual representation of the system;
- Interoperability identification: considerations towards dependencies;
- Emergent behaviour evaluation: coupling systems for a new purpose.

[1] https://www.enisa.europa.eu/topics/threat-risk-management/risk-management/current-risk/risk-management-inventory/rm-ra-methods/m_octave.html
[2] https://www.nist.gov/privacy-framework/nist-sp-800-30

Understanding of the operational environment is a crucial attribute in a complex decision-making and dynamic environments where IoT devices get frequently replaced, added, removed, updated and moved around [11]. In such environments the emphasis is given to the collection and projection of various contextual factors, as well as time and space specific data analysis. Any of the above-mentioned models are based only on the periodic assessment with limited system knowledge and lack of dynamic adaptation of the evolving situational risks, where the user is a key element in future risks projections.

Several works exist in the domain of the RA aiming to address this shortcoming with partial automation achievement. For instance, risk assessment framework [12] in IoT systems had been developed with periodic risk assessment. The main reasoning for such an approach is the limitation on the system knowledge and dynamic adaptation due to the lack of understanding of risk propagation and dependencies between different assets. MS STRIDE and DREAD application for risk modelling described by Nurse, Creese, and De Roure [8], widely used in RA, attempts to solve the automation characteristic, yet still relies on a completely manual approach. This is where automated RA, so called Risk Assessment Engine (RAE) is bringing its innovation capacity and implements theoretical advancements into real-life applications [13]. It not only integrates the human-centric aspects into the risk assessment model, but also empowers its users to perform dynamic near-to-real-time risk assessment on constantly evolving situational risks.

IoT, which has attracted considerable attention during the last decade, presents a huge opportunity for many industrial and business stakeholders in various domains [14]. As an emerging technology, IoT is prone to cyber attacks. The demands for the countermeasures for the protection of such ecosystems are constantly growing. The heterogeneity and diversity of the "Things", as well as new lightweight communication protocols appropriate for IoT technology, create new challenges for the protection of such systems. Emergence of this field inspired the development of the European Union Horizon 2020 Research and Innovation funded project – GHOST – Safe-Guarding Home IoT Environments with Personalised Real-time Risk Control (https://www.ghost-iot.eu/). Being a major contributor and concept architect of this project, permitted to establish the initial basis of this thesis dissertation topic and research direction. Most importantly, we aimed at developing a reference architecture for securing the smarthome's IoT ecosystem. The multi-layer solution integrates traditional cybersecurity countermeasures, while it introduces new mechanisms for the efficient defence of common IoT threats through the establishment of the dynamic real-time risk assessment.

1.3 Problem

Generic smarthome classification, presented in [15], gives an excellent overview of the variety of the IoT devices that can be present in any smarthome installation. The difficulty here lies in the heterogeneous nature of IoT devices to be monitored and analysed during RA. The spectrum of the threats for smarthomes is twofold, privacy and security related. Very often it

is difficult to draw a strict line for the classification, as the attacks are targeting to exploit both vectors. Unfortunately, primitive security settings, such as keeping default passwords, are still being ignored by unaware users [16]. According to [17], multiple occurrences of the performed attacks with the help of IoT devices were aiming to disrupt the operational performance of large organisations, exploiting primitive attack vectors such as the use of default passwords or weak communication protocols. More advanced attacks were performed by research in [18], able to distribute and propagate a worm using simple Philips Hue smart lamp while driving in a car past 70 metres away from the target building. The most powerful example was the appearance of the Mirai botnet [19], taking over at least 100 000 IoT devices. It is evident that regular citizens have no way to gather a full picture of the potential risks involved in the smarthome they are living in, and that an automatic monitoring solution is essential.

The main goal of the DRA is to provide real-time security and privacy risk assessment of the ongoing activities on the network. It validates the current communication by assessing any anomaly detected and deviations in the IoT device's normal behaviour through device profiles. DRA performs real-time risk assessment by continuous evaluation and monitoring of various risk levels at multiple stages of data processing. To control the behaviour of devices and corresponding payload exchanges, permitted risk levels of ongoing network activity are dynamically calculated for each network activity event, practically determining the required decision to be taken. RA is also establishing the risk controls for the users' privacy and making them aware of the associated risks. DRA integrates multi-faceted anomaly detection analysers and risk receptors to support behaviour deviation detection, involving deep understanding of risk propagation and interdependencies within the network. For this purpose, it leverages the existing open threat modelling tools (eg the Open cyber threat intelligence platform) to integrate a network entity correlation ontology. Furthermore, an initial set of expert values for the risk estimation was established to allow comparison of associated impacts in various risk situations.

- RQ1: Can a generic ontology be developed to capture complex relationship between heterogeneous IoT properties to encapsulate vulnerabilities, attack attribution, impact evaluation, and mitigation strategies?
- RQ2: Can a unique risk scoring be developed to eliminate context dependency? How the initial setting of the expert values for the risk assessment offer a valid approach and whether these values are generally applicable in a standard installation?
- RQ3: What are the limitations on the automated decision-making for risk assessment in dynamic environments, such as smarthomes, where deployed IoT devices constantly evolve (get replaced, updated and moved)?
- RQ4: To what extent DRA can be performed in real-time considering evolving multidimensional situational risks, such as human behaviour change, emerging attack vectors and a dynamic IoT ecosystem?

1.4 Significance

Putting efforts towards generation of the easily available knowledge, understanding and control over the IoT objects activities in privacy crucial environments, such as our homes, is of topmost priority. The establishment of a generic ontology on the threats-risks-objects relation will be of notable importance for the future research on the DRA in any kind of IoT enabled system, ranging from smarthomes to smartcities. It will enable open and standardised approach on describing the existing knowledge of cybersecurity experts to extract concise, repeatable and explainable results on risk projection and mitigation. Furthermore, by advancing on the formation of the unique risk scoring system, we generalise the initialisation process of the RA framework deployment in any kind of Information Technology (IT) environment. Proving that automation is possible and demonstrating it in real-life deployments, we advance the general awareness of the citizens towards cybersecurity and privacy problems. We equip them with a tooling capable of returning their freedom on the choice, where daily sacrifices will slowly fade in the past.

1.5 Limitations

The main limitation of the research performed under the scope of this thesis is the lack of the data for the developed framework validation.

Due to COVID-19 situation and raised constraints to execute trials, the scope of the validation was reduced to deliver the GHOST project's results on time. The responsible partners were only able to create and collect a minimalistic data set, due to privacy limitations on the smarthome deployment. The main limitations of the created data set are: reduced number of the deployed IoT devices, reduced variation on the communication protocols used, the reduced duration of the experiment execution time, possibility to perform controlling executions with improvements on the algorithms and expert values for risk evaluation and decision automation measuring.

As a countermeasure, in-house simulation and replay of the collected traffic data was implemented. The resulting experiment data was analysed and fine-tuned for the experiment control execution, as well as adjustment of the expert value data. Furthermore, the adjustment of the RAE for another project (nIoVe[3]) in a similar setting was performed. For this purpose the underlying attack model was readjusted to the automotive environment (instead of smarthome), to be more specific for the Connected Automated Vehicles (CAVs) and their infrastructure (eg public transport with a fleet of automated shuttles). However, similar problems were encountered with the availability of the real-life data. It should be underlined how critical it is to have proper data sets, including benign data as well as annotated data from a diverse set of attack scenarios. While both projects failed in this regard, ongoing additional efforts are made to establish these data sets.

[3]H2020 nIoVe - Cordis https://cordis.europa.eu/project/id/833742

A secondary data limitation we have encountered relates to the input data type and granularity. From the RA automation perspective, more granular reporting on the potential threats and anomalies is required if aiming to achieve higher intelligence in the decision-making. We have addressed this limitation by developing various Reporting Strategies (described in detail in Section 6.4) as an attempt to normalise incoming reports for DRA. However, it remains a subject to external technical factor.

1.6 Methodology

We have followed the adaptation of the Design Science Research (DSR) methodology [20], which mostly focuses on the qualitative and experimental research approaches. The following research phases were executed to ensure the scientifically sound framework application:

1.6.1 Phase 1: Problem Identification

During this phase we have identified current research gap in the availability of the DRA for general audience in the real-life environments. A thorough literature review was performed to identify all relevant advancements in the domain of the smarthomes. Then an analytical comparison of the derived results laid the foundation for the development of the reference architecture, capable to support the envisioned model of RA. Technical, functional and user requirements were analysed to define the potential workflow of the data through the system, starting with possible data inputs and controlled by the output requirements. More specifically, an extensive list of the requirements was compiled in collaboration with the project's partners under GHOST project, which served as a common base for multiple concepts definition, published as a publicly available deliverable[4]. Such approach enabled us to identify relevant RQs and have partial anticipation of the envisioned results.

1.6.2 Phase 2: Theoretical Model Development

Once the limitations were clearly identified, we have proceeded with the fine-grained RA model definition, mostly focusing on RQ1 and RQ2. This is when we have developed our concept of IoT Stack and expert values definition (described in detail in Chapter 6). For this purpose we continued the extraction of the available published data to understand and gather knowledge on the granular composition of the risk properties. This served as a base for the initial risk weight assigning.

1.6.3 Phase 3: Technological Artefact Implementation

Once initial version of the Theoretical Model was developed, we proceeded with an actual implementation of the Technological Artefact. This was achieved under the umbrella of the

[4]GHOST - D3.17 https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5bd4ff219&appId=PPGMS

GHOST project execution, where developed framework was integrated with the multitude of additional tools, supporting the input data feed to the RAE. The process of Theoretical Model development and Technological Artefact was iterative in its nature, constantly complimenting one another and ensuring the improvements on both sides.

1.6.4 Phase 4: Results Validation

Most critical phase of the followed methodology. Once the framework was fully integrated into singular solution, it was deployed in the real-life trials. However, several obstacles were encountered. First, the execution of the real attacks to verify the feasibility of the integral framework was considered to be unethical and therefore was replaced with the voluntarily gathered data, equivalent to the captured IoT data traffic from one of the smarthome installations. This was further replaced with the testbed simulated data, were partial solutions were deployed due to time and available resources constraints. Finally, we have attempted to replicate all experiments reply in our premises, establishing a testbed with local attack execution to validate the applicability of the used RAE parameters.

1.7 Contributions summary

This thesis is based on published articles already presented at scientific journals and conferences. Below is a brief summary of each article. Finally, Table 1.1 shows the summary of scientific contributions.

1.7.1 Article I [21]

Collen, A., Nijdam, N.A., Augusto-Gonzalez, J., Katsikas, S.K., Giannoutakis, K.M., Spathoulas, G., Gelenbe, E., Votis, K., Tzovaras, D., Ghavami, N. and Volkamer, M. (2018, February). *GHOST – Safe-guarding Home IoT Environments with Personalised Real-time Risk Control.* In International ISCIS Security Workshop (pp. 68-78). Springer, Cham.

In this article we present the European research project GHOST, (Safe-guarding home IoT environments with personalised real-time risk control), which challenges the traditional cybersecurity solutions for the IoT by proposing a novel reference architecture that is embedded in an adequately adapted smarthome network gateway, and designed to be vendor-independent. GHOST proposes to lead a paradigm shift in consumer cybersecurity by coupling usable security with transparency and behavioural engineering.

1.7.2 Article II [22]

Augusto-Gonzalez, J., **Collen, A.**, Evangelatos, S., Anagnostopoulos, M., Spathoulas, G., Giannoutakis, K.M., Votis, K., Tzovaras, D., Genge, B., Gelenbe, E. and Nijdam, N.A. (2019, September). *From Internet of Threats to Internet of Things: A Cybersecurity Architecture for*

Smarthomes. In 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD) (pp. 1-6). IEEE.

The H2020 European research project GHOST – Safe-Guarding Home IoT Environments with Personalised Real-time Risk Control – aims to deploy a highly effective security framework for IoT smarthome residents through a novel reference architecture for user-centric cybersecurity in smarthomes providing an unobtrusive and user-comprehensible solution. The aforementioned security framework leads to a transparent cybersecurity environment by increasing the effectiveness of the existing cybersecurity services and enhancing system's self-defence through disruptive software-enabled network security solutions. In this article GHOST security framework for IoT-based smarthomes is presented. It is aiming to address the security challenges posed by several types of attacks, such as network, device and software. The effective design of the overall multi-layered architecture is analysed, with particular emphasis given to the integration aspects through dynamic and re-configurable solutions and the features provided by each one of the architectural layers. Additionally, real-life trials and the associated use cases are described showcasing the competences and potential of the proposed framework.

1.7.3 Article III [13]

Pandey, P., **Collen, A.**, Nijdam, N., Anagnostopoulos, M., Katsikas, S. and Konstantas, D. (2019, July). *Towards Automated Threat-Based Risk Assessment for Cybersecurity in Smarthomes.* In European Conference on Cyber Warfare and Security (pp. 839-XVII). Academic Conferences International Limited.

Cybersecurity is a concern of each citizen, especially when it comes to novel technologies surrounding us in our daily lives. Fighting a cyber battle while enjoying your cup of coffee and observing gentle lights dimming when you move from the kitchen to the sitting room to review your today's running training, is no longer science fiction. A multitude of the cybersecurity solutions are currently under development to satisfy the increasing demand on threats and vulnerabilities identification and private data leakage detection tools. Within this domain, ubiquitous decision-making to facilitate the life of the regular end-users is a key feature here. In this paper we present a Risk Assessment Model (RAM), originating from Negative to Positive approach, to automate the threat-based RA process, tailored specifically to the smarthome environments. The calculation model application is demonstrated on derived threat-triggered evaluation scenarios, which were established from analysing the historical evidence of data communication within the smarthome context. The main features of the proposed RAM are identification of the existing risks, estimation of the consequences on possible positive and negative actions and embedding of the mitigation strategies. The application of this modelling approach for automation of RA would lead to a deep understanding on the extent to which decision-making could be automated while tracking and controlling the cyber risks within the end-user's accepted risk level. Through the proposed RAM, common factors and variables are extracted and integrated into a quantified risk model before being embedded in the automated decision-making process. This research falls within the GHOST (Safe-Guarding

Home IoT Environments with Personalised Real-time Risk Control) project, aiming to provide a cybersecurity solution targeted at the regular citizens.

1.7.4 Article IV [23]

Collen, A. and Nijdam, N.A., 2022. *Can I Sleep Safely in My Smarthome? A Novel Framework on Automating Dynamic Risk Assessment in IoT Environments.* Electronics, 11(7), p.1123.

Fully automated homes, equipped with the latest IoT devices, aiming to drastically improve the quality of lives of those inhabiting such homes, is it not a perfect setting for cyber threats? More than that, this is a fear of many regular citizens and a trending topic for researchers to apply Cyber Threat Intelligence (CTI) for seamless cybersecurity. This paper focuses on the RA methodology for smarthome environments, targeting to include all types of IoT devices. Unfortunately, existing approaches mostly focus on the manual or periodic formal RA, or individual device-specific cybersecurity solutions. This paper presents a Dynamic Risk Assessment Framework (DRAF), aiming to automate the identification of ongoing attacks and the evaluation of the likelihood of associated risks. Moreover, DRAF dynamically proposes mitigation strategies when full automation of the decision making is not possible. The theoretical model of DRAF was implemented and tested in smarthome testbeds deployed in several European countries. The resulting data indicate strong promises for the automation of decision making to control the tightly coupled balance between cybersecurity and privacy compromise in terms of the embedded services' usability, end-users' expectations and their level of cyber concerns.

1.7.5 Article V [24]

Spathoulas, G., **Collen, A.**, Pandey, P., Nijdam, N.A., Katsikas, S., Kouzinopoulos, C.S., Moussa, M.B., Giannoutakis, K.M., Votis, K. and Tzovaras, D. (2018, July). *Towards Reliable Integrity in Blacklisting: Facing Malicious IPs in GHOST Smart Contracts.* In 2018 Innovations in Intelligent Systems and Applications (INISTA) (pp. 1-8). IEEE.

The European research project GHOST challenges the traditional cybersecurity solutions for the IoT sector by exploiting novel technologies, such as blockchain, to provide resilience and integrity of decision-making on the communication exchange in a smarthome context. When it comes to novel cybersecurity solutions for extremely heterogeneous environments like IoT and smarthomes, the key focus is typically given to the understanding of network activities and elimination of suspicious traffic. The GHOST project adds an extra dimension to this approach by integrating blockchain technology at its core decision mechanism. On a daily basis, each GHOST installation is encountering malicious behaviour and suspicious IoT communications, where easy information sharing with other installations, as well as decentralised decision-making, are mandatory features for the efficient protection of the end-user. GHOST's Smart Contracts (SC) are designed to tackle in an easy, yet productive

way, the reporting on suspicious Internet Protocol (IP) addresses which the IoT devices in a smarthome are trying to communicate with. Two variations of blacklisting smart contracts are presented in this paper, covering a diverse spectrum of possible attack vectors while closely following the Privacy by Design (PbD) principles. A reputation scoring scheme for malicious IPs reporting is integrated in the SC, uncovering the implementation details on the penalisation of existing entries in case of malicious behaviour of reporting devices.

1.7.6 Article VI [25]

Collen, A., Szanto, I-C., Benyahya, M., Genge, B. and Nijdam, N.A. (April 2022). *Integrating Human Factors in the Visualisation of the Usable Transparency for Dynamic Risk Assessment.* Information (under 1st review, to be published).

Modern technology and digitisation era accelerated the pace of the data generation and collection for the various purposes. The orchestration of such data is a daily challenge faced by even experienced professional users in the context of IoT-enabled environments, especially when it comes to the cybersecurity and privacy risks. This article presents an application of user-centric process for the visualisation of the automated decision-making security interventions. The User Interface (UI) development was guided by the iterative feedback collection from the user studies for the visualisation of the DRA-based security solution for regular lay users. The methodology we applied starts with the definition of the methodological process to map possible technical actions to related usable actions. The definition and refinement of the UI was controlled by the survey feedback loop from the end-user studies on their general technological knowledge, experience with smarthomes, cybersecurity awareness and privacy preservation needs. We continuously improved the visualisation interfaces for configuring cybersecurity solution and adjusting usable transparency on the control and monitoring of the DRA. For this purpose we have designed, developed and validated a decision tree workflow and showed the evolution of the interfaces through various stages of the real-life trials executed under European H2020 project GHOST.

Table 1.1 Scientific Contribution.

	Article I	**Article II**	**Article III**	**Article IV**	**Article V**	**Article VI**
Title	GHOST – Safe-guarding Home IoT Environments with Personalised Real-time Risk Control	From Internet of Threats to Internet of Things: A Cybersecurity Architecture for Smarthomes	Towards Automated Threat-Based Risk Assessment for Cybersecurity in Smarthomes	Can I Sleep Safely in My Smarthome? Automating Dynamic Risk Assessment on Cyber Attacks Impact	Towards Reliable Integrity in Blacklisting: Facing Malicious IPs in GHOST Smart Contracts	Understanding Human Factors in the Visualisation of the Usable Transparency for Dynamic Risk Assessment
Authors	**Collen, A.**, Nijdam, N.A., Augusto-Gonzalez, J., Katsikas, S.K., Giannoutakis, K.M., Spathoulas, G., Gelenbe, E., Votis, K., Tzovaras, D., Ghavami, N., Volkamer, M	Augusto-Gonzalez, J., **Collen, A.**, Evangelatos, S., Anagnostopoulos, M., Spathoulas, G., Giannoutakis, K.M., Votis, K., Tzovaras, D., Genge, B., Gelenbe, E., Nijdam, N.A	Pandey, P., **Collen, A.**, Nijdam, N., Anagnostopoulos, M., Katsikas, S., Konstantas, D.	**Collen, A.**, Nijdam, N.A.	Spathoulas, G., **Collen, A.**, Pandey, P., Nijdam, N.A., Katsikas, S., Kouzinopoulos, C.S., Moussa, M.B., Giannoutakis, K.M., Votis, K., Tzovaras, D	**Collen, A.**, Szanto, I-C., Benyahya, M., Genge, B., Nijdam, N.A.
Stages	Problem Identification; Theoretical Model; Artefact Implementation	Problem Identification; Theoretical Model; Artefact Implementation	Problem Identification; Theoretical Model; Artefact Implementation; Result Validation	Problem Identification; Theoretical Model; Artefact Implementation; Result Validation	Artefact Implementation; Result Validation	Artefact Implementation; Result Validation
Methods	Literature Review; Requirements Analysis	Literature Review; Workshops	Literature Review; Use Case Studies	Expert opinions; Experiments; Trials	Experiments; Trials; Use Case Studies	Surveys Feedback; Trials; Experiments
Type	**Conceptual research:** -Theories Development; -Reference architecture definition	**Conceptual research:** -Reference architecture definition; -Methods definition	**Applied research:** -Theories development; -Methods definition	**Empirical research:** -Prototype development; -Methods definitions; -Empirical findings	**Applied research:** -Prototype development; -Theories development	**Applied research:** -Methods definition; -Prototype development
Impact	IS: 0.48 h-Index: 51 SJR: 0.16	IS: 1.78 h-Index: 11 SJR: 0.307	IS: 0.31 h-Index: 9 SJR: 0.141	IS: 3.02 h-Index: 36 SJR: 0.36	IS: 1.25 h-Index: 6.5 SJR:0.207	IS: 2.38 h-Index: 28 SJR: 0.349
Publication	*Book series* CCIS, Springer, 2018	*Conference Proceedings* CAMAD, 2019	*Conference Proceedings* ECCWS, 2019	*Journal* Electronics, 2022	*Conference Proceedings* INISTA, 2018	*Journal* Information, 2022
Status	Published	Published	Published	Published	Published	Under 1st review
Contributions	Own contribution: 90% Concept: A.C, N.N Methodology: A.C Writing: A.C, N.N Visualisation: A.C, N.N Supervision: N.N	Own contribution: 85% Concept: A.C, N.N Methodology: A.C, J.AG Writing: A.C, S.E Visualisation: A.C, N.N Supervision: N.N	Own contribution: 45% Concept: P.P, A.C Methodology: P.P, A.C Analysis: A.C, P.P Writing: P.P, A.C Visualisation: N.N, A.C Supervision: N.N, S.K	Own contribution: 95% Concept: A.C, N.N Methodology: A.C, N.N Software: A.C Validation: A.C Writing: A.C Visualisation: N.N, A.C Supervision: N.N	Own contribution: 50% Concept: G.S, A.C, N.N Methodology: G.S, A.C Software: G.S, A.C Validation: G.S Writing: G.S, A.C Visualisation: G.S, N.N Supervision: N.N, S.K	Own contribution: 80% Concept: A.C, S.C, B.G Methodology: A.C Software: A.C Validation: A.C Writing: A.C Visualisation: N.N, A.C Supervision: N.N, B.G

1.8 Structure

The overall structure of this thesis and its content correlation to the RQs, methodology phases, published works, and scientific methods used is outlined in Figure 1.1.

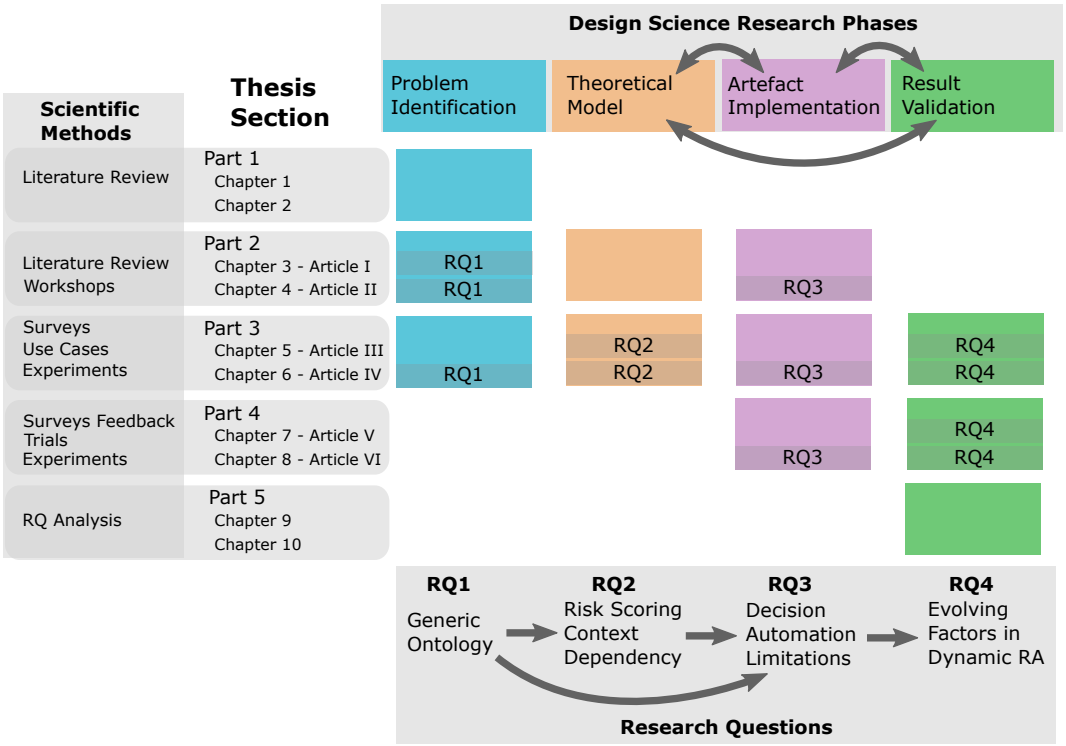


Figure 1.1 Thesis Structure.

- Part 1 – Introduction: includes the definition of the research context, motivation and problem statement. It is further supported with the description of the methodology applied in the framework of this thesis in Chapter 1. Additionally, we provide short revised Relevant Work in Chapter 2 for the core topics we believe had significant advancement since the publication of the associated articles. The chapters in this part of the thesis are corresponding to the definition of the scientific problem guiding this thesis (Phase 1).

- Part 2 – Reference Architecture: focused on the definition of the background knowledge, generic architecture, identification of the core research questions and general technological constraints influencing the envisioned framework. It's primarily goal is to provide the answers to RQ1 and is supported by the Article I as Chapter 3 and Article II as Chapter 4. Also the initial efforts towards RQ3 were set as developing reference architecture also initiated the technical limitation identification process. The chapters in this part of the thesis are corresponding to the definition of the theoretical model (Phase 1), input/output identification for the practical implementation of theoretical model (Phase 2), and the definition of the validation protocol (Phase 3).

- Part 3 – Risk Assessment: represents the core of the thesis work, aiming to address RQ2 and RQ3 through Articles III as Chapter 5 and Article IV as Chapter 6. Here we present in detail the implementation of the designed framework, the risk model utilised, as well as validation of the preliminary results. It also contributes to the RQ1 as a work where IoT Stack concept was grounded by its technical implementation. The real-time performance aspects raised by RQ4 are also partially covered in Chapter 6. This part of the thesis followed on its own a complete cycle of the research methodology applied in the thesis. Starting from the identification of the problem on the DRA (Phase 1) to design and conceptualisation of the Risk Model (RM) (Phase 2). Consecutively, Chapter 6 corresponds to the thesis most crucial advancements, presented as the implementation of the DRA model (Phase 3) and its validation (Phase 4).
- Part 4 – Collective Resilience: Outlines the challenges of DRA related to the human factors, external threat intelligence and decentralised approaches for resilience improvement. It focuses mostly on RQ4 and is supported by the Articles V as Chapter 7 and Article VI as Chapter 8. Article VI also contributes the RQ3 from human factor related limitations on the decision automation. The chapters in this part of the thesis correspond to the improvements of the DRA core implementation, aiming to enhance user-centric aspects of the final framework (Phase 3). Finally, the definition of the user-centric interface improvement is considered as a validation of the results RA decision-making usability strategy (Phase 4).
- Part 5 – Conclusions: Concludes the thesis with the supplementary discussion (Chapter 9) and final remarks (Chapter 10).

Disclaimer

This thesis is based on already published papers, which are included as chapters. However, the included text was further homologised to align the writing style throughout the thesis.

Chapter 2

Related Work

Chapter Contents

2.1 Introduction

Generic smarthome classification, defined by Schiefer [15], gives an excellent overview of the variety of the IoT devices that can be present in any smarthome installation. The spectrum of the threats for smarthomes is twofold, privacy and security related. Very often it is difficult to draw a strict line for the classification, as the attacks are targeting to exploit both vectors. Unfortunately, the biggest problem still relies on primitive security settings that are ignored by unaware users, such as keeping default passwords. According to the [17], multiple occurrences of the performed attacks with the help of IoT devices were aiming to disrupt the performance of the large organisations, exploiting primitive attack vectors such as the use of default passwords or weak communication protocols. More advanced attacks were performed by researches in [18], able to distribute and propagate the worm using simple Philips Hue smart lamp while driving in a car past 70 metres away from the target building. The most powerful example was appearance of the Mirai botnet [19], taking over at least 100 000 IoT devices. It is evident that a regular citizen has no way to gather a full picture of the potential risks involved in the smarthome she is living in, and that an automated monitoring solution is essential.

There are two main approaches widely used in security monitoring: network monitoring and host analysis. While the first one is applicable to any existing infrastructure installation, the second one has to be implemented at the design stage or additional software has to be installed on the device of interest. The network monitoring in turn is further based on signature extraction, log events correlation and behaviour analysis. Signature extraction is useful for identification of known malicious instances of the software. Log events correlation requires existence of network-level security mechanisms able to provide these logs. Behaviour analysis can be applied directly on any existing network at the router/gateway entry/exit point of any smarthome installation. In terms of the approaches used in behaviour analysis, Machine Learning (ML) is the most common method used for anomaly detection and extraction. [26] successfully identified malicious behaviour on the network by comparing application of several existing ML classifiers: Nearest Neighbours Classifier (NNC), Linear Support Vector Machine (LSVM), Artificial Neural Network (ANN), Gaussian Based Classifier (GBC), Naive Bayes Classifier (NBC). LSVM, ANN and NNC appears to be the most efficient ones, but unfortunately none of them is capable to detect the novel or zero-day threats, especially in the real-time regime. Their consecutive work [27], expanded the existing method with the use of the decision trees, using Reduced Error Pruning algorithm, allowing zero-day detection of the involvement in botnet activities. The framework proposed by [28], aimed at detecting malware, is using behaviour graphs, improving the accuracy and false positive detection by adding the graph features, which specifically influence the automated classification and provide necessary abstraction of network flows with their dependencies.

In terms of the risk assessment, the closest survey on the risk analysis, vulnerability and mitigation techniques identification can be found in [29], providing risk overview for smart

cities. To develop an appropriate and precise risk modelling framework adapted to smarthome environment, five main vulnerability categories must be taken into account:

- Weak software security and data encryption: directly applicable to smarthome and various IoT devices deployed in smarthome installations;
- Use of insecure legacy systems and poor ongoing maintenance: applicable to any IoT devices plugged into the smarthome installation;
- Many interdependencies and large and complex attack surfaces: directly applicable to smarthome due to heterogeneous nature of deployed IoT devices;
- Cascade effects: applicable to the extent of having interdependencies between various groups of IoT devices deployed in the smarthome;
- Human error and deliberate malfeasance of disgruntled (ex)employees: directly applicable to the smarthome installations.

Almohri et al. [30] and Martins et al. [31] suggest incorporating threat modelling for risk assessment directly at the IoT device design stage, distinguishing three main approaches: attacker-, system- and asset-centric. The attacker-centric model focuses on identifying the possible attacker and evaluates their goal and the ways to achieve them. System-centric modelling is considering all possible attacks that target all different components of the system. Asset-centric approach focuses on the individual assets the system has access to. They also conclude that current advancements in anomaly detection in Cyber-Physical System (CPS) can be split in two directions: physical process (system states modelling based on immutable laws of physics) and cyber models (characterising expected software behaviour).

Often, the user experience is improved by eliminating any explicit need of user interaction in the context of increased trust among IoT devices. However, this leads to unawareness of the user on the data flows between the devices in his home, and therefore decreases the overall privacy and security [32]. Many IoT devices use externalised cloud-based services to process in real-time their data. The question of data ownership is raised directly in these situations but are often not clearly answered. As per [33], a risk analysis for a smarthome automation system was performed. Most notably, Information Security Risk Analysis (ISRA) was used with a manual evaluation of risk probabilities and associated impact on a five level scale (1–5). Their main conclusions are (i) the need for real use-cases on risk evaluation in smarthome environments; (ii) a stronger involvement of security into the design; and (iii) finer classification of the user personal data and associated risks is required. They highlight an urgent challenge to find effective ways to equip regular users with tools to comprehend the whole picture of their IoT home deployed infrastructure together with data flow sensitivity indicators, where automated risk analysis functionality is integrated. In [34] a framework is described for automatically controlling the Personally Identifiable Information (PII) in Web of Objects (WoO) within the smarthome environment. The automation is performed

based on consents collection of user PIIs to various service providers. Also the notion of risk prediction is completely neglected. The key difference with DRAF proposed solution is a lack of a visualisation component which would allow the regular users to acquire insights of their private data and take an action upon it to resolve related issues.

Another approach is the use of the digital identity inheritance principle [35]. It not only detects a new context but also deals with a fine definition of the rules and their correlation in the different context. The design and implementation is done using model based security toolkit named SecKit which supports policy management and enforcement at all layers of the infrastructure proposed by the iCore project. The limitation, however, of their proposed framework is that it does not address potential ambiguity and quality of the data collected by the sensors while analysing the context.

2.2 Privacy perception and awareness raising

As presented by Burghardt [36], two main factors affect the trust relationship between online platform providers and users in terms of data protection practices. These are national or international legal regulations and Privacy Enhancement Technologies (PET). Platform for Privacy Preferences (P3P) [37] was a protocol standardised by the World Wide Web Consortium (W3C), which enables websites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P user agents allow users to be informed of site practices (in both machine- and human-readable formats) and to automate decision-making based on these practices when appropriate. Thus, users do not need to read the privacy policies at every site they visit. Nevertheless, P3P was only supported by Microsoft's Internet Explorer and Edge, while it has been discontinued and is currently not in practical everyday use.

Furthermore, Weitzner et al. [38] worked on creating a policy-aware web. The authors presented a rule-based policy management system that can be deployed openly and distributed on the World Wide Web. They argue for the necessity of such infrastructure, also presenting its required features. Accordingly, the Semantic Web rules language (N3) with a theorem prover designed for the Web (Cwm) is used, allowing the Hypertext Transport Protocol (HTTP) to provide a scalable mechanism for the exchange of rules and, eventually proofs, for access control on the Web. The TOR Project[1] is based on a network of "onion routers" that anonymize the users' IP address in Internet by routing traffic through multiple routers and encrypting it each step of the way. Accordingly, a user that participates in the TOR network, is visible towards the service provider only as the IP address of a randomly chosen onion router, masking this way both his geographical position and web page visits. Usage might be cumbersome, although the available TOR browser allows for a wider user base, while TOR cannot prevent that other browser plugins reveal private information. Moreover, TOR does not provide encryption between the web server and the last router, thus other

[1] https://www.torproject.org/

potential privacy threats arising from the use of cookies or active scripting must also be considered. Additional privacy focused search engines and browsers include DuckDuckGo[2], Qwant[3], Fireball[4], and Search Encrypt[5]. Examples of other PET include (i) additional communication anonymisers, which can be protocol specific (for use with specific services, such as email and FTP) or protocol independent; (ii) obfuscation mechanisms for misleading automated precision analytics processes; and (iii) sharing of online accounts or even public verification keys through Enhanced privacy ID.

In addition to the introduction of PET, the cyber habits of users are critical in respect to their privacy exposure. Yong Jin [39] examined the impact of three dimensions of digital literacy on privacy-related online behaviours: (i) familiarity with technical aspects of the Internet; (ii) awareness of common institutional practices; and (iii) understanding of current privacy policy. This study showed a strong predictive power between user knowledge, as indicated by these dimensions, on privacy control behaviour. Additionally, Ariu et al. [40] have worked on studying the level of awareness and perception of IT security amongst University students, paying particular attention to mobile devices. Their report analyses the answers given by 1012 students from over 15 Italian Universities to a multiple-choice questionnaire. The analysis shows that students' perceptions of their knowledge is generally wrong, and that they are unaware of the risks arising from their behaviours. Multiple similar studies focused on the analysis of user behaviour and the effect of usage patterns or user backgrounds to online privacy exposure. Moreover, multiple meta-studies, such as the one by Kushzhanov and Aliyev [41], highlighted the importance of cybersecurity awareness and protection, discussing obstacles and solutions towards developing an effective digital society by addressing cybersecurity challenges in digital life.

2.3 Risk scoring and exposure measurement

Privacy is often neglected by citizens, as they don't see themselves being affected directly by massive scale Personal Data (PD) collection, as defined by Parkinson [42]. For instance, research studies [43], were performed to identify how easily a typical user is willing to let someone collect his/her PD. The results were surprising, users were installing the application from the research team even if there was no practical use of it, without even taking an option to enter the draw to win a prize. It is evident from that experiment that very often users are not paying attention to any warning messages at installation time, or that they don't fully understand them.

One of the first to provide an intuitive methodology for computing users' privacy scores was Liu and Terzi [44] with their work focused on online social networks. The calculated score was aimed to indicate users' potential risk caused by their activities in the network.

[2] https://duckduckgo.com/
[3] https://www.qwant.com/?l=en
[4] https://fireball.com/
[5] https://www.searchencrypt.com/

The methodology is relying on the two main properties: (i) sensitivity of the data; and (ii) visibility of the data within the network. Furthermore, their modelling was advantageous to be platform independent, therefore comparable between different networks. However, its design was considering only one source of information.

Aghasian et al. [45] took their work further and developed the Privacy Disclosure Score (PDS) measurement, taking into consideration users' data shared across multiple social networking sites. One source of data may not disclose a wide range of information of a user that can pose a privacy risk, but when the information is combined from different sources, it can be risky and dangerous. While the majority of other methods are utilising a dichotomous approach, where data is either publicly or privately available, the PDS model uses a polytomous approach. One of the major limitations of this work is a lack of inclusion of technical difficulty for data extraction parameters.

An empirical framework on the measurement of the perceived privacy risk was proposed by Bhatia and Breaux [46]. Their framework targets privacy analysts, software engineers or user interface designers to have a broad factor coverage when measuring privacy risks. They found out that people's decisions about privacy are inconsistent, and they might behave differently in actual versus hypothetical situations, especially when they perceive the benefits of data sharing.

Li et al. [47] considered accessibility, extraction difficulty, reliability, and privacy awareness in comparison to other works. The simplified half-suppressed fuzzy C-means based algorithm permits the calculation of the visibility together with sensitivity to obtain a final privacy score. They argue that the best solution for privacy awareness is to provide a method to quantify the privacy of individuals, transform the virtual concept of privacy into a visible physical space, help users accurately recognise the state of their privacy, and help users improve their privacy.

Part II

Reference Architecture

Chapter 3

Article I: GHOST – Safe-guarding Home IoT Environments with Personalised Real-time Risk Control

Relevance

This article partially examines RQ1 and RQ4. Here we present the first version of the envisioned reference architecture suitable for the deployment of the DRA framework in the smarthome, IoT enabled environment. It serves as a basis on the identification of a sparse requirements matrix, mainly to comprehend the technical and functional constraints imposed by such system design and deployment environment. At the time this article was presented, a first prototype of the RAE was realised, described in detail in Chapter 7, in the context of an experimental setup, demonstrating the full data flow from threat detection to risk identification and intervention propositions for the end-user to mitigate encountered risks.

Context

This article was presented at the First International ISCIS Security Workshop 2018, Euro-CYBERSEC 2018, London, UK, February 26–27, 2018. The article was published as part of the Communications in Computer and Information Science book series (CCIS, volume 821), Springer. According to the Resurchify portal[1], the Impact Score (IS): 0.48, h-Index: 51 and SCImago Journal Rank (SJR): 0.16.

Own Contribution

Being the lead author of this paper and also the main conceptualist of the GHOST project's proposal preparation, my contribution to this work is problem definition, framework conceptualisation, related work analysis, layered architecture design and validation process definition. Majority of the paper content was edited by me, further reviewed by other co-authors.

[1] https://www.resurchify.com/about

Chapter Contents

3.1 Introduction

According to [48], the average IoT device was attacked once every two minutes in 2016. Unfortunately, such botnets as Mirai are taking advantage of the fact that security is still not a priority for device manufacturers, leading to the lack of possibility of automatic firmware upgrades, exposing the devices to simple attacks such as account enumeration and open ports scanning up to unpatched vulnerabilities presence and their exploitation to gain full control.

In addition to forcing the integration of security aspects into IoT devices at the manufacturer level, it is evident that a monitoring solution is essential to protect the end-users. glsiot devices are often completely closed, not standardised or openly developed. Hence, the user does not have a clear idea of the potential risks involved. On top of the purely technological and operational cybersecurity challenges, the end-user behaviour becomes a determinant factor, with the human typically portrayed as the weakest link in security. Indeed, consumers tend to exhibit low tolerance and fatigue in using sophisticated cybersecurity solutions or practices, while the cybersecurity industry often addresses usability as at trade-off on security rather than as a security enhancing component. Thus, combining or integrating usability and security requirements is a major research challenge, which recently has been brought forward [49, 50], while turning end-user behaviour in favour of cybersecurity remains a field with a promising exploitation potential [51].

This paper gives an overview of the European Union Horizon 2020 Research and Innovation project GHOST (https://www.ghost-iot.eu/). GHOST aims to increase the level and the effectiveness of automation of existing cybersecurity services and to enhance system self-defence while prioritising the opening up the cybersecurity 'blackbox' to consumers and building trust through advanced usable transparency tools derived from end-users' mental models.

The rest of the paper is structured as follows. Section 3.2 discusses related work. Section 3.3 presents the GHOST system, whilst Section 3.4 presents the GHOST validation process. Finally, conclusions are summarised in Section 3.5.

3.2 Related work

In traditional cybersecurity, Intrusion Detection Systems (IDS) are taking the main role in detecting any anomalous activity on the network. Best known solutions are Snort [52], Suricata [53] and Bro [54]. While Snort and Suricata are based on pattern matching detection, Bro is relying on semantic matching of the network events. However, these solutions are designed for professional use and are not explicitly aimed at the IoT environment in terms of protocol analysis availability. Global scale architecture with distributed data storage and correlation for Intrusion Detection Systems (IDS) was proposed in [55]. While taking advantage of novel technologies and providing wide coverage of monitored data for expert users, this system is not adapted for smarthome installation where regular citizens have to understand the usage of this tool. Graphical representation of attack and threats scenes was greatly advanced

in [56]. These works are targeting professional analysts with the deep technology knowledge though. Modelling uncertainties in the cyber threat arena was presented in [57], Grey theory application for threat prediction was analysed in [58] and a framework assessing the impact of cyber attacks was described in [59]. Once again, all these advancements are focusing on the expert users, not regular citizens.

3.2.1 Advancements in IoT Cybersecurity Monitoring

Similarly to traditional cybersecurity IoT ecosystem is vulnerable to the analogous issues as in web, sensor and mobile communications networks, with particular focus on privacy, authentication and access control network configuration, information storage and management, standardisation and data integrity. The most complete classification of the IoT attack vectors is described in [60], referring to the IoT ecosystem as a Web3 or Web of Things phenomenon, where four main categories are provided: Device, Application Service, Network, Web Interface and Data integrity. Developing a cybersecurity solution targeting to protect all of the identified vectors is a very challenging and crucial task. [61] is raising a necessity to apply the Negative selection and Danger Theory to traditional IDS, to cover ubiquitous nature of the IoT devices and target all attack vectors specified above. Such systems, however, encounter serious limitations in terms computational power and storage requirements. An overview of the Real-time IoT security solutions was provided in [62]. The authors conclude that existing approaches can be divided into two major classes: hardware and software based security. Alternative to IDS approach is described in [63], where Security Information and Event Management (SIEM) system for IoT environment is proposed.

3.2.2 Smarthome Cybersecurity Frameworks

The authors of [64] analyse existing architectures of smarthomes from the security perspective, concluding that gateway architectures are the most suitable to provide key technologies for cyber protection: auto-configuration and automatic update installation. An overview of existing tooling for the implementation of cyber protection in smarthomes is also included in their work, however, all these tools are applicable only for newly designed devices to be included in a future smarthomes. On contrary, the IDS framework [65], based on Anomaly Behaviour Analysis, approaches this problem for existing and hardly changeable smarthome installations. Their focus is given to measuring the activities of installed sensor devices a smart house is equipped with, and detecting any anomalies in the quantity and quality of the collected measurements. The limitation of their work relies in the ability to apply their analysis only on the primitive IoT devices without direct internet access. Similarly to GHOST, traffic monitoring and inspection solution IoTGuard, based on Bro, is presented in [66]. The main drawback of their framework is the requirement to forward all router's traffic to IoT Controller and link each IoT device with the IoT Watchdog. On the contrary, GHOST provides all-in-one

solution to be deployed in the existing smarthome installations with key focus given to user's experience and understanding of a cybersecurity solution.

The great interest of developing smarthome cybersecurity solutions is also given by the commercial entities. Already a wide selection of the commercial products is available on the market: F-Secure SENSE [67], Luma smart Wi-Fi router [68], Dojo [69], CUJO [70], Bitdefender [71], Norton Core [72].

3.3 The GHOST System

The GHOST system is being realised by analysing existing technical infrastructure and existing software components corresponding to the aims of the project. Usability studies have been defined with the aim to establish mental models of the end-users. This allows systematical and effective addressing of the human factor with the aim to facilitate end-users' proper decision making in relation to security and privacy issues and adequate usage of the GHOST solution. It also allows the definition of a first set of end-user requirements, which in turn facilitate better specification of the development and integration of core technologies. Since human participants will be involved in the evaluation phase of the project and personal data will be collected, special emphasis is given on elaborating a data management plan for respecting privacy related issues according to national and EU legislation. It should be noted that the access to the collected data will be provided only to the members of the consortium for development and demonstration purposes.

3.3.1 Development Approach

To keep up with cybersecurity issues and threats GHOST not only follows guidance documents, best practices and standards (issued by international, European and national stakeholders) at all stages of design and development, but it also scans for emerging threats/issues. To this end, it makes use of security intelligence available within the consortium and outside (eg through mining insightful security blogs), as well as related information collected directly from the end-users and the smarthome pilots. The development of GHOST follows an iterative approach. Three iterations have been specified for the implementation of the technical components of the infrastructure. These will be evaluated through real-life trials and feedback will be reflected back for further refinement and acceptance, according to the validation process discussed in Section 3.4.

3.3.2 GHOST Software Architecture

GHOST's conceptual design involves advanced data flow analysis on a packet basis to build the context of communication. From this context, data are classified into user and device profiles, which in turn are used in the automated real-time risk assessment. The assessment is based on evaluation, comparison and matching with safe data flow patterns, utilising a

self-learning approach. Data analytics and visualisation techniques are deployed to ensure enhanced user awareness and understanding of the security status, potential threats, risks, associated impacts and mitigation guidelines.

The architecture of the GHOST system, shown in Figure 3.1, follows a layered approach that allows independent development of the separate components, while preserving a high interdependency within the framework. A brief outline of each layer and its main functionality is presented in this subsection.

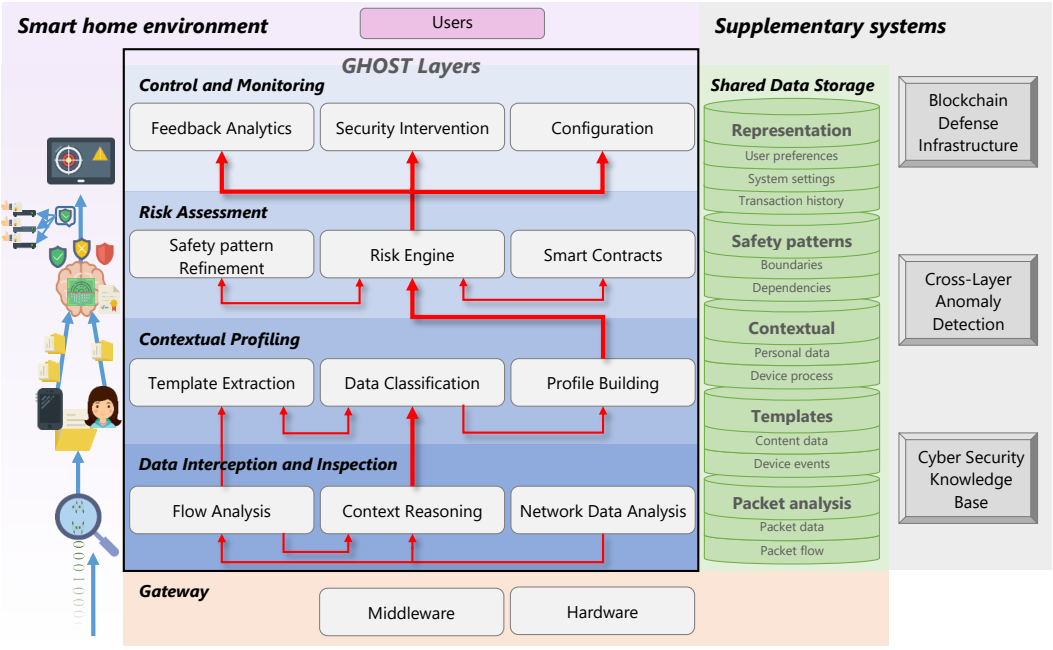


Figure 3.1 GHOST architecture.

3.3.3 Core Layers

Data Interception and Inspection (DII) Data related to traffic of all network interfaces in a smarthome environment is gathered directly from the network. This data is analysed and stored in order to be used by GHOST components. Significant data extracted from traffic packets is stored to a shared data storage. Additionally traffic packets are aggregated into groups related to specific communications or actions. These groups of packets are also analysed to extract information of a higher abstraction level and store it along with the information produced by single packets analysis. Additionally context information is extracted from traffic data. Recurring patterns of traffic are detected and the causes they are produced by are identified and an initial classification of the data type of traffic is performed. The network traffic may be correlated to actions of people or events in the smarthome and the data in the packets are categorised accordingly as personal data or device data.

Contextual Profiling (CP) The classification templates and actual profiles of the typical devices' behaviour are built in this layer, by extracting valuable data from the local network

communication already prepared by Data Interception and Inspection (DII) profiles for the normal behaviour of the devices are built in a tree based format for further processing by the risk assessment component. This layer also monitors the communications occurring between any combination of devices including the gateway, along with the status of each device and the status of the gateway. Monitoring is learning based, and models are trained to recognise the normal status of devices and the normal status of communication between them. Random Neural Networks are employed for each pair of devices and reinforcement learning is used to update them through time.

Risk Assessment (RA) This a core layer, which gathers information about the current risks and analyses in real-time current network traffic flows. It correlates device activity on the network with the profiles available from Contextual Profiling (CP) layer. The automatic decision making of the Risk Engine presents transparency of the cybersecurity solution, informing the end-user only about urgent decisions. Its capabilities is enhanced with the use of SC to ensure the reliability and trustworthiness of decisions. RA is also designed for controlling users' privacy and making them aware of the associated risks.

Control and Monitoring (CM) Three types of the user interfaces are forming this layer: Feedback Analytics (advanced professional-alike interfaces, Security Intervention (daily decision-making support tooling) and Configuration. The input data include historical and current packet flow behaviours, risk levels, device profiles, packet classification score, etc. The layer provides visual and intuitive presentations and reports of the smarthome security status, including visualisations of packet features through time, visual monitoring and distinction of packet behaviours, and visual identification of potential anomalies and vulnerabilities. The appropriate visualisation and human-machine interaction mechanisms are put in place to allow users to effortlessly and effectively review security issues and take key decisions that affect their privacy and security.

3.3.4 Supplementary Layers

Blockchain Defense Infrastructure (BDI) GHOST uses blockchain technology and SC for ensuring data and code integrity. At this layer the decisions made by RA are verified according to commonly agreed SC, turning the decision making into a truly decentralised and resilient system against intrusions. The integrity of the code running on smarthome gateways can be certified by the use of blockchain technology. Additionally valuable security related information can be stored at a blockchain infrastructure in order to be shared between smarthome gateways.

Cross Layer Anomaly Detection Framework (CLADF) Cross layer anomaly detection framework integrates existing open source solutions for traditional cybersecurity features. The main purpose is to collect, correlate, combine, and provide a unified output to other components in terms of possible events that require further analysis.

Cyber Security Knowledge Base (CSKB) A common cloud based knowledge repository is integrated with GHOST to collect anonymised security intelligence and insights from external

web-sources to enhance the automatic decision making and improve end-user visual experience within the Control and Monitoring (CM) layer. It will maintain list of malicious actors and properties (IP Addresses, Domains, Uniform Resource Locator (URL)s, File Hashes).

Shared Data Storage (SDS) The data structures defined by each of the components are normalised and unified within a single storage framework. A combination of relational and non-relational databases is used to satisfy the needs of all components. There is distinction between local and cloud based storage, as some components will perform off-site analytics.

3.3.5 GHOST Hardware Platforms

GHOST is based on the existence of a communication gateway with network monitoring capabilities, in which GHOST modules capture and analyse the different traffic patterns by devices and users. This gateway is a trustable and secure-by-design device as far as it is located inside the home network and it has two main responsibilities: (i) to provide connectivity capabilities for the devices inside the network; and (ii) to run the different algorithms and mechanisms for ensuring the security and privacy of the user data. Having these in mind, this element of the GHOST solution must accomplish market requirements related with size, weight and objective cost, among others. Therefore, it is needed to find a trade-off between the different features and capabilities of the gateway, resulting in a device that can be defined as constrained node [73]. The main restrictions that a constrained device can have are the following:

- maximum code complexity and size,
- size of the memory of the system,
- processing power that the device can offer in a certain period of time,
- allowed energy consumption or battery duration,
- communication methods and interfaces of the system,
- user interfaces and accessibility to the system in deployment.

Several techniques has been proposed in the literature to keep these set of constraints controlled in different environments and specific solutions [74, 75, 76], including the for security and privacy applications [77, 78].

GHOST is being developed and tested using two resource-constrained platforms: a proprietary IoT gateway, and a Raspberry Pi (with some expansion modules for IoT networks). The use of both devices allows several different IoT protocols, such as 802.11, Bluetooth Low Energy, Z-Wave and 802.15.4 to run on GHOST. Differences do exist between these two devices, but there are also some similarities regarding their constraints as regards processing power; memory; communications; and energy efficiency. These constraints pose a number of research challenges.

3.4 GHOST Validation Process

The validation strategy defined for GHOST is based on a three-fold vision that combines a complete set of robustness and laboratory testing; the specific definition of realistic testbeds; and real-life trials or pilots. First, the laboratory testing will be done with the objectives of reducing the number of possible bugs and functional errors and of checking the stability of the hardware. Therefore, unit tests will be performed over each specific GHOST module and an acceptance test plan will be defined and tested, including both software and hardware stability testing. After this first stage, two already functional testbeds will be used to deeply test the functionality of the GHOST solution in a controlled environment.The testbeds designed for two specific smarthome demonstrators include more than 15 different types of devices, involving up to 25 devices that will be simultaneously connected and monitored by the GHOST suite. In order to have a broad view of the possible services and solutions, devices like smart locks, biomedical devices, companion robots or smart lights based on several communication solutions (like 802.11, 802.15.4, Z-Wave or Bluetooth Low Energy) have been included in the testbeds. Potential threats against the smarthome can be categorised into [79]: (i) Physical attacks; (ii) Unintentional damage (accidental); (iii) Disaster (natural/environmental); (iv) Damages or loss of IT assets; (v) Failures/malfunctions; (vi) Outages; (vii) Eavesdropping / interception / hijacking; (viii) Nefarious activity/abuse; and (ix) Legal. Of these, relevant to GHOST are groups (ii), (iv), (vii), and (viii). Each of these groups includes a number of threats that can exploit relevant vulnerabilities by launching different attacks. The response of GHOST when faced with those amongst the above attacks that lead to higher risks and/or are most prevalent will be assessed in the controlled environment of the GHOST testbeds.

In addition to the testbeds, a set of pilots in real scenarios (homes of end-users) in three different countries (Spain, Romania and Norway) and with complementary use cases related with telecare, eHealth, home security and home automation will be carried out. The real-life trials have been designed to cover a varied set of application and services. Four different use cases have been defined: Ambient assisted living in smarthomes for older people in Galicia, Spain; Continuous health monitoring for adult people in Galicia, Spain; Regular private homes (smart-home solutions) in Norway and Regular private homes (smart-home solutions) in Romania.

Each use case has their own set of devices to be installed and a complete test plan is being developed to simulate the possible results of specific attacks (previously validated and performed in the testbeds) to capture the response of the users to the GHOST behaviour.

3.5 Conclusions

GHOST brings professional security tools down to regular home users. The strategic outcome of GHOST is threefold: increased resilience of existing cybersecurity solutions for smarthomes and the IoT; a leap forward to usability and automation in cybersecurity; and a boost in the

competitiveness of European ICT security industry in the advent of the IoT in the connected world. From a user perspective, GHOST will help end-users to increase their control over their smart-home IoT devices and it will provide an option for smart-living service providers to use its security services to ensure that they respect the security and privacy needs of their clients.

Future work includes the iterative implementation, testing and validation of GHOST in existing laboratory testbeds and in real-life and scale pilots in three European countries, using appropriately designed use case scenarios. Related work from the GHOST project can be also found in [80, 81, 82, 83].

Chapter 4

Article II: From Internet of Threats to Internet of Things: A Cybersecurity Architecture for Smarthomes

Relevance

This article further establishes the path towards RQ1, RQ3 and RQ4. Here we present the revised version of the implemented reference architecture suitable for the deployment of the DRA framework in a smarthomes environment, as an improvement of the previous article (see Chapter 3). This article further explores the user-centric aspects of the envisioned solution and provides the integration and validation strategy definition. At the time this article was presented, we had executed the first stage of the real-life trials of the GHOST project (see Section 3.4), with a second prototype version of the RAE. This version had a small subset of analysers feeding the RAE, limiting the validation of the framework only to the conceptual level, further elaborated in the Chapter 6.

Context

This article was presented at the IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Limassol, Cyprus. The article was published in IEEE Explore proceedings of 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). According to the Resurchify portal, the IS: 0.1.78, h-Index: 11 and SJR: 0.307.

Own Contribution

Being one of the lead authors of this paper, my contribution to this work is leading the architecture revision, definition and implementation of the integration and validation strategy. Majority of the paper content was edited by me, further reviewed by other co-authors.

Chapter Contents

4.1 Introduction

IoT, which has attracted considerable attention during the last decade, presents a huge opportunity for many industrial and business stakeholders in various domains. According to [84], by the year 2020 approximately 50 billion connected devices will be deployed and the total IoT revenue is expected to outreach more than one trillion euros.

As an emerging technology, IoT is prone to cybersecurity attacks and demands for countermeasures for the protection of such ecosystems are constantly growing. The heterogeneity and diversity of the "Things", as well as new lightweight communication protocols appropriate for IoT technology, create new challenges for the protection of such systems.

GHOST – Safe-Guarding Home IoT Environments with Personalised Real-time Risk Control (https://www.ghost-iot.eu/) – is European Union Horizon 2020 Research and Innovation funded project, aiming at developing a reference architecture for securing smarthomes IoT ecosystem. The multi-layer solution integrates traditional cybersecurity countermeasures, while it introduces new mechanisms for the efficient defence of common to IoT threats. This paper presents the detailed architecture of the solution, and discusses the integration and validation strategy followed for the delivery of the framework on the real-life deployments.

The rest of the paper is organised as follows: Section 4.2 introduces the related work of security frameworks in IoT and in smarthome environments specifically. Section 4.3 describes in detail the proposed security framework and its technical architecture, while Section 4.4 presents the integration and validation strategy followed for its successful implementation. Section 4.5 concludes the paper with the discussion on the results gained and provides possible future directions.

4.2 Related work

Internet-enabled smarthome is one of the most well-known applications of the IoT since heterogeneous devices are networked together to provide smart services to the occupants of the smarthome, offices and surrounding environment. The raise of the automation technology and the ubiquitous computing together with the constant growth of lightweight and low-energy devices have made smarthomes more technology dependent and, thus, more complex in terms of security to handle.

4.2.1 Motivation

Meng et al. [85] demonstrate the challenges and security concerns in smarthome installations and they argue that threats against a smarthome environment can lead to security breaches and put at risk the safety and privacy of the unaware users residing in it. Bugeja, Jacobsson, and Davidsson [86] attribute these security issues to the heterogeneous, dynamic and Internet connected nature of an IoT environment. To this direction, Jacobsson and Davidsson [87]

propose a model that integrates security and privacy into the design of smarthome services and systems. According to the authors, the security design principles and technologies of a smarthome environment should incorporate security-enhancing technologies to protect the user information and provide resilience against malicious actions. Furthermore, Lin and Bergmann [64] suggest that a gateway architecture is most suitable to provide cyber protection to resource constrained devices. They also deduce that existing tools for the implementation of cyber protection in smarthomes are applicable only for newly designed devices to be included in a future smarthome installation.

4.2.2 Security Frameworks

A plethora of competing security frameworks for smarthomes have emerged in order to tackle the security attacks that threaten the privacy and safety of the smarthome residents. For instance, Park et al. [66] present a traffic monitoring and inspection solution, called IoTGuard. In their work, the authors utilise Bro IDS to detect abnormal behaviours in an IoT environment. The main drawback of their framework though is the requirement to forward all router's traffic to IoT Controller and link each IoT device with the IoT Watchdog in order to target and monitor particular IoT devices using device-specific IoT protocols.

The IDS framework by Pacheco and Hariri [65], based on Anomaly Behaviour Analysis, tries to provide security for existing and hardly changeable smarthome installations. Their focus is given to measuring the activities of sensor devices installed in a smart house, and detecting any anomalies in the quantity and quality of the collected measurements. The limitation of their work relies in the ability to apply their analysis only on the primitive IoT devices without direct internet access.

The work by Rafferty et al. [88] is based upon the Agent-based modelling, where agents inside the smarthome environment make observations and implement intended behaviour. This model requires minimal engagement by the user and it is focused on threat detection. However, it neglects the detection of vulnerable devices within the smarthome. Furthermore, the reasoning process, namely the process of deciding what actions to perform to reach a goal, is taking place in the Cloud layer. Finally, the aforementioned framework was not tested against live data, ie operating real-time.

A more user-intrusive approach for network security is presented by Habibi et al. [89]. There, the authors propose a whitelist-based intrusion detection technique specific for IoT devices. The proposal aims to prevent IoT devices to get entangled in botnets activities, so it blocks at the gateway level DNS lookups to malicious sites. However, this solution is only applicable for IP-based IoT devices and networks.

Similarly, DeMarinis and Fonseca [90] state that a network-layer architecture is required for the protection of a smarthome against external threats and the mitigation of attacks from compromised devices. The authors recommend the implementation of a policy-based framework to restrict malicious traffic. The adopted policies will follow a white-listing approach based on the observed and predictable patterns in network traffic of the IoT devices. The

main drawback of this proposal, however, is that each different purpose IoT device exhibits distinct patterns, requiring a monitoring period of the legitimate usage for each IoT device to construct its network pattern. Nevertheless, the work by DeMarinis and Fonseca is in preliminary stage and presents only considerations for designing a novel security layer.

Serror et al. [91] follow a rule-based approach, where every IoT device is allowed a specific behaviour, namely specific set of allowed connections, in order to fulfil its intended functionality. In this work the gateway enforces these rules with traffic filtering and anomaly detection techniques. An apparent drawback is the required definition of the communication rules, whereas in the case of the lack of which by the manufacturer or a certification authority, should be provided by the end-users.

A different approach is followed by Dorri et al. [92], where the authors propose a blockchain-based solution for decentralised security and privacy in a smarthome environment. Specifically, they utilise a local and private blockchain to control and audit the communications internal and external to a smarthome. This way an access control policy to the IoT devices and their data is enforced. However, the proposed mechanism exhibits a relative large overhead regarding traffic, processing time and energy consumption, as it requires each smarthome to be equipped with a high-resource miner for the administration of the blockchain.

4.2.3 Emerging advancements

Nowadays, the absence of IoT standards and the intrinsic complexity demand for proper security layers constitute the need of holistic IoT security solutions imperative. Apart from some notable exceptions, such as [93] where the authors propose a methodology to validate and certify different technological solutions in large-scale conditions and [94] where the cybersecurity aspect regarding the communication between IoT devices and external entities is addressed, there is still a long way until the total armour of the IoT.

Several research papers, derived by the work done in various EU funded projects, exist in the literature mainly focusing on crucial aspects of the IoT domain, such as the interoperability in different IoT environments [95] and for heterogeneous testbeds [96], privacy [97] in terms of authorisation and sensitive information handling, cloudification [98] and smart applications and services towards an open IoT ecosystem [99], just to name a few. Nevertheless, there are numerous issues left open for further discussion, with the most prominent one being the security in IoT.

GHOST project aims to close this security gap by providing a generic, hardware agnostic, security solution for smarthome installations. It takes into account multiple different protocols and monitors the behaviour of all installed IoT devices along with the activity of the smarthome gateway. The system automatically handles detected security events, while self defending mechanisms have also been employed to ensure its normal operation. It requests user intervention only when this is absolutely required, while a lot of effort has been concentrated on the usability of the interfaces used for user interaction. Additionally, GHOST solution has been designed upon the restriction that it should be functional while running on limited

hardware resources, an evident constraint for smarthome gateways. The developed algorithms are performance efficient and require minimal resources. In a few cases where additional hardware resources are required, a lot of attention has been given to preventing sensitive personal data of smarthome inhabitants leave the gateway, and thus any privacy implications are eliminated. Finally, while blockchain technology has been employed, there is no requirement for significant hardware resources. A modular architecture has been implemented, that enables the blockchain related components, to either connect to external blockchain nodes or run a local lightweight node inside the smart-home gateway.

4.3 System architecture

The conceptual design of the GHOST architecture [21] relies on the thorough identification of all crucial elements of the wide attack vector applicable for the smarthome environments. Due to the numerous constraints depicted in related works, GHOST follows a network monitoring and anomaly detection approach, allowing to preserve the existing heterogeneity of the IoT devices deployed in the smarthome and focusing on the analysis of the generated network activity. The conceptual design was further enhanced through the functional requirements extraction process directly from the end-user need analysis and advancements in the research on a security intelligence available within the consortium. To this end, GHOST pursues a layered system architecture approach, allowing independent development of the separate components, while preserving a high inter-dependency within the framework. This section describes in details GHOST's system logical layers depicted in Figure 4.1.

4.3.1 Gateway

This layer focuses on linking an already existing gateway software environment with GHOST solution, mostly composed of the Interoperability Middleware (IM). Its main goal is to provide a uniform access to the gateways managing IoT devices in the smarthomes. This is where the actual traffic packet capture is realised, ensuring consistent implementation across different protocols supported. An important effort is devoted to gaining a near real-time performance, and part of the research is to strike a balance between having real-time and a low as possible resource use impact. Furthermore, this layer is responsible for the sensitive operations at the core of the system, such as modifications to the *iptables*, aggregation of device events for inclusion in the risk assessment and direct control on IoT devices. For example, *iptables* cannot be used for Bluetooth communication, it is therefore not possible to block a certain Bluetooth communication flow. Instead the whole device can be excluded if necessary through the control commands through the unified Application Programming Interface (API), exposed to the internal modules.

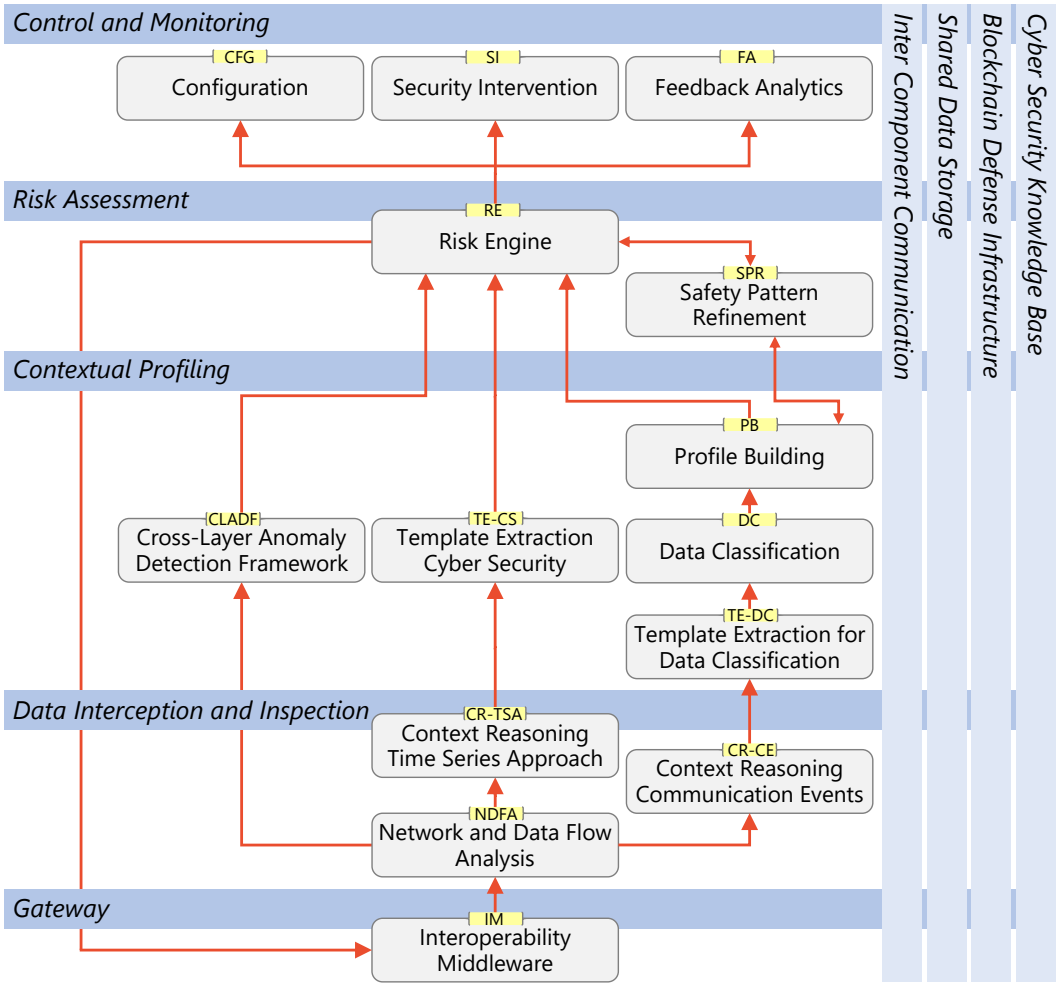


Figure 4.1 GHOST system architecture.

4.3.2 Data Interception and Inspection

Responsible for the direct network data gathering and extraction, this layer is composed of three modules: Network and Data Flow Analysis (NDFA) Context Reasoning Time Series Approach (CR-TSA) and Context Reasoning Communication Events (CR-CE). The NDFA component takes incoming network traffic that is going through the IM and extracts 'valuable' data, to be utilised by other components afterwards for anomaly detection. For all supported protocols (IP, Bluetooth, Z-wave, RF869) full packet data are being retained for a certain time-frame in the Shared Data Storage (SDS). Consecutively, a data release strategy, based on time interval & size, is applied. Whenever a new packet or flow is detected, the NDFA directly propagates this event through the Inter Component Communication (ICC) to all subscribed components, which in turn will access the SDS for its related data.

The CR-CE component extracts meaningful context information (*generic metrics*) and CR-TSA extracts metrics specific for attack detection (*cybersecurity metrics*). The context information is further utilised by the upper layer's components to identify user data, with a special focus on the privacy monitoring. Its focus lies on knowledge about similarities and repetition of similar events, and deduction of the reasoning for each particular communication. Using the data flow information prior processed by NDFA, the components process these data to identify the communications related to distinct events occurring for smarthome devices.

4.3.3 Contextual Profiling

This layer provides current state of data identification and related behaviour of the IoT devices' generated network data and is relying on the performance of several components: Template Extraction Cyber Security (TECS), Template Extraction for Data Classification (TEDC), Data Classification (DC), Profile Building (PB) and Cross Layer Anomaly Detection Framework (CLADF). TEDC is using the context information (*generic metrics*) to create templates according to the communication patterns of the devices. For example, a motion sensor template can contain the type of packets being sent to the Gateway (GW), their frequency, the number of device or personal packets sent during a day. TECS utilises the *cybersecurity metrics* which are based on *cloud trained Artificial Intelligence (AI)* to detect attacks with a probability which is then fed to the Risk Engine (RE) through the SDS.

The DC component is responsible for classifying data as content data (user-related data) and device events. It applies a variety of algorithms together with the templates from the TEDC to process incoming traffic. A certain probability is given to each captured packet or flow and its classification. The communication between the TECS and the DC follows a reinforcement learning scheme. Information regarding the classification quality, when ground truth class labels are available through user feedback, is provided back to the TEDC, to probe different configurations for more accurate classifications.

The PB component is responsible for building behavioural profiles of devices. Its purpose is to guide the detection of abnormal behaviours in communications and of new devices that

enter the network. It incorporates the results from the DC and the metrics from the NDFA for building *behaviour graphs*. Its self-improvement mechanisms are relying on the feedback intelligence distributed by the Safety Pattern Refinement (SPR) from the Cyber Security Knowledge Base (CSKB).

The main purpose of the CLADF is to utilise existing cybersecurity features and combine them in a unified output. The output of the CLADF is used by the RE directly to perform a risk assessment and to provide visual support in the representation of any reported event by the Feedback Analytics (FA). A significant aspect targeted by CLADF is the correlation and combination of different events. Even though a particular cyber attack may trigger several alarms in different tools, it is important to understand the semantics of the events and to find a possible correlation by combining several apparently distinct events into a unified output, including data on the possible source, time, magnitude, and severity on the event.

4.3.4 Risk Assessment

This layer combines intelligence reports regarding noticeable alerts and performs real-time risk assessment. It is composed of RE and SPR. The RE is responsible for the risk assessment for any communication, correlating device activity on the network with the prior established metrics and profiles available from the CP layer. The communication is being processed through a variety of analysers, handling different aspects of the incoming data, such as the *behaviour* support by the PB component, the *payload* related to the DC and NDFA, the *blocking rules* interfacing with the IM component and the *alert processor* for handling detected attacks/anomalies by the CLADF and the TECS. In addition, it is enhanced with the Blockchain Defense Infrastructure (BDI) protection to ensure the reliability and trustworthiness of the automated decisions.

SPR is optimising the performance and efficiency of the PB and RE components and acts as an intermediate for providing safety insights to the RE's operations and strengthening the PB device profiles. These insights are foremost coming from the CSKB and indirectly from the RE and PB themselves. The information from the CSKB is based on the *user feedback* decisions on blocking communications aggregated and analysed from all connected GHOST installations.

4.3.5 Control and Monitoring

Visualisation of the available risk reports through user friendly interactions is ensured by components of this layer: Configuration (CFG), Security Intervention (SI) and FA.

The CFG provides the means and functionality to the user to control and configure the GHOST platform. It has two stages, the first being the *first time usage* and second the *regular usage* of the configuration pages. All configuration options are stored in the SDS and are mostly directly related with the RE component itself, calibrating the risk assessment parameters. A key focus is given to the effortless and usable design of the configuration setup

process and further settings review and fine-tuning of applied configuration policies. These include, *authentication management* and *factory reset* at its highest level and control over the *blocking rules*, personalising the *risk levels* and *awareness automation* directly related to the risk assessment.

The SI serves for user friendly visualisation of the risk tracking and evaluation results. The appropriate visualisation and human-machine interaction mechanisms are put in place to allow users to effortlessly and effectively review security issues and *take key decisions* that affect their privacy and security. Visualisations are fed by data analysis and results from the RE component, tailored in relation to the usability studies and results from real-life trials. A set of scenarios have been extracted in which a certain mitigation action by the RE is realised.

The FA component is responsible for providing high-level monitoring and analysis of data originating from the DII, CP and RA layers. The input data include historical and current packet flow behaviours, risk levels, device profiles, packet classification scores and any metrics available and deemed suitable for display. The input data are used in order to provide visual and intuitive presentations and reports of the smarthome security status, including visualisations of packet features through time, visual monitoring and distinction of packet/flow behaviours, and visual identification of potential anomalies and vulnerabilities. For this purpose, existing visualisation techniques, employing multimodal graph-based visualisations, will be adapted to the data generated in the home environment.

4.3.6 Blockchain Defense Infrastructure

In order to ensure the integrity of the data exchanged among devices for central decision making for risk assessment, GHOST uses a variation of the default blockchain approach.

Public blacklisting The RE is able to assess the risk imposed by the connections and communications between internal with external end points, which may result in a mitigation action to push a malicious IP onto the blockchain. Here, a list of malicious IPs is collaboratively created and maintained by different installations.

Forms of consent As part of the operation of the BDI network, records of transactions are hashed by the miner nodes in an encrypted format, including potentially sensitive user data, such as records from medical devices. Informing the users about the operating principles of the network as well as to request the acceptance of the principles by the users is done by digitally signing a *Form of Consent*.

Software integrity The BDI network can utilise a new firmware update scheme, based on a synergy between the Blockchain network and a BitTorrent network. That way, the version of the firmware can be checked securely, its correctness can be validated and the installation of the most up-to-date firmware on all the devices of the network can be ensured.

4.3.7 Cyber Security Knowledge Base

The CSKB is a cloud-based knowledge repository, which collects anonymised security intelligence and insights from external web sources and other GHOST instances. It maintains a list of malicious actors and properties (IP addresses, domains, URLs, file hashes). These data are produced by feedback from the users through the FA and SI components; scraped regularly from open online research; collected from specific commercial feed publishers; generated by correlation triggers and malware analysis engine. The information is further analysed and propagated to the SPR for enhancing each individual GHOST platform.

4.3.8 Shared Data Storage

At the core of the SDS lies a PostgreSQL database and a service that provides easy access to it. All internal GHOST components have the option to directly access the database, whereas a more secured interaction is provided by the service for *external* access. For example, for the GW itself and the exchange of configuration data or information regarding the devices which are subscribed to the GW. Additionally, the data encryption mechanisms are put in place to comply with security requirements.

4.3.9 Inter Component Communication

The ICC component is the glue between all the components for direct communication and, based on ZeroMQ[1], it offers two exchange patterns:

- Request/Reply, a client connects to a service and performs a *request*.
- Publish/Subscribe, a client(service) sends data to a set of *subscribed* clients, with the possibility to set an intermediate broker.

Furthermore, messages between the components are encoded by Protocol Buffer[2], efficient method to serialise structured data.

4.4 Integration and validation strategy

GHOST project tackles several challenges in terms of integration of the solution due to its modular and interrelated architecture presented in previous section. Therefore, the development process that has been followed relies on the use of a funnel approach, both in terms of innovation and development.

[1] http://zeromq.org/
[2] https://developers.google.com/protocol-buffers/docs/overview

4.4.1 Innovation and development approaches

GHOST follows a user-centred methodology where several experiments have been defined to lead the road from the conceptual idea to the final market solution. To this end, the initial stages of the project have involved the user through online questionnaires and focus groups with potential users of the system. These experiments led to the conclusions that there was a lack of awareness of the potential risks associated to cybersecurity [100] and the need of assistance in the configuration and management of the system (tips and tricks, baseline guides, etc.). During the lifetime of the project, the realisation of trials in real-life installations, involving up to 200 people is envisioned. In this way, the GHOST's functional design is being continuously evaluated to find the most useful and usable features and characteristics through the continuous involvement of users within the development process, fostering the innovation of the solution and the route to the market.

4.4.2 Cybersecurity validation

In addition to the innovation path, technical feasibility and robustness of the system, GHOST validation strategy is also focused on the demonstration of its capabilities to detect and prevent cyber threats. To this end, three possible types of attacks have been defined and analysed from the GHOST perspective:

Physical attacks related to physical actuation over one or several devices leading to malfunctioning of these devices. This category is formed by attacks such as physical damage caused by the removal of the battery, shut down of the proper device or physical breaking of the device, injection of an actual device with malicious objectives in the network or mechanical exhaustion of physical buttons or triggers that creates, in the long term, malfunctioning of the device. GHOST addresses these type of attacks through the detection of changes in the communication patterns between the devices and the gateway, where the rate of communication increases, decreases or becomes absent all together.

Network attacks related to direct actuation over the network traffic to cause malfunctioning of the system or capturing relevant information. This category includes well-known traditional attacks normally based on IP protocols (such as network scanning and enumeration techniques as TCP/IP and UDP related scan, Denial-of-service (DoS) or Distributed denial-of-service (DDoS)), device impersonation attacks where the attacker injects packets captured from an authorised device to trigger other devices or sniff communications, or artificial creation of network activity causing battery drains by eliminating, for example, idle times of devices. GHOST solution is dealing with these attacks through the analysis of changes in the communication templates characterising each device in the network and monitoring the evolution of flow based communication profiles between different devices within the network.

Software attacks based on gaining access to a device within the network and using or altering its software to provoke malfunctioning of the specific device or of the service. In addition to the well-know traditional software-based attacks that implies the exploitation of software flaws through the use of viruses, worms or malicious script executions, this category also analyses specific to IoT attacks, such as software compromising in the gateway or in the devices (where the attacker gains access to the software inside the device and modifies somehow its behaviour), the injection of unexpected commands or communications between two devices on the same network by utilising gateway legitimate communication channels for malicious purposes or sleep deprivation of the device leading to battery drains (where the attacker is able to change the logic of the device forcing longer wake up times than usual and affecting directly the service). The specific blockchain-based integrity checking mechanisms defined for the GHOST solution in combination with the network monitoring and analysing tools of the system are key assets for protecting the smarthome against this type of attacks.

The GHOST architecture has been designed to cover a broad set of attacks due to the monitoring of critical parameters, probably affected by any attack designed, covering possible omitted attack vectors. The combination of data analysis (for extracting templates of devices and for analysing changes in the data within the network), blockchain (for protecting the integrity of the firmware of devices) and other security-related technologies enables a multi-sided cybersecurity tool for smarthomes. GHOST detection and prevention capabilities against above identified attacks will be tested in specific testbeds in the partners' facilities, to avoid leakage or burden risks to the end-users participating in the real-life trials.

4.4.3 Integration methodology

Layered and multimodal architecture of the GHOST, the device-agnostic concept and the use of state-of-the-art tools and technologies raise the complexity of the integration process. To overcome this complexity a combination of development methodologies to make the most of the development efforts was used. At early stages a waterfall approach was followed until the release of the first prototype. Additionally, it was combined with an iterative agile approach for continuous feature improvement. The waterfall approach in early stages of the project ensured the clarification of the basic requirements of the solution and the coherence in the development in spite of the interdependence of modules. However, it showed limited effectiveness for managing the continuous integration and, therefore, the continuous change of the requirements when incorporating the feedback of the end-users.

To solve this lack of flexibility, GHOST adopted a SCRUM-based approach (OpenProject[3]). Designating a product owner for each GHOST module, based on the interrelation between modules. The product owner is the main user of the information generated by each module and/or the interfaces, where the product owner is in direct contact with the end-users (incorporating the user's feedback in the technical development).

[3]https://www.openproject.org/

The product owner creates monthly sprints by prioritising the tasks in the backlog of each of the modules according to the needs of the user. Consecutively, monthly releases of the GHOST platform are remotely uploaded to the gateways within the trials and directly presented to the end-users.

4.5 Discussion and future work

This paper proposes a cybersecurity reference architecture, tailored for smarthomes consisting of Internet of Things devices. The work performed under the European research project GHOST targets efficient countermeasures for defencing cyber attacks on lightweight smarthomes gateways. The design of the functional elements of the architecture was derived from thorough analysis of the IoT infrastructures and particularities of smarthomes environments, further enhanced by the specific needs of the end-users. The different architectural layers presented, despite their high inter-dependencies, cover different scopes for the detection and mitigation of attacks, from network analysis and reasoning to security intervention and analytics.

The followed user-centred approach and the continuous involvement of the end-users in the design and evaluation phase of the project, made the validation and integration more complex. To overcome the complexity of these tasks, a detailed validation plan has been developed, while an agile development and integration approach has been adopted, providing the required flexibility to the project in comparison to waterfall based approaches.

The real-life trials, executed in three phases, where the actual deployment of the system on smarthomes will take place, will offer significant feedback on the usability and validation of the system to realistic cybersecurity incidents. Continuous improvements, through the agile approach and the iterative real-life trials, will construct a solid and effective solution for the protection of smarthomes using lightweight gateways.

Part III

Risk Assessment

Chapter 5

Article III: Towards Automated Threat-based Risk Assessment for Cybersecurity in Smarthomes

Relevance

This article addresses RQ2 by designing the quantifiable RAM to automate the RA process. The presented model is supplemented with a use-case study demonstrating an application of the calculation model on a threat-triggered scenario. This article showcases identification of the risks based on historical evidence from the network activity, estimation of the possible impacts and derivation of the most optimal mitigation measure.

Context

This article was presented at the 18th European Conference on Cyber Warfare and Security (ECCWS 2019), Coimbra, Portugal and published in the conference's proceedings. According to the Resurchify portal, the IS: 0.31, h-Index: 9 and SJR: 0.141.

Own Contribution

Being one of the lead authors of this paper, my contribution to this work is definition of the threat-triggered scenarios and the calculation model with an input parameter definitions and boundary value estimations leading to a sample that served as an exemplification of the calculation model.

Chapter Contents

5.1 Introduction

The goal of the GHOST project [21] is to provide a cybersecurity solution targeted at the non-expert citizens by raising their awareness and understanding of the security risks associated with all aspects of cybersecurity from threats and vulnerabilities identification and personal data leakage detection up to making informed decisions affecting their cyber-physical smarthome. GHOST aims to transform smarthome occupants' decisions into reliable automated security service, promoting user-friendly end-user habits through usable security. The RA is a central functionality of the GHOST software implementation focused on the context-aware real-time threat protection. It gathers information about the current risks, analyses in real-time current network traffic flows and correlates them with the normal behaviour of the smarthome. RA is responsible for determining at multiple stages in the processing of the data what the current Risk Level (RL) is. This RL is associated with a particular action a device or an end-user is about to take. RA validates real-time communication context using device behaviour profiles, entailing the processing of the communication context properties. The fusion of the RLs accepted according to user preferences and of typical behaviour stored in security patterns allows an automatic decision making, where RLs matching and comparison indicates the appropriate security action: allowing or blocking the whole communication stream, or propagating the intervention to the user interface for the end-user's approval or correction. The structure of this paper is as follows. The recent advancements in the field of Behaviour Analyser (BA), Risk Prediction and Estimation (RPE) and Mitigation Techniques (MT) are presented in Section 5.2. Section 5.3 explains the RAM approach, whereas the calculation of the RLs is demonstrated in Section 5.4. The application of RAM in a selected scenario is presented in Section 5.5. Finally, conclusions and directions for further work are summarised in Section 5.6.

5.2 Related Work

Schiefer [15] demonstrates the challenges that RA poses in a smarthome installation due to the heterogeneous nature of the IoT devices. The spectrum of the threats for smarthomes is twofold, namely privacy and security related. However, in most cases, the attacks are targeting both aspects. Unfortunately, the biggest problem still relies in primitive security settings that are ignored by unaware users. According to Sivaraman, Habibi Gharakheili, and Fernandes [17], multiple security incidents involving IoT devices exploit primitive attack vectors, such as the use of default passwords or weak communication protocols. The most notorious example is the break out of the Mirai botnet [19], that took over at least 100 000 IoT devices. From the above, it is evident that a non-expert user has no way to perceive the full picture of the potential risks involved in the smarthome she is living in, and that an automatic security risk monitoring solution is essential.

5.2.1 Behaviour Analysis

One of the approaches widely used in proactively managing security incidents is BA. In the case of smarthome security, BA can be applied directly on any existing network at the router/gateway entry/exit point of any smarthome installation. In terms of the approaches used in BA, Machine Learning is the most common method used for anomaly detection. For example, Saad et al. [26], successfully identified malicious behaviour on the network by comparing application of several existing ML classifiers. Zhao et al. [27] expanded the existing method with the use of the decision trees, allowing zero-day detection of the involvement in botnet activities. The framework proposed by Nari and Ghorbani [28], aimed at detecting malware, is using behaviour graphs, improving the accuracy and false positive detection by incorporating graph attributes.

5.2.2 Risk Prediction and Estimation

In Kitchin and Dodge [29] provide a risk overview for the case of smart cities. This survey can be considered the closest on the risk analysis, vulnerability and MT identification in the field of CPS security. There, the authors determine five main vulnerability categories: (i) Weak software security and data encryption; (ii) Use of insecure legacy systems and poor ongoing maintenance; (iii) Many inter-dependencies and large and complex attack surfaces; (iv) Cascade effects; and (v) Human error. The same categories are also applicable to the case of a smarthome environment. Furthermore, Almohri et al. [30] suggest to incorporate threat modelling for RA directly at the IoT device design stage, distinguishing three main approaches: attacker-, system- and asset-centric [31]. Rao et al. [12] present a very promising approach, based on the execution time of the processes in a CPS environment. This approach is the closest to the work in GHOST, in terms of dynamic real-time RA.

5.2.3 Mitigation Techniques

Current research in the glsmt does not spread much further than providing generic recommendations for formal risk evaluation processes. The closest work presented in [29], provides guidelines for smart cities environment. The authors recognise three main categories of glsmt: (i) Security by design; (ii) Traditional security mitigation; and (iii) Formation of the core security teams within the administrative staff supporting infrastructure installations. However, no further dynamic and automatic solutions are presented in the relevant literature.

5.3 Proposed Risk Assessment Model

The approach taken for RA in GHOST involves the use of predefined RL. “Negative to Positive Model” [101] was adapted for RL definition relying on four-dimensional correlation between values and activities. This model assesses risk on the basis of the cost (or benefit) associated with the option to either proceed with an action or not and turns negative values (cost) to

positive (yield/return). Use of this model in our case results into the definition of four RLs, as shown in Table 5.1.

Table 5.1 Risk Level Definitions.

	Question	Example
RL_1	What will the positive value be if an activity is done?	Compliance with privacy laws thus at the lowest level of risk in failing the compliance
RL_2	What will the positive value be if an activity is not done?	Collecting anonymised user information thus at a slightly higher level of risk in the event of failure of anonymisation technique and/or data theft
RL_3	What will the negative value be if an activity is done?	Collecting personal information and sharing the data with unauthorised third party
RL_4	What will the negative value be if an activity is not done?	Not anonymising the user data and paying penalty for the misuse of the data

The Basic Value Model (BVM) [101] is used to estimate the positive or negative value involved in each RL. The principle of BVM which is based on three different characteristics is shown in Figure 5.1.

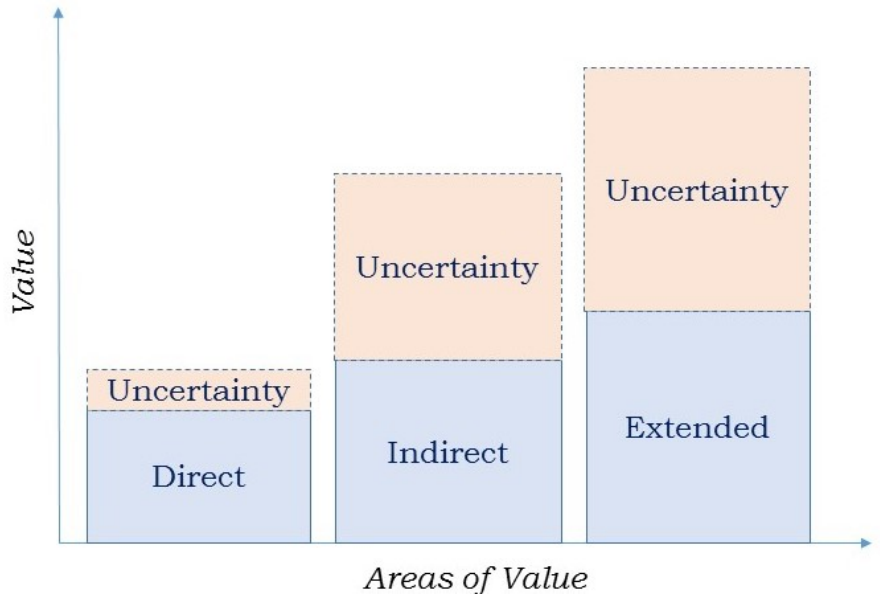


Figure 5.1 Principle Basic Value Model.

With reference to the BVM, the following definitions apply:

- *Direct Values* direct economic values, such as failure of a device, or direct investment based on an occurrence which could be active or passive.
- *Indirect Values* the additional and more intangible values gained or lost, having a greater uncertainty and as such they can be within ranges. For example, the unavailability of services due to DDoS attacks or increased administrative tasks.

- *Extended Values* reflect the values affected by the direct and indirect values and can be significantly huge and are also affected by other factors, such as impact on society and/or the GHOST network as a whole. Extended values of items such as brand or reputation are often difficult to quantify. Extended values are mostly negative but may also be positive as a consequence when information security is applied.

Addressing the four RLs and corresponding questions (Table 5.1) in combination with the principle BVM, led to the creation of a balance board to assure coverage of all risk relevant aspects. The potential duplication of the values related to the same activity is consecutively handled by using a simple balance table as shown in Table 5.2. Note that the factor C (Cost) is not applicable to the formulation used, but is part of the original BVM model.

Table 5.2 Balance Table for Net Values.

	Base	Activity	Positive Value Activity Done	Positive Value Activity Not Done	Negative Value Activity Done	Negative Value Activity Not Done	Net
Ref			A	B	C	D	
1	A possible activity to change the current situation	Activity "XY" done	Value	Not Applicable	Cost	Not Applicable	A1-B1
2	The possible activity not done	Activity "XY" not done	Not Applicable	Value	Not Applicable	Cost	B2-D2

5.4 Risk Exposure Calculation

Estimation of risk exposure at different RLs is based on incorporating a multitude of Influence Factors (IF). Their listing along with the current integration status in the RAM is outlined in Table 5.3.

Table 5.3 Types of Influencing Factors.

Type of IF	Description	Status	Reasoning
Physical	Sum of the tangible assets that comprise the GHOST network	Yes	Devices, sensors, or any IoT assets in a smarthome
Customer/ User	Smarthome residents/owner	No	Perception factor, to be quantified
Societal	Perception that the society in general has about an appliance/device in the GHOST network and network as a whole	No	Perception factor, to be quantified
Reputational	Perception that competitors, suppliers, customers, government and other stakeholders have about the devices in the network and services provided by the GHOST network	No	Perception factor, to be quantified
Intangible/ Logical	Intangible assets handled by the GHOST network such as user data, forms of consent, blacklisted IP addresses, software integrity	Yes	Information/ data and services generated/ available in a smarthome
Legal/ Regulatory	Potential sanctions and/or penalties that might result from a breach	Yes	Data protection regulations, service contracts and legal obligations

The calculation model for RLs is defined as follows and is based on the balance table (Table 5.3).

$RL_1 = T \times (V_1 \times A)$; $RL_2 = T \times (V_2 - AC_1)$; $RL_3 = T \times C$; $RL_4 = T \times (AC_1 + AC_2)$

Where T = Time period, V_1= Value created by taking an action, A = Risk reduction as a result of action taken, V_2 = Value created by not taking an action, AC_1 = Additional internal cost, C = Cost associated with an action, AC_2 = Additional external cost. Steps determining the RL in relation to an action taken:

1. If the action is completed, then go to step 2 else go to step 3.
2. If $RL_1 > RL_3$, then RL = RL_3 else RL = RL_1
3. If $RL_2 > RL_4$, then RL = RL_4 else RL = RL_2

5.5 Demonstration and Evaluation

We use a scenario based approach, a common practice in Design Science Research Method (DSRM) [102] for ongoing work, to demonstrate and evaluate the application of the proposed RAM in the given scenario.

5.5.1 Example scenario - A to B communication

Internal IoT device A (Table 5.4) is sending data to malicious entity B (malware.com). B is already blocked by GHOST firewall (eg iptables).

Table 5.4 Device Exposure Vectors.

Device	Exposure	Data
IP static camera	Wi-Fi connection, Motion detection, Remote control, Night vision, Video & sound capturing, Face recognition	System status, Configuration data, Video frames, Credentials, Facial profiles

Possible GHOST actions to take on this suspicious situation are listed in Table 5.5.

Table 5.5 Action and Consequence Correlation.

Action	Positive Consequences	Negative Consequences
Block outgoing communication from device A to B	Controlled traffic, Avoiding privacy infringement of data sent to malware.com, Avoiding ransomware attack	Partial service disruption, User discomfort as no alert is received
Block all outgoing communication from device A	Controlled traffic, Avoiding ransomware attack	Full service disruption, Exposure to theft
Allow outgoing communication from device A to B	Continuous monitoring of sick (elderly) person, Physical security monitoring	Remote control by unauthorised party, Privacy violation, Involvement in DDoS, Potential danger in extreme scenario, GDPR regulatory fine, Ransomware

5.5.2 Application of Proposed Model

The proposed RAM is applied to the above-mentioned scenario, and few assumptions are made for the data used in the calculations below to demonstrate the positive and negative values of doing or not doing the required action.

RL_1: Positive Value – Activity Done

Let us assume that by removing the device from the network, we gain a positive value of EUR 5000 (from the positive consequences as listed in outlined scenario). Time period under consideration is 1 day. Risk reduction for the GHOST network in the given home is 90%.

Hence, $T = 1$, $V_1 = 5000$, $A = 90\%$. Therefore, $RL_1 = 1 \times (5000 \times 0.9) = 4500$. *$RL_2$: Positive Value – Activity Not Done*

Let us assume that by not removing the device from the network, we gain a positive value of EUR 3000 (from the positive consequences as listed in outlined scenario and annotated with (+)). Further, there is an additional cost associated with the unwanted data flow between A to B, which we assume as EUR 1000.

Hence, $T = 1$, $V_2 = 3000$, $AC_1 = 1000$. Therefore, $RL_2 = 1 \times (3000 - 1000) = 2000$. *$RL_3$: Negative Value – Activity Done* Let us assume that the negative consequences are critical in nature and by applying a method like Cyber Value-at-Risk (CVaR) for the above consequences as listed in outlined scenario and annotated with (–), we get an estimated cost (negative consequence) of EUR 8000.

Hence, $T = 1$, $C = -8000$. Therefore, $RL_3 = 1 \times (-8000) = -8000$. *$RL_4$: Negative Value – Activity Not Done* Since the device is not removed, the associated external cost is estimated by using a method like Single Loss Expectancy (SLE) for the above-mentioned negative consequences as listed in outlined scenario and annotated with (–). Let us assume that by applying SLE we get EUR 10000.

Hence, $T = 1$, $AC_1 = 1000$, $AC_2 = -10000$. Therefore, $RL_4 = 1 \times (1000 + (-10000)) = -9000$. Based on the output values at the respective risk levels for the given scenario, the user can take an appropriate risk management decision whether or not to take the underlying action.

5.6 Conclusion and Future Work

The RAM presented in this paper is currently an ongoing research and development effort and is at the heart of the GHOST solution for RA. Deployed at the network traffic capture level, the incoming data is constantly monitored and fed into several distinct analysers. The resulting output is a set (zero or more) of risk related properties. Further grouped into identified risks, they serve as a base for the exposure value calculation. Various RLs at multiple stages of data processing are evaluated and monitored to ensure permitted RLs of current activity at each case, practically determining the required action to be taken. Experimental evaluation of the risk boundaries is enabling further fine-tuning of the calculation model to achieve automatic risks assessment. It is envisioned to perform several iterations of the model values refinement

through the data obtained during the trials. Furthermore, a process on effective allocation and association of the mitigation actions should be identified. The current prototype relies on the hard-coded set of the actions extracted from the set of predefined attack scenarios.

Chapter 6

Article IV: Can I Sleep Safely in My Smarthome? A Novel Framework on Automating Dynamic Risk Assessment in IoT Environments

Relevance

This article is a fundamental work to investigate RQ2 and RQ3. Here we present all theoretical concepts designed for DRA, their implementation and validation of the resulting framework. The contribution of this work towards the finalisation of RQ1 is threefold. First, we incorporate into the deployment environment the realisation of the IoT Stack ontology. Secondly, the DRA is implemented according to the reference architecture defined in Chapter 4. Finally, execution of the decision automation evaluation in the smarthome context for RA proves the applicability and appropriateness of the IoT Stack in IoT enabled environments.

Context

This article was published in the Electronics journal's special issue New Challenges on Cyber Threat Intelligence. According to the Resurchify portal, the IS: 3.02, h-Index: 36 and SJR: 0.36.

Own Contribution

Being the lead author of this paper, my contribution to this work is conceptualisation, methodology definition and execution, software implementation, validation and experiment execution, and overall visualisation.

Chapter Contents

6.1 Introduction

The IoT has attracted considerable attention during recent decades and still presents a significant opportunity for many industrial and business stakeholders in various domains [14]. Smarthomes are adopting IoT as an emerging technology to provide specialised services to control household appliances, automate specific tasks and, in general, improve quality of life. These devices take all forms and shapes, varying from a smart lightning or heating system to a smart fridge. Moreover, they are embedded in our daily appliances, making it less and less transparent what data is going in and out of our homes, leveraging existing and creating new cyber threats. Each of these devices, capable of communicating with one another and with external services accessible through the Internet, creates unmonitored data flows, unknowingly exposing us, regular users, to all kinds of cyber threats. The heterogeneity and diversity of the 'Things', as well as new lightweight communication protocols appropriate for IoT technology, create new challenges for the protection of such systems [103]. This gives a rise to the need for tools that provide visibility into the cyber risks and threats in smarthomes in an easy and understandable way, tailored towards people without a deep knowledge of Information Communication Technology (ICT) in general. Such tools will empower their users to take control and make appropriate decisions regarding home cybersecurity and their privacy.

Despite the acknowledged need for RA for smarthomes [33, 11], the risk awareness solutions available for regular users with the purpose of understanding ongoing risks evolving on a daily basis are still very immature. Several academic works aim to provide policy based solutions [104] and formal periodic RA frameworks [33, 105]. Other works demonstrate limited system knowledge and a lack of dynamic adaptation [11], lack of risk propagation and dependencies understanding [106] or lack of usability for regular citizens [107]. However, to the best of our knowledge, none of these works are capable of establishing a holistic approach to include the identification of risks and automation of mitigation measures in a dynamic manner.

The main goal of the proposed framework is to provide a real-time security and privacy RA of the ongoing activities on the network. It validates current communications by assessing any anomaly detected and deviations in the IoT device's normal behaviour through device profiles. DRAF performs real-time RA by continuous evaluation and monitoring of various risk levels at multiple stages of data processing. To control the behaviour of devices and corresponding payload exchanges, permitted risk levels of ongoing network activity are dynamically calculated for each network activity event, practically determining the required decision to be taken. RA also involves establishing risk controls for the users' privacy and making them aware of the associated risks. DRAF integrates multi-faceted anomaly detection analysers and risk Receptors to support behaviour deviation detection, involving deep understanding of risk propagation and inter-dependencies within the network. For this purpose, it leverages the existing open threat modelling tools (eg the open cyber threat intelligence platform) to integrate a network entity correlation ontology. Furthermore, a set of expert values for

risk estimation was established to allow the comparison of associated impacts in various risk situations.

This work aims to reply to the following research questions:

RQ 1 *Can a generic ontology be developed to capture complex relationships between heterogeneous IoT properties to encapsulate vulnerabilities, attack attribution, impact evaluation, and mitigation strategies?*

RQ 2 *Can a unique risk scoring be developed to eliminate environment context dependency? How does the initial setting of the expert values for the RA offer a valid approach, and are these values generally applicable in a standard installation?*

RQ 3 *What are the limitations on the automated decision making for RA in dynamic environments, such as smarthomes, where deployed IoT devices constantly evolve (get replaced, updated and moved)?*

Our contributions can be summarised as follows:

- Development of a generic ontology for the representation of the IoT objects to encapsulate vulnerabilities, attack attribution, impact evaluation, and mitigation strategies in a smarthome environment.
- Presentation of the DRAF, encapsulating a risk scoring methodology based on the expert opinion settings. This framework was validated in real-life testbeds deployed in three European countries and demonstrates the potential of automated RA.
- Investigation into the limitations of the automated decision making for RA in a smarthome environment and their potential adaptation for similar dynamic environments, such as autonomous driving.

This work is organised as follows. In Section 6.2, main concepts are presented, necessary to understand the scope of this research, together with our analysis on the ongoing research efforts. Section 6.3 outlines the methodology we have followed for the development of the theoretical models. In Section 6.4, the implementation of the proposed approach framework on dynamic RA is described. Then, in Section 6.5, the map of the performed experiments, the deployment setup and the results are presented, followed by the analysis of the research questions and identified challenges in Section 6.6. Finally, Section 6.7 concludes the paper and outlines directions of future work.

6.2 Related Work

6.2.1 Background

RA is a holistic process of measuring the probability and severity of various effects on a system in question. In general, it is based on many unknowns, yet our concern should be even

higher when operating in a setting where we know very little. It is the only way to enable any involved stakeholders to make pragmatic decisions on the system when those risks will eventually occur. It is a multidisciplinary approach and can be broken down into five stages:

1. Risk Identification
   - What can go wrong?
2. Risk Modelling, Quantification and Measurement
   - Assessing likelihood;
   - Modelling relationships between risks and impacts.
3. Risk Evaluation
   - Trade-offs in terms of costs, benefits and risks;
   - Multi objective analysis.
4. Risk Acceptance and Avoidance
   - Decision making through level of risk acceptability;
   - How safe is safe enough?
5. Risk Management
   - Execution or actual implementation of decision making.

6.2.1.1 Traditional Approaches

RA is a process enabling the identification, estimation and prioritisation of risks associated with different dimensions: activity, operation, subject, environment.

Risks can be evaluated for each dimension in isolation or in a mixed setup, where various combinations are possible. RA sets the foundations for the following risk evaluation stages [7]:

- Acceptance: acknowledgement of the possibility of the risk to occur in a specific setup, and taking the responsibility of dealing with the caused consequences;
- Mitigation: taking actions to limit the exposure of the risk and its consequences by controlling and limiting its occurrence;
- Transfer: delegation or propagation of the risk occurrence to a third party capable of taking responsibility and liability of the risk's consequences; and
- Avoidance: ignorance of the risk occurrence likelihood and assumption of risk non-existence, as evidence of its occurring is too low or the associated cost of mitigation and transfer is too high.

When evaluating the risk stages and associated measurements costs, the RA relies on five main variables:

- Assets: any items of value (infrastructure or reputation);
- Vulnerabilities: how to exploit assets;
- Threats: action to exploit vulnerability (deliberate or accidental);
- Attack likelihood: probability of threat; and
- Impact: estimation of the attack consequence.

The significance and weight of the previously mentioned variables produce variations of the RA approaches and models. The asset-centric models, such as OCTAVE [106], evaluate the impact of the risk occurrences. The threat-centric models, such as NIST SP800–30 [108], are focused on the feasibility of the risk occurrence.

Furthermore, risks can be measured in two ways: qualitatively and quantitatively [109]. While the first method appears to be very simple, time- and cost-effective, it is also known to be not precise and without impact measurement, as it uses non-numeric values as descriptive results. On the contrary, quantitative methods give a numeric probability, enabling easy measurement of the impact. However, its complex modelling relies on the historical data and, therefore, cannot provide values at loss at a particular time, especially for the risks that never occurred before.

The hybrid RA methodologies aim to address those shortcomings by including user-centric concepts in traditional RA models, where the following properties are included [110]:

- Human system integration: visual representation of the system;
- Interoperability identification: considerations towards dependencies; and
- Emergent behaviour evaluation: coupling systems for a new purpose.

6.2.2 Dynamic Risk Assessment

Understanding the operational environment is crucial in complex decision making and dynamic environments [11], where the collection and projection of various contextual factors, as well as time- and space-specific data collection takes place. The formal RA models presented in Section 6.2.1 are based only on periodic assessment with limited system knowledge and lack of dynamic adaptation of the evolving situational risks, where the user is a key element in the future risk projections.

Several works exist in the domain of RA aiming to address this shortcoming with partial automation. For instance, the RA framework [12] in IoT systems was developed with periodic RA. The main reasoning for such an approach is the limitation on the system's knowledge and

dynamic adaptation due to the lack of understanding of risk propagation and dependencies between different assets.

The MS STRIDE and DREAD application for threat modelling, described in [8] and widely used in RA, attempts to solve automation characteristics, yet still relies on a completely manual approach. The application of OCTAVE methodology provided in [106] presents the top 10 risks. Nevertheless, this method once again lacks dynamic properties and is subject to one expert opinion for eventual risk score calculation. Furthermore, the linking between threats and assets is unclear.

Atlam et al. [111] proposed a dynamic risk calculation method, but only in the domain of access control for IoT devices, incorporating the real-time contextual data, such as user attributes related to the surrounding environment. It is our understanding that this work is the closest achievement towards dynamic RA in a real-time situational monitoring network flow. Fuzzy logic and expert judgement risk estimation approaches were fused together to enhance crucial aspects of risk model applications, such as dynamism and usability. Nevertheless, this work stays only at the theoretical level, with future projected validation with security expert interviews. Furthermore, a fuzzy logic enabled system implies natural language based operation. While this can be beneficial in the setting where input is taken directly from the end-users, in network monitoring solutions such an approach has severe limitations, especially in terms of scalability. Alali et al. [112] also proposed the use of fuzzy logic for the impact assessment of criminal activities. The authors developed a RA process with the application of Mamdani and Sugeno fuzzy methods and compared RA models in a simulation environment. While the viability of the proposed approach is sound, its main limitation is that RA is performed on static objects only.

Gonzalez-Granadillo et al. [107] realised a dynamic risk management framework for critical infrastructure systems. Their work was based on the fusion of three models: attack modelling via attack graph generation, RA via threat risk quantification of observed network events, and response assistance via evaluation of all possible mitigation actions and safety scoring assignments. The main advancement of this work is the integration of security action impacts into the final mitigation response. However, the whole framework operates on a dynamically loaded set of configuration files: network topology, abstract privacy policies, authorised mitigation actions, vulnerability inventory and a reachability matrix. The final reports generated are for the risk management experts, not regular users, and still need to be deployed in the operational environment after an expert analyses them and chooses the most suitable action.

This is where our advancement in automated RA, called DRAF, brings its innovation capacity and implements theoretical advancements in real-life applications. The main differentiating factors of our framework is its capability to execute RA on IoT assets in a dynamic and constantly evolving smarthome environment. DRAF addresses the limitation of existing frameworks in performing periodic and static RA by operating on the currently observed data. Furthermore, not only does it integrate human-centric aspects into an RA

model, but it also empowers its users to perform dynamic near-to-real-time RA on constantly evolving situational risks through the integration of a constant feed of the external intelligence.

6.2.3 Attack Classification

One of the most broad surveys attempting to enumerate a full attack landscape in an IoT ecosystem also proposed a classification system based on a layered approach: physical objects, protocols, data and software [59]. However, the main drawback of this approach is its overlapping attack attribution based on a singular property of any IoT object. For example, an object jamming attack, which is classified under physical object attacks, lacks annotation of the belonging property of the involved protocol characteristics.

Adat and Gupta [113] also propose a taxonomy of IoT based on the classical architecture of IoT, composed of four layers: perceptual, network, support and application. Yet, the explicit link between those architecture layers and the proposed threat classification is lacking, as only a subset of layers can be explicitly annotated. Physical issues are linked to the perceptual layer; data link, network and transport issues can be attributed to the network layer. However, support and application layers are omitted all together. Nevertheless, an in-depth review of commercial network protection solutions is presented.

Similarly, Chen et al. [114] presented an attack taxonomy based on the IoT ecosystem architecture: perception, network, middle-ware and application. Further classification of application domains was provided: industry (automobile or mining), urban infrastructure (smart grid, transportation, logistics), healthcare (medical devices) and smart environment (smarthome, smartphone, wearables).

A three-level attack classification approach was suggested by the authors of [115], where the level attribution was dictated by the severity of the security issue, ranging from low to high. The low level class was mostly correlated with physical and data link layers of communication. The intermediary level mostly concerned communications, routing and session management. Finally, the high level was applicable to applications executed in the IoT devices. In contrast, a three-dimensional taxonomy of attacks was proposed by [116], where the security landscape was analysed based on connectivity, actual device specification characteristics and the application domain. Furthermore, the attacks were classified into eight categories.

However, any attempt to uniquely attribute the attack to a specific category is prone to fail, as any IoT ecosystem, due to its heterogeneous nature, will always have a multitude of device specific properties, which will dictate attack scope. These include the affection vector, mitigation possibilities, likelihood impact, and cascading effects.

On the contrary, by focusing on the risk analysis, one can merge several attacks into the same risks, which matter the most to the end-user. Regular citizens are more keen to know the result of the attack and how it affects them rather than the technical cause at the root of the problem. For instance, the technology specific attack attribution provided in [117] can be used to form a multidimensional IoT attack correlation model, which, in turn, serves as the basis of the generic threat model and RA. The need for the development of a standardised

representation of the diversified properties of a generic IoT object is continuously increasing as we move towards attack attribution issue, especially in the context of raising end-users' awareness. Our proposal is to move towards embedding attack classification directly in the RA process, where the end-user will be presented with comprehensible information on what the possible consequences of the detected threat are.

6.2.4 Threat Modelling and Ontologies

An excellent identification of the current research gap in the domain of threat modelling and its standardisation adoption is provided in [118], outlining potential solutions to narrow the gap. A wide adaptation and referencing to the same (standardised) threat modelling framework would be highly beneficial to enable the possibility of comparative studies, where the reasoning on the model selection and attack prioritisation would be possible. Unfortunately, such a recommendation remains an unrealistic target due to the competitive nature of security research in the IoT environment. Most of the existing frameworks either come from similar, yet relevant, domains, or have a dedicated focus on some specific attributes.

Doynikova, Fedorchenko, and Kotenko [119] presented an ontology on metrics for cybersecurity assessment, as well as comparative studies on existing ontologies for security management. Unfortunately, none of them combine all necessary aspects of IoT object specific properties, correlated security issues and associated mitigation actions.

Semantic-based approaches were widely proposed in the recent academic literature. One of the most complete and promising approaches is based on the Semantic Web Rule Language (SWRL), which is designed specifically for smarthome safety services [120]. However, the main limitation of their approach is that devices' data are redirected to the middleware on the remote server, which is further uploaded to the database to map with the ontology and propagates risk related information to a safety manager. Furthermore, their ontology does not incorporate cybersecurity related attributes, such as vulnerabilities, threats and PD data exposure, constraining its capacity for attack and risk correlation.

The closest taxonomy of threats in the smarthome domain belongs to [121]. It incorporates impacts on the system and home's occupants and their lives. However, explicit links between attack classification and an IoT object taxonomy is not provided. This classification serves as a base for the definition of our IoT Stack, which takes any existing and freely available ontology one step further by incorporating the functionality of vulnerability-attack-risk association in real time. This innovative step bypasses the static evaluation of the possible attack attribution and impact prediction, making the RA process truly dynamic.

6.3 Materials and Methods

6.3.1 Attack and Risk Mapping

In a classical setting, when observing network traffic and detecting an ongoing or recent attack, one of the most important challenges is to identify the source of the threat. Once the origin of the threat is known, it can be addressed adequately.

To identify an ongoing attack, one needs to have a deep understanding of the numerous properties that each attack can possess. This task becomes even more challenging when dealing with passive network traffic observation. One way to approach this is to create a generic attack taxonomy relevant to a specific environment. However, this implies creation of static definitions and classification of known attacks existing up to today only. This means that tomorrow's attacks will not be included when using such approach. An alternative method is to start from the entities that can be deployed on the network and create an ontology supporting a generic description of properties belonging to those objects. This, in turn, enables reverse association from an IoT device property to a specific technological risk, which in turn will serve as a descriptor of attack association.

6.3.1.1 IoT Stack

For this purpose, we have developed IoT Stack ontology. As already stated in Section 6.2.4, our work is based on the existing taxonomy of the smarthome domain. With the aim to further extend it with the attack attribution properties, we have performed literature reviews and conducted several interviews with the expert groups. Furthermore, through the establishment of the reference architecture for securing smarthomes [22, 21], a comprehensive set of the generic IoT object properties was derived. Each IoT device is characterised by multidimensional properties at three layers that we have identified: hardware, software and data. An attack can not be attributed by only one property defining the IoT stack. This is why attack association with the risk always overlaps with various IoT stack properties.

A smarthome is a system that can not be secured by isolating its components. Instead, we break it apart with the help of the IoT stack, extract smaller properties, attribute them to the associated attacks and risks, and monitor the smarthome as a whole. With this approach, we can also slice the observing system when a more detailed view is required, for instance by looking at a specific IoT device as a whole system on its own.

Furthermore, multiple properties of IoT devices dictate the affection vector, mitigation possibilities, likelihood and impact vector, and cascading vectors. During our studies, we have extracted key properties allowing such a correlation, which incorporates substantial variation of the IoT objects's properties at the identified layers.

The overall concept of allowing attributions from the attack to the affected IoT device, and from specific characteristics of the IoT device back to the associated potential attack is depicted in Figure 6.1, represented with many-to-many relationship. Furthermore, an

illustrative example of the multidimensional properties specification of the IoT object by utilising the IoT Stack concept is presented in Figure 6.2.

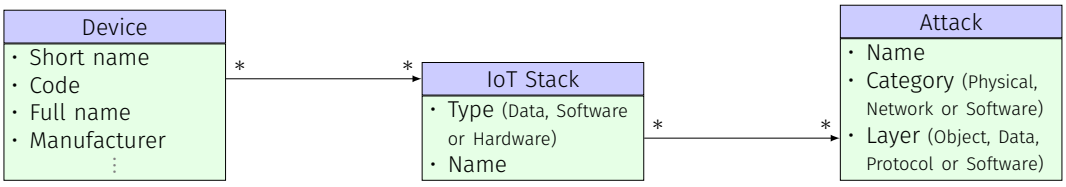
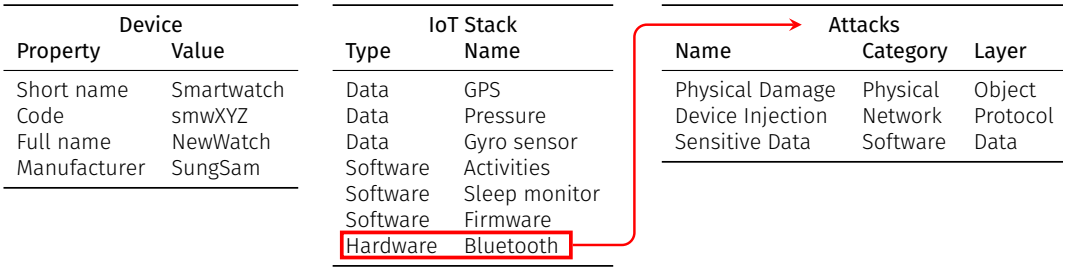


Figure 6.1 IoT Stack with many-to-many relationship concept.



Device

Property	Value
Short name	Smartwatch
Code	smwXYZ
Full name	NewWatch
Manufacturer	SungSam

IoT Stack

Type	Name
Data	GPS
Data	Pressure
Data	Gyro sensor
Software	Activities
Software	Sleep monitor
Software	Firmware
Hardware	Bluetooth

Attacks

Name	Category	Layer
Physical Damage	Physical	Object
Device Injection	Network	Protocol
Sensitive Data	Software	Data

Figure 6.2 IoT Stack example.

Such an approach enables each property to be linked to the risk through attack association, targeting the inclusion of all types of IoT objects that can be deployed in the smarthome.

6.3.2 Risk Definition

Extending the original idea of a system concept, we consider the smarthome environment as a holistic system, where risks can be measured quantitatively by utilising formal methodology on impact assessments when making decisions in critical situations.

Our initial analysis was inspired by the formal methodology widely used in the risk identification stage of the RA, called Hierarchical Holographic Modelling (HHM) [122]. This unique methodology enables the versatile aspects and dimensions of a system to be captured from the systemic modelling perspective, corresponding to our vision on the system's slicing approach, where the object in question can be observed from a different granularity view (see Section 6.3.1.1). More specifically, the definition and application of the HHM framework will serve as a basis for the smarthome risk identification, answering the prime RA question: 'What can go wrong?'. The graphical representation of the final definition of the HHM for the smarthome environment is depicted in Figure 6.3.

The main sources of cyber risks in the smarthome environment can be classified into four categories:

- Manufacturer related, such as human factors, industrial process influence, operations, specifications, and protocols utilised, manufacturing costs, prioritisation of the customer requests, time constraints and legal compliance and obligations;

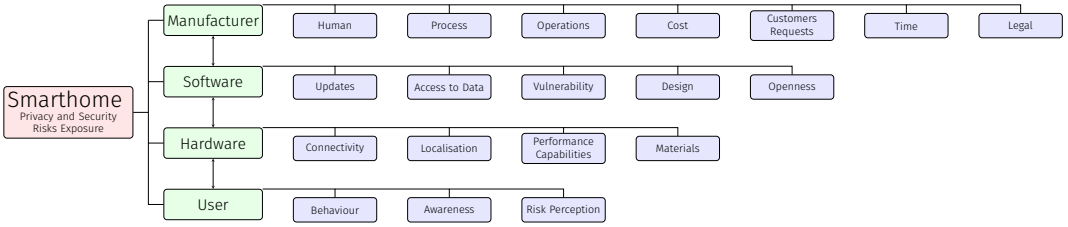


Figure 6.3 HHM for smarthome risk identification.

- Software related, such as the capability to perform and frequency of the updates, potential access to various kinds of data (personal and non-personal), vulnerability exposure, the inclusion of security and privacy threat modelling in the design process and source code openness for independent auditing;
- Hardware related, such as connectivity capabilities (direct Internet access or proxy mediator), the localisation of the object, the power and performance constraints and the actual materials used for object fabrication; and
- User related, such as behaviour patterns, general cyber awareness and human risk perception.

The visualisation and deep understanding of these perspectives set the foundation of the development of the risk model, defined in the next section.

6.3.3 Risk Model

While developing the IoT Stack concept, we have observed that, similarly to the specification of IoT objects' multidimensional properties, generic risk can also be represented by various descriptors.

The smallest piece of information describing the risk descriptor, which can also be observed directly in the network traffic, is defined as an Artefact. The Artefact itself can be defined by several properties: origin, name and message. An Artefact which is assigned a certain weight then forms a so-called Receptor. Each Artefact with different weight values will create a new Receptor. Finally, risk is composed of numerous Receptors. The main concept here comes from the idea that each Receptor can be associated with many risks, and each Artefact is accepted by many Receptors. In order for the risk to become active, ie identified to be caused by an Attack, certain Receptors should be activated to outweigh the threshold of the risk. Figure 6.4 outlines the full Object Role Modelling (ORM) schematic for the risk model from the implementation perspective. More details on the technical implementation of this model are given in Section 6.4.1.

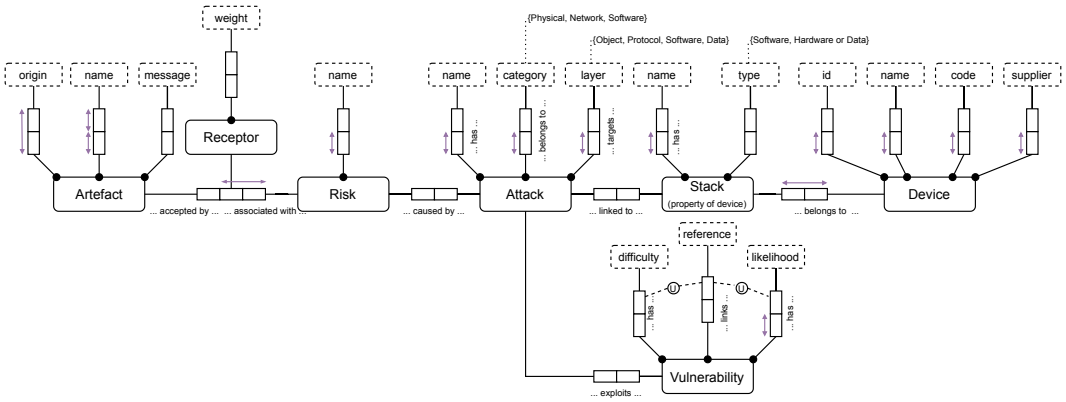


Figure 6.4 ORM for the risk model.

6.3.3.1 Risk Calculation

The main question that we want to answer here is: 'What is the likelihood that Risk X will happen, given X,Y and Z circumstances'? Let us assume that we are given R1: a PD leak. For the IoT device, iKettle, we have the following properties:

- Software: v23.6.4;
- Data: username; risk weight: 0.3; and
- Hardware: BLE; risk weight: 0.4.

In this case, the overall R1 = Medium, as the device potentially can leak only username. For the IoT device Fitbit watch, we have the following properties:

- Software: PROM24js;
- Data: age; risk weight: 0.7;
- Data: address: risk weight: 0.6;
- Hardware: WiFi; risk weight: 0.7; and
- Hardware: BLE; risk weight: 0.4.

In this case, the overall R1 = High, as the device has access to PD data directly and has direct connectivity to the network to transmit the data. Hence, the R1 for the Fitbit watch is higher than the R1 for the iKettle. However, the question is then raised of how to determine such initial weights. In this specific example, it is rather intuitive, but we need to build a scientifically sound model to support such a comparison.

6.3.3.2 Methodology on Risk Mapping Derivation

We have established the following methodology to identify the starting weight values for the IoT stack properties association with risks. First, we performed an extensive literature review to collect all relevant approaches and risks estimations in the domain of smarthomes and intelligent buildings. We have searched all indexed scientific repositories with the following search words: risk estimation, RA, smarthome, IoT, cyber physical systems, intelligent buildings, formal RA, and risk identification. This resulting list of papers was screened first for the relevance of the content. Then, in-depth analysis of most relevant works was performed to identify the list of the most appropriate risks appearing frequently in the scientific literature. The exhaustive list was aggregated with initial severity classification and potential cause attribution. In the second phase, we started from the attack classification landscape gathering, where the taxonomy of smarthome specific attacks was developed. It consisted of three main categories: Physical, Network and Software attacks. Each group was further classified into specific attacks. For each attack, we then derived a list of associated risks, based on the expert opinion survey. In the third phase, we cross-correlated risk lists from two previous phases and finalised it into a single list with manually assigned risk attribute weights. An extract of the data can be seen in Table 6.1.

Table 6.1 Risks and attacks association.

Risk Name	**Risk Shortname**	**Attack Association**
Physical Damage	PD	Physical Damage
Trigger Fake Events	TFE	Malicious Device Injection
Flood Network with Fake Events	FNFE	Mechanical Exhausting
Absence of Service	AS	DoS Participation
Sniff Traffic	ST	Device impersonation
Battery Exhausting	BE	Battery Attack
Unauthorised Control	UC	Malware
Leaking Data	LD	Sensitive Data
Gateway Abnormality	CA	Gateway Misbehaviour
Malicious Destination	MD	Malicious Destination

6.3.3.3 Expert Values for Receptor Weights

We have applied the same methodology as for the risk mapping (Section 6.3.3.2) to establish expert defined values for the initial Receptors' weights. Based on the literature review and expert opinion surveys, we have assigned averaged values for each identified risk Receptor. As described in Section 6.3.3, each risk can be triggered by the Receptor provided by the network Artefact. Therefore, we had to decompose all identified risks into associated Receptors and

attribute the values for linked Artefacts, as shown in the Table 6.2. We have marked in green the Receptors with different weights that can trigger the same 'Unauthorised Control' risk.

Table 6.2 Snippet of expert values for Receptors.

Artefact Name	Risk Shortname	Expert Value for Weight
SUSPICIOUS_TRAFFIC	UC	0.5
	ST	0.4
	LD	0.4
UNKNOWN_TRAFFIC	UC	0.2
	LD	0.2
	FNFE	0.2
SUCCESSFUL_DOS	UC	0.8
	AS	0.8
	FNFE	0.2
*BATTERY_ATTACK	BE	0.8
	ST	0.4
	AS	0.2
	PD	0.2
NEW_EXTERNAL_IP_ADDRESS	TFE	0.4
	ST	0.4
	UC	0.4
	LD	0.2

Furthermore, we have created a logical structure to calibrate the weights at run-time, depending on the anomaly report's type (Section 6.4.4.3) and the feedback collected directly from the end-users of the proposed solution (Section 6.4.5.3).

6.4 Implementation

Inspired by the Immune Theory, which arose from the biology domain [123, 124], we have conceptualised and implemented DRAF. This section illustrates in detail the internal architecture of the DRAF with the entire flow of the RA, as shown in Figure 6.5.

Data coming in from the several underlying modules is interpreted and allocated to the several distinct Analysers for further processing. These Analysers will each output a set of zero or more risk related events, so-called Artefacts.

These Artefacts are entities describing an element of the threat risk and an associated probability. This probability is the confidence level of the existence of that element of risk,

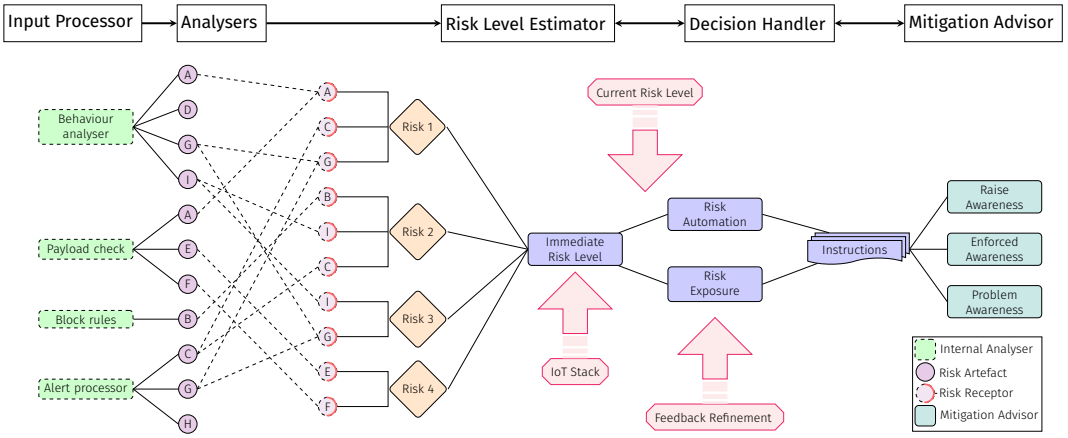


Figure 6.5 DRAF architecture with data flow and risk propagation.

which is thereafter reflected against a set of predefined risk scenarios for identification. A risk, as seen by DRAF, can be defined as a set of risk Receptors, which are the Artefacts with threshold and weight values. Each induced Artefact from the Analysers comes with a probability high enough to overcome the threshold for activation of risk and is then multiplied by the Artefacts internal weight attribution. The sum of all activated risk Receptors gives an overall risk likelihood. Using the matching and threshold comparison, a resulting set of risks with their likelihoods are then set as candidates for the calculation of the Immediate Risk Level (IRL) and the Current Risk Level (CRL). This produces a sparse matrix of the risks and their severity in conjunction with current in-place automation, possible (user) actions and outcomes (see [13] for more details on the theoretic model application for the definition of the risk levels). The Risk Automation and Exposure recommendation and related statistical data are parsed through producing a set of Instructions for the Mitigation Advisor.

6.4.1 Architecture and Workflow

The main goal of the DRAF is to provide real-time security and privacy RA of the ongoing activities on the network. It validates the current communication by assessing any anomaly detected and deviations in the IoT device's normal behaviour through device profiles. RA performs real-time RA by continuous evaluation and monitoring of various risk levels at multiple stages of data processing. To control the behaviour of devices and corresponding payload exchanges, permitted risk levels of ongoing network activity are dynamically calculated for each network activity event, practically determining the required decision to be taken. RA also establishes the risk controls for the users' privacy and makes them aware of the associated risks. DRAF integrates multi-faceted anomaly detection analysers and risk Receptors to support behaviour deviation detection, involving a deep understanding of risk propagation and inter-dependencies within the network. For this purpose, it leverages the existing open threat modelling tools (eg the Open cyber threat intelligence platform(https://www.opencti.io/en/ accessed on 31 March 2022)) to integrate a network entity correlation ontology.

6.4.2 Input Processor

The main controller component for the DRAF is depicted in Figure 6.6. It receives and manages incoming Anomaly reports that trigger the creation of the jobs to be executed. It is composed of a Launcher and Scheduler, and is constrained by the Reporting Strategies (RSs) for conform and standardised anomaly reporting inputs.

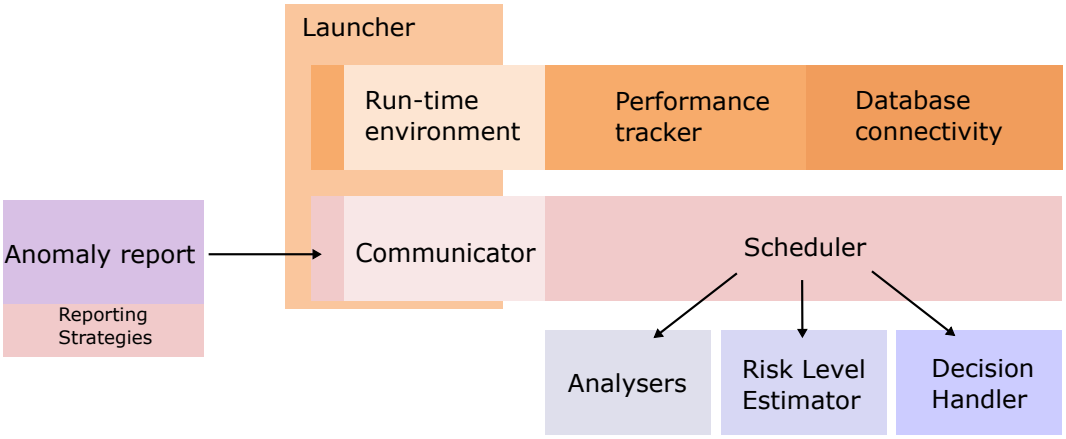


Figure 6.6 Input Processor.

6.4.2.1 Launcher

The Launcher is a wrapper component responsible for the run-time environment configuration of the DRAF's essential internal components, such as performance tracker, access to database repositories, risk estimator, communication handler, decision making, analysers launchers and the actual job scheduler. It serves as an initiator of the main execution process and exists in four variations to permit necessary integration into different deployment environments: a stand-alone command line Java virtual machine (JVM)-enabled environment, Jsvc (http://commons.apache.org/proper/commons-daemon/jsvc.html accessed on 31 March 2022) and OSGi (https://www.osgi.org/ accessed on 31 March 2022) wrappers for specific gateway environments and a dedicated trials launcher for controlled testing of the DRAF automation features. It also initiates a Communicator component, responsible for listening to all incoming messages to be further dispatched and normalised for the risk analysis, as well as to handle remote call execution for asynchronous communications.

6.4.2.2 Reporting Strategies

To process incoming network messages coming from an external anomaly reporting module, a structured formulation of its content is required. The goal is to get the right information processed into the correct outputting Artefacts with minimal overhead. The network messages describe a detected anomaly report through the means of several predefined attributes:

- Severity: the degree the impact of the detected anomaly can have on the device or home environment;
- Priority: the importance of having the anomaly resolved or investigated. This attribute relates mostly to the end-user point of view of how an anomaly may affect them (eg private data leakage vs. a non-functional smart lamp);
- Reliability: the confidence factor from the anomaly reporting module;
- Attack attribution: a reporting module can indicate a potential attack classification for a detected anomaly;
- Attack probability: the confidence of the reporting module on the attack's class attribution;
- Source/Destination: depending on the reporting module field of view within the network, these two fields contain the corresponding source and destination identifiers. These identifiers are provided in several granular levels, namely at the smarthome level, indicating there is something wrong in the network, at the interface of a device (eg gateway endpoint with device ID or interface channel) or generic subnetwork type (eg Zigbee, Bluetooth, Z-wave), or precisely to the device by providing its IP or Media Access Control address (MAC);
- Target recipient: similar to source/destination, looking at the communication channel can precisely narrow down the scope of the anomaly. Depending on the type of analysed communication protocol, this can be indicated as broadcast, unicast or multicast destination points; and
- Reasoning: a descriptive field in which the reporting module may provide additional information through the means of a fixed set of acceptable values.

While restricting and enforcing conformance to the input structure reduces the complexity, it does not entirely cover the vast differences that the reporting modules can exhibit. This means, for example, that a reporting module may report, in particular, a certain attack detection, while another reporting module operates at a higher level of granularity and reports on overall changes in the network communication behaviour. As a result, not all anomaly occurrences contain information on every attribute, nor is it required. Taking into account that any reporting module can differ and is residing somewhere in the network, eg on an IoT device, intermediate gateway or a home server, it is important to take into account their purpose. To illustrate this further we exemplify this with some questions:

- *What information needs to be transmitted?* A reporting module can encompass a broad range of monitoring functionalities that can describe a variety of anomalies or focus on very specific issues. Thus, the information that reporting modules can provide can vary strongly and distinctly from one another.

- *When and in what form?* Many factors influence when a report will be generated, due to the intended functions of the module, the observed parameters, its operating context (on edge device, on the gateway or even external as a service in the Internet). The DRAF has to handle asynchronously the incoming reports from multiple reporting modules, which report in an irregular (sparse) manner.

- *Why and how?* As functions differ, their intent for providing a report can be misleading. For example a module monitoring the absence of data reporting, eg a life beat packet from a smart smoke detector, may indicate that the battery has been depleted or the device is malfunctioning. On the other hand, a bed sensor may also report absence of data due to the person not being at home, while both stating the same report for absence of data.

Although most of these questions are covered by the attributes defined previously, we further systemise reporting modules into RSs by identifying their prime attributes. We defined three anomaly RSs that DRAF is fully integrated with. Each RS focuses on a subset of the variables as shown in Table 6.3.

- Aggregated Prioritisation (AG): As the name implies, the primary attribute is priority. The anomaly reports incorporate intelligence on the aggregation of anomalies output based on the priority of the individual elements and provide their final outcome as the conclusion of the aggregation process (eg threshold, time window, batch size) with a priority score.

- Behaviour Deviation (BD): Leaning towards the severity and reliability as primary attributes, there are anomaly reports that report device and non-device behaviour deviations and higher level (application layer) events caused by a smarthome habitants.

- Attack Attribution (AA): Represented by anomaly reports that provide attack attribution data, either in the report as a dynamically filled attack identifier or as a fixed identifier defined by the scope of a specific attack detection use case.

Table 6.3 Correlation matrix: inclusion of anomaly reports' attributes per RS.

Strategy	Priority	Severity	Reliability	Attribution	Probability	Src/Dst	Recipient	Reasoning
AG	✓					✓	✓	✓
BD		✓	✓		✓	✓		✓
AA				✓	✓	✓		

Based on the identified strategy for an incoming reporting message, the DRAF then applies weight adjustments that influence the internal risk analysers' assessment of the report (see Section 6.4.4.3).

6.4.2.3 Scheduler

The main DRAF workflow control unit, Scheduler, is in charge of incoming reports propagation from the the input reception and the risk analyser job dispatching to the result collector for the Decision Handler (DH) when the risk exceeds the configured acceptable risk threshold. It contains several functions/classes:

- Job creator: for each packet an encapsulated job is launched, enabling the control and monitoring of the analysis of the packet;
- Parameter application: an internal interface to get and set DRAF parameters (eg number of threads, sleep time settings, priorities);
- Selection of the checker for the performance control and monitoring. For example, Block Rules (BR) are always verified, but BA or Payload Check (PC) verification depends on the metrics that influence the choice of not running a checker; and
- Tracker: responsible for the main pipeline monitor. It creates estimations of how long jobs take to complete based on the historical effort log. For example, a huge number of profiles per device can influence the time it takes to process a packet from a specific device. Additionally, the type of the data might affect the effort for the PC.

6.4.3 Risk Analysers

Scheduler coordinates and runs a set of Analysers producing risk Artefacts, which in turn trigger risk Receptors. The correlation between internal components is depicted in Figure 6.7.

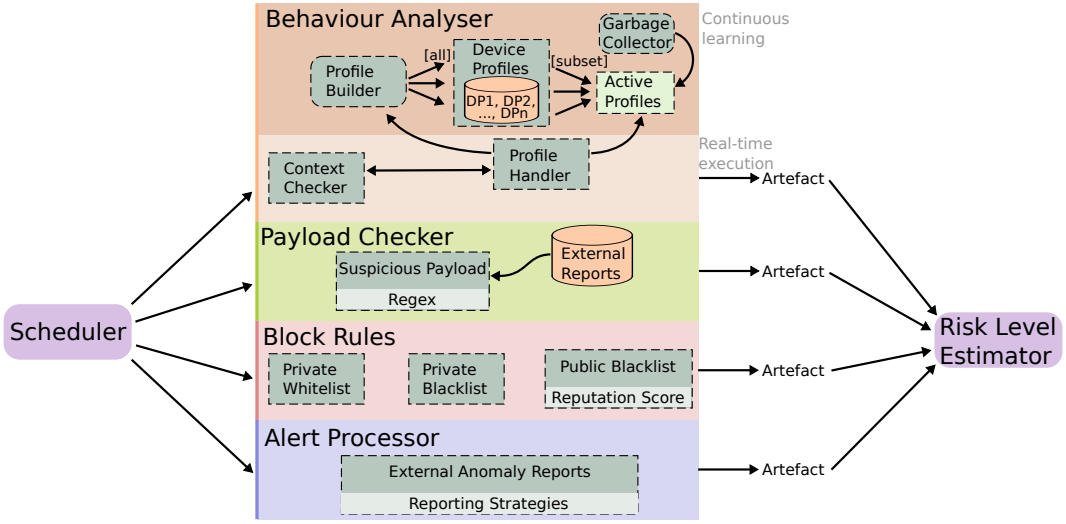


Figure 6.7 Risk analysers.

6.4.3.1 Behaviour Analyser

The aim of the BA is to see if the data flow from a device is within the nominal operations of non-faulty day-to-day usage data behaviour. The difficulty of operating and judging device behaviour roots in many unknowns of its operating environment.

Other than operating within a smarthome, nothing much is known, and its behaviour is established based on historic data flows, for which the PB is responsible. For each device, the PB creates a graph of connected data flows, also known as the Device Profile, linking the flows together chronologically. as nodes. Each node then has the observed attributes of the data flows (accumulating, aggregating and merging of similar data flows). Each Device Profile is specific to the device and associated user's interaction with it.

The Profile Handler (PH) component is responsible for the management and communication with the PB for retrieving and restructuring the Device Profiles to be used by the Analyser.

The main logic of the BA, which is responsible for comparing the current packet with any existing profile, is encapsulated in the functionality of the Contextual Checker (CC). Whenever the BA receives a processing job from the Scheduler, it will request from the PB the device profile linked to the job. The profile is then internally loaded through the PH in a new data structure enhanced for several operations. From the job, the properties and metrics are translated into a comparison query that can be given to the PH. Consecutively, PH will compare the query parameters with the nodes in the graph and output a set of pointers to nodes that are closest to the given query data. The CC then analyses these nodes on several aspects, with examples including:

- A comparison to the historical observations between the last identified pointers and accepted pointers;
- A comparison to the forecasted behaviour for the predicted pointers;
- Query matching thresholds with weights per identified property (eg 5 out of 10 properties matched);
- Per matched property evaluation of the threshold margins of the comparison per node (eg flow size property matched 80%); and
- Analysis of the system and user set configuration and preferences, forming exception or inclusion rules for properties, eg 'ignore flow size'.

Finally, CC then determines if the behaviour is in line with its expectations of that specific device and marks the identified pointers as accepted pointers. The output will be an Artefact of misbehaviour with a given weight, as previously described in Section 6.3.3.

The Active Profile (AcP) is a profile that was matched in the recent past with incoming packets to the DRAF and is currently being tracked. It is a cursor in the profile's tree, indicating at what stage the communication flow is, as well as the depths of the built tree.

The Garbage Collector (GC) is a separate process that monitors AcPs and checks on their activity and progression heuristics (eg time constraints on the next node in the tree).

We assume that several communications are going on at the same time. Additionally, there can be several matching profiles for a packet/flow, which should decrease over time, either by new packets and their matching probability or checking in the tree for the next node timeouts by GC. Therefore, we have built a structure to support the buffer of all active profiles.

A separate process for monitoring the absence of any communication was also envisioned as part of the device presence behaviour. It keeps track of profiles' activity by counting how many times the profile was executed partially and fully.

Finally, the BA has a secondary behaviour verification algorithm running independently from its main function. In this case, it periodically requests a device profile from the PB and deduces in which time period it is expected of the device to communicate on the network. This algorithm further verifies if this has happened, and if not, it emits an Artefact of absence.

6.4.3.2 Payload Check

The PC uses the gathered information to inspect the data of the flow for any suspicious data patterns, the scope of which is predefined by user configuration. In addition, when the data stream is encrypted, it tries to verify the certificate data related to it. The internal logic of this component includes the following features:

- RegEx matching: a set of regular expressions to detect private data;
- Secure Socket Layer (SSL) check: verify the SSL certificate and see if proper SSL packet is observed; and
- Suspicious payload confidence level.

6.4.3.3 Block Rules

The BR are re-verified and extended to a broader view on the target destination to see if there could be any reason if a certain communication should be blocked. This includes the current rules in place from the IPtables, which are the *raw* rules as used by the underlying Operating System (OS) provided through the interoperable middleware or complemented by user-added information from the user configuration or public blacklisting. It verifies the authenticity of the destination (eg *who is* scraping from several sources). The gathered information is used to produce threat/risk levels for three categories: direct destination (eg IP), domain name (eg the Domain Name System (DNS) name) and connectivity (neighbouring IPs) in relation to their reputations (eg previously reported hosting malware and/or phishing sources and gathered by the central intelligence repository).

The current implementation of the BR operates on various interface types (IP, Bluetooth, Zigbee, Z-Wave, RF869, and PPP). However, in the case of IP traffic, we check only the

destination, while in all other cases, both the destination and source are verified. BR is supported by the following use cases:

- Public Blacklist: based on the data retrieved from the official external threat intelligence by means of the scoring system;
- Private Whitelist: as the name indicates, it enables the public blacklist to be bypassed by adding the destination point to the whitelist; and
- Private Blacklist: a personalised blacklist of selected addresses.

BR performs external checks only once a dedicated alert is sent and executes the following procedure:

- Prioritisation: check the local whitelist, private blacklist and then copy of the public blacklist;
- Data refresh: if not blacklisted, then check with the external repository. If there is a reply time out, assume that IP is not present in the central intelligence repository (as the local copy of the public blacklist was synchronised earlier); and
- Score comparison: if the address is not blacklisted, check the score through the external resilience infrastructure. If the reply times out, assume that IP is safe. Otherwise, perform the reputation scoring routine [24].

6.4.3.4 Alert Processor

The Alert Processor (AP)'s primary function is to act on messages coming from the externally plugged reporting modules. These components provide a risk/anomaly analysis themselves and 'merely' inform the DRAF. It is up to the AP to handle the provided information in an appropriate way, meaning that the information given may not always be conclusive and the AP will try to consult with the other analyser outputs and historic data to relate the presented risks with ongoing events. Furthermore, it will apply aggregation and merging strategies to the incoming messages, as the underlying components may continuously emit these messages, and it would not be appropriate to overwhelm either the DH or the Mitigation Advisor (MA), and thus, ultimately the user, with it.

6.4.4 Risk Level Estimator

The overall workflow of the Risk Level Estimator (RLE) is illustrated in Figure 6.8. It is the most crucial component of the DRAF responsible for the risk estimation of an ongoing event and its impact on the overall risk level.

The risk model is composed of a collection of identified risks atomically split in the risk Receptors (as outlined in Section 6.3.2) and serves as a main controlling input for the RLE. Risk Receptors identify the set of currently applicable risks by incorporating the threat level, consequential exposure and the automation strategy.

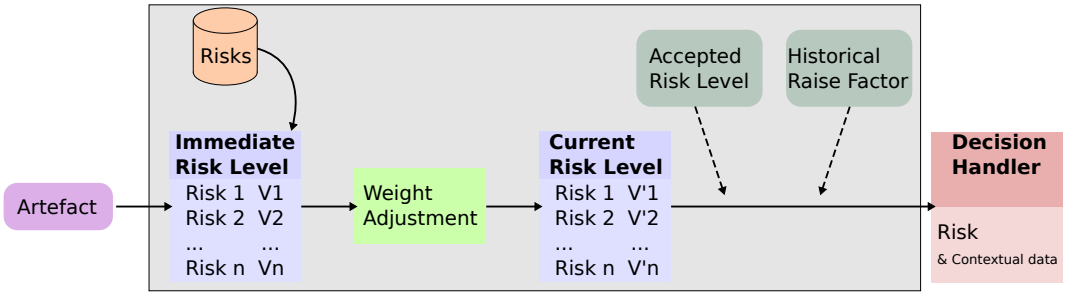


Figure 6.8 Risk Level Estimator.

6.4.4.1 Immediate Risk Level

One or more produced Artefacts with an established Receptor's weights, initialised with expert weight values, can trigger one or more risks. For the risk to become active, an overall cumulative weight of all Receptors should overpass the established Receptor threshold. This threshold is controlled by the end-user and depends on his personal risk perception in terms of the compromise between acceptable risk tolerance and smarthome functionality. The currently processed Artefact produces a value called IRL. It initialises the corrective course of actions towards DH, described in detail in Section 6.4.5.

6.4.4.2 Current Risk Level

CRL is responsible for tracking all recently active risks by maintaining an in-memory table on the latest risk estimation scores. Once a risk value from IRL is processed by the weight adjustment, a comparison can be performed by CRL to identify whether any of the active risks exceed the risk threshold. Risk estimation is calculated by summing up CRL with the risk brought by the observed Artefact. This approach enables the sequential processing of the multiple Artefacts, as any of them can contribute to the detection of a risk. When Receptors weights are below the threshold, the average value is used for summing up risk estimation. When the latest Receptor observed is above the threshold, its value is used to report the detected risk accident.

The risk threshold value is controlled by two elements: the Accepted Risk Level and the Historical Raise Factors. The Accepted Risk Level is a structure to support the user definition of Accepted Risk Levels per device and is used for the overall smarthome installation. It has a dynamic property, as it evaluates over the time the initial setting by means of the feedback pipeline from the user interface. For example, if the user prefers to perform corrective action on the automated decision, then the Accepted Risk Level will be modified accordingly. This is implemented through the Historical Raise Factors feature, which reviews the risk history to reveal when and what risk events occurred, what was handled automatically, and which event requested a decision from the user.

6.4.4.3 Weight Adjustments

As previously mentioned, to calibrate expert weights assigned to the Receptors, we have implemented a weight adjustments strategy for each type of the RS. More specifically, we have integrated 180 different types of external anomaly reporting inputs, which can be categorised into three major RSs, as was described in Section 6.4.2. Below we provide examples of the weight adjustment per RS type through an example of a specific anomaly report integrated in the DRAF.

AG Weight Adjustment This report example of type AG provides aggregated reports on IP-related anomalies of the format presented in Listing 6.1, including its own classification on the alert priority.

The maximum number of the occurrences provided in the list of `FlowAlert` can change from one deployment environment to another. Therefore the controlling variable for giving indicative value on the maximum number of accumulated `alert_counts` is externalised into the configuration settings of DRAF.

```
message AGReport {
    repeated AttackAlert attack_alert;
}
message AttackAlert {
    required AttackClass alert_class;
    required string alert_description;
    required Priority alert_priority;
    repeated FlowAlert flow_alert;
}
message FlowAlert {
    required string src_ip;
    required string dst_ip;
    required uint32 src_port;
    required uint32 dst_port;
    required uint32 alert_count;
}
```

Listing 6.1 AA alert format extract.

For each attributed Artefact, the Receptor's weight for the associated risks is adjusted in the following way:

1. Prioritisation: `alert_priority` is given higher impact, and the sum of `alert_count` is given lower impact;

2. Normalisation: a higher priority with a higher count should be more crucial to report, yet a higher priority with a low count is more valuable than low priority with a high count;

3. Score adaptation:

$$(priority + 1) * significanceFactor - occurrenceCount$$

   where `significanceFactor` is loaded from application properties and corresponds to the forecasted maximum number of occurrences (as described above);

4. The final score adaptation strategy is calculated based on two variables:

   - stepSize: $adaptationScore/significanceFactor$;
   - maxNumberOfSteps: number of possible priorities scores, equal to 4 for the AG type of report;

5. During the risk evaluation phase, the Artefact's weight is adjusted according to this score, taking into consideration the observation of other ongoing anomalies.

For example, for an Artefact with the Receptor weight 0.8, an alert of HIGH priority (highest value) and 10 occurrences will give a final weight of 0.8. For an Artefact with a Receptor weight of 0.8, an alert with VERY_LOW priority and 1000 occurrences will give a final weight of 0.6. The adjusted weight is always less or equal to the predefined weight value.

BD Weight Adjustment This module provides reports on the devices' behaviour deviation anomalies with corresponding severity and reliability scores, summarised in Listing 6.2.

```
1 message AnomalyDetection {
2     required uint32 severity_score;
3     required uint32 reliability_score;
4     required DeviceInfo device;
5     required string timestamp;
6     required string reason;
7 }
```

Listing 6.2 BD alert format extract.

The `severity_score`'s value range is $[0, 10]$, where 0 is a safe score and 1 to 10 is a threat, where 10 is the most severe score. The score 0 is used for the 'safe' reports, while values in the range $[1, 10]$ are used for 'threat' reporting. The `reliability_score`'s value range is $[0, 10]$, where 0 is the lowest confidence and 10 is the highest confidence of the algorithm for its response. Furthermore, the severity-reliability relationship for 'threat' is linearly dependent (eg 1 severity score corresponds to 1 reliability score), and severity-reliability for 'safe' depends on the stability of the cluster.

The controlling variables for ensuring configurable and dynamic adaptations for anomaly reporting are externalised into the configuration settings of DRAF, setting the margins for the threat detection sparsity.

For each attributed Artefact, the Receptor's weight for the associated risks are adjusted in the following way:

1. Prioritisation: `severity_score` is given higher impact, while `reliability_score` is given lower impact;
2. Normalisation: higher severity with higher reliability should be more urgent to report, yet higher reliability with lower reliability is more valuable than low severity with high reliability;
3. Score adaptation:

$$severity * significanceFactor + reliability$$

   where `significanceFactor` is a multiplication product of the value ranges of `severity` and `reliability`, loaded from application properties;
4. The final score adaptation strategy is calculated based on two variables:
   - stepSize: $adaptationScore/significanceFactor$;
   - maxNumberOfSteps: number of possible priority scores, equal to 10 for the BD type of report;
5. During the risk evaluation phase, an Artefact's weight is adjusted according to this score, taking into consideration the observation of other ongoing anomalies.

For example, for an Artefact with a Receptor weight of 0.7, an alert with a severity score of 7 (highest value) and a reliability score of 7 will give a final weight of 0.6. For an Artefact with a Receptor weight of 0.8, a BD alert with a severity score of 1 and a reliability score of 1 will give a final weight of 0.4. The adjusted weight is always less than or equal to the predefined weight value.

AA Weight Adjustment This module provides reports on detected attack classifications for devices based on various metrics, where attribution probabilities are incorporated for the most likely detection attribution, as summarised in Listing 6.3.

```
message CybersecurityStatus {
    required InterfaceType if_type;
    required InterfaceId if_id;
    required int32  id_slot;
    required float attack_proba;
    required double start_time;
    required double end_time;
    repeated AttackClassification attack_classification;
}
message AttackClassification {
    required AttackAttribution attack_class;
    required float probability;
    optional DeviceId device_id;
```

```
14 }
```

Listing 6.3 AA alert format extract.

The `attack_proba`'s value range is $[0, 1]$, where 1 corresponds to the highest likelihood. Furthermore, each `AttackClassification` provides an additional `probability` for attack classification to a specific attack class. The controlling variables for ensuring configurable and dynamic adaptations for anomaly reporting are externalised into the configuration settings of DRAF, setting the margins for the threat detection sparsity and the threshold for integrating the report result into the overall RA.

For each AA attributed Artefact, the Receptor's weight for the associated risks are adjusted in the following way:

1. Prioritisation: overall `attack_proba` is given higher impact, than individual `probabilities` for each classified attack;
2. Normalisation: the bigger the difference between two variables, the less impact will be propagated to the RA;
3. Score adaptation:

$$attackProbability * significanceFactor + diffFactor$$

   where `significanceFactor` is a scaling product of `probabilityScale`, loaded from application properties, and `diffFactor` is the normalised difference of the overall attack probability and probability of attack class attribution;
4. The final score adaptation strategy is calculated based on two variables:
   - stepSize: $(1 - threshold) * scale$;
   - maxNumberOfSteps: number of possible priority scores, considering the threshold;
5. During the risk evaluation phase, the Artefact's weight is adjusted according to this score, taking into consideration the observation of other ongoing anomalies.

6.4.5 Decision Handler

Depending on the user preference settings and resulting risk identification, DH either propagates a mitigation strategy to the end-user or performs automated actions with the possibility for the user to retract decisions made through a feedback loop.

6.4.5.1 Automated Decision Making

Utilising resulting data on exposure and automation together with user preferences, a course of Mitigation Advisory is extracted. Furthermore, a tracking mechanism is initiated for the follow-up of the scenario actively presented to the end-user.

The developed algorithm, demonstrated in Algorithm 1, outlines its main logic with the following abbreviations utilised:

- SCI corresponds to Special Case Intervention, where no automation is possible at all; only Mitigation Advisory is provided;
- AA corresponds to Automated Action, where a decision was made in accordance with user's desirable risk level settings;
- SI corresponds to Security Intervention, where a decision should be made by the user with a recommendation.

As can be observed from the above algorithm, a mandatory user interaction is required only in the case of SCI. In the other two instances, an automated decision will be executed, which can be verified by the end-user. Guided by the validation methodology established at the architecture level of the overall GHOST solution [22], a set of automatable technical actions was established, an extract of which is depicted in Table 6.4. The purpose of this example is to show the variety of possible technical actions that we could automate and to demonstrate the variation of the composed variables for the action selection. For example, different Receptors can cause the same risk, leading to different actions (ID1 and ID2). The same action can be executed in case of different risks triggered by different Receptors (ID2 and ID3). The same risk can lead to the same action, even if caused by different Receptors (ID3 and ID4). Additionally, the same Receptors causing different risks will lead to different actions (ID5 and ID8). This is due to the underlying risk model, which enables the diverse mapping of the Receptors, risks and associated attacks.

Table 6.4 Automatable action mapping.

ID	Technical Action	Triggered Risk	Final Receptor
1	Verify physical integrity	Physical Damage	BEHAVIOUR DEVIATION
2	Verify battery	Physical Damage	FREQUENCY ANOMALY
3	Verify battery	Battery Exhausting	BATTERY ATTACK
4	Verify battery	Battery Exhausting	BATTERY SILENT
5	Block device temporarily	Flood Network with Fake Events	NEW EXTERNAL IP ADDRESS
6	Block device permanently	Unauthorised Control	UNREGISTERED DEVICE
7	Drop packets for flow temporarily	Sniff Traffic	NETWORK SCAN
8	Drop packets for flow permanently	Sniff Traffic	NEW EXTERNAL IP ADDRESS
9	Drop packets for source temporarily	Leaking Data	STRING DETECT
10	Drop packets for source permanently	Leaking data	TROJAN ACTIVITY

Algorithm 1: Decision Handler.

```
Result: send(notification)
loading user configuration;
if isMitigation then
    notification = SCI;
else
    getDecisionsCommands;
    if isSafeRiskLevel then
        executeCommand;
        notification = AA;
    else
        if isNotificationDisabled then
            executeCommand;
            notification = AA;
        else
            notification = SI;
        end
    end
end
```

6.4.5.2 Rendering Mediator

This component is responsible for the translation of the risk parameters into an end-user-friendly explanation and acts as mediator between the Automated Decision Making and Web Interpreter. More specifically, it provides the reasoning on the automated action performed by DRAF or the Mitigation Advisory provided to the end-user. It is composed of:

- The attribution level, eg device, interface, gateway;
- The last triggered Receptor, which enabled acceptable risk threshold over-passing;
- The associated risk, controlled by the attack vector scope;
- The identifier for the translation key; and
- The mitigation advisory.

Furthermore,to follow-up on the pending user decisions or to correlate user decisions with reoccurring risks, a tracking mechanism is implemented.

The sample output is provided in Table 6.5, outlining the partial selection of the resulting variables. It should be emphasised that the ability of the DRAF to perform automated decisions is tightly coupled with the end-user preferences linked to their risk perception. For instance, for row ID1, we can see that the Mitigation Advisory value is set to n/a, while in row ID2, Mitigation Advisory is set to Block. This happens when the end-user is selecting a high tolerance risk acceptability level, meaning allowing maximum automation of DRAF, and does not receive notifications for each automated action (in case of ID1). However, when

the end-user settings indicate that all automated actions should be reviewed, we provide the Mitigation Advisory. The final text displayed to the end-user corresponds to "*Private data has been detected, coming from a blood pressure measurement device*". A similar differentiation can be observed in the rows ID3 and ID4. In both cases, the risk of communication with a malicious destination was addressed by continuous blocking of the traffic on a specific interface with the blacklisted address. However, in the row ID3, the end-user's settings permitted the maximum automation level, hence, no Mitigation Advisory was provided. The final end-user text for the row ID4 corresponds to "*Communication to a known malicious destination, botnet.com, was detected again on your network. Please contact the manufacturer to replace the malfunctioning device*". Finally, the row ID6 demonstrates the output of the DRAF when only mitigation by the end-user is possible due to technical limitations of the automation aspects. The final end-user text corresponded to the following: "*No measurements were detected for the sleeping sensors. Please indicate how you resolved the issue: (i) I removed the device; (ii) I checked the device and behaviour is normal; and (iii) Contact Manufacturer.* ".

Table 6.5 Output sample.

ID	Type	Automated Decision	Alternatives	User Action	Mitigation Advisory	Attribution Level	Triggered Risk
1	AA	Block	Keep blocking, Allow	n/a	n/a	Device	Leaking Data
2	SI	Block	Keep blocking, Allow	n/a	Block	Device	Leaking Data
3	AA	Keep blocking	Keep blocking, Allow	n/a	n/a	Interface	Malicious Destination
4	SI	Keep blocking	Keep blocking, Allow	n/a	Manufacturer	Interface	Malicious Destination
5	SI	Block	Keep allowing, Block	Allow	Block	Device	Leaking Data
6	SCI	n/a	Removed, Checked, Manufacturer	Removed	n/a	Device	Absence of Service

6.4.5.3 Feedback Refinement

To include the direct feedback of the end-users in the automated decision-making process, a set of user interfaces with different input options was created, as described in Section 6.4.5.1. Together with the Historical Raise Factors, this approach served to allow the fine-tuning and rectification (if desired) of the automated decisions to be overruled by the end-users. An example of such an action is demonstrated in Table 6.5, row ID5, where the output of the

automated decision-making was refined directly by the end-user from continuing to block the traffic to allow it instead.

6.5 Results

We have executed several experiments to validate the proposed solution. Table 6.6 shows the correlation between each experiment, RQ relevance and utilised method for the validation of achieved results.

Table 6.6 Experimental validation mapping.

Objective	**Relevance**	**Setup**	**Method**
Performance overhead	RQ 3 (Section 6.5.1.1)	Real-life trials	Run-rime monitoring
Workflow validation	RQ 1 (Section 6.5.1.3)	Testbed	Real attack execution
Expert values	RQ 2 (Section 6.5.1.4)	Testbed	Replay of real attack execution
Risk coverage	RQ 1 (Section 6.5.1.5)	Real-life trials	Statistical analysis

6.5.1 Experimental Validation

The DRAF was deployed in more than 80 smarthomes as part of real-life trials in three European countries (Spain, Norway and Romania) in the period from June 2019 to April 2020 under the umbrella of the GHOST research project https://cordis.europa.eu/project/id/740923 (accessed on 31 March 2022). Furthermore, explicit validation and the calibration of the DRAF workflow was implemented with the help of the *GHOST-IoT-dataset* [125] in a testbed environment, fully replicating the smarthome setup. This dataset was collected in anticipation of the ethical constraints for attack simulation in real-life deployments faced in the GHOST project. For this purpose, on the voluntary basis, a full smarthome setup was deployed in one of the project participant's apartments to capture two types of smarthome network traffic: normal behaviour and attack simulation. Availability of this data is a significant contribution to the functional validation of the IoT enabled environment under the execution of a cyber threat.

The purpose of all executed experiments is threefold:

- Measure the performance overhead of DRAF on the gateway;
- Validate the workflow capability to detect ongoing risks and apply an appropriate mitigation strategy in a real smarthome environment; and
- Validate the expert values' correctness and their independence from the user profile settings, such as acceptable risk levels, automation optimisation and IoT devices' profiles.

6.5.1.1 Deployment Setup

A typical installation of the smarthome was composed of the following devices:

- Smarthome gateway (eg Raspberry Pi 3 single-board computer, CareLife smart IoT gateway (https://cordis.europa.eu/project/id/740923 accessed on 31 March 2022));
- Zigbee sensors (eg presence detector, door aperture detector);
- Bluetooth enabled medical devices (weight scale, blood pressure meter);
- Z-wave sensors (eg motion, door and window opening, smoke and flood sensors); and
- Z-wave devices (eg smart plug, smart dimmer).

All IoT devices deployed in the smarthome environment communicated through the gateway, on which DRAF was deployed.

The real-life trials were performed in three stages, where different combinations of the RSs were used. The outline is provided in Table 6.7. Such consequential inclusion methodology allowed close monitoring of the performance overhead measurement.

Table 6.7 Inclusion of RSs in the real-life trials.

Reporting Strategies	Trial I	Trial II	Trial III
AG	✓	✓	✓
AA		✓	✓
BD			✓

Statistically, we have observed a 3.2% Central Processing Unit (CPU) overhead and 61.22 MB of memory consumption when deploying DRAF on the gateways. More details are depicted in Table 6.8. Considering that a smarthome gateway has, on average, at least 1.2 GHz CPU and 1 GB memory, our solution creates minimal overhead, completely acceptable for real-life deployment environments.

Table 6.8 DRAF overhead statistics.

Parameter	Minimum	Average	Maximum
CPU	3.2%	3.49%	3.9%
Memory	56.96 MB	61.22 MB	63.32 MB

6.5.1.2 Ethical Constraints

In order to validate the correctness of the DRAF workflow and perform the necessary calibration of the expert values, one has to test the solution with cyber attacks. Due to the ethical approach applied throughout the project, it was agreed with the National Data Protection

Authorities involved in the real-life trials that any form of real attack execution in the smarthome environments was not possible, and only simulated approaches in a controlled environment were allowed (https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ceacdf7d&appId=PPGMS accessed on 31 March 2022). Furthermore, dedicated debriefing of the participants after simulations was required, limiting the effectiveness of technological and functional testing of the framework. Therefore, any validation involving the execution of attacks was constrained to be performed only in the testbed environments.

6.5.1.3 Workflow Validation in the Testbed Environments

In alignment with the methodology on risk modelling, as described in Section 6.3.3.2, the technical functioning of the proposed framework was validated in the testbeds. The extract of the tested DRAF components is presented in Table 6.9. Here, the focus was given to the validation of the workflow correctness by triggering a specific incoming anomaly report. Simulation of the attack in isolation primarily ensured the correctness of the invoked analyser capable of identifying applicable risks and application of the appropriate automated Decision (D) and Mitigation Action (MA). Secondly, this setting also permitted the validation of the expert weights' initialisation values, as different testbeds equipped with different IoT devices were utilised. As can be observed, only in cases when dealing with BD RSs the automation of mitigation actions was not possible. This is explained due to the technical constraint on the nature of the associated risks, where additional feedback from the end-users was required.

Table 6.9 Automation status of DRAF for decision making and mitigation.

Analyser	Attack	Risk	Automated D/MA
BA	Physical Damage	Physical Damage	✓/ ✓
PC	Sensitive Data	Leaking Data	✓/ ✓
BR	Malicious Device Injection	Trigger Fake Events	✓/ ✓
AP (AA)	Sensitive Data	Leaking Data	✓/ ✓
AP (AA)	Battery Attack	Battery Exhausting	✓/ ✓
AP (AA)	Malware	Unauthorised Control	✓/ ✓
AP (BD)	Gateway Misbehaviour	Gateway abnormality	✓/ ✗
AP (BD)	DoS Participation	Absence of Service	✓/ ✗
AP (AG)	Malware	Unauthorised Control	✓/ ✓
AP (AG)	DoS Participation	Absence of Service	✓/ ✓
AP (AG)	Sensitive Data	Leaking Data	✓/ ✓

6.5.1.4 Alert Fusion and Receptors Verification

The final and most challenging validation experiment was designed as a simultaneous reporting of various attack scenarios to verify that the DRAF assigned risk Receptor weights used at the initialisation are adequate and the overall behaviour of risk detection is correct. For this purpose, we used the *GHOST-IoT-dataset*, which was replayed in the testbed's gateway with several anomaly reporting components running together with DRAF. This setup simulated, as close as possible, an authentic smarthome but in a controlled manner and with no real users involved to avoid ethical constraints. The testbed further permitted simultaneous triggering of the AA and BD Reporting Strategies and invocation of the AP, BA and BR analysers.

We replayed seven attack scenarios:

- Attack 1: The battery of the living room sensor was removed;
- Attack 2: A total of 10 consecutive wrong blood pressure measurements were made;
- Attack 3: The emergency button was activated 20 times in a row;
- Attack 4: The door opening sensor of the entrance was uninstalled, and several detection triggers were forced;
- Attack 5: The battery of the door opening sensor of the entrance was removed;
- Attack 6: The bedroom sensor was moved to a nearer position; and
- Attack 7: A connection from an unknown device to the WiFi network.

The DRAF outcomes are outlined in Table 6.10 for each attack (Nr). As one can note, the verification of the DRAF behaviour was not possible in all cases due to the missing underlying anomaly reporting for Attacks 2, 4 and 6. In all other cases, the risks were identified in an acceptable time-frame, ranging from 56 to 295 ms. In the cases of Attacks 1 and 5, two Artefacts were observed, both triggering the Receptors for the same risk of Physical Damage, which finally was flagged as 'over-passed the threshold'. For Attack 3, three distinct Artefacts were observed, where weight adjustments were applied due to the underlying RSs. More specifically, the risk of Physical Damage was lowered down for the BEHAVIOUR_DEVIATION Artefact, coming from AP:BD RSs. However, the same risk was enhanced by the BEHAVIOUR_ANOMALY Artefact processed by BA and UNKNOWN_TRAFFIC processed by AP, resulting in the final risk of Physical Damage being detected in the shortest time-frame observed during the experiment. Finally, Attack 7 caused the creation of two Artefacts, both triggering the Sniff Traffic risk. However, despite the fact that initial risk weights did not surpass the set threshold, the prioritisation of the BR analysers permitted propagation of the risk estimation to finally be triggered with a minor delay of less than 0.3 s.

In conclusion, this experiment also showcased that Receptors' weight initialisation based on expert's opinion and their further run-time adjustment presents an accurate approach, as the

risk detection was executed in a required time. Furthermore, the applied weight adjustment enables faster detection time, as shown in the case of Attack 3.

Table 6.10 Detection of attacks, receptors, and risks.

Nr	Analyser/Observed Artefacts	Expert Weight (Adjusted)	Triggered Risk	Time	D/MA
1	AP: BEHAVIOUR_DEVIATION	AC 0.3 (0.26) PD 0.8 (0.68)	Physical Damage	157 ms	✓/✓
	BA: BEHAVIOUR_ANOMALY	UC 0.8 PD 0.2 TFE 0.2			
2	No reports	n/a	n/a	n/a	n/a
3	AP: BEHAVIOUR_DEVIATION	AC 0.3 (0.24) PD 0.8 (0.64)	Physical Damage	56 ms	✓/✓
	BA: BEHAVIOUR_ANOMALY	UC 0.8 PD 0.2 TFE 0.2			
	AP: UNKNOWN_TRAFFIC	UC 0.2 (0.12) PD 0.2 (0.12) FNFE 0.2 (0.12)			
4	No reports	n/a	n/a	n/a	n/a
5	AP: BEHAVIOUR_DEVIATION	AC 0.3 (0.26) PD 0.8 (0.72)	Physical Damage	123 ms	✓/✓
	BA: BEHAVIOUR_ANOMALY	UC 0.8 PD 0.2 TFE 0.2			
6	No reports	n/a	n/a	n/a	n/a
7	BR: NEW_EXTERNAL_IP_ADDRESS	TFE 0.4 ST 0.4 UC 0.4 PD 0.2	Sniff Traffic	295 ms	✓/✓
	AP: TCP_CONNECTION	ST 0.2 (0.18) LD 0.1			

6.5.1.5 Risk Coverage Analysis

Finally, we performed an analysis on the coverage of the risks in relation to the attack association and reported Artefacts. Table 6.11 shows an extract of the generated distribution of the Artefacts utilised in the validation setup, grouped by the RSs. For each risk shown in the table, we can see the overall number of Artefacts potentially capable of triggering the risks and their distribution per RS. Furthermore, for each risk shown in the table, a specific RS appears to be more reliable for the risk-triggering mechanisms, eg for the risk of Physical Damage, the BD RS provided 29 potential Artefacts, while the AA and AG RSs provided only 2 potential Artefacts. This can be explained by the technological nature of the risk in question. The risk of physical damage implies an actual change in the behaviour of the IoT object, and therefore, reports containing data on the behavioural aspects will be of greater relevance. Some of the risks exhibit a very low coverage by RSs, such as Malicious Destination and Behaviour Deviation. This can be explained by the actual setup of the experiment executions. DRAF heavily relies on the external anomaly reporting inputs. These two reports were integrated in the deployed testbed environment, providing additional inputs for the DRAF for very specific cases, which were requested in the final iteration of the projects' development cycle to support the integration of externalised threat intelligence.

Table 6.11 Risks and Artefacts per RS distribution.

Risk Name	**Artefacts Total**	**Reporting Strategies**		
		AA	**AG**	**BD**
Physical Damage	33	2	2	29
Trigger Fake Events	13	2	10	1
Flood Network with Fake Events	5	1	4	0
Absence of Service	123	21	3	99
Sniff Traffic	10	3	6	1
Battery Exhausting	3	2	0	1
Unauthorised Control	38	3	33	2
Leaking Data	25	3	22	0
Gateway abnormality	104	0	0	104
Malicious Destination	1	1	0	0
Behaviour Deviation	1	1	0	0
Overall	356	39	80	237

6.6 Discussion

Our research goal was threefold: study the identified research questions, set the foundation for the generalised dynamic RA framework and provide validation on the implementation of our solution.

6.6.1 RQ1: Generic Ontology

RQ 1 was the prime motivation of our study. The main goal was to develop a generalised ontology to encapsulate fine-granular descriptors of vulnerabilities, threat vectors, risk mitigation strategies, impact evaluation, and cascading effects. This was successfully demonstrated through the development of the IoT Stack concept, which was included in the core of the DRAF (see Section 6.3.1.1). This enabled efficient bi-directional linking from the IoT device to the potential attack attribution and associated risks. Furthermore, such modelling permits the generalisation of the concept being applied to any IoT domain, ranging from smarthomes and connected vehicles to industrial smart factories and smart city infrastructures. The same concept comprises the notion of the granular slicing of the system into atomic components, identified as IoT objects. This approach grants the possibility of the identification of the cascading effects due to the inter-dependencies of object properties and their communications and coexistence in the same environment.

6.6.2 RQ2: Risk Calculation and Context Dependency

Quantification is the most crucial aspect in any RA system. Our methodology, applied throughout various risk modelling steps (see Section 6.3.2), eliminated most common problems for quantification of the risk weight values, namely high dependency on the available historical and current data. By utilising the initial expert values for risk weight allocation, we could advance the development of the framework from a prototyping environment into the real-life environment. The dynamic adaptation of the risk weights for the Receptors, based on the individual and collective decision making feedback, permitted the fine-tuning and validation of the DRAF in the real-life deployments (see Section 6.5.1). Furthermore, it was observed that the expert values are context independent, as no calibration was required for the utilised values in different smarthomes, where the actual smarthome setup was different from one place to another.

6.6.3 RQ3: Limitations on Dynamic Risk Assessment

The automated decision making in the context of the RA is bound to the limitation of the available data granularity of the underlying anomaly reporting, addressing entirely RQ 3. DRAF was proven to be efficient in the automation of the decision making when quality reports were provided (see Section 6.5.1.4). For example, in cases when external reports being fed into the system did not have granular data on the affected device, no automation

of mitigation actions was achieved. This was due to the technical limitation of attributing a specific device causing the anomaly observed throughout various RSs. Another important factor is the risk perception of the actual end-users. The less the users cared for cybersecurity and privacy risks, the more they were willing to compromise the decision-making efficiency towards smarthome functionality features.

6.6.4 Challenges and Limitations

Any theoretical solution can only be validated with technical implementation and the data quality. The original plan of this study was based on the availability of the IoT data to be generated and collected directly from the real smarthome installations. Unfortunately, this approach faced several obstacles, most importantly the ethical and privacy related restrictions when monitoring and analysing the network data. As a result, we had to readjust the scope of our experiments and rely mostly on testbed environments to mitigate the imposed delays. We have faced the same source of another challenge regarding the validation of expert values, originally planned to be performed during real-life trials by simulating attacks. However, due to the ethical constraints (see Section 6.5.1.2), these experiments had to be substituted by testbed simulations and voluntary IoT dataset collection.

Several technological challenges were also confronted, including, for instance, the integration of the external threat intelligence and decentralised resistance directly into the enhancement of the automated decision making of DRAF. The first technology targeted the inclusion of the openly available data from the Internet to contribute to the identification of risks and attribution to the known attacks. This was achieved by means of the development of the sub-component responsible for feeding DRAF the external data and incorporating known threats reporting into the BR decision making logic. On the same note, we have also anticipated the decentralised resilience of the DRAF through reputation scoring integration, enabling other instances of the system running in other smarthomes to share zero-day threats and identify the misbehaviour of the IoT devices [24].

While the methodology utilised in our study already had a dedicated research question on the limitations of DRAF's application in the smarthome environment (Section 6.6.3), we have also observed direct restrictions related to the study's execution. More specifically, the input data type and its granularity affects the quality of the automation aspects of the RA. More granular reporting on the potential threats and anomalies is required for higher intelligence and automation in the decision-making process. We have addressed this limitation by developing various RS (described in detail in Section 6.4.2.2) as an attempt to normalise incoming reports for DRAF. However, it remains subject to external technical factors.

6.7 Conclusions and Future Work

The presented solution is a complete framework that successfully demonstrates the feasibility of decision-making automation in the RA domain in a dynamic environment, such as a smarthome.

As a background, the domain of RA has been introduced, highlighting the five stages for its definition, followed up by traditional approaches on asset-centric and threat-centric models, with newer models taking the hybrid approach and also including user-centric concepts. Our framework leverages the RA by focusing on the attacks and risk association through the addition of an ontology and a unique methodology on binding elementary attacks or anomaly properties to risks. The full RA model has been illustrated from several angles, including the ORM outlining the conceptual objects with their relations and attributes, a technical flow diagram addressing each of the elements on their utility, and describing their actual implementation from the development viewpoint. The framework emphasises the interoperability with external reporting by proposing a structured API for anomaly reports, which thereafter are handled by DRAF's own internal analysers. This provides flexibility to the types of smarthome monitoring and its locality (eg monitoring on IoT devices themselves, whereas our solution runs on the home gateway). Our approach has demonstrated that a unique risk scoring can be developed to eliminate context dependency. While initial expert weight values are required, by introducing weight adjustment strategies to calibrate the values to a given environment, we have shown that they are transferable to other environments without per-site modifications. Thus, we can conclude that the approach utilised is context-independent and generally applicable to any standard installation on a home gateway. Furthermore, our theoretical model was implemented and tested in smarthome testbeds and real-life environments in several European countries, demonstrating its potential for technological adoption.

An analysis of the presented research questions was provided, evaluating corresponding achievements and shortcomings. Our scientific contribution is notable in terms of the successful illustration of the IoT Stack model, the dynamic adaptations of risk scores and the automated decision making in RA for the smarthome environment. It is of the utmost priority not only in academic research, but also for regular citizens to be provided with the tools enabling them to understand and to have control over the IoT objects' activities in privacy-crucial environments, such as smarthomes. Furthermore, we generalised the initialisation process of the DRAF deployment in the IoT-enabled ecosystem by advancing the formation of the CTI system based on the risk scoring model. The demonstration and validation of the automation aspects of our framework in real-life deployments indirectly pushes forward the frontier of the general awareness of the citizens regarding cybersecurity and privacy problems.

As a future work, we are currently working on an extension of the risk coverage mapping, as already indicated in Section 6.5.1.5, by integrating a greater variety of external reports, more specifically in the domain of the connected and automated vehicles. Future work will put more emphasis on the privacy compliance aspects for the refinement of the currently used risk model. This will also be achieved through already ongoing efforts in the certification domain with the purpose of developing a standardised interfacing for easier integration of any external anomaly reporting in the DRAF.

Part IV

Collective Resilience

Chapter 7

Article V: Towards Reliable Integrity in Blacklisting: Facing Malicious IPs in GHOST Smart Contracts

Relevance

This article examines RQ4 from the decentralised resilience perspective and integration of external knowledge for DRA. We present the support mechanism for enhancing the decision-making of the RAE. More specifically, a reputation scoring scheme is designed, implemented and validated in a simulated environment for reporting malicious IP addresses in a GHOST-enabled deployment environment.

Context

This article was presented at the Innovations in Intelligent Systems and Applications (INISTA), Thessaloniki, Greece. The article was published in IEEE Explore proceedings of 2018 Innovations in Intelligent Systems and Applications (INISTA). According to the Resurchify portal, the IS: 1.25, h-Index: 6.5 and SJR: 0.207.

Own Contribution

Being one of the lead authors of this paper, my contribution to this work is design and implementation of the decentralised reporting and integration of the input into DRA decision making process.

Chapter Contents

7.1 Introduction

Traditional cybersecurity solutions mostly offer server-centric design, causing difficulties in decision making distribution and information sharing. Such setups are prone to a system's single point of failure. The use of the blockchain technology exploits the true benefits of the decentralised approach by replicating valuable data on each network's node, thus enforcing dynamism and reliability of the whole system. In the context of smarthome security, various problems arise from the unmonitored and unknown data exchange of IoT devices that are installed in smarthome solutions, with external parties.

The European research project GHOST (https://www.ghost-iot.eu/) challenges the traditional cybersecurity solutions for the IoT sector by proposing a novel reference architecture that is embedded in an adequately adapted smarthome network gateway, and is designed to be vendor-independent. It proposes to lead a paradigm shift in consumer cybersecurity by coupling usable security with transparency and behavioural engineering. The GHOST project is exploiting novel technologies, such as blockchain, to provide resilience and integrity of decision making on the communication exchange.

In particular, the blockchain technology integrated in GHOST aids to counteract the most modern attacks by decentralising its core risk assessment decision making engine, *RE*. The SC assimilated in GHOST aim to facilitate data sharing on the malicious IPs reported by individual installations and retrieved from the online collective intelligence, *CSKB*. The detailed architectural setup and overview of the complete blockchain integration into the cybersecurity solution are discussed in our previous works [21, 80].

This paper presents a novel SC use case, that aims to provide decentralised security for the protection against data exchange with malicious nodes external to the smarthome. Each GHOST smarthome installation collaboratively creates and maintains a blacklist of malicious IP addresses, by sharing RE produced data from the evaluation of the risks imposed by specific connections between the external IPs and the gateway, an external IP and the IoT devices in the smarthome, and the actual behaviour of the IoT devices and their network communication profiles. Blacklisted IPs are stored on the private blockchain with the help of SCs. The implementation of the blacklist implies operation with data of different levels of sensitivity and trustworthiness and, therefore, can be split into two scenarios: public and private. The main contribution is the integration of a SC with reputation scoring method, further exploited by the RE to perform dynamic risk metric calculations in order to characterise the likelihood of an external IP being malicious.

The rest of the paper is structured as follows. Section 7.2 outlines the research method followed for producing the results presented in this paper. Related work is described in Section 7.3. Section 7.4 presents the Blacklisting IP use case, and open issues are discussed in Section 7.5. Our conclusions are summarised in Section 7.6.

7.2 Research Method

The DSRM [20] is followed in this paper. Design science research relies on the creation of "knowledge and understanding of a design problem, and its solution is acquired in the building and application of an artefact" [102]. DSRM also involves "analysis of the use and performance of designed artefacts to understand, explain and very frequently to improve the behaviour of aspects of information systems" [126]. Johannesson and Perjons presented a DSRM Framework consisting of five main activities as shown in Figure 7.1 [20]. These activities and their coverage by corresponding paper sections are presented in Table 7.1.

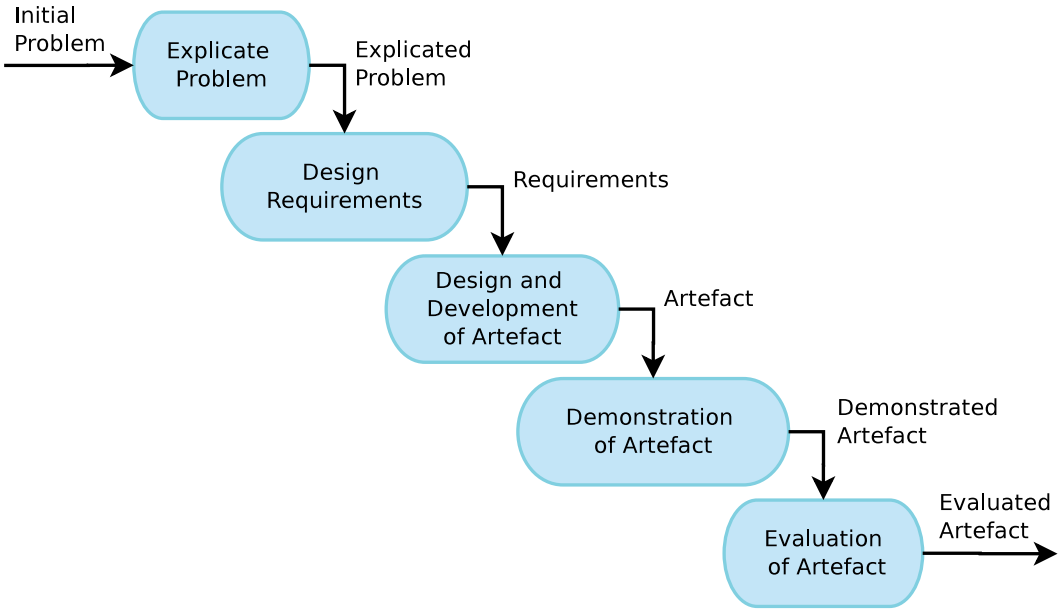


Figure 7.1 Design Science Research Method.

Table 7.1 DSRM and Paper's correlation.

Design Science Activity	Corresponding Section in this Paper
Explicate Problem	Introduction
Design Requirements	Introduction and Related Work
Design and Development of Artefact	Use case description
Demonstration of Artefact	Implementation
Evaluation of Artefact	Implementation

Offermann et al. [127]studied the artefacts designed and developed in the area of information technology and information systems, and classified the artefacts into eight categories: System Design, Method, Language/Notation, Algorithm, Guideline, Requirements, Pattern, and Metric. According to that classification of artefacts, this paper presents a "Method - Defines the activities to create or interact with a system" and "Algorithm - An executable description of behaviour of a system" types of artefacts.

Vaishnavi and Kuechler presented a set of methods for evaluation and validation of design science artefacts as shown in Table 7.2 [126]. This paper adopts a mix of "Demonstration" and "Logical Reasoning" forms of evaluation according to this classification of evaluation (validation) methodologies.

Table 7.2 Design Science Evaluation Methods.

Method	Description
Demonstration	The demonstration method is the weakest form of validation. This method is appropriate for the solutions that are novel and solve a problem for which no other solution exists.
Logical Reasoning	The strength of the logical reasoning form of evaluation depends on the strength and preciseness of the arguments and assumptions. This form of evaluation is an alternate method for experiment and simulation methods.
Experiment and Simulation	This method is useful when the problem is complex and not receptive to a mathematical proof.
Using Metrics	The use of metrics is useful in experiments, simulation, and mathematical proof methods. It is valuable in quantification of the claims about the artefact.
Benchmarking	The benchmarking evaluation method is a weaker form of using metrics method. This method is useful in comparing the results obtained from the experiments and simulations method. This method is useful when there is a lack of suitable metrics to validate the artefact's claims.
Mathematical Proof	The mathematical proof form of evaluation is the strongest form of validation.

7.3 Related Work

The blockchain technology has been recently applied to multiple domains to test its efficiency and to evaluate whether it can be an adequate solution to various problems. One such application domain is the IoT. The main reason behind this choice is the structure of the IoT, that is currently evolving to the Internet of Everything, where numerous devices are going to be interconnected and will interact with each other [128, 129]. These devices are going to be mostly autonomous and they shall communicate on an equal basis, without any central node to provide the required trust. Using the blockchain technology, which mainly

provides trust between nodes, seems to be an effective approach in order to facilitate the future underlying infrastructure for IoT. On the other hand, serious limitations hinder such implementations, as the main approach for consensus, proof of work, demands significant processing resources, which are probably non existent in the IoT ecosystem. Some of the most interesting approaches to the combination of blockchain technology and the IoT paradigm are discussed in this subsection. Huckle et al. [130] present scenarios in which IoT and blockchain can be used in order to enable sharing economies of different assets. One issue with applying blockchain in sharing economy applications is the interaction of the assets with the blockchain infrastructure. By using smart IoT devices it is possible to automatically restrict or grant access to assets like vehicles or buildings according to rules implied by smart contracts without the need for any human intervention. Another important problem regarding IoT is security and one of the factors making this problem worse is the fact that IoT devices often function with outdated firmware. Lee and Lee propose to use the blockchain to certify IoT devices running on the latest and most secure firmware [131]. This is an interesting approach, that could also employ the creation of an open market between manufacturers, end-users, validators and penetration testers. However, there are limitations with respect to the application of the proper rules in the devices to provide the initiative for all stakeholders to push for more secure firmware installed on deployed devices. Supply chain monitoring has been one of the first domains for the application of blockchain technology. Specifically, blockchain allows to monitor the whole process of producing an end product from the moment when the initial ingredients are produced, through the whole manufacturing and exchange process and until the moment the end-user buys the product. This approach can have a significant effect on the quality of products sold. The IoT devices are needed to monitor the process in the most secure and non intrusive way [132].

The use of the blockchain technology also recently gained interest in the cybersecurity domain. In particular, the use of smart contracts was incorporated in the design and implementation of a DDoS defence mechanism in [133]. The use of the existing public infrastructure of Ethereum to advertise blacklisted IPs suspected to be involved in ongoing DDoS attacks is fully exploited in this work. However, despite the advantageous reuse of existing infrastructure and the addition of a novel security mechanism for easy data sharing, a necessity of a central element for proper system functioning still remains at its core. A smart contract *Registry* is required for issuing certificates of the ownership of IP addresses, when submitting new data to the smart contracts. The use of timeout notions while creating white/black/grey lists of IP addresses in the light of the ChainGuard firewall functionality and nodes classification is discussed in [134]. Any node is eventually placed in the greylist, and on the timeout expiration or inactivity from the application itself, it is moved to the blacklist. Alternatively, the nodes are transitioned to the whitelist upon granted connection permission. The greylist capacity is fixed, and once it is full, the ChainGuard generates short-lived flow entry with instruction to drop similar packets. This eventually serves as a DDoS attack monitoring mechanism. Another interesting approach of using blacklists is discussed in [135],

in the field of Vehicular Wireless Networks. When a node is accused of one or more malicious acts, the reporting party prepares a submission of the blacklist data on the blockchain network, including information on the vehicle unique identification, a timestamp of when the ban was issued, a timestamp of when it will expire, and the identities of the source and destination nodes, together with their public keys. These entries are stored on each node locally with distributed access ensured by a blockchain network.

There are some recent research attempts to use the blockchain technology in order to implement reputation or scoring systems for various use cases. Dennis and Owen propose a blockchain based reputation system [136]. After the interaction between two nodes, each node can commit a transaction consisting of the reputation score, a timestamp, and a hash of the interaction. The authors state that the proposed methodology may solve many problems of current reputation systems, such as unfair ratings attack, collusion attack, sybil attack and re-entry attack. Zhao and others discuss the use of blockchain in order to build a decentralised system capable of emulating the functioning of traditional publish-subscribe systems [137]. Publishers publish a topic on the blockchain and subscribers specify an interest message by making a deposit to subscribe to the topic. Then, if there is a match, the publisher transmits encrypted content to the blockchain, subscribers decrypt it, and mark the publisher as its reputation. Finally, the publisher receives the payment from the subscriber. An implementation of the protocol on Ethereum is also demonstrated in this work. Schaub and others describe a decentralised reputation scheme for e-commerce [138]. They use a blockchain structure to store ratings of service providers submitted by their customers. Each new rating is accompanied by a reference to the block of the previous rating for the specific provider, to enable fast summation of all ratings of a provider. Additionally, the authors propose the use of blind signatures to protect the anonymity of the customers, while ensuring that only the customers that should do so submit ratings.

The approach presented in this paper describes a self-maintained cybersecurity mechanism for the IoT. Knowledge produced by RE components, which are placed in different installations, is automatically combined in order to collaboratively access the risk of communicating with specific IPs. Technically the collaboration between different nodes is achieved through the use of SC, in order to ensure the integrity and the availability of the submitted information. The reputation of each IP is directly calculated by a formula that is capable to restore automatically the reputation score to higher levels by incorporating various factors (eg reporting time and quantity). To the best of our knowledge, our approach is the only blockchain based security mechanism for the IoT, which does not require manual intervention to function.

7.4 Smart Contract for Blacklisting IPs

The GHOST network will consist of a full-scale blockchain deployment: Smarthome installations, Smart Sensors and Ethereum nodes. Apart from the main, live network though, a second experimentation private Ethereum [139] blockchain network will be maintained

during the development phase of the project. The experimentation blockchain network will act as a testbed for the development and testing of SC prior to their release on the live network.

Ethereum is probably the most open and flexible among established blockchain implementations. It enables the developer to construct smart contracts, which are more or less equivalent to normal traditional programs in terms of functionality, while they are executed in a completely decentralised way. An Ethereum network is constructed from multiple Ethereum nodes, which are running instances of one of the Ethereum clients. For the GHOST blockchain network Geth, the most popular client, will be used.

For the purposes of the GHOST project, a node has to be installed in each smarthome. This node will be installed at the smarthome gateway middleware. Because of limitations in hardware resources, these nodes are going to be light nodes and will not conduct mining. Additionally, in order to make the Ethereum blockchain functional, multiple full nodes are required. Hence, these are going to be deployed by the partners of the project on more capable hardware. The light smarthome nodes and the full partners' nodes will altogether make up a fully functional private Ethereum blockchain network. The access to this private network though will be restricted only to GHOST nodes, for security reasons.

Due to the limited number of nodes, the GHOST Ethereum network may be vulnerable to 51% attacks, where a group of mining nodes concentrate more than half of the total hashing power. Apart from the full nodes maintained by partners, only the smarthome gateway or middle-ware devices will be allowed to connect to the network. This means that multiple owners of GHOST smarthomes must collaborate in order to combine their light nodes to gather more hashing power than what if the servers were combined, in order to theoretically be able to execute a 51% attack. By using enough hardware resources, one can always ensure that the GHOST Ethereum network is secure.

Additionally, in terms of Ether, full nodes will have income from mining while light nodes will have a balance of zero. Ether in the private GHOST Ethereum network do not correspond to real world value. They can be used though to regulate the network usage. Full nodes may equally distribute all or part of their balance to certified GHOST installations nodes, to enable them to commit transactions.

The experimentation network will consist of 10 different nodes; 4 full nodes and 6 light nodes. The full nodes, that act as miners, will have an Intel quad-core 2.2Ghz cpu each, 32GB of memory and 100GB of storage. From the 6 light nodes, 4 will be installed on Raspberry PIs with an assortment of connected Smart Sensors while 2 will be set up on the GHOST gateway development hardware. The proposed infrastructure setup is depicted in Figure 7.2.

7.4.1 Blacklisting IPs in a Distributed Sharing

Two types of smart contracts, intended for private and for public blacklisting are envisaged.

For the Blacklisting IPs use case, there are two relevant entities: CSKB and RE. RE will create blocking rules, based on two input vectors: user and CSKB. An example of a blocking rule is a list of IP addresses maintained and classified as *whitelist* and *blacklist*. The

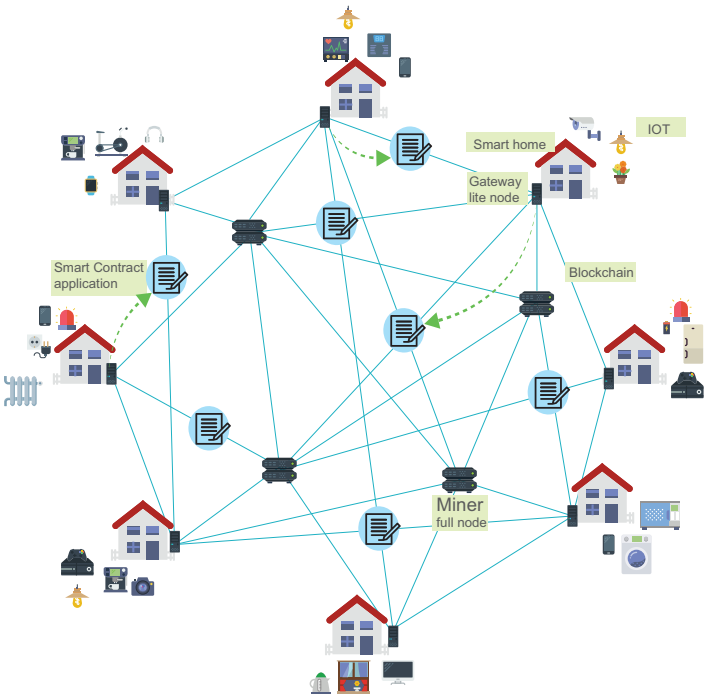


Figure 7.2 Infrastructure for SC implementation.

whitelist will contain the IP addresses marked as safe and stored locally on the gateway, further distributed on the SC. The blacklist will contain the IP addresses that are suspected to be malicious. Those among them that come from an online gathered intelligence (public IP addresses) will be added to the SC by CSKB, while the rest will be added to the blockchain by RE through the end-user's configurations and automatic RE decisions. The local copy will be always available on the gateway, and continuous integrity checks will be performed by SC.

7.4.1.1 Public Blacklisting

For the public blacklisting data, a shared and publicly available knowledge of potentially malicious IP addresses will be established. The main storage of the contract will contain a list of records, each one corresponding to the event that a client in a specific installation is reported to the GHOST blockchain as malicious. These records cannot be maliciously altered or deleted, as such an action would need to alter data already stored in the blockchain, which is extremely hard. In other words, given that all GHOST nodes act according to the predefined procedure, the information stored in the SC is genuine and may be used from GHOST installations to enhance their risk assessment function.

In terms of the mechanism used to calculate a score (bad reputation) for each external IP, two factors are considered. The first one is the number of existing records related to the specific IP and the second one is the age of such records. It is common for a legitimate host to get infected from malicious software or to be utilised by a remote attacker as intermediary hop in the execution of an attack plan. In these cases the administrator responsible for the

host usually discovers the problem and resolves the issue. In such cases, a blacklisting service must be able to remove an IP. In this notion, the SC returns for a specific IP a score (bad reputation), that is calculated according to the number and the age of records for this IP. External hosts that are being reported by multiple GHOST installations will be characterised by high scores. Additionally, when hosts that have been previously reported as malicious stop appearing in the subsequent records, the corresponding score will eventually decrease, and will finally be set equal to zero after a specific time period.

In order for such a mechanism to be functional through an Ethereum SC, specific refinements are required. Too old records, which according to the calculation formula do not have an effect on scores currently produced, are of no practical use. These should be discarded, in order to limit the storage resources required for the deployment of the contract and the processing resources required for the execution of the contract. To avoid completely disregarding past information, the discarded records can be stored in a higher level of information. For example a list containing a predefined number of the worst (the ones with highest scores) IPs for each day may be maintained. Such a supplementary layer of information would enable the maintenance of longer history for malicious IPs, without requiring significant resources. In this way it would be possible to additionally penalise old malicious IPs, if they appear again in the future.

The formula that calculates a bad reputation score for each IP is depicted in Equation 7.1. By dividing the time in discrete time frames or steps it is easier to implement a scheme that takes into account more recent values with a higher weight. The score is calculated for a specific time period, a specific length of time steps denoted as t_p. If the current time step it t_n, then the score is:

$$score = \frac{\sum_{t=t_n-t_p}^{t=t_n} -\ln(cf) sr_t (\lambda)^{t_n - t}}{(\frac{t_p}{trr})^{2r}} \tag{7.1}$$

The sr_t is equal to 1 if there is a record for the IP in time step t and equal to 0 otherwise. The summation in the nominator does not accumulate values for the time steps at which no record exists for the specific IP.

The λ factor is a decay parameter that takes values in the range (0,1). The higher the value of λ is the strongest the memory of the scheme is. Lower λ values mean that the scoring scheme penalises old values in a more heavy way.

The factor trr in the denominator is the total number of requests that have been made for the specific IP in the time window $[t_n - t_p, t_n]$. The more these requests are, the smaller the denominator is, so the score increases.

Parameter r denotes if a removal request for the specific IP has recently been submitted. If there has been one, then $r = 1$, otherwise it is equal to zero, $r = 0$. In practice, if a removal request has been submitted, then the r factor limits the score. If no recent request exists, then the denominator is equal to 1 and the score equals the nominator.

Finally, the cf parameter stands for the cardinality factor and penalises the case where all records come from the same submitting address. It is equal to the percentage of records for the specific IP existing in the time window $[t_n - t_p, t_n]$ and have been submitted from the address being examined. It practically protects the reputation of IPs from spamming accounts, that would want to harm the owner of an IP by repeatedly submitting blacklisting records for this IP.

In order to present how the formula works, some simulation tests have been conducted, the results of which are depicted in Figure 7.3.

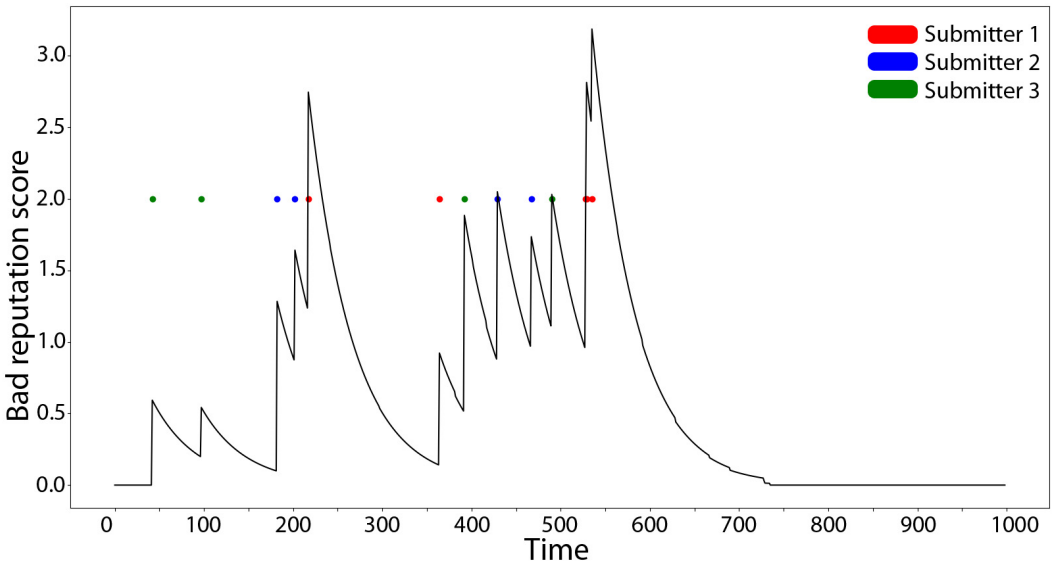


Figure 7.3 Reputation score for a specific IP.

This figure shows the calculated reputation score for a particular malicious IP, given that the relevant reports for this IP are shown by the coloured dots. The colour of each dot represents the unique id of the submitter of the report. For instance, all red dots are representing submissions of Submitter 1, all blue dots come from Submitter 2 etc.

It is evident in the graph that the scoring scheme values reports according to how recent they are. This is why the score starts to decrease with time after a specific report, at least until a new report is submitted. The rate at which the scheme phases out the past reports is dictated by the decay parameter λ.

Additionally, not all reports contribute the same value to the total reputation score for a particular IP. If a submitter keeps sending reports for the same IP, then every new report is weighted less. This is evident in the case of the first two reports by Submitter 3, ie the first two green dots, or in the case of the first two submissions by Submitter 2, ie the first two blue dots approximately at $t = 200$. In contrast, when Submitter 2 keeps quiet for a period of time, his records start again to be valued more, at $t = 470$.

7.4.1.2 Private Black/Whitelisting

The private blacklisting and whitelisting of the IP addresses is a variation of the public blacklisting, where the malicious IP addresses have influence only on a per installation basis. Despite any public recommendation (ie Public blacklisting), a user still can have personalised settings and a set of rules. Another set of SCs will be put in place, where a private list of rules is recorded. Each rule in turn is encrypted together with a state indicating to which list it belongs (ie blacklist, whitelist or none for the purpose of resetting the state).

To retrieve the rules, a RE communicates with the SC, which has to aggregate all the rule entries linked with the owner and return a list of the encrypted rules with their entry dates that are stored with the SC. The GHOST client can then decrypt and use the lists to perform several actions, eg to restore its internal black and white lists.

7.4.2 Implementation Details

There are three main features implemented in the SC for both variations of the blacklisting. The schematic capture of these components is presented in Figure 7.4, that demonstrates the communication exchange between RE and SC.

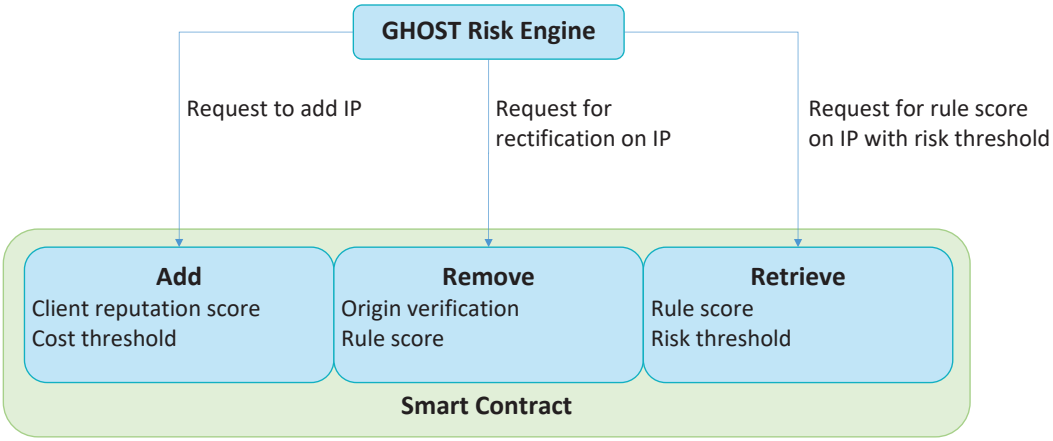


Figure 7.4 Smart Contract function outline.

7.4.2.1 Reporting Malicious IP

Upon submission of a request to add a new IP to the blacklist, SC will initially calculate the current reputation score of a submitting party, so called *Client*. The SC will also calculate the cost of fulfilling the *Client*'s request. Knowing a consecutive *Client*'s reputation score upon positive outcome, will provide the basis for the cost threshold comparison. If the *Client* recently reported the same IP address, there is no need to spend more Ethers on committing the transaction, as well as modifying the related reputation score of the *Client*. Finally, SC will return the outcome of the evaluation (positive or negative) together with a new reputation score of the *Client*, if applicable.

7.4.2.2 Rectifying Erroneous Report

It is possible that *Client* can report the IP address being malicious by mistake as well as by a faulty automated or manual decision. In such a case, a request to remove an IP address from the blacklist can be sent. Such a request should be first verified by the origin, to ensure that only the reporting *Client* can rectify an erroneous action. Once verification is successful, SC will calculate the reputation score of the existing entry, so called *Rule*. If the resulting score exceeds the predefined threshold, a resulting outcome (positive or negative) is sent back to the *Client* together with the new *Rule*'s score if applicable. The threshold is used to avoid situations when committing a new transaction is no longer valuable, as *Rule*'s reputation score won't be significantly improved due to the expiration of the entry's freshness.

7.4.2.3 Gathering Common Dataset

On a regular basis, RE will make requests to the SC to retrieve the *Rule*'s reputation score for the IP address of interest. Very often a user's device will initiate communication with an unknown destination. In this case an acceptable risk level, defined by the GHOST user, will be incorporated for deciding on the threshold of the *Rule*'s reputation score. Upon receiving the request on blacklist retrieval, SC will calculate *Rule*'s reputation score and will compare it with the user's defined risk threshold. The result will include only those blacklisted IPs that are below the risk threshold.

The implementation of the smart contracts is outlined in Listing 7.1 for Public Blacklisting and in Listing 7.2 for Private Black/Whitelisting respectively.

```
1  pragma solidity ^0.4.16;
2
3  contract PublicBlacklisting {
4
5      struct Record {
6          bytes4 ip_address;
7          address submitter;
8          uint time;
9      }
10
11     Record[] public records;
12
13     function submitRecord(bytes4 ip) public {
14     .....
15     }
16
17     function RemoveRecord(bytes4 ip) public {
18     .....
19     }
20
21     function evaluateIPScore(bytes4 ip) public {
22     .....
```

```
23      }
24  }
```

Listing 7.1 Public Blacklisting smart contract.

```
1  pragma solidity ^0.4.16;
2
3  contract PrivateBlackWhitelisting {
4
5      struct Record {
6          bytes4 ip_address;
7          uint time;
8      }
9
10     Record[] public whiteRecords;
11     Record[] public blackRecords;
12
13     function submitRecordWhite(bytes4 ip) public {
14     .....
15     }
16     function submitRecordBlack(bytes4 ip) public {
17     .....
18     }
19     function removeRecordWhite(bytes4 ip) public {
20     .....
21     }
22     function removeRecordBlack(bytes4 ip) public {
23     .....
24     }
25     function retrieveWhiteList() public view returns(Record[]){
26     .....
```

Listing 7.2 Private Black/Whitelisting smart contract.

7.5 Open Issues

This section discusses the open problems encountered during the design and implementation life-cycle of the Blacklisting IPs use case.

7.5.1 Reputation Scoring and Anonymity

The PbD framework employed in the GHOST project dictates specific principles to be applied at all stages of the GHOST solution lifetime. Therefore, during the analysis and design stage, privacy protection measures were considered. First, an anonymous public blacklist submission was considered to prevent GHOST blockchain network neighbours from exploiting the full discovery of blacklisted parties. However, this approach was directly leading to an even bigger issue. If one of the network neighbour nodes would happen to have malicious

intentions, it could blacklist any node in the network or targeted destination party, thereby compromising the targeted node's communication capabilities for any incoming data from that device. Secondly, the reputation scoring approach currently employed in the SC would not be possible due to the inability to identify the maliciously reporting party and to verify its trustfulness.

7.5.2 Spamming on the Blockchain

The nature of the possible spamming within the GHOST blockchain network is similar to malicious blacklisting. One could think of introducing a pre-analysis stage prior to the submission of the blacklist data. Reputation scoring is also providing certain countermeasures to address this problem. However, the introduction of transaction costs for blacklist data submission appears to be most reasonable. In particular, prior to data submission by RE, the associated cost is evaluated based on several criteria: *reporting frequency*, *data freshness* and *common intelligence*. Several strategies are put in place for the dynamic cost calculation, for example making transactions cheaper in case of reporting blacklisting for the party that is publicly known to be malicious (eg identified as part of botnet network or being targeted by DDoS attack).

7.5.3 Storage on Ethereum

Another important challenge attempted to be addressed is a limitation on the storage capacity of the data on the GHOST private blockchain network. The infrastructure deployment has limitation on the light nodes bound by hardware. This problem is not specific to the GHOST installation only, but is a generic problem in the blockchain community (eg technology enthusiasts, developers, business pioneers) and was attempted to be solved by few recent works [140, 141]. To further evaluate realistic limitations and implement appropriate solution, an in-depth analysis of the possible lifetime is currently performed. In link with the privacy profile typology presented in [142], several nodes' profiles are considered, based on possible activity levels and smarthome setup diversity: *fundamentalists*: preserved and sceptical users leaning towards maximum security and minimal exposure, low participation; *unconcerned*: the opposite users, happy to discover new devices and having high trust in technology, high participation; and *pragmatists*: majority of users, continually weighting the benefits between secure exposure and novelty of features provided, medium participation.

7.5.4 Costs of Data Submission

As discussed earlier, a need for legitimate strategy on costs and rewards is identified in link with the open issues in Section 7.5.2 and in Section 7.5.3. One can think of rewarding the nodes on a submission of the blacklist data correlated with publicly available intelligence.

7.5.5 Lifetime Availability of the Data and Novel Security

An important generic problem applicable to the whole blockchain community is the potential decryption of the data stored on the blockchain network in the near future. This may happen not only due to technology acceleration and the possibility to decrypt data with stronger hardware or quantum computers, but also due to the high likelihood of vulnerability discovery in the protocols and software in general. It is therefore possible at some point in the future, for an attacker to discover the black/whitelist data associated to a particular installation and the user. However, by following PbD principles [143], it is always a priority to expose as little personally identifiable data as possible. In particular, even though blacklist reporters can be uniquely identified within GHOST network of installations, the data stored in the blockchain network won't be linking to a unique user globally identified. First an attacker will have to learn the associated unique identifier and real users correlation within the GHOST network, before further being capable of unique tracing of the associated blacklist data.

7.6 Conclusions

Given the kind of attention that blockchain technology has received from academia and industry, it appears to be an effective approach to facilitate the future underlying infrastructure for the IoT. In this paper we proposed a smart contracts-and-blockchain-based mechanism to mitigate some of the potential risks associated with the identification and tracking of malicious IP addresses. One of the main contributions of this paper is the use case description and implementation method for two types of smart contracts designed to blacklist malicious IP addresses: publicly and privately. The blacklist will contain the IP addresses suspected to be malicious. For the public blacklist of IP addresses, shared and publicly available knowledge of potentially malicious IP addresses will be used. On the other hand, for the private blacklist of IP addresses, the list will contain the malicious IP addresses that are likely to have adverse influence only on per installation basis. Despite any public recommendation (ie Public blacklisting), a smarthome user can personalise the settings and override public rules by whitelisting the party of interest. The second main contribution of this paper is the proposal of a scoring mechanism that serves as the reputation score for the *Rules*. The mechanism calculates a score (bad reputation) for each external IP with the help of two main factors. The first factor is the number of existing records related to the specific IP and the second one is the age of such records. The third main contribution of this paper is the identification of relevant open issues such as reputation scoring and anonymity, spamming on the blockchain, storage on Ethereum, costs of data submission, and lifetime availability of data and novel security. These open issues set the directions for future research, including the validation of the proposed smart contracts by simulation and pilot run.

Chapter 8

Article VI: Integrating Human Factors in the Visualisation of the Usable Transparency for Dynamic Risk Assessment

Relevance

This article examines RQ3 and RQ4 from the human factor perspective while applying usable security principles for the visualisation of the DRA in the context of the platform configuration and decision-making for risks mitigation. This approach was designed through user studies under GHOST project, further implemented into user interfaces for the RAE.

Context

This article was submitted to the journal Information and at the current stage is under 1st revision by the reviewers According to the Resurchify portal, the IS: 2.38, h-Index: 28 and SJR: 0.349.

Own Contribution

Being the lead author of this paper, my contribution to this work is the design and implementation of the usable security interfaces and support for the user studies to validate the interfaces.

Chapter Contents

8.1 Introduction

IoT is a powerful emerging technology that has been ascertained to make home environment smarter, more secure, connected and automated [14]. Though, the technologies supporting the smarthomes functionalities ushered new daunting cybersecurity and privacy challenges [144]. While security became one of the first priorities in software development, multitude of challenges were identified by the developers experiencing difficulties in integrating security principles into the designs and structures of their implementations [145]. Consequently, an eager need for tools providing visibility into the cyber risks and threats has been raised in parallel with the deployment of the cutting edge smarthomes [33]. For that matter, DRA based tools are foreseen to allow the smarthomes users to take control and make appropriate decisions regarding the existing cybersecurity and privacy risks [23]. Such technology intends to automate the threats identification and to provide control and monitoring features for mitigation of detected risks. However, the DRA's prevalence depends on how its UI is tweaked towards users feedback and involvement.

The professional cybersecurity analytics tools evolved tremendously in the last decade. Previously, many tools were developed for security analysts, but mainly containing tables and text, intended for a professional use only. However recent advancements demonstrate that data visualisation tools for Information Retrieval (IR) have completely changed security, as they enable users and analysts to understand information about their network security more easily [146].

Nevertheless, the usability aspects are often neglected. Poor usability of cybersecurity solutions tends to be the effect of security constraints. Finding the right trade-off between usability and security or preferably integrate usability and security requirements is part of a major research challenge, which recently has been raised by scholars [50]. For instance, user-centred approaches are recommended as a means to accomplish usable security [147], while the definition of objectives for both security and usability is suggested as a way to decide on the right balance between the two [148]. Understanding the security and usability collectively is recognised as a critical factor for the successful development, implementation and usage of information systems, according to Andriotis et al. [149]. As far as the IoT is concerned, usability is among the major research challenges identified [150]. Consecutively, we observe a growth in privacy concerns, as IoT device manufacturers for the smarthome are acquired by large corporations, such as Google [151]. Most recent research suggests new usable security frameworks particularly for modelling security and privacy risks in smarthomes at consumer level. For example, the framework presented in [152] aims to support home users with a highly usable security decision support tool. However, it still needs to address improvements on usability and scalability and validate real utility offered to the user.

Existing cybersecurity solutions tend to provide an increased protection at the expense of usability [148, 153], a choice that typically backfires in practice because of end-user demotivation due to poor usability leading to an even weaker protection. On the contrary, our

implementation targets to achieve a substantial increase in usability with minimal security trade-offs. To realise this ambition, we have adopted a threefold strategy that builds upon extensive automation (minimising security related user interactions), user motivation and building trust.

This article aims to reply to the following research questions:

RQ 1 *What are the limitations to automate technical actions in case of detected risk exposure in the scope of the threat landscape specific to smarthome environments?*

RQ 2 *How these technical actions can be translated to the lay user, assuring high usability and efficient cybersecurity?*

RQ 3 *Can we have equally engaged lay users with different security and privacy perception and risk acceptance?*

Our added value and contribution can be summarised as follows:

- Definition and demonstration in practice of the methodological process on technical actions mapping to the usable and automatable actions for cybersecurity mitigation.
- Based on the constant end-user feedback through an interactive user-centric approach, we tested our model in a smarthome environment.
- Development of a decision tree conceptualisation by exploring the perception of the end-users security, privacy related risks and associated IR methods in the context of usable security.
- Translation of the decision tree model into a final set of decision-making monitor and control UIs.

The remainder of this paper is structured as follows: in Section 8.2, a literature review of the two main concepts of cybersecurity visualisation are presented. Section 8.3 outlines the methodology for user actions mapping that we have followed. Section 8.4 extends the first step of our process by exploring the potential threat vectors that are applicable to the smarthome ecosystem. Section 8.5 provides a mapping of the predefined threats to relevant technical actions aiming to countermeasure the attacks' risks. Section 8.6 shows how the DRA is incorporated to our process to dynamically evaluate the cybersecurity risks. Section 8.6 outlines the two final steps of our process through a proposition of a decision tree conceptualisation that is translated into two types of UI afterwards. Finally, Section 8.8 present an analysis on research questions and conclusions respectively.

8.2 Related Work

The primary focus of our work is in the development of the reusable framework to map technical risk mitigation actions to usable UI. However such mapping implies a deep understanding

of all intermediary steps from attack identification and mitigation to usability aspects and human factor inclusion in the design of the UI. Therefore, the focus lies on four interrelated fields of research, starting from the existing usability recommendations in security domain, their applicability in terms of a lay user risk perception, tailoring of those risks into threats consequences mitigation, and, finally, inclusion of the user in the whole process of developing the final solution.

8.2.1 Security Usability Guidelines

The foundation work on the guidelines for secure interaction design widely applied in real-life products was provided by Yee [154]. It is grouped into three pillars, each providing more fine-grained recommendations on the design of the interfaces:

- General principles: relies on the path of least resistance and appropriate boundaries;
- Actor-ability state maintenance: achieved by explicit authorisation, visibility, revocability and expected ability; and
- Communication with the user: accomplished with trusted path, identifiability, expressiveness and clarity.

However, usable security is offering much more than graphical interfaces. It addresses how people think about and use computer systems. The authors of [155] attempted to analyse three organisational case studies, aiming at improving their security products. They have observed that usability is never a priority for the development of the new products, and mostly was introduced due to customer's complaints. Furthermore, usable security is not seen as a quality improvement property and its definition differs from one organisation to another. In addition, there are no clearly defined evaluation criteria for usability, and, therefore, developers are not capable to deliver usability, as they lack understanding on its impact on end-users' performance. Based on these findings, the authors raise an open question *'Does risk-based security make security more usable than compliance-based security?'*.

Balfanz et al. [156] provided lessons extracted from their work, while building usable security solution. First, one can not retrofit usable security, instead the interaction principles should be changed at the core of the product. This hinders the need of end-user inclusion in the whole design and development process of the security product. Second, developers should be careful about front-end UI layers of the application design. If a security feature prevents the user from accomplishing a certain task without being exposed to what is happening within the system of their device, then that security feature is likely to be turned off completely due to the lack of comprehension. To that end, transparent and understandable control functionality of the security software solution is highly needed. Finally, security expertise should not out-weight the end-user's needs.

In terms of the usability measurements techniques, Atzeni, Faily, and Galloni [157] proposed to consider the following characteristics: effectiveness, satisfaction, accuracy, efficiency,

memorability and knowledge. However, the authors stressed out that such measurement should not be utilised on its own as it stands, but in a comparative manner from the improvement perspective. This leads to the need for iterative design process and thus to continuously improve the usability.

8.2.2 Security and Privacy Risks Perception

Gerber, Reinheimer, and Volkamer [158] performed a study on privacy risk perception in three distinct domains: online social networks, smarthomes and smart healthcare. Their findings pointed the common lack of understanding of the consequences on the risks exposure. More specifically, they evaluated the differences between perception of the abstract and specific risk scenarios related to the cyber risks. A proposition of the design of a risk perception intervention to raise security and privacy threats attitude specifically in the smarthome environment, was advanced in the follow-up work [159]. Their proposed intervention targets the vision enhancement for a smarthome as a whole.

Another study was run by Barbosa, Zhang, and Wang [160] to understand the smarthome device adoption limitation factors. The researchers identified three clusters of consumers who go forwards with the purchasing the devices for the installation in their homes. For the first category, affordability-oriented, the primarily criteria was the actual price of desired devices. The second cluster's representatives, privacy oriented, prioritised the personal data preservation politics. Final group of consumers, reliability oriented, were guided mostly by the assessment of the devices functionality. Similarly, Emami-Naeini et al. [161] studied IoT security and privacy label by utilising layered approach for the information visualisation in the context of the smarthome devices acceptability barriers by lay users. Interestingly, they confirmed the known phenomenon on *Privacy Paradox*: observation of the discrepancy between privacy concerns and actions taken to mitigate those concerns.

The end-user perception on the IoT data field practices and associated risks was analysed in the interview-based study [162]. The authors claim that the results indicate differences in the lay user mental models risks and protection behaviour, highly depending on their background and experience. This implies the need for designing adoptable UI to suit the diverse needs and support distinct perceptions of the end-users. More specifically, a set of recommendations is derived from this work, highlighting the need for transparent controls and educating regular citizens about involved risks.

8.2.3 Risk assessment for threat mitigation as a usability improvement

As outlined by the Bugeja, Jacobsson, and Davidsson [86], the smarthome domain is fulfilled with the security and privacy related challenges dictated by three conceptual sources: devices, communication channels and services provided. This survey also served as a basis for the derivation of the threat landscape applicable to our methodology, further described

in Section 8.4. The authors also stressed out the need for the empirical risk evaluation methods to facilitate and improve usability of cybersecurity solutions for lay users.

To have a better understanding of the current threat landscape in IoT-enabled systems, a honeypot-based environment was deployed by researchers to capture the most recent snapshot of existing risks [163]. The outcome of their six months experiment confirms the relevance of the traditional cyber attacks to smarthome environments.

An interview based research was accomplished by Haney, Furman, and Acar [164] to identify currently available and used in practice by non-technical end-users mitigation techniques for security and privacy risks in smarthome environments. The resulting conclusion clearly demonstrates the lack of available tools, even for more technology-savvy users. A set of guidelines was derived from this study, pointing the emerging need for data collection and cybersecurity transparency, privacy and security controls and general assistance availability for lay users.

8.2.4 User-centric Approaches

Experience-centred approach in privacy and security technologies was coined by Dunphy et al. [165]. The value and necessity of proposed approach are argued to be encouraged by the changing context of where the technology to be deployed. More specifically, authors have demonstrated three use-cases, each outlining a different approach, but permitting to establish findings that other methods would not be able to capture:

- Collage building: enables the end-user to be in charge of the engagement method and extent of their contribution and experience sharing;
- Questionable concepts: facilitates the expression of the opinions, as provocative concepts are proposed by the designers; and
- Digital portraits: permits to establish trust between participating parties.

All three methods are applicable at the solution design phases and emphasise a secondary place to the technology itself, spotlighting the desire to use the technology by the lay user.

A plethora of user-centric approaches was proposed afterwards in the academic research in the domain of security, privacy and trust design. Collard and Briggs [166] analysed a range of the relevant tool-kits and assessed their effectiveness through the execution of series of the workshops. The methods they have utilised are based on story-telling, visual and 3D modelling, improvisation and role-play, games and cards and, finally, on problem setting and mapping. Each of the applied methods showed the potential on discovering versatile results at the design phases. A visualisation of the outgoing network traffic of a smarthome through the application of the participatory design was presented in recent work of Victora [167], *IoTGuard*. While not targeting directly the provision of the cybersecurity risks exposure mitigation, *IoTGuard* targets to provide transparency and control options to lay user to improve the understanding of the smarthome cyber risks.

Awareness campaigns in cybersecurity have raised researchers' attention with the purpose to study the influencing factors on the online behaviour change for a lay user. The main hypothesis of Bada, Sasse, and Nurse [168] on failing awareness campaigns relied on the fact that security interfaces as they are developed, are often too difficult to be used by a lay user. They have proposed generic classification of the influence factors into two domains: personal factors and cultural and environmental factors. While personal factors are recognised to be formed by an individual's knowledge, skills, personal motivation and experience, the cultural and environmental factors stem from the collective phenomenon of self-perception. The authors of the same work also identified that techniques used for persuasion of the behavioural change rely on versatile elements such as fear, humour, expertise, repetition and scientific evidence. However, none of them shown to be more effective than the others.

We have identified a general lack of the user-centric methods applicable directly to the cybersecurity of smarthome domain, especially with the focus on the RA. One of the closest works we have found was on the topic of User Experience (UX) in design of smarthome devices. While not addressing directly the question on providing a methodology for development of the add-hoc cybersecurity solution for the smarthome, the authors suggested guidelines to improve the data protection in smarthomes through a series of the interviews [169]. Their work shows a persisting need for a methodological design processes to deliver security solutions for smarthome environments.

In the same context, Feth, Maier, and Polst [170] presented a user-centred design model for usable security systems, relying on four iterative pillars:

1. Context of use: defined by the users, tasks and environment;
2. System awareness: ensures the system is understandable for the end-user by mapping conceptual model to the user's mental model;
3. System design: selects UI patterns to support envisioned functionality of the final system; and
4. Design evaluation: iterative cycle through feedback collection and analysis.

While the proposed framework remains an abstract concept, the authors demonstrate its theoretical application to IoT device deployment in a smarthome by a fictitious user.

Finally, a thorough survey was presented by [171] in the domain of human-centric cybersecurity. Their work defines a wide perspective as a generic framework applicable to cybersecurity products encompassing user, usage and usability - *3U's*. Such taxonomy enables efficient positioning of any existing methods in the domain of usable security.

8.3 Methodology for User Actions Mapping

To address the emerging need for usability in cybersecurity, we have defined methodology on the user-centric development of usable user actions to ensure transparent monitor and

control of detected risks. The main phases of our methodology are depicted in the Figure 8.1. Our work was performed under the umbrella of the GHOST project, offering a smarthome cybersecurity software solution (https://cordis.europa.eu/project/id/740923 accessed on 31 March 2022).

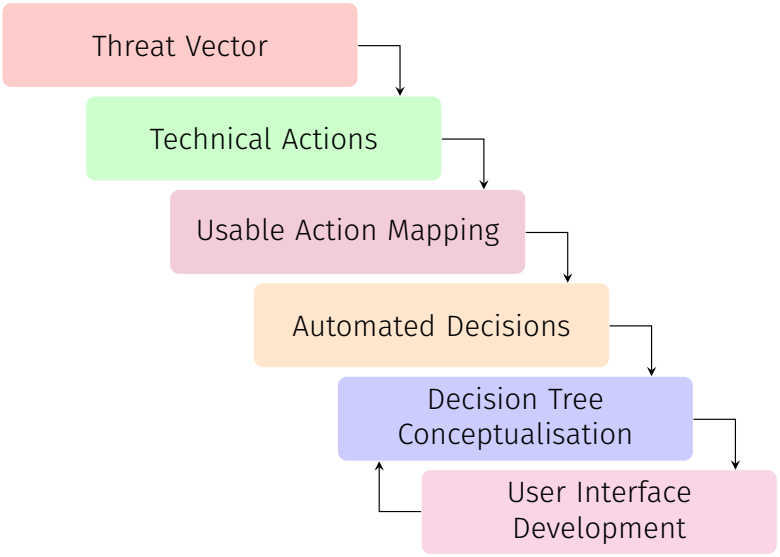


Figure 8.1 Methodology on User Actions Mapping.

We started our process with the identification of the relevant threat vectors applicable to the smarthome environment. Guided by the deduced attack taxonomy, operational context and deployment constraints, we have established a list of attacks, which are technically feasible to simulate in a safe setting, without endangering the end-user to be exposed to the real cyber risks (see detailed analysis in Section 8.4). The Table 8.1 provides a summarised overview of the selected attacks, their demonstration methodology and associated tooling for attack replication.

Table 8.1 Summary of attacks.

	Attacks	ID	Validation methodology	Software Tools
Physical	Damage	P1	Remove battery, shutdown	N/A
	Device injection	P2	Device registration, sniffers	N/A
	Mechanical exhaustion	P3	Trigger device operation	N/A
Network	Traditional	N1	Scanning and enumeration	nmap, Scapy, tcpreplay
	Device impersonation	N2	Packet injection	Scapy, tcpreplay, tcprewrite
	Side-channel	N3	Hardware/software sniffers	Wireshark, tcpdump
	Battery Attacks	N4	Packet injection, sniffers	Scapy, tcpreplay, tcprewrite
Software	Traditional	S1	Traffic replay	PCAP files, tcpreplay, tcprewrite
	Software compromise	S2	Alter module behaviour	Module-specific software
	Command injection	S3	Inject legitimate commands	Specially crafted software
	Mechanical exhaustion	S4	Inject legitimate commands	Specially crafted software
	Sleep deprivation	S5	Inject legitimate commands	Specially crafted software

The next step was the identification of the Technical Actions applicable to each category of attacks, outlined in details in Section 8.5. Those represent the techniques that can be used to address the attacks. A summary of the possible actions and their potential on the automation is provided in Table 8.2.

Table 8.2 Technical Actions on Attacks.

ID	Description	Automatable
T1	Verify physical integrity	No
T2	Verify battery	No
T3	One-way sandboxing	Yes
T4	Two-way sandboxing	Yes
T5	Permit	Yes
T6	Block device temporarily	Yes
T7	Block device permanently	Yes
T8	Drop packets for flow temporarily	Yes
T9	Drop packets for flow permanently	Yes
T10	Drop packets for source temporarily	Yes
T11	Drop packets for source permanently	Yes
T12	Restart GHOST	Yes
T13	Restart module	Yes
T14	Disable module temporarily	Yes
T15	Disable module permanently	Yes
T16	Send update request	Yes

The next step is the definition of possible Usable Actions, to enable lay user to control and monitor their smarthomes from cybersecurity and risk evolution perspective. Table 8.3 shows the final mapping between identified and technical actions. This mapping served as a base for the Automated Decision derivation and corresponding conceptualisation of the Decision Tree, described in details in Section 8.6 and Section 8.7 respectively.

Table 8.3 Technical Actions on Attacks.

ID	Description	Linked technical actions
U1	Allow	T5, T16
U2	Block	T3, T4, T6, T7, T8, T9, T10, T11
U3	Ignore	T3, T4, T5, T6, T7, T8, T9, T10, T11, T16
U4	Remind	T3, T4, T5, T6, T7, T8, T9, T10, T11, T16
U5	Advisory	T1, T2, T12, T13, T14, T15

The final step of the actual development of the user interfaces was continuously improved upon the execution of the user studies running in the scope of the GHOST project, mostly based on the user surveys for their feedback collection [100, 158]. The outline of the interactive

user-centred approach is detailed for two types of the interfaces CFG and SI in Section 8.7.2 and Section 8.7.3 respectively.

8.4 Threat Vector Landscape

We have analysed the existing threat landscape applicable to the smarthome environments to define an initial set of applicable attacks. The most notable works providing the taxonomy of smarthome specific attacks and threats were provided by Heartfield et al. [121] and [86]. With cross-correlation performed in the scope of the risks identification under GHOST project [23], we have further reduced the initial listing by applying criteria on the attack simulation feasibility. The sections hereafter outline each category of attacks included in our analysis along with scenario definition, implemented demonstration and validation methodology.

8.4.1 Physical attacks

8.4.1.1 Physical damage

Scenario: Physical damage to an IoT device may be caused by various means: remove battery, shut down the device, physically break the device (the communication component, complete physical destruction), etc. However, irrespective on the actual cause, the result will be the same: communications between the device and the IoT gateway will be interrupted. An attacker may pursue the effective physical destruction of IoT devices for various reasons. For example, the attacker may break into a house and (physically) disable sensors/actuators in order to avoid alarms being triggered.

Demonstration/validation methodology: According to the specific functioning of each device available at the deployment site, the attack is demonstrated via:

- Removing the battery: Z-Wave and Zigbee devices.
- Shutting down the device: Wi-Fi devices.

8.4.1.2 Malicious device injection

Scenario: The 'injection' of a new device may happen in various scenarios. This might not necessarily represent an attack scenario, since a user may just want to extend the set of IoT devices with new ones. On the other hand, an attacker may attempt to add a device to an existing network, in which case the DRA should signal this attempt to the end-user. An attacker may pursue the injection of new devices for various reasons. For example, the attacker may wish to indirectly trigger events in other IoT devices (eg trigger fire alarms). Conversely, the attacker may want to alter the behaviour of an installation by flooding it with packets, valid requests, which may lead to effective DoS attack by filling the gateway's disk with event logs, by flooding the user interface with alerts and valid events.

Demonstration/validation methodology: Two scenarios are demonstrated to replicate the attack:

- Following the typical procedures for adding a new device to the IoT installation.
- Use of a Z-Wave sniffer to demonstrate the sniffing of events via a new device.

8.4.1.3 Mechanical exhaustion

Scenario: The mechanical exhaustion implies that an attacker with physical access to an IoT device is able to repeatedly trigger the mechanics of a particular device. For example, in the case of an IoT switch the attacker may repeatedly turn ON/OFF the device (at a higher rate than expected). The objective of the attack may be diverse, however, the attacker may try to cause a mechanical exhaustion, which, in time, may lead to malfunctioning and/or physical damage(s). The attacker may also exploit the physical access to a device to indirectly trigger other devices. Accordingly, in an IoT scenario it is common an event originating from a particular device to trigger actions on other devices. Therefore, the attacker may indirectly trigger repeatedly the other devices as well, which, in turn, may also be damaged.

Demonstration/validation methodology: The attack is demonstrated by repeatedly physically triggering an IoT device (eg a switch/relay). The triggering needs to be performed at a higher rate than it would be expected from the device's normal operation.

8.4.2 Network attacks

8.4.2.1 Traditional attacks

Scenario: Traditional network attacks imply the exploitation of the operation of traditional IP-based protocols. In this category we find the well-known network scanning and device enumeration techniques (TCP/IP-, and UDP-related scans), DoS and DDoS attacks. Given that the smarthome is expected to use IP-based protocols to communicate with remote sites, it is therefore important that the DRA is aware of traditional attacks that exploit IP-based protocols. The attack recreates the typical steps that would be performed by an attacker in the attempt to discover IoT devices and gateways, and to enumerate the available services.

Demonstration/validation methodology: Traditional tools will be used to mount attacks against the IoT gateway, including:

- `nmap`: SYN scan, XMAS scan, full scan.
- `tcpreplay, tcprewrite, Scapy`: DoS attack, DDoS attack.

8.4.2.2 Device impersonation

Scenario: By leveraging specially crafted hardware, an attacker may use a registered device's identifier and key in order to inject packets, to trigger other devices, and to sniff the

communication network. Indeed, the attack requires a higher level of technical knowledge and a deep understanding on the underlying infrastructure, and of communication protocols. Furthermore, in case communications are encrypted, the attacker needs to get into the possession of the encryption key. This may be possible due to flaws in cryptographic algorithms and/or cryptographic protocols, or due to flawed implementations. This attack will enable device impersonation, and the attacker may attempt to trigger other devices to open doors, trigger alarms, simulate presence. The attacker may also attempt to get into the possession of alerts issued by other devices. To this end, sensitive data such as user presence may be revealed by other devices, which may by used to break into the user's house.

Demonstration/validation methodology: The demonstration methodology for this attack will not attempt to recreate the full attack vector. Instead, we emulate a device impersonation by specially forging new data packets (specific to the IoT protocol, eg Bluetooth, Z-Wave) and by injecting these packets directly to the network capture.

8.4.2.3 Side-channel attacks

Scenario: Side-channel attacks try to extract information indirectly from the behaviour of a particular system. In the case of network attacks, side-channel attacks attempt to infer sensitive data on users, and devices based on the network protocol implementation, and not on its contents. The attack may be used, for example, to infer user presence or particular events (eg alarm triggered, switch triggered). The attack shall evaluate the level of sensitive information that could be inferred by an attacker that passively sniffs and analyses network traffic. An attacker may attempt to infer sensitive user/device information by sniffing communications (which may be encrypted), and by analysing the unencrypted packet headers (data link, IP, TCP/UDP, Z-Wave, Bluetooth). This way, the attacker may learn that the user is not home or that the door/window is opened/closed. This information may then be used for breaking into the user's house.

Demonstration/validation methodology: The demonstration focuses on IoT device traffic and gateway traffic, thus on both IP-based and non IP-based traffic. In particular, for specific protocols the following tools will be considered:

- Bluetooth: packets may be sniffed by leveraging the device's capability (eg Android has a built-in sniffer).
- Z-Wave: a special device is needed to register into the network and to sniff the Z-Wave packets.
- Wi-Fi/Ethernet: traditional network sniffing tools such as Wireshark/tcpdump may be used.

8.4.2.4 Unusual activities and battery depleting attacks

Scenario: A wide variety of attacks may trigger additional processing in IoT devices. Therefore, ultimately, the presence of ongoing attacks and, in general, and 'unusual activity', may be detected by monitoring battery consumption. An attacker may simply attempt to exhaust the systems energy, or may carry out activities on or around the gateway for undetermined reasons. Such activities can signal various forms of malfunctions, anomalies and even attacks, and can constitute a simple means of monitoring the health and safety of the gateway.

Demonstration/validation methodology: The validation exploits the gathered traffic rates within the gateway, and between the gateway and the outside world, recording normal traffic levels over extended periods. The validation will also measure energy consumption in various devices and the rates of battery depletion on a daily and if possible hourly basis. These metrics will be used to determine normal traffic rates, and normal energy consumption, and can be used to determine anomalies and attacks.

8.4.2.5 Traditional attacks

Scenario: Similarly to traditional network attacks, the case of traditional software attacks implies the exploitation of traditional software flaws. Here, we find traditional attacks including virus infection exploits, worms, malicious script executions, etc. In this context we observe that the smarthome installation, especially the gateway and the user devices that may be communicating with the gateway (eg smartphone, laptop, Smart TV), are built on traditional software systems that require periodic maintenance and security updates. In such a diverse technological ecosystem, it is easily conceivable that in a particular IoT installation there may be out of date software with vulnerabilities that may be exploited by malicious actors. The attack showcases the execution of typical software exploits in the attempt to gain access to sensitive (IoT-specific) information. In this case the attacker may exploit the vulnerabilities of services installed on the gateway or on one of the user devices (eg laptop, smartphone) to obtain sensitive information.

Demonstration/validation methodology: Various attacks will be mounted against the gateway by emulating software flaws via network traffic that contains malware traces. The following tools will be used for this purpose:

- Packet Capture (pcap) files containing malware traces: malware traces shall be used from public sources (https://zeltser.com/malware-sample-sources/ accessed on 12 April 2022).
- `tcpreplay`, `tcprewrite`: these two tools are used to edit and to replay the malware pcap files against the gateway.

8.4.2.6 Compromised software attacks

Scenario: It is conceivable that cybersecurity solution deployed in the smarthome may not function properly, or it may get compromised by an attacker (as a consequence of software flaws and attack exploits). Consequently, the gateway will exhibit a different behaviour, like stopping sending anomaly alerts, or simply, they may be stopped. Therefore, the user may be notified that the software is not functioning properly, and that actions should be taken in order to ensure that attacks are detected. By exploiting software flaws an attacker may succeed in changing one of smarthome software modules. The attacker may succeed in injecting new and malicious code that alters the solution's behaviour.

Demonstration/validation methodology: The attack demonstrates the awareness of the DRA to the malfunctioning/compromise of its own modules. This entails that each module detects the malfunctioning of the other modules it depends on. Obviously, the mechanisms implemented for detecting the changes in the behaviour of the own software need to be as lightweight as possible and must build on simple decision algorithms and computations. Furthermore, given the overhead of more complex computations and the limited hardware resources, only the most predictive behaviours shall be taken into account. In particular, the following expected operational behaviour shall be monitored:

- Network analysis: the periodic generation of pcap files.
- Anomaly reporting: the periodic alerts issued.
- Configuration manager: the status of each OS process; the consumption of CPU and memory for each of gateway OS processes.

8.4.2.7 Command injection

Scenario: Considering the software-oriented construction of IoT system's inbuilt logic, it is reasonable to presume that, by exploiting software flaws, malicious software may be hosted by one of the IoT system's devices (eg smartphone, gateway). Consequently, the malicious software may open legitimate communication channels for injecting commands to IoT devices. By exploiting software flaws an attacker may succeed in running new software alongside cybersecurity solution enabling sending of the forged commands to IoT devices sensors, and communicate this data with an outside 'command and control' server.

Demonstration/validation methodology: The undertaken procedure includes additional software scripts hosted on the gateway and on an external device (eg smart phone), from where the test, through legitimate commands, triggers relays, and request the status information on sensor devices.

8.4.2.8 Mechanical exhaustion

Scenario: Contrarily to the mechanical exhaustion attack included in the list of physical attacks, the software-oriented mechanical exhaustion presumes a malfunctioning software

module and/or a new software that sends, at a high rate, ON/OFF commands to switching devices (eg relays). Considering that, inadvertently, a device has an upper limit in terms of mechanical switching, once that upper limit is exceeded, devices may malfunction, and they can be physically damaged.

Demonstration/validation methodology: The undertaken procedure shall include additional software scripts hosted on the gateway and on an external device (eg smart phone), from where the tests, through legitimate commands, shall trigger relays at a higher rate than expected from the normal operation.

8.4.2.9 Sleep deprivation

Scenario: It is commonly known that, in order to save energy, battery-powered IoT devices reduce their energy consumption by entering into 'sleep mode'. Once an event is detected, the devices then resume their normal operation and forward the alert(s) to their associated controllers (eg the IoT gateway). However, the normal behaviour of IoT devices can be exploited by malicious actors in order to prevent devices from activating their energy saving mode (ie the 'sleep mode'). An attacker can use software solutions to remotely launch this type of attack via legitimate commands. The commands may not necessarily need to cause the mechanical triggering of devices, or a change of state. It would suffice to periodically request status information, or to send a specifically forged packet that would prevent the device from entering its energy saving mode.

Demonstration/validation methodology: The undertaken procedure shall include additional software scripts hosted on the gateway and on an external device (eg smartphone). From these locations, the developed software, shall repeatedly request the status of the devices, which, in most cases, should prevent the device to enter 'sleep mode'.

8.5 Technical Actions

8.5.1 Physical attacks actions

In the case of the physical attacks the supported actions are detailed in Table 8.4. Here, we observe that, in case of physical damage, the GHOST solution can mainly suggest some manual actions that the user should perform. Accordingly, the user should verify the physical integrity, he/she should verify the battery level. However, other actions may also be taken, which may be automatable. To this end, 'sandboxing' may be used to isolate the device and its traffic from other devices. In this respect, 'one-way sandboxing' refers to application-level commands not routed to the IoT software's other modules (this is only applicable when there are also automated scripts available that logically link devices between them). A more restrictive scenario is the 'two-way sandboxing', in which case the application-level commands (or notifications) will be blocked from being sent out to the specific device. Lastly, all traffic may be blocked to/from that particular device [172].

Next, in the case of device injection (ie a new device is detected), all actions may be automated. The device may be simply be allowed to be added. However, the device may be sandboxed, where the meaning of the 'sandboxing' term is the one given above. Lastly, the device may be temporarily, or permanently blocked.

In the last case of physical attack, we find the mechanical exhaustion. For this type of attack, the action may be simply permitted, or sandboxing can be used to limit its effects. Lastly, the device can be simply blocked (ie by dropping all traffic related to that particular device's flows).

Table 8.4 Technical Actions on Physical Attacks.

ID	Attack name	Action	Action ID
P1	Damage	Verify physical integrity	T1
		Verify battery	T2
		One-way sandboxing	T3
		Two-way sandboxing	T4
		Block device (temp/perm)	T6,T7
P2	Device injection	Permit	T5
		One-way sandboxing	T3
		Two-way sandboxing	T4
		Block device (temp/perm)	T6,T7
P3	Mechanical exhaustion	Permit	T5
		One-way sandboxing	T3
		Two-way sandboxing	T4
		Block device (temp/perm)	T6,T7

8.5.2 Network attacks actions

In the case of the network level attacks the supported actions are detailed in Table 8.5. Here, we observe that all actions can be automated by the GHOST solution. In case of traditional network attacks, packets may be dropped for a particular flow, or, more dramatic measures may be taken in case of higher risk levels. To this end, all flows associated to a particular source may be dropped. The actions may be taken temporarily, or permanently, depending on the attack impact and its execution in time.

Next, in the case of device impersonation, one-way sandboxing is an effective measure in order to keep monitoring the flows associated to a device, but ensuring that the application-level data does not reach other devices. Conversely, two-way sandboxing may be used to continue the monitoring of flows, but to ensure that messages are not forwarded to the possibly malicious

device. Lastly, in this case as well the device may be blocked, which means that all traffic to/from that particular device is dropped.

In terms of the side-channel attack, since this type of attack is performed at design-time, in order to analyse and to infer the information that may be leaked from the GHOST solution, no actions are suggested.

Lastly, in the case of the battery attacks, the same type of actions are defined as in the case of the traditional network attacks: drop the packets for a specific flow or for a specific source. Both actions are automatable.

Table 8.5 Technical Actions on Network Attacks

ID	Attack name	Action	Action ID
N1	Traditional	Drop packets for flow (temp/perm)	T8,T9
		Drop packets for source (temp/perm)	T10,T11
N2	Device impersonation	One-way sandboxing (temp/perm)	T3
		Two-way sandboxing (temp/perm)	T4
		Block device (temp/perm)	T6,T7
N3	Side-channel	– (design-time test only)	–
N4	Battery attacks	Drop packets for flow (temp/perm)	T8,T9
		Drop packets for source (temp/perm)	T10,T11

8.5.3 Software attacks actions

In the case of the network level attacks the supported actions are detailed in Table 8.6. Here, we observe that all actions can be automated by the GHOST solution. As expected, in the case of traditional malware attacks, packets associated to flows, or to a particular source may be dropped temporarily or permanently. In the case of software compromise, the particular module/the complete GHOST solution may be restarted. In the same scenario, a module may be temporarily or permanently disable, and update requests may be issued.

Next, in the case of command injection, mechanical exhaustion and sleep deprivation attacks, the GHOST solution may, once again, drop the packets associated to the particular flows/sources.

Table 8.6 Technical Actions on Software Attacks.

ID	Attack Name	Action	Action ID
S1	Traditional	Drop packets for flow (temp/perm)	T8,T9
		Drop packets for source (temp/perm)	T10,T11
S2	Software compromise	Restart module/ GHOST	T12,T13
		Disable module (temp/perm)	T14,T15
		Send update request	T16
S3	Command injection	Drop packets for flow (temp/perm)	T8,T9
		Drop packets for source (temp/perm)	T10,T11
S4	Mechanical exhaustion	Drop packets for flow (temp/perm)	T8,T9
		Drop packets for source (temp/perm)	T10,T11
S5	Sleep deprivation	Drop packets for flow (temp/perm)	T8,T9
		Drop packets for source (temp/perm)	T10,T11

8.6 Decision Automation in the Risk Assessment

The DRA at its core is relying on three key innovation areas:

- real-time risk assessment,
- decision automation and
- security usability.

These advances are fused together within the implementation of DRA [23], which relies on the probabilistic traffic profiling, the intelligence on the threat and vulnerability likelihood and associated risk's severity, which permits to identify network anomalies and coordinate the selection of the appropriate mitigation action.

The DRA incorporates a variety of analytic algorithms, called *Analysers*, each responsible for distinct features listed below.

- BA: the main purpose of this analyser is to detect any deviation from the device normal behaviour
- PC: a set of defined rules aiming to detect the presence of the user sensitive data within the traffic flow
- BR: responsible for verifying the destination maliciousness from personalised settings and common shared intelligence
- AP: alert extraction analytics from external input for anomalies detection

The resulting scores are incorporated together for predictive risk forecasting. Triggered by so called *Risk receptors*, a current probability of identified risk is estimated in link with the user defined risk levels tolerance.

Table 8.7 contains the identified attacks and the associated analytic algorithms used for the risk evaluation.

Table 8.7 Attack & analysers mapping.

Attack ID	**Analyser**	**Rationale**
P1	BA	Absence or change in behaviour of the communication
P2	BA & AP	No behaviour profile present and alert propagation on non-registered device
P3	BA & AP	Absence or change in behaviour of the communication and alert propagation on anomalous traffic
N1	AP	Alert propagation on threat detection
N2	BA & AP	Absence or change in behaviour of the communication and alert propagation on anomalous traffic
N3	BA, PC & AP	Absence or change in behaviour of the communication, presence of sensitive data and alert propagation on anomalous traffic
N4	BA & AP	Absence or change in behaviour of the communication and alert propagation on anomalous traffic
S1	AP	Alert propagation on threat detection
S2	BA, PC & BR	Absence or change in behaviour of the communication, presence of sensitive data and attempted communication with malicious destination
S3	BA & BR	Absence or change in behaviour of the communication and attempted communication with malicious destination
S4	BA & AP	Absence or change in behaviour of the communication and alert propagation on anomalous traffic
S5	BA & AP	Absence or change in behaviour of the communication and alert propagation on anomalous traffic

We have identified three different communication profiles allowing the end-user to customise the automatic decision making when exceeding risk expectations as per user preferences. The choice between three awareness modes is proposed through the CFG interface. These modes are recapped in Table 8.8.

Table 8.8 Awareness preferences.

Name	Description	Informed on any decision	Allow risk controlled automation
Raise Awareness	Stay informed on any decisions that GHOST made by displaying a corresponding notification. The GHOST system will automatically block any suspicious communication as soon as maximum risk level is exceeded.	Yes	Yes
Enforced Awareness	Stay informed on any decisions that GHOST made by displaying a corresponding notification. GHOST will not perform any automatic decisions when the maximum risk level is exceeded. You will be constantly prompted to review suggested actions and make decisions yourself.	Yes	No
Problem Awareness	Stay informed on decisions that GHOST made only when exceeding maximum risk level by displaying a corresponding notification. GHOST will automatically block any suspicious communication as soon as maximum risk level is exceeded.	No	Yes

Finally, we have derived several decision scenarios, which served as a base for the conceptualisation of the Decision Tree, presented in Section 8.7.

- Missing communication (Absence): Absence of the device communication in relation to its normal behaviour
- Whitelisting: New communication was neither blacklisted, neither whitelisted before by the user or the GHOST itself
- Data type (Privacy): Private data leakage is tracked and the user is informed on the violation of the policies defined
- Frequency: Suspicious situation is detected in terms of the communication frequency: too often or not frequently enough
- Time (Timing): Pattern of the communication stays within the safe profile derived, but the actual timing is suspicious

- Blacklisting: Known illegitimate communication taking place, but it is generating activity on the internal network despite being blocked for further external propagation

The sparse automation matrix is presented in Table 8.9, mapping the above defined awareness modes, decision scenarios and automation feasibility dictated by risk acceptance agreement.

Table 8.9 Sparse decision automation matrix.

	Absence	**Whitelisting**	**Privacy**	**Frequency**	**Timing**	**Blacklisting**
Raise	-	Automatable	Automatable	Automatable	Automatable	Automatable
Enforced	-	-	-	-	-	Automatable
Problem	-	Automatable	Automatable	Automatable	Automatable	Automatable

8.7 Decision Tree Conceptualisation

The scenarios from the initial scenarios definition were further transformed into a decision tree, focusing on the essential interaction with the end-user and aiming at the enlightening comprehension and in-depth involvement of their feedback through user studies on the end-users' mental models. The overall view of the decision tree implemented for the first prototype is depicted in Figure 8.2. The distinct decision branches have been grouped by colour within the tree and are described in the section hereafter.

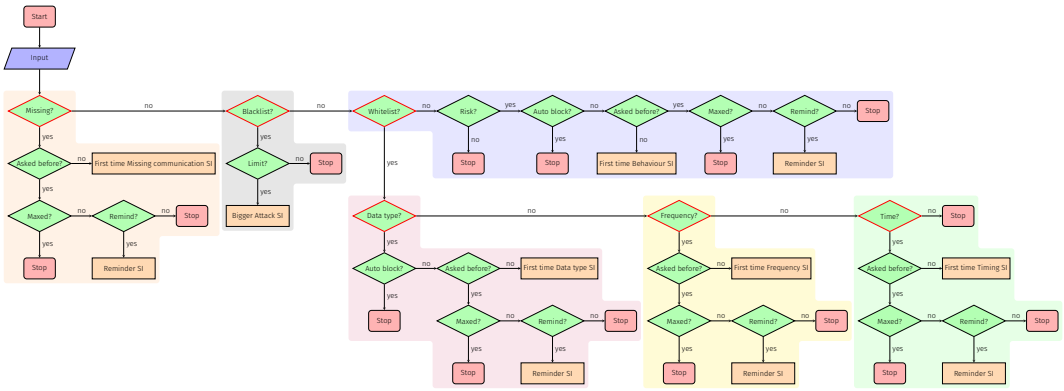


Figure 8.2 Decision tree conceptualisation, with coloured highlighting of the main branches.

8.7.1 Decision branches

8.7.1.1 Missing communication

The first group of the interfaces is covering the case when the DRA detects the absence of the device communication in relation to its normal behaviour. The close-up of the decision tree is depicted in Figure 8.3. This module includes a standard SI interface with a possible reminder

on the pending user's decision allowing her to ignore the situation, until automated decision is made to treat this case as safe.

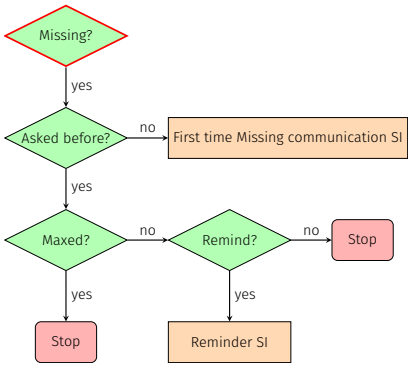


Figure 8.3 Missing communication.

8.7.1.2 Whitelisting

The second group of the interfaces refers to the case when a new communication was neither blacklisted, neither whitelisted before by the user or GHOST itself. The DRA will make a risk-based decision based on the likelihood of the maliciousness of the destination party, the actual profile of the communicating device and also user risk acceptability. This user interface is also composed of an initial SI interface and a reminder on a pending decision in case when automatic decision can not be made (according to the user preference).

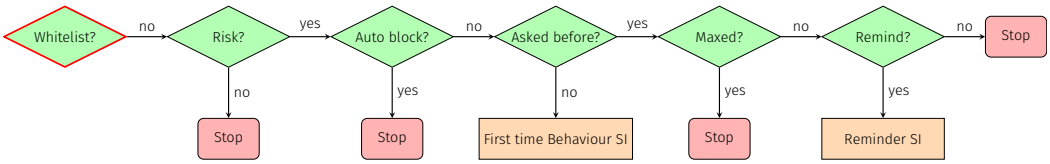


Figure 8.4 Whitelisting.

8.7.1.3 Data type

The third type of the SI interface is coming into play in the light of the privacy controls. Depending on the end-user configuration, certain private data leakage will be tracked and the user will be informed on the violation of the policies defined. This interface is also composed of a first time SI and a consecutive reminder if necessary.

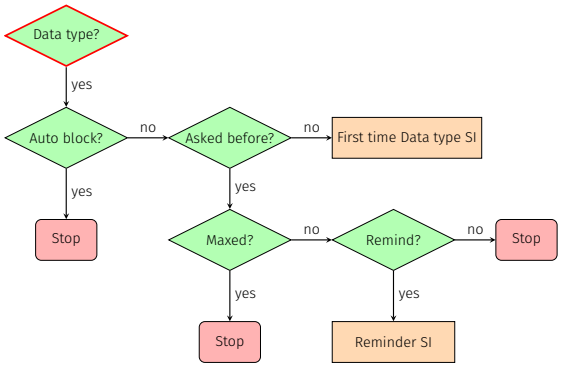


Figure 8.5 Data type.

8.7.1.4 Frequency

The fourth type of the user interface is focused on making decisions when suspicious situation is detected in terms of the communication frequency: too often or not frequently enough. Once again, this interface is composed of first time SI and the reminder for the pending decision. Once the limit is overpassed, the notification will be completely discarded.

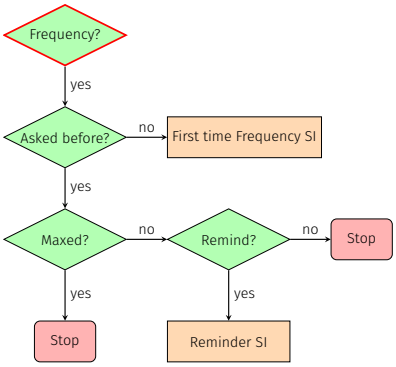


Figure 8.6 Frequency

8.7.1.5 Time

The fifth group of SI interfaces is covering the case when the pattern of the communication stays within the safe profile derived, but the actual timing is suspicious. In this case a notification will be sent to the user to confirm if this timing is appropriate.

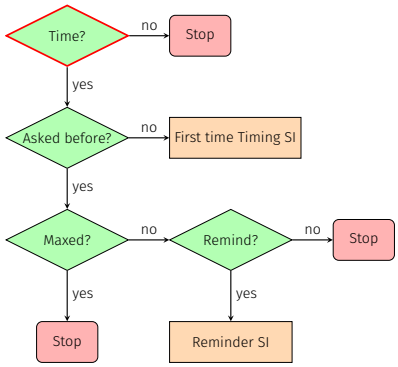


Figure 8.7 Time

8.7.1.6 Blacklisting

The last group of the interfaces is preserved for the situations with mitigation propositions in cases where it is known that communication is illegitimate, but it is generating activity on the internal network despite being blocked for further external propagation. In such case we suspect an IoT device to be part of the bigger attack like DDoS, and possibly affecting the performance of the actual device.

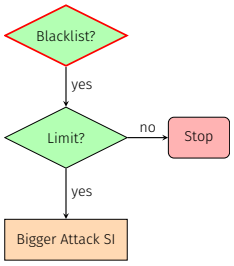


Figure 8.8 Blacklisting

8.7.2 Configuring DRA

The main focus of this type of the interface is given to the effortless and usable design of the configuration setup process and further settings review and fine-tuning of applied configuration policies, called CFG. This design and implementation was performed in four iteration cycles, each being stipulated by the feedback received from the end-users and fed back to the decision tree conceptualisation.

1. The development approach was based on the requirements derived from literature research and results from the first set of user studies. Furthermore, the categorisation of the initial set of the navigation pages was developed, outlined in Table 8.10.

2. For the second iteration, the CFG interfaces were improved in terms of their usability with a mobile device form factor and a new menu was developed for easier access to the

different CFG sections. Furthermore, the colour theme was updated to match with the official project theme.

3. The third iteration of the Configuration was refined based on the results from the second set of user studies and the derived requirements. The updated specifications for the categorisation of the initial set of the navigation pages developed are shown in Table 8.11.

4. 4_{th} cycle was mostly based on the perception of the risk, requiring clarification on the associated impact. The final look of the CFG interfaces is depicted in Figure 8.9, Figure 8.10 and Figure 8.11.

Table 8.10 Configuration interface - Initial Setup.

Category	**Specification**
Welcome	Initial welcome screen
User registration	Initial user registration and authentication setup
Dedicated device registration	User's interfacing device registration
Smarthome environment	Configuration of the smarthome, aiming to provide settings for: - Adding and removing IoT devises - Configuration of "unknown" devices - Naming and custom identification of IoT devices
Mode selection	Three configuration modes are envisioned to target different user profiles: - Manual: based on predefined settings - Assistant: step by step configuration - Delegation: configuration by trusted 3rd party
Step by step configuration	Detailed configuration of the GHOST main features: - Blocking rules: Customisation for black/ whitelisting the communication destination parties - Acceptable risk level: Definition of the permitted risk levels for security and privacy settings, defining a threshold for DRA automated decisions - Privacy monitor: Selection of private data categories for tracking - Awareness requirements: Configuration of the desired intervention and involvement level into decision making

Table 8.11 Configuration interface - Updates to the Initial Setup.

Category	Specification
Mode selection	The three configuration modes are updated to differentiate between different target user profiles: - Manual: based on predefined settings - Delegation: configuration by trusted 3rd party - Advanced: for expert level configurations (eg whitelist/ blacklist)
Step by step configuration	Detailed configuration of the GHOST main features: - Security preferences: Definition of the accepted risk levels for security after which all communications will be blocked - Privacy preferences: Selection of private data categories not to be transferred to the Internet - Notification preferences: Configuration of the corresponding preference for security and privacy notifications

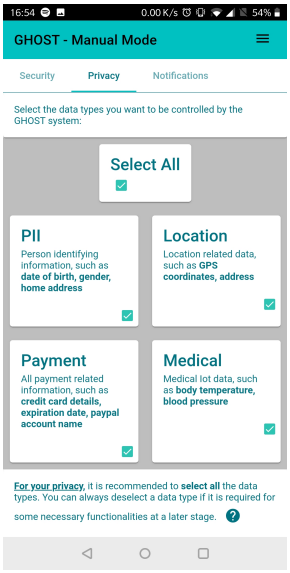


Figure 8.9 Privacy Preferences

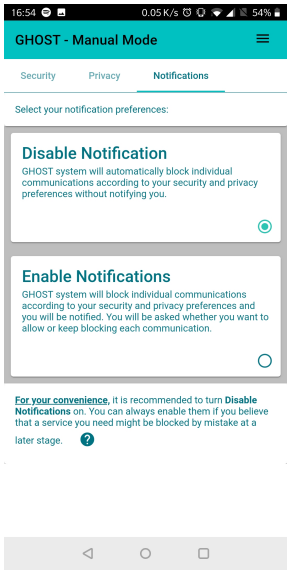


Figure 8.10 Notification Preferences

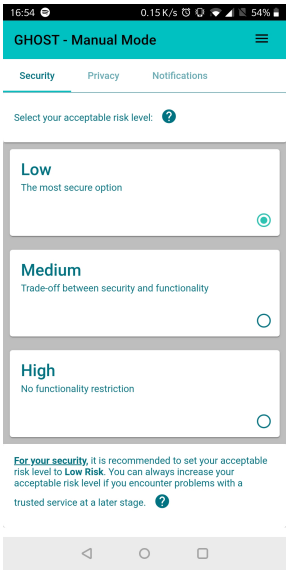


Figure 8.11 Security Preferences

8.7.3 Monitoring Automated DRA

SI component is aiming to develop type of user interfaces to serve for user friendly visualisation of the risk tracking and risk evaluation results and will enable the end-user to make informative decisions. The overall technological selection and implementation refinement process is closely aligned with the CFG interfaces. Five iteration cycles were implemented, each being stipulated by the feedback received from the end-users and fed back to the decision tree conceptualisation.

1. The first iteration of the security intervention interface was developed based on the requirements from initial literature review. Furthermore, the initial listing of possible interactions with the GHOST was created. This outline is summarised in Table 8.12.

2. To provide a more fluent and unified experience, the SI notifications and related frontend were included in the same Angular web application package as the CFG interface.

3. The second prototype was developed based on the input from the user trials, in particular, by attempting to provide a more understandable and actionable input for user's decisions.

4. The existing SI was further fine-tuned in preparation for the 3rd trials, where attack simulations to be performed. For this purpose, not only additional menu items were added, but also generic system flow was amended to reassure the end-user and provide additional information on the ongoing evaluation and feedback gathering.

5. The final iteration was mostly concerned with the proper naming of Awareness preferences, previously outlined in Table 8.8. The final look of the SI interfaces is depicted in Figure 8.12 to 8.14.

Table 8.12 Security Intervention - Interaction Identification.

Category	**Specification**
Mismatching device behaviour	- Absence of communication or its decrease - Increase of communication or frequency change - Extra 'steps' in communication
Communication with blocked src/dst	Attempt to initiate communication with blocked party
'New' src/dst (unknown)	- Fetching contextual information on the new party, eg 'whois' info - Consecutive configuration update
Device parameters anomalies	An overview of the device profile and its activity (eg battery power)
Payload related	- Security related: clear text password of credit card - Privacy related: PII data (date of birth, home address) - Secured transmission security check - Masking of data in encrypted communication channels
Mitigation action	Hypothesis presentation with possible suggestion on the mitigation recommendation
Current status	Display of the risk level current status in relation to defined accepted level for security and privacy risks
Predicted status	A risk level estimation representation after the evaluation of the current communication associated risk
Impacts (text)	Possible impact score value
Recommendations	Mitigation action to support the decrease of the raised risk level
History of risk assessments	Interface link to historical data visualisation
History on suspicious activity	Interface link to historical data visualisation

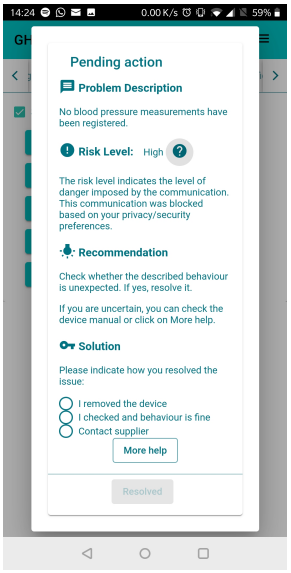


Figure 8.12 Pending Action

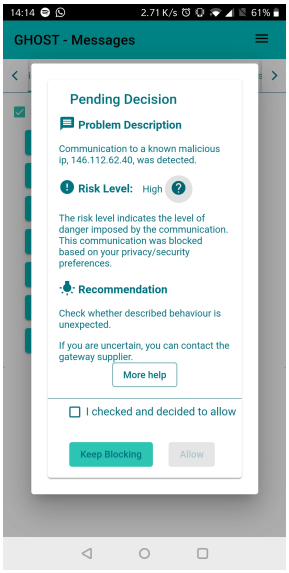


Figure 8.13 Pending Decision

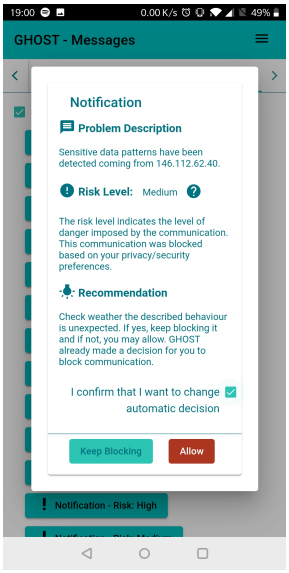


Figure 8.14 Notification on Automated Decision

8.8 Discussion and Conclusions

This section presents a summary of our findings through a dedicated analysis on each research question.

RQ1 *Limitations on automation:* We have started our research with the identification of possible technical actions to mitigate the exposure to the smarthome threats. As outlined in Section 8.5, for each category of the attacks we derived a list of technical actions further linked to the automation feasibility as presented in Table 8.2. As it can be observed, only actions of physical nature (like physical verification of the device integrity or battery state) were not possible to be automated. However, we were able to address it through the inclusion of the actions into mitigation advisory, enabling the guidance to the user.

RQ2 *Usable actions translation:* Guided by the user studies and continuous feedback collection, we were able to derive a short set of the usable actions to be presented as part of the final UI. The process we have followed transformed significantly the initial interfaces with minimum information into detailed and fine-tuned textual descriptions to have a maximum user engagement. Limiting the number of usable actions had a positive effect on the usability aspects of the final interfaces, which was showcased during real-life pilots deployments with an average System Usability Scale (SUS) score [173] increase throughout the project's lifetime.

RQ3 *Perception linkability to engagement:* The differences and impact of personal risk perception was a key challenge that we have addressed in this research. For this purpose

we have developed a Decision Tree concept and three types of awareness preferences (outlined in Section 8.6 and Section 8.7 respectively).

In this work we have presented a methodological framework applied in the design and implementation of usable interfaces for the DRA-enabled smarthome environment. Guided by the definition of the threat vectors, we identify a set of technical actions, suitable for the threat prevention. Those are further translated into usable actions, meaning understandable digital decisions which an end-user of the smarthome with different technological knowledge can make. This flow, from threat vector, to technical actions, then translated into usable actions, sets the basis for the decision tree concept, which further translates the usable actions, in an iterative manner, into two types of UI: CFG and SI.

Part V

Conclusions

Chapter 9

Discussion

Chapter Contents

RA is a complex process comprising a multitude of influencing factors. The nature and origin of these factors range from the availability of historical data and knowledge representation to the adaptation for dynamic analysis and the inclusion of the human perception and behaviour aspects. Each of these influencing factors posed a challenge that required its own methodology, analysis and a bespoke integration process into the DRA process to develop a scientifically sound framework supported with the real-life field experiments to validate the final solution. The scientific artefacts produced as a result of this thesis work are the following:

- Reference Architecture (Article I & Article II)
- IoT Stack Model (Article IV)
- Risk Level Calculation Model (Article III)
- Expert Values for the Risk Weight Assignment (Article IV)
- DRA Prototype (Article IV)
- Automation of DRA (Article IV)
- User-centric Mapping Methodology from Attack to Action (Article VI)
- Reputation Scoring for Decision Decentralisation (Article V)

This chapter goes back to identified RQs and provides explicit analytical answers for each of them. Finally, the findings are also categorised in the foreseen application domains.

9.1 Generic Ontology

RQ1: Can a generic ontology be developed to capture complex relationship between heterogeneous IoT properties to encapsulate vulnerabilities, attack attribution, impact evaluation, and mitigation strategies?

Answer: This question was explored in Chapter 3, Chapter 5 and Chapter 6. We have started with the identification of the threat landscape applicable to the smarthome environments in particular. This permitted to compile a comprehensive knowledge on the associated risks, caused by the threats' exploitation. For each compiled risk we performed the analysis with an expert group to build a relationship map between threats, risks, mitigation actions and possible cascading effects. The resulting work was a conceptualisation of the IoT Stack model, allowing to embed at a fine-grained level all heterogeneous properties of the objects composing smarthome environments. The visual representation of this model is depicted in Figure 9.1. This model focused on the representation of the IoT properties for vulnerabilities and attack attribution encapsulation. In parallel, we have also conceptualised a theoretical reference architecture on the full workflow of data propagation in the smarthome environment to feed

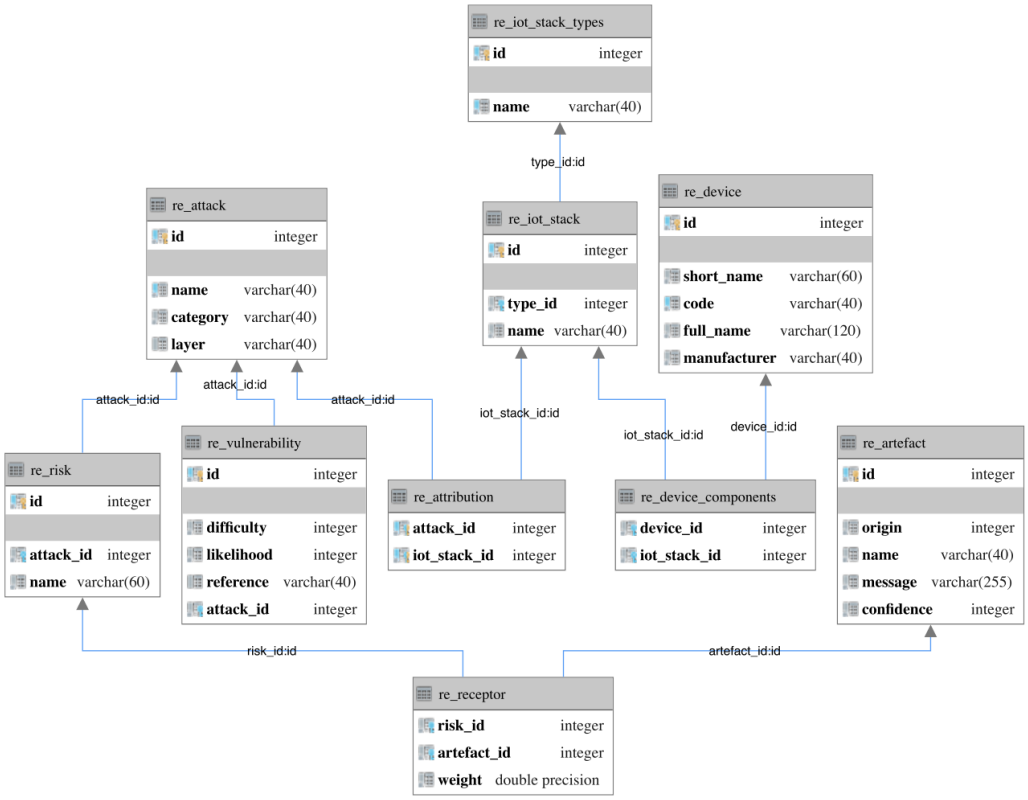


Figure 9.1 IoT Stack.

the RAE and visualise in a user-friendly manner the risks analysis. This reference architecture was then further improved during the setup of the GHOST project, which resulted in the works described in Chapter 3 and Chapter 4. This permitted to expand our theoretical model into a practical implementation with four distinct layers: network analysis for data capture, device profiling for anomaly detection, risk assessment for threat correlation and visualisation for decision-making. The final version of this architecture is shown in Figure 9.2. The implementation of the established reference architecture enabled the improvement of the DRA model by including association of the IoT Stack with the impact evaluation and selection of the mitigation strategies associated to the risks.

The defined architecture served as a reference point for the RAE input definition, which evolved over the project lifetime, shifting the original design towards generalised processing of the anomaly input data, which is described in detail in Chapter 6. Through the validation experiments we have demonstrated that the final IoT Stack ontology encapsulates all required relationships to enable correlation from threat detection to mitigation strategy proposition. Furthermore, the developed model permits the generalisation of the concept being applied to any IoT domain. As a continuation of this work, we are currently integrating RAE in the automotive domain. Such integration is mostly limited by the actual threat vector data to be fed into the ontology developed (see Section 1.5), such as applicable vulnerabilities and possible mitigations to define relevant risks.

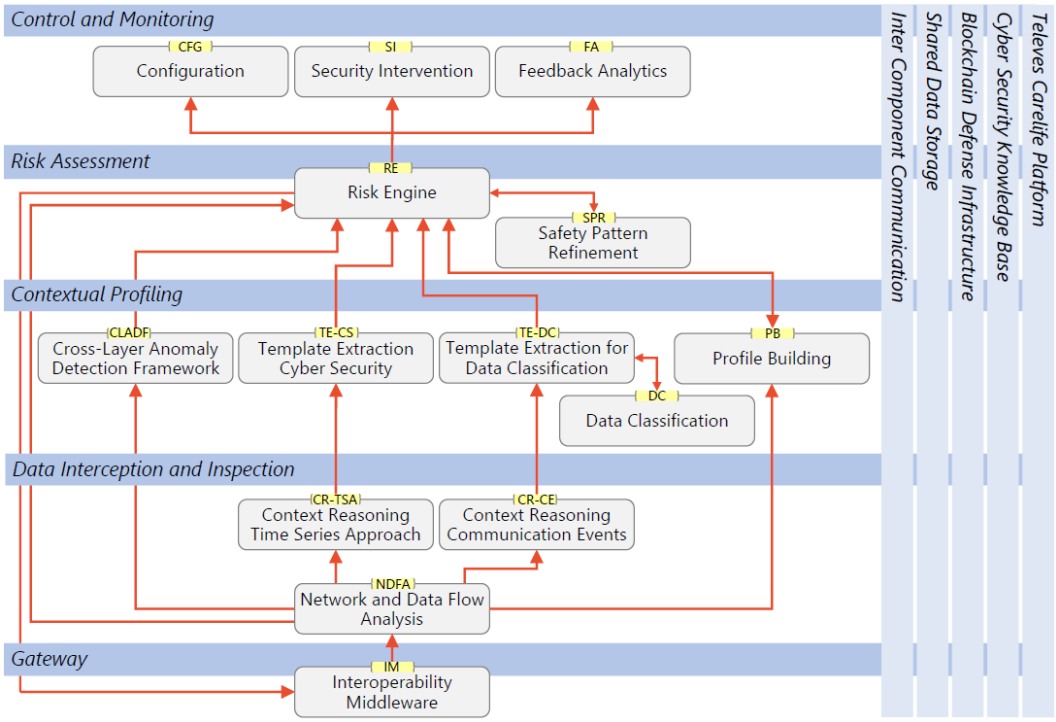


Figure 9.2 Final Reference Architecture.

9.2 Risk Scoring

RQ2: Can a unique risk scoring be developed to eliminate context dependency? How the initial setting of the expert values for the risk assessment offer a valid approach and whether these values are generally applicable in a standard installation?

Answer: This question was explored in Chapter 5 and Chapter 6. We have started with the development of the threat evaluation scenarios to understand how the decisions affect the risk. With those scenarios we have also attempted to quantify the risk data for the RA process. As described in details in Chapter 5, quantification is a fundamental problem as in general it relies on the historical data availability. With the developed Positive-to-Negative model for RL calculation, we have concluded that comparing one IoT object to another is a challenge on its own, even if we would have all historical data available. In a way, the results of this work set us back to reflect on the envisioned approach, to attempt the elimination of the context dependency while dealing with the risks' quantification.

This is how the idea of the initial setting of the expert values and their further readjustment at the runtime took up. First, having constraints on how DRA can obtain and process the input data, flowing from the network observation and analysis, we have developed the risk data propagation model, inspired by the Immune Theory, outlined in details in Chapter 6. Such methodological framework enabled the possibility of incorporating the expert knowledge directly in the RM. On a similar note, we generalised the input data by originating the concept

of *Reporting strategies* and associated *Weight Adjustment Strategies*. Advancing on these concepts in the implementation of the RAE, granted the possibility to verify our theoretical assumption in the real-life deployment environments. The experiments demonstrated that the same risk scoring values are equally effective in different smarthome environments, with no need to readjust initial expert weights. This was possible through the integration of the dynamic weights adaptation by each individual instance and the collective decision-making feedback, supplementing the RAE. Finally, the approach of the expert values for the initial setting demonstrated the context independence for a standard installation.

9.3 Decision Automation

RQ3: What are the limitations on the automated decision-making for risk assessment in dynamic environments, such as smarthomes, where deployed IoT devices constantly evolve (get replaced, updated and moved)?

Answer: This question was explored in Chapter 6, Chapter 5 and Chapter 8. The answer to this question is very much linked with the RQ2. As explained in previous section, the initial result indicated a significant limitation of the automated decision-making for the RA process in both, static and dynamic environments – comparison of the same risk between different IoT objects. To make it more explicit, lets imagine the following:

- Risk1: Personal data exposure to a 3rd party.
- Object1: Smart Kettle with Bluetooth Low Energy (BLE) connectivity, having access to the username
- Object2: Smartwatch with Wi-Fi connectivity, having access to the age and address

How in this setting the Risk1 for Object1 can be compared to the Risk1 for Object2? IoT Stack enables us to compare the risk associated with the BLE in general to be lower than Wi-Fi, and access to age and address to be of higher impact than username. However, combining both properties already becomes problematic in a very simple setting where both objects are static. On top of this comparison, we have to also include the impact projection in case we allow or do not allow activity in question on the network. It all brings us to the conclusion that automated decision-making is bound to the granularity of the data available in the IoT Stack, as also observed in Chapter 6.

Another limitation is the risk perception of the end-user. This is why the initial works of the risk levels determination, described in Chapter 5, was redirected into different approach to include the end-user feedback and adjustments on automated decisions through the SI and CFG interfaces setting, outlined and demonstrated in Chapter 8. It was observed during the real-life trials that the perception of the risk often led to the compromise of the privacy and security over the functionality of the smarthome devices. End-users had a general preference

to set the options of the highest automation possible on decision-making with a minimal compromising of the operation of the smarthome services. Having a functional smarthome is of higher priority than being exposed to the possible consequences of the threats.

Finally, the effectiveness of the RA automation depends on the quality and variety of the underlying reporting data fed into RAE. As observed in the simulation experiments (Chapter 6), different Reporting Strategies were providing different automation levels on decision-making and mitigation. While the automation of the decision-making is a binary setting (possible or not possible), the mitigation advisory automation is highly dependent on the granularity of the underlying reports. The technical limitation here relies on the lack of the attribution of the attack to a specific device or, at least, the interface. Even if certain contextual data can be deduced from the IoT Stack (eg out of 20 devices only one is with BLE connectivity and, hence, anomaly report containing only interface data can be attributed to the only device present on the network with this type of connectivity), its automation is rather limited and constraint with boundary cases. As discussed in Section 6.6, the automation of the reasonable mitigation execution was not possible in such settings, as it was enclosing the information relevant for the whole environment.

9.4 Dynamic Assessment

RQ4: To what extent DRA can be performed in real-time considering evolving multidimensional situational risks, such as human behaviour change, emerging attack vectors and a dynamic IoT ecosystem?

Answer: This question was explored in Chapter 7 and Chapter 8. Several challenges were encountered and addressed successfully in due course of this work. Most important one, the human factor, already related as one of the identified limitations of the DRA, is generic audience interest and awareness in cybersecurity and privacy. The RA process tightly couples both notions and relates them back to the user, as its central element. We have attempted to design the decision-making and tracking visualisation following a user-centric approach, as described in detail in Chapter 8. It is not only the perception of the cyber risks that plays a crucial role in the usability design, but also the actual flow of the interfaces, which we developed through a Decision Tree visualisation. The design and implementation of this flow assisted in the expansion of the existing knowledge on the human behaviour change influence on the RA in general. Guided by the project's operational setting, we were able to iteratively improve the user experience for a greater system acceptance. For this purpose we have included the end-users feedback loop into automated decision-making and enabled different system perspectives for the risk monitoring. We believe we have initiated the future research direction, by merging the heavy technically-driven cyber risk domain with security, privacy, usability and user psychology.

Another challenge on the constantly emerging attack vectors and the dynamic nature of the IoT system in general, was the inclusion of continuous provision of external knowledge for improving the accuracy of the decision-making, and the distributed collective resilience of the RAE to eliminate internal malicious actors. For this purpose we have integrated in the core of the automation functionality of RA the external intelligence data feed, presented in details in Chapter 7. To address the potentially intentional malicious behaviour of involved actors, a scheme for the reputation scoring was also implemented. The simulation experiments demonstrated the efficiency of such an approach to deal with the encountered challenge.

Finally, real-time execution of DRA will remain a challenge and we continue our efforts on improving not only its internal components, but overall solution as a framework to be deployed in various IoT environments. The main constraint is based on the technological limitation of analysing traffic data in real-time, detecting potential threats through behaviour deviation, and only then performing the actual RA. In the framework of this thesis we have demonstrated near-to-real-time performance of the developed system.

9.5 Future Application Domains

ICT systems are steadily expanding their presence in all facets of industry and consumer lives facilitating many of the menial or high volume tasks and enable citizens, consumers and manufacturers to interact with the digital world. As these systems are integrated into the infrastructure and our lives, malicious enterprises and hackers have greater capacity to exploit security and privacy vulnerabilities to disrupt cities and networks[1], control critical infrastructures[2], invade our privacy[3] or otherwise misuse our data[4]. The digital assets and ICT systems used by the majority of enterprises are deployed with minimal security and privacy assessment on individual composing elements or the resulting ecosystem. Currently, the security of ICT systems must be assessed on an individual basis, and no single tool can be used to automate this process. There is a need for standardisation of verification methodologies of all compound elements of any system to facilitate the automation of cybersecurity assessment in a trusted manner. This should include the full life-cycle of digital assets, from the manufacturing of individual components and their assembling into a whole unit up to operational deployment in the network ecosystem. Recognition and wide adoption of such certification will have a significant impact on the reliable security and trust management of the entire IoT ecosystems.

We anticipate DRA will have a practical impact on a several application domains, such as automotive driving, critical infrastructures, smart factories. The field of the cybersecurity in general will benefit from the adaptation of several findings from this work, namely the IoT

[1] https://www.wired.com/story/sensor-hubs-smart-cities-vulnerabilities-hacks/
[2] https://www.herjavecgroup.com/wp-content/uploads/2019/12/Healthcare-Cybersecurity-Report-2020.pdf
[3] https://www.forbes.com/sites/thomasbrewster/2020/05/19/easyjet-hacked-9-million-customers-and-2000-credit-cards-hit/
[4] https://blog.emsisoft.com/en/36303/ransomware-statistics-for-2020-q1-report/

Stack and DRA framework. Both, in conjunction with the usability aspects, can be deployed as an integral part of the Intrusion Prevention Systems (IPS) in any digital ecosystem.

Chapter 10

Closure

In this work, a complete framework on the DRA has been presented. This thesis has begun with the definition of the theoretical foundation for the establishment of the RA and its application in IoT environments. We have demonstrated the evolution of the reference architecture from conceptualisation to implementation in one of the IoT application domains – smarthomes. Guided by the constantly evolving user and functional requirements, we have designed a complete workflow from data capture and network analysis to anomaly detection and operational DRA framework, completed with the usability focused visualisation of the user interfaces for control and monitoring to support the decision-making process. Constant evolution of those requirements also shaped the input and output interfacing of the DRA shifting the initial focus of behaviour comparison to anomaly processing integration to the RA process.

The first innovative step of this work is the development of the generic IoT ontology – IoT Stack – to enable the representation of the knowledge on the RA indispensable relationships, including threats, vulnerabilities, mitigation and actual risks definition in any IoT enabled environment. With our approach we have demonstrated its successful application in the domain of the smarthome cybersecurity.

The next challenge addressed was the development of the RAM, aiming to automate the decision-making process in the RA setting. We have started with the definition of the RL and their calculation methods. However, the resulting work indicated serious limitations bound to the risk perception of the individuals and lack of the historical data for the quantification of the risks' impacts. This resulted in the innovation of creating DRA framework based on the expert's risk weights initialisation, with their further adjustment at the runtime of the developed framework. Inspired by the Immune Theory, we have extended IoT Stack into RM permitting automated decision-making in the near-to-real time.

Validation of developed framework, also tackled in depth in this work, is a subject of a strong limitation on the data availability. DRA operating on top of cybersecurity incidents analysers, which in turn operate on the network captured events in a very specific environment, restrict the number of channels where such data can be gathered. Furthermore, due to the ethical constraints, generating attacks in a real-life settings is simply inappropriate and was

bound by the GDPR regulations in the bodywork of the GHOST project. Therefore, we had to adjust our validation strategy to validate operational criteria of the DRA by using a simulated environment with the replay of the gathered data from one voluntarily installation under the umbrella of the GHOST project.

Despite this serious limitation, we were able to demonstrate and validate several hypothesis set by this thesis. Namely, we have confirmed the context independence in the deployment environment and applicability of the expert values setting in a standard installation. Furthermore, we have also anticipated and integrated additional support mechanisms into DRA to enhance resilience of the final system from numerous threats: privacy preservation, malicious intentions in data reporting, singular point of failure in the decision-making process.

Our work is not finished with the presentation of this thesis. We are expanding this work in different domains, industrial manufacturing, automotive driving and privacy preservation in general with the ultimate purpose to make this work valuable for regular citizens. The ultimate goal of this framework is to provide the tool for understanding, monitoring and addressing the risks encountered in the digital arena of our lives.

References

[1] Christina Ruse. “Oxford Dictionary”. In: *Oxford Dictionary* 3.2015 (1999), pp. 2–5.

[2] Inc. The Gale Group. *West’s Encyclopedia of American Law, edition 2*. 2008.

[3] NIST. “NIST SP 800-30 Revision 1”. In: *Risk Management Guide for Information Security* 29.September (2012), p. 95. ISSN: 00986283. URL: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:NIST+Special+Publication+800-30#0.

[4] E. Zio. “The future of risk assessment”. In: *Reliability Engineering & System Safety* 177 (Sept. 2018), pp. 176–190. ISSN: 09518320. DOI: 10.1016/j.ress.2018.04.020. URL: https://doi.org/10.1016/j.ress.2018.04.020%20https://linkinghub.elsevier.com/retrieve/pii/S0951832017306543.

[5] Terje Aven. “Risk assessment and risk management: Review of recent advances on their foundation”. In: *European Journal of Operational Research* 253.1 (Aug. 2016), pp. 1–13. ISSN: 03772217. DOI: 10.1016/j.ejor.2015.12.023. URL: http://dx.doi.org/10.1016/j.ejor.2015.12.023%20https://linkinghub.elsevier.com/retrieve/pii/S0377221715011479.

[6] K Patel and Keyur. “Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges.” In: *Universidad Iberoamericana Ciudad de México* May (2016), pp. 6123, 6131. URL: http://www.opjstamnar.com/download/Worksheet/Day-110/IP-XI.pdf.

[7] Evan Wheeler. *Security Risk Management*. Elsevier, 2011. ISBN: 9781597496155. DOI: 10.1016/C2010-0-64926-1. URL: https://linkinghub.elsevier.com/retrieve/pii/C20100649261.

[8] Jason R.C. Nurse, Sadie Creese, and David De Roure. “Security Risk Assessment in Internet of Things Systems”. In: *IT Professional* 19.5 (2017), pp. 20–26. ISSN: 1520-9202. DOI: 10.1109/MITP.2017.3680959. URL: http://ieeexplore.ieee.org/document/8057728/.

[9] Stefan Taubenberger et al. “Problem analysis of traditional IT-security risk assessment methods - An experience report from the insurance and auditing domain”. In: *IFIP Advances in Information and Communication Technology* 354 AICT (2011), pp. 259–270. ISSN: 18684238. DOI: 10.1007/978-3-642-21424-0{_}21.

[10] Duncan Ki-Aries et al. “From Requirements to Operation: Components for Risk Assessment in a Pervasive System of Systems”. In: *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*. IEEE, Sept. 2017, pp. 83–89. ISBN: 978-1-5386-3488-2. DOI: 10.1109/REW.2017.36. URL: http://ieeexplore.ieee.org/document/8054834/.

[11] Mookyu Park, Haengrok Oh, and Kyungho Lee. “Security risk measurement for information leakage in IoT-based smart homes from a situational awareness perspective”. In: *Sensors (Switzerland)* 19.9 (2019). ISSN: 14248220. DOI: 10.3390/s19092148.

[12] Aakarsh Rao et al. “Probabilistic Threat Detection for Risk Management in Cyber-physical Medical Systems”. In: *IEEE Software* 35.1 (2018), pp. 38–43. ISSN: 0740-7459. DOI: 10.1109/MS.2017.4541031. URL: http://ieeexplore.ieee.org/document/8239935/.

[13] Pankaj Pandey et al. “Towards automated threat based risk assessment for cyber security in smart homes”. In: *Proceedings of the 18th European Conference on Cyber Warfare and Security (ECCWS 2019), Coimbra, Portugal*. Vol. 2019-July. 2019, pp. 4–5. ISBN: 9781912764280.

[14] Maggi Bansal, Inderveer Chana, and Siobhán Clarke. “A Survey on IoT Big Data”. In: *ACM Computing Surveys* 53.6 (Nov. 2021), pp. 1–59. ISSN: 0360-0300. DOI: 10.1145/3419634. URL: https://dl.acm.org/doi/10.1145/3419634.

[15] Michael Schiefer. “Smart Home Definition and Security Threats”. In: *Proceedings - 9th International Conference on IT Security Incident Management and IT Forensics, IMF 2015* (2015), pp. 114–118. ISSN: 21581339. DOI: 10.1109/IMF.2015.17.

[16] Peter Mayer and Melanie Volkamer. “Addressing misconceptions about password security effectively”. In: *Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust*. 2018, pp. 16–27.

[17] Vijay Sivaraman, Hassan Habibi Gharakheili, and Clinton Fernandes. *Inside job: Security and privacy threats for smart-home IoT devices*. Tech. rep. Sydney: Australian Communications Consumer Action Network, 2017. URL: https://www.runnersworld.com/running-gear/inside-job.

[18] Eyal Ronen et al. “IoT Goes Nuclear: Creating a ZigBee Chain Reaction”. In: *Proceedings - IEEE Symposium on Security and Privacy* (2017), pp. 195–212. ISSN: 10816011. DOI: 10.1109/SP.2017.14.

[19] Elisa Bertino and Nayeem Islam. “Botnets and Internet of Things Security”. In: *Computer* 50.2 (Feb. 2017), pp. 76–79. ISSN: 0018-9162. DOI: 10.1109/MC.2017.62. URL: http://ieeexplore.ieee.org/document/7842850/.

[20] Paul Johannesson and Erik Perjons. *An introduction to design science*. Vol. 9783319106. Springer International Publishing Switzerland, 2014, pp. 1–197. ISBN: 9783319106328. DOI: 10.1007/978-3-319-10632-8.

[21] A. Collen et al. “GHOST - Safe-guarding home IoT environments with personalised real-time risk control”. In: *Communications in Computer and Information Science*. Vol. 821. 2018, pp. 68–78. ISBN: 9783319951881. DOI: 10.1007/978-3-319-95189-8{_}7. URL: http://link.springer.com/10.1007/978-3-319-95189-8_7.

[22] J. Augusto-Gonzalez et al. “From Internet of Threats to Internet of Things: A Cyber Security Architecture for Smart Homes”. In: *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. IEEE, Sept. 2019, pp. 1–6. ISBN: 978-1-7281-1016-5. DOI: 10.1109/CAMAD.2019.8858493. URL: https://ieeexplore.ieee.org/document/8858493/.

[23] Anastasija Collen and Niels Alexander Nijdam. “Can I Sleep Safely in My Smarthome? A Novel Framework on Automating Dynamic Risk Assessment in IoT Environments”. In: *Electronics* 11.7 (Apr. 2022), p. 1123. ISSN: 2079-9292. DOI: 10.3390/electronics11071123. URL: https://www.mdpi.com/2079-9292/11/7/1123.

[24] Georgios Spathoulas et al. “Towards Reliable Integrity in Blacklisting: Facing Malicious IPs in GHOST Smart Contracts”. In: *2018 Innovations in Intelligent Systems and Applications (INISTA)*. Vol. 2018. 3. IEEE, July 2018, pp. 1–8. ISBN: 978-1-5386-5150-6. DOI: 10.1109/INISTA.2018.8466327. URL: https://ieeexplore.ieee.org/document/8466327/%20https://doi.org/10.1016/j.jnca.2019.102481.

[25] Anastasija Collen and Niels A Nijdam. “Understanding Human Factors in the Visualisation of the Usable Transparency and Cybersecurity and Privacy Configuration for Dynamic Risk Assessment”. In: *Information* (2021).

[26] Sherif Saad et al. “Detecting P2P botnets through network behavior analysis and machine learning”. In: *2011 Ninth Annual International Conference on Privacy, Security and Trust*. IEEE, July 2011, pp. 174–180. ISBN: 978-1-4577-0584-7. DOI: 10.1109/PST.2011.5971980. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5971980%20http://ieeexplore.ieee.org/document/5971980/.

[27] David Zhao et al. “Botnet detection based on traffic behavior analysis and flow intervals”. In: *Computers & Security* 39.8 (Nov. 2013), pp. 2–16. ISSN: 01674048. DOI: 10.1016/j.cose.2013.04.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404813000837.

[28] Saeed Nari and Ali A Ghorbani. “Automated malware classification based on network behavior”. In: *2013 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, Jan. 2013, pp. 642–647. ISBN: 978-1-4673-5288-8. DOI: 10.1109/ICCNC.2013.6504162. URL: http://ieeexplore.ieee.org/document/6504162/.

[29] Rob Kitchin and Martin Dodge. “The (In)Security of Smart Cities: Vulnerabilities, Risks, Mitigation, and Prevention”. In: *Journal of Urban Technology* 0.0 (2017), pp. 1–19. ISSN: 14661853. DOI: 10.1080/10630732.2017.1408002. URL: https://doi.org/10.1080/10630732.2017.1408002.

[30] Hussain Almohri et al. “On Threat Modeling and Mitigation of Medical Cyber-Physical Systems”. In: *Proceedings - 2017 IEEE 2nd International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2017* (2017), pp. 114–119. DOI: 10.1109/CHASE.2017.69.

[31] G Martins et al. “Towards a systematic threat modeling approach for cyber-physical systems”. In: *Resilience Week (RWS), 2015* (2015), pp. 1–6. DOI: 10.1109/RWEEK.2015.7287428.

[32] Ahmad-Reza Sadeghi, Christian Wachsmann, and Michael Waidner. “Security and privacy challenges in industrial internet of things”. In: *Proceedings of the 52nd Annual Design Automation Conference on - DAC ’15*. Vol. 17. New York, New York, USA: ACM Press, 2015, pp. 1–6. ISBN: 9781450335201. DOI: 10.1145/2744769.2747942. URL: http://eceasst.cs.tu-berlin.de/index.php/eceasst/issue/view/24%5Cnhttp://dl.acm.org/citation.cfm?doid=2744769.2747942%20http://dl.acm.org/citation.cfm?doid=2744769.2747942.

[33] Andreas Jacobsson, Martin Boldt, and Bengt Carlsson. “A risk analysis of a smart home automation system”. In: *Future Generation Computer Systems* 56 (Mar. 2016), pp. 719–733. ISSN: 0167739X. DOI: 10.1016/j.future.2015.09.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167739X15002812%20http://dx.doi.org/10.1016/j.future.2015.09.003%20http://linkinghub.elsevier.com/retrieve/pii/S0167739X15002812.

[34] Muhammad Ansar Latif et al. “User Privacy Framework for Web-of-Objects based Smart Home Services”. In: *International Journal of Smart Home* 9.5 (May 2015), pp. 61–72. ISSN: 19754094. DOI: 10.14257/ijsh.2015.9.5.07. URL: http://www.sersc.org/journals/IJSH/vol9_no5_2015/7.pdf.

[35] Ricardo Neisse et al. “Dynamic context-aware scalable and trust-based iot security, privacy framework, Chapter in Internet of Things”. In: *Chapter in Internet of Things Applications-From Research and Innovation to Market Deployment, IERC Cluster Book* January (2015), pp. 199–224. URL: https://www.researchgate.net/profile/Ricardo_Neisse/publication/271371042_Dynamic_Context-Aware_Scalable_and_Trust-based_IoT_Security_Privacy_Framework/links/54c8a8330cf238bb7d0e1014/Dynamic-Context-Aware-Scalable-and-Trust-based-IoT-Security-Privacy-Fram.

[36] Thorben Burghardt. “Why do privacy-enhancement mechanisms fail, after all? a survey of both, the user and the provider perspective”. In: *Workshop W2Trust, in ...* (2008). URL: http://www.ipd.uni-karlsruhe.de/~buchmann/pdfs/burghardt08whyPetFail.pdf.

[37] World Wide Web Consortium et al. “The platform for privacy preferences 1.0 (P3P1. 0) specification”. In: (2002).

[38] Daniel J Weitzner et al. “Creating a policy-aware web: Discretionary, rule-based access for the world wide web”. In: *Web and information security*. IGI Global, 2006, pp. 1–31.

[39] Yong Jin Park. “Digital Literacy and Privacy Behavior Online”. In: *Communication Research* 40.2 (2013), pp. 215–236. ISSN: 00936502. DOI: 10.1177/0093650211418338.

[40] Davide Ariu et al. “Security of the Digital Natives”. In: *Available at SSRN 2442037* (2014).

[41] N V Kushzhanov and U Zh Aliyev. “Digital space: changes in society and security awareness”. In: *Bulletin of the national academy of sciences of the Republic of Kazakhstan* 1 (2018), pp. 94–101. ISSN: 1991-3494.

[42] Parkinson. “Personal Data: Definition and Access”. PhD thesis.

[43] Yuan Niu et al. *One experience collecting sensitive mobile data*. 2010. URL: http://www2.parc.com/csl/members/eshi/docs/users.pdf.

[44] Kun Liu and Evimaria Terzi. “A framework for computing the privacy scores of users in online social networks”. In: *Proceedings - IEEE International Conference on Data Mining, ICDM*. Vol. 5. 1. 2009, pp. 288–297. ISBN: 9780769538952. DOI: 10.1109/ICDM.2009.21.

[45] Erfan Aghasian et al. “Scoring Users’ Privacy Disclosure Across Multiple Online Social Networks”. In: *IEEE Access* 5 (2017), pp. 13118–13130. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2720187. URL: http://ieeexplore.ieee.org/document/7959160/.

[46] Jaspreet Bhatia and Travis D. Breaux. “Empirical Measurement of Perceived Privacy Risk”. In: *ACM Transactions on Computer-Human Interaction* 25.6 (Dec. 2018), pp. 1–47. ISSN: 10730516. DOI: 10.1145/3267808. URL: http://dl.acm.org/citation.cfm?doid=3300063.3267808.

[47] Xuefeng Li et al. “A Privacy Measurement Framework for Multiple Online Social Networks against Social Identity Linkage”. In: *Applied Sciences* 8.10 (Oct. 2018), p. 1790. ISSN: 2076-3417. DOI: 10.3390/app8101790. URL: http://www.mdpi.com/2076-3417/8/10/1790.

[48] Kavitha Chandrasekar et al. “ISTR April 2017”. In: *Internet Security Threat Report - Symantec* 22.April (2017), p. 77. URL: https://digitalhubshare.symantec.com/content/dam/Atlantis/campaigns-and-launches/FY17/Threat%20Protection/ISTR22_Main-FINAL-JUN8.pdf?aid=elq_.

[49] Jason R C Nurse et al. “Guidelines for usable cybersecurity: Past and present”. In: *Proceedings - 2011 3rd International Workshop on Cyberspace Safety and Security, CSS 2011* (2011), pp. 21–26. DOI: 10.1109/CSS.2011.6058566.

[50] Paulo C. Realpe et al. “Towards an Integration of Usability and Security for User Authentication”. In: *Proceedings of the XVI International Conference on Human Computer Interaction*. Vol. 07-09-Sept. ACM. New York, NY, USA: ACM, Sept. 2015, pp. 1–6. ISBN: 9781450334631. DOI: 10.1145/2829875.2829912. URL: https://dl.acm.org/doi/10.1145/2829875.2829912.

[51] Terrence August, Robert August, and Hyoduk Shin. “Designing user incentives for cybersecurity”. In: *Communications of the ACM* 57.11 (Oct. 2014), pp. 43–46. ISSN: 00010782. DOI: 10.1145/2629487. URL: http://dl.acm.org/citation.cfm?doid=2684442.2629487.

[52] M Roesch. “Snort: Lightweight Intrusion Detection for Networks.” In: *LISA ’99: 13th Systems Administration Conference* (1999), pp. 229–238. ISSN: 15437221. DOI: http://portal.acm.org/citation.cfm?id=1039834.1039864. URL: http://static.usenix.org/publications/library/proceedings/lisa99/full_papers/roesch/roesch.pdf.

[53] Open Information Security Foundation (OISF). *Suricata*.

[54] Vern Paxson. "Bro: a system for detecting network intruders in real-time". In: *Computer Networks* 31.23-24 (Dec. 1999), pp. 2435–2463. ISSN: 13891286. DOI: 10.1016/S1389-1286(99)00112-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S1389128699001127.

[55] Samuel Marchal et al. "A big data architecture for large scale security monitoring". In: *Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014*. IEEE, June 2014, pp. 56–63. ISBN: 9781479950577. DOI: 10.1109/BigData.Congress.2014.18.

[56] Hideki Koike, Kazuhiro Ohno, and Kanba Koizumi. "Visualizing cyber attacks using IP matrix". In: *IEEE Workshop on Visualization for Computer Security 2005, VizSEC 05, Proceedings* (2005), pp. 91–98. DOI: 10.1109/VIZSEC.2005.1532070.

[57] Peng Xie Peng Xie et al. "Using Bayesian networks for cyber security analysis". In: *Dependable Systems and Networks DSN 2010 IEEEIFIP International Conference on* (2010), pp. 211–220. ISSN: 1530-0889. DOI: 10.1109/DSN.2010.5544924. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5544924.

[58] Lai Jibao, Wang Huiqiang, and Zhu Liang. "Study of Network Security Situation Awareness Model Based on Simple Additive Weight and Grey Theory". In: *2006 International Conference on Computational Intelligence and Security* Vol. 2 (2006), pp. 1545–1548. DOI: 10.1109/ICCIAS.2006.295320. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4076226.

[59] G Jakobson. "Mission cyber security situation assessment using impact dependency graphs". In: *14th International Conference on Information Fusion*. July 2011, pp. 1–8.

[60] Samuel Tweneboah-Koduah, Knud Erik Skouby, and Reza Tadayoni. "Cyber Security Threats to IoT Applications and Service Domains". In: *Wireless Personal Communications* 95.1 (July 2017), pp. 169–185. ISSN: 0929-6212. DOI: 10.1007/s11277-017-4434-6. URL: http://link.springer.com/10.1007/s11277-017-4434-6.

[61] Marin Emilov Pamukov. "Application of artificial immune systems for the creation of IoT intrusion detection systems". In: *Proceedings of the 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2017*. Vol. 1. IEEE, Sept. 2017, pp. 564–568. ISBN: 9781538606971. DOI: 10.1109/IDAACS.2017.8095144.

[62] Chien Ying Chen, Monowar Hasan, and Sibin Mohan. *Securing real-time internet-of-things*. May 2018. DOI: 10.3390/s18124356. URL: http://arxiv.org/abs/1705.08489.

[63] Peter Zegzhda. "Safe Integration of SIEM Systems with Internet of Things : Data Aggregation , Integrity Control , and Bioinspired Safe Routing". In: *Proceedings of the 9th International Conference on Security of Information and Networks SIN '16* (2016), pp. 81–87. DOI: 10.1145/2947626.2947639.

[64] Huichen Lin and Neil W Bergmann. "IoT Privacy and Security Challenges for Smart Home Environments". In: *Information* 7.3 (2016), p. 44.

[65] Jesus Pacheco and Salim Hariri. "IoT Security Framework for Smart Cyber Infrastructures". In: *2016 IEEE 1st International Workshops on Foundations and Applications of Self* Systems (FAS*W)*. IEEE, Sept. 2016, pp. 242–247. ISBN: 978-1-5090-3651-6. DOI: 10.1109/FAS-W.2016.58. URL: http://ieeexplore.ieee.org/document/7789475/.

[66] Younghee Park et al. "IoTGuard: Scalable and agile safeguards for Internet of Things". In: *MILCOM 2016 - 2016 IEEE Military Communications Conference*. Nov. 2016, pp. 61–66.

[67] F-Secure. *F-Secure SENSE router*. https://www.f-secure.com/en/web/home_global/sense.

[68] Luma Home Inc. *Luma*. https://lumahome.com/.

[69] BullGuard. *Dojo by BullGuard*. https://dojo.bullguard.com/.

[70] CUJO. *CUJO LLC*. https://www.getcujo.com/.

[71] Bitdefender B O X 2. *Bitdefender*. https://www.bitdefender.com/box/.

[72] Norton Core™. *Symantec Corporation*. https://us.norton.com/core.

[73] C Bormann, M Ersue, and A Keranen. *Terminology for Constrained-Node Networks*. Tech. rep. Internet Engineering Task Force (IETF), May 2014. DOI: 10.17487/rfc7228.

[74] Sparsh Mittal. "A survey of techniques for improving energy efficiency in embedded computing systems". In: *International Journal of Computer Aided Engineering and Technology* 6.4 (2014), p. 440. ISSN: 1757-2657. DOI: 10.1504/IJCAET.2014.065419.

[75] Zhengguo Sheng et al. "Lightweight Management of Resource-Constrained Sensor Devices in Internet of Things". In: *IEEE Internet of Things Journal* 2.5 (Oct. 2015), pp. 402–411. ISSN: 2327-4662. DOI: 10.1109/JIOT.2015.2419740.

[76] Heng Wang et al. "A Lightweight XMPP Publish/Subscribe Scheme for Resource-Constrained IoT Devices". In: *IEEE Access* (2017). ISSN: 21693536. DOI: 10.1109/ACCESS.2017.2742020.

[77] Mohit Sethi et al. "Secure and low-power authentication for resource-constrained devices". In: *2015 5th International Conference on the Internet of Things (IOT)*. IEEE, Oct. 2015, pp. 30–36. ISBN: 978-1-4673-8056-0. DOI: 10.1109/IOT.2015.7356545.

[78] Pawani Porambage et al. "Secure end-to-end communication for constrained devices in IoT-enabled Ambient Assisted Living systems". In: *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*. IEEE, Dec. 2015, pp. 711–714. ISBN: 978-1-5090-0366-2. DOI: 10.1109/WF-IoT.2015.7389141.

[79] D Barnard-Wills, L Marinos, and S Portesi. *Threat Landscape and Good Practice Guide for Smart Home and Converged Media*. Tech. rep. ENISA, 2014.

[80] Charalampos S Kouzinopoulos et al. "Using Blockchains to Strengthen the Security of Internet of Things". In: *Communications in Computer and Information Science*. Vol. 821. Lecture Notes CCIS No. 821, Springer Verlag, 2018, pp. 90–100. ISBN: 9783319951881. DOI: 10.1007/978-3-319-95189-8{_}9. URL: http://link.springer.com/10.1007/978-3-319-95189-8_9.

[81] Erol Gelenbe and Yasin Murat Kadioglu. "Energy Life-Time of Wireless Nodes with Network Attacks". In: *Proceedings of the 2018 ISCIS Security Workshop, ICL*. Lecture Notes CCIS No. 821, Springer Verlag, 2018.

[82] Olivier Brun et al. "IoT Attack Detection with Deep Learning". In: *Proceedings of the 2018 ISCIS Security Workshop, ICL*. Lecture Notes CCIS No. 821, Springer Verlag, 2018.

[83] Olivier Brun et al. "Deep Learning with Dense Random Neural Network for Detecting Attacks against IoT-connected Home Environments". In: *Submitted for Publication*. 2018.

[84] Hrishikesh Jayakumar et al. "Powering the Internet of Things". In: *Proceedings of the 2014 International Symposium on Low Power Electronics and Design*. ACM. 2014, pp. 375–380. ISBN: 9781605586366. DOI: 10.1145/2491185.2491191.

[85] Yan Meng et al. "Securing Consumer IoT in the Smart Home: Architecture, Challenges, and Countermeasures". In: *IEEE Wireless Communications* 25.6 (2018), pp. 53–59. ISSN: 1536-1284. DOI: 10.1109/MWC.2017.1800100.

[86] Joseph Bugeja, Andreas Jacobsson, and Paul Davidsson. "On Privacy and Security Challenges in Smart Connected Homes". In: *2016 European Intelligence and Security Informatics Conference (EISIC)*. IEEE, Aug. 2016, pp. 172–175. ISBN: 978-1-5090-2857-3. DOI: 10.1109/EISIC.2016.044. URL: http://ieeexplore.ieee.org/document/7870217/.

[87] Andreas Jacobsson and Paul Davidsson. "Towards a model of privacy and security for smart homes". In: *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*. IEEE, Dec. 2015, pp. 727–732. ISBN: 978-1-5090-0366-2. DOI: 10.1109/WF-IoT.2015.7389144. URL: http://ieeexplore.ieee.org/document/7389144/.

[88] Laura Rafferty et al. "Intelligent multi-agent collaboration model for smart home IoT security". In: *Proceedings - 2018 IEEE International Congress on Internet of Things, ICIOT 2018 - Part of the 2018 IEEE World Congress on Services*. 2018, pp. 65–71. ISBN: 9781538672440. DOI: 10.1109/ICIOT.2018.00016.

[89] Javid Habibi et al. "Heimdall: Mitigating the Internet of Insecure Things". In: *IEEE Internet of Things Journal* 4.4 (2017), pp. 968–978. ISSN: 23274662. DOI: 10.1109/JIOT.2017.2704093.

[90] Nicholas DeMarinis and Rodrigo Fonseca. "Toward Usable Network Traffic Policies for IoT Devices in Consumer Networks". In: *Proceedings of the 2017 Workshop on Internet of Things Security and Privacy - IoTS&P '17*. 2017, pp. 43–48. ISBN: 9781450353960. DOI: 10.1145/3139937.3139949.

[91] Martin Serror et al. "Towards In-Network Security for Smart Homes". In: *Proceedings of the 13th International Conference on Availability, Reliability and Security - ARES 2018*. 2018, pp. 1–8. ISBN: 9781450364485. DOI: 10.1145/3230833.3232802.

[92] Ali Dorri et al. "Blockchain for IoT security and privacy: The case study of a smart home". In: *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. March. IEEE, Mar. 2017, pp. 618–623. ISBN: 978-1-5090-4338-5. DOI: 10.1109/PERCOMW.2017.7917634. URL: https://ieeexplore.ieee.org/document/7917634/.

[93] Salvador Perez et al. "ARMOUR: Large-scale experiments for IoT security & trust". In: *2016 IEEE 3rd World Forum on Internet of Things, WF-IoT 2016*. 2017, pp. 553–558. ISBN: 9781509041305. DOI: 10.1109/WF-IoT.2016.7845504.

[94] Charalampos S. Kouzinopoulos et al. "Implementing a Forms of Consent Smart Contract on an IoT-based Blockchain to promote user trust". In: *2018 Innovations in Intelligent Systems and Applications (INISTA)*. IEEE, July 2018, pp. 1–6. ISBN: 978-1-5386-5150-6. DOI: 10.1109/INISTA.2018.8466268. URL: https://ieeexplore.ieee.org/document/8466268/.

[95] G. Carrozzo et al. "Interoperation of IoT Platforms in Confined Smart Spaces: The SymbIoTe Smart Space Architecture". In: *2018 Fifth International Conference on Internet of Things: Systems, Management and Security*. IEEE, Oct. 2018, pp. 38–45. ISBN: 978-1-5386-9585-2. DOI: 10.1109/IoTSMS.2018.8554894.

[96] Rachit Agarwal et al. "Unified IoT ontology to enable interoperability and federation of testbeds". In: *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. IEEE, Dec. 2016, pp. 70–75. ISBN: 978-1-5090-4130-5. DOI: 10.1109/WF-IoT.2016.7845470.

[97] Juan Hernandez-Serrano et al. "Privacy risk analysis in the IoT domain". In: *2018 Global Internet of Things Summit (GIoTS)*. IEEE, June 2018, pp. 1–6. ISBN: 978-1-5386-6451-3. DOI: 10.1109/GIOTS.2018.8534534.

[98] Congduc Pham and Mamour Diop. "Demo: WAZIUP, an Open and Versatile Long-range IoT Framework to Fully Take Advantage of the Cloudification of the IoT". In: *2018 3rd Cloudification of the Internet of Things (CIoT)*. IEEE, July 2018, pp. 1–2. ISBN: 978-1-5386-4629-8. DOI: 10.1109/CIOT.2018.8627084.

[99] Antonio F. Skarmeta et al. "IoTCrawler: Browsing the Internet of Things". In: *GIoT 2018 proceedings , IEEE Communications Society* (2018). DOI: 10.1109/GIOTS.2018.8534528.

[100] Nina Gerber, Benjamin Reinheimer, and Melanie Volkamer. "Home Sweet Home? Investigating Users' Awareness of Smart Home Privacy Threats". In: *An Interactive Workshop on the Human aspects of Smarthome Security and Privacy (WSSP 2018), Symposium on Usable Privacy and Security (SOUPS) 2018* (2018), pp. 1–3. URL: https://spice.sice.indiana.edu/files/2018/07/wssp2018-paper2.pdf%0Ahttps://spice.sice.indiana.edu/wssp/.

[101] *ISO/IEC TR 27016:2014 Information technology – Security techniques – Information security management – Organizational economics ISO/IEC*. Tech. rep. 1. Geneva, CH: International Organization for Standardization, 2014, p. 31.

[102] Olusola Samuel-Ojo et al. "Meta-analysis of design science research within the IS community: Trends, patterns, and outcomes". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 6105 LNCS. Springer, 2010, pp. 124–138. ISBN: 3642133347. DOI: 10.1007/978-3-642-13335-0{_}9.

[103] Waqar Ali et al. "IoT based smart home: Security challenges, security requirements and solutions". In: *2017 23rd International Conference on Automation and Computing (ICAC)*. IEEE, Sept. 2017, pp. 1–6. ISBN: 978-0-7017-0260-1. DOI: 10.23919/IConAC.2017.8082057. URL: http://ieeexplore.ieee.org/document/8082057/.

[104] Amir Rahmati et al. "Tyche: Risk-Based Permissions for Smart Home Platforms". In: *arXiv preprint arXiv:1801.04609* (Jan. 2018). URL: http://arxiv.org/abs/1801.04609.

[105] Jason R.C. C Nurse et al. "If you can't understand it, you can't properly assess it! The reality of assessing security risks in internet of things systems". In: *IET Conference Publications* 2018.CP740 (2018), pp. 1–9. DOI: 10.1049/cp.2018.0001.

[106] Bako Ali and Ali Ismail Awad. "Cyber and physical security vulnerability assessment for IoT-based smart homes". In: *Sensors (Switzerland)* 18.3 (2018), pp. 1–17. ISSN: 14248220. DOI: 10.3390/s18030817.

[107] G Gonzalez-Granadillo et al. "Dynamic risk management response system to handle cyber threats". In: *Future Generation Computer Systems* 83 (2018), pp. 535–552. ISSN: 0167739X. DOI: 10.1016/j.future.2017.05.043.

[108] Richard Caralli et al. *Introducing OCTAVE Allegro: Improving the Information Security Risk Assessment Process*. Tech. rep. CMU/SEI-2007-TR-012. Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2007. URL: http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=8419.

[109] Keyun Ruan. "Introducing cybernomics: A unifying economic framework for measuring cyber risk". In: *Computers and Security* 65.2017 (2017), pp. 77–89. ISSN: 01674048. DOI: 10.1016/j.cose.2016.10.009. URL: http://dx.doi.org/10.1016/j.cose.2016.10.009.

[110] Antonio Colella. "Cyber security and ubiquity: an human-centric approach". In: (2017). DOI: http://dx.doi.org/10.14273/unisa-1038.

[111] Hany F. Atlam et al. "Fuzzy Logic with Expert Judgment to Implement an Adaptive Risk-Based Access Control Model for IoT". In: *Mobile Networks and Applications* (2019). ISSN: 15728153. DOI: 10.1007/s11036-019-01214-w.

[112] Mansour Alali et al. "Improving risk assessment model of cyber security using fuzzy logic inference system". In: *Computers and Security* 74 (2018), pp. 323–339. ISSN: 01674048. DOI: 10.1016/j.cose.2017.09.011. URL: https://doi.org/10.1016/j.cose.2017.09.011.

[113] Vipindev Adat and B. B. Gupta. "Security in Internet of Things: issues, challenges, taxonomy, and architecture". In: *Telecommunication Systems* 67.3 (2018), pp. 423–441. ISSN: 15729451. DOI: 10.1007/s11235-017-0345-9.

[114] Kejun Chen et al. "Internet-of-Things Security and Vulnerabilities: Taxonomy, Challenges, and Practice". In: *Journal of Hardware and Systems Security* 2.2 (2018), pp. 97–110. ISSN: 2509-3428. DOI: 10.1007/s41635-017-0029-7.

[115] Minhaj Ahmad Khan and Khaled Salah. "IoT security: Review, blockchain solutions, and open challenges". In: *Future Generation Computer Systems* 82 (2018), pp. 395–411. ISSN: 0167739X. DOI: 10.1016/j.future.2017.11.022. URL: https://doi.org/10.1016/j.future.2017.11.022.

[116] Md Mahmud Hossain, Maziar Fotouhi, and Ragib Hasan. "Towards an Analysis of Security Issues, Challenges, and Open Problems in the Internet of Things". In: *Proceedings - 2015 IEEE World Congress on Services, SERVICES 2015* (2015), pp. 21–28. DOI: 10.1109/SERVICES.2015.12.

[117] Ivan Cvitić, Miroslav Vujić, and Siniša Husnjak. "Classification of security risks in the iot environment". In: *Annals of DAAAM and Proceedings of the International DAAAM Symposium* 2015-Janua.2016 (2015), pp. 731–740. ISSN: 17269679. DOI: 10.2507/26th.daaam.proceedings.102. URL: http://www.daaam.info/Downloads/Pdfs/proceedings/proceedings_2015/102.pdf.

[118] Peter Aufner. "The IoT security gap: a look down into the valley between threat models and their implementation". In: *International Journal of Information Security*. Vol. 19. 1. Springer Berlin Heidelberg, 2020, pp. 3–14. ISBN: 1020701900. DOI: 10.1007/s10207-019-00445-y. URL: https://doi.org/10.1007/s10207-019-00445-y.

[119] Elena Doynikova, Andrey Fedorchenko, and Igor Kotenko. "Ontology of metrics for cyber security assessment". In: *ACM International Conference Proceeding Series* (2019). DOI: 10.1145/3339252.3341496.

[120] Xiaoci Huang et al. "A semantic approach with decision support for safety service in smart home management". In: *Sensors (Switzerland)* 16.8 (2016). ISSN: 14248220. DOI: 10.3390/s16081224.

[121] Ryan Heartfield et al. "A taxonomy of cyber-physical threats and impact in the smart home". In: *Computers & Security* 78 (Sept. 2018), pp. 398–428. ISSN: 01674048. DOI: 10.1016/j.cose.2018.07.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404818304875.

[122] Yacov Y Haimes. "Hierarchical Holographic Modeling". In: *IEEE Transactions on Systems, Man and Cybernetics* 11.9 (1981), pp. 606–617. ISSN: 21682909. DOI: 10.1109/TSMC.1981.4308759.

[123] Michael Meisel, Vasileios Pappas, and Lixia Zhang. "A taxonomy of biologically inspired research in computer networking". In: *Computer Networks* 54.6 (Apr. 2010), pp. 901–916. ISSN: 13891286. DOI: 10.1016/j.comnet.2009.08.022. URL: https://linkinghub.elsevier.com/retrieve/pii/S1389128609003740.

[124] Vishwa Teja Alaparthy and Salvatore Domenic Morgera. "A Multi-Level Intrusion Detection System for Wireless Sensor Networks Based on Immune Theory". In: *IEEE Access* 6 (2018), pp. 47364–47373. ISSN: 21693536. DOI: 10.1109/ACCESS.2018.2866962. URL: https://ieeexplore.ieee.org/document/8449076/.

[125] Marios Anagnostopoulos et al. "Tracing your smart-home devices conversations: A real world iot traffic data-set". In: *Sensors (Switzerland)* 20.22 (Nov. 2020), pp. 1–28. ISSN: 14248220. DOI: 10.3390/s20226600. URL: https://www.mdpi.com/1424-8220/20/22/6600.

[126] Vijay K. Vaishnavi and William Kuechler. *Design science research methods and patterns: Innovating information and communication technology*. 2nd. CRC Press, 2007, pp. 1–227. ISBN: 9781420059335. DOI: 10.1201/9781420059335.

[127] Philipp Offermann et al. "Artifact Types in Information Systems Design Science – A Literature Review". In: *Lecture Notes in Computer Science*. Springer, 2010, Vol 6105. ISBN: 978-3-642-13335-0. DOI: 10.1007/978-3-642-13335-0{_}6.

[128] Andrew Whitmore, Anurag Agarwal, and Li Da Xu. "The Internet of Things—A survey of topics and trends". In: *Information Systems Frontiers* 17.2 (Apr. 2015), pp. 261–274. ISSN: 15729419. DOI: 10.1007/s10796-014-9489-2.

[129] Ala Al-Fuqaha et al. "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications". In: *IEEE Communications Surveys and Tutorials* 17.4 (2015), pp. 2347–2376. ISSN: 1553877X. DOI: 10.1109/COMST.2015.2444095.

[130] Steve Huckle et al. "Internet of Things, Blockchain and Shared Economy Applications". In: *Procedia Computer Science*. Vol. 58. Elsevier, Jan. 2016, pp. 461–466. DOI: 10.1016/j.procs.2016.09.074.

[131] Hosub Lee and Alfred Kobsa. "Privacy preference modeling and prediction in a simulated campuswide IoT environment". In: *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, Mar. 2017, pp. 276–285. ISBN: 978-1-5090-4327-9. DOI: 10.1109/PERCOM.2017.7917874. URL: http://ieeexplore.ieee.org/document/7917874/.

[132] Feng Tian. "A supply chain traceability system for food safety based on HACCP, blockchain & Internet of things". In: *14th International Conference on Services Systems and Services Management, ICSSSM 2017 - Proceedings*. IEEE, June 2017, pp. 1–6. ISBN: 9781509063697. DOI: 10.1109/ICSSSM.2017.7996119.

[133] Bruno Rodrigues et al. "A blockchain-based architecture for collaborative DDoS mitigation with smart contracts". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Ed. by Daphne Tuncer et al. Vol. 10356 LNCS. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 16–29. ISBN: 9783319607733. DOI: 10.1007/978-3-319-60774-0{_}2.

[134] Mathis Steichen, Stefan Hommes, and Radu State. "ChainGuard-A firewall for blockchain applications using SDN with OpenFlow". In: *2017 Principles, Systems and Applications of IP Telecommunications, IPTComm 2017*. Vol. 2017-Septe. IEEE, Sept. 2017, pp. 1–8. ISBN: 9781538613221. DOI: 10.1109/IPTCOMM.2017.8169748.

[135] John Tobin, Christina Thorpe, and Liam Murphy. "An Approach to Mitigate Black Hole Attacks on Vehicular Wireless Networks". In: *IEEE Vehicular Technology Conference*. Vol. 2017-June. IEEE, June 2017, pp. 1–7. ISBN: 9781509059324. DOI: 10.1109/VTCSpring.2017.8108460.

[136] Richard Dennis and Gareth Owen. "Rep on the block: A next generation reputation system based on the blockchain". In: *2015 10th International Conference for Internet Technology and Secured Transactions, ICITST 2015*. IEEE, Dec. 2016, pp. 131–138. ISBN: 9781908320520. DOI: 10.1109/ICITST.2015.7412073.

[137] Yanqi Zhao et al. "Secure Pub-Sub: Blockchain-Based Fair Payment with Reputation for Reliable Cyber Physical Systems". In: *IEEE Access* 6 (2018), pp. 12295–12303. ISSN: 21693536. DOI: 10.1109/ACCESS.2018.2799205.

[138] Alexander Schaub et al. "A trustless privacy-preserving reputation system". In: *IFIP Advances in Information and Communication Technology*. Vol. 471. Springer, Cham, May 2016, pp. 398–411. ISBN: 9783319336299. DOI: 10.1007/978-3-319-33630-5{_}27.

[139] Gavin Wood. "Ethereum: a secure decentralised generalised transaction ledger". In: *Ethereum Project Yellow Paper* (2014), pp. 1–32. ISSN: 1098-6596. DOI: 10.1017/CBO9781107415324.004.

[140] Ittay Eyal and Emin Gün Sirer. "Majority is not enough: Bitcoin mining is vulnerable". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 8437. 2014, pp. 436–454. ISBN: 9783662454718. DOI: 10.1007/978-3-662-45472-5{_}28. URL: http://link.springer.com/10.1007/978-3-662-45472-5_28.

[141] Bruno Rodrigues, Thomas Bocek, and Burkhard Stiller. "Enabling a Cooperative , Multi-domain DDoS Defense by a Blockchain Signaling System". In: *Semantic Scholar* (2017).

[142] Ab Pedersen. "Usability of authentication in web applications–a literature review". In: *July* (2010), pp. 1–33. URL: http://www.andersbp.dk/studier/dat/HCISEC/projekt.pdf.

[143] Ann Cavoukian and Claudiu Popa. "Embedding Privacy Into What's Next: Privacy by Design for the Internet of Things". In: *Ryerson University Privacy & Big Data Institute* April (2016), pp. 1–10. URL: http://www.ryerson.ca/content/dam/pbdce/papers/Privacy-by-Design-for-the-Internet-of-Things.pdf.

[144] Zahrah A. Almusaylim and Noor Zaman. "A review on smart home present state and challenges: linked to context-awareness internet of things (IoT)". In: *Wireless Networks* 25.6 (2019), pp. 3193–3204. ISSN: 15728196. DOI: 10.1007/s11276-018-1712-5. URL: https://doi.org/10.1007/s11276-018-1712-5.

[145] Hala Assal and Sonia Chiasson. ""Think secure from the beginning": A survey with software developers". In: *Conference on Human Factors in Computing Systems - Proceedings* (2019), pp. 1–13. DOI: 10.1145/3290605.3300519.

[146] Bar Haim et al. "Visualizing insider threats: An effective interface for security analytics". In: *International Conference on Intelligent User Interfaces, Proceedings IUI* (2017), pp. 39–42. DOI: 10.1145/3030024.3038264.

[147] Sören Preibusch. "Privacy behaviors after Snowden". In: *Communications of the ACM* 58.5 (Apr. 2015), pp. 48–55. ISSN: 0001-0782. DOI: 10.1145/2663341. URL: https://dl.acm.org/doi/10.1145/2663341.

[148] Gurpreet Dhillon et al. "Deciding between information security and usability: Developing value based objectives". In: *Computers in Human Behavior* 61 (Aug. 2016), pp. 656–666. ISSN: 07475632. DOI: 10.1016/j.chb.2016.03.068. URL: https://linkinghub.elsevier.com/retrieve/pii/S0747563216302394.

[149] Panagiotis Andriotis et al. "A study on usability and security features of the Android pattern lock screen". In: *Information & Computer Security* 24.1 (Mar. 2016), pp. 53–72. ISSN: 2056-4961. DOI: 10.1108/ICS-01-2015-0001. URL: https://www.emerald.com/insight/content/doi/10.1108/ICS-01-2015-0001/full/html.

[150] Jong Hyouk Lee and Hyoungshick Kim. "Security and Privacy Challenges in the Internet of Things [Security and Privacy Matters]". In: *IEEE Consumer Electronics Magazine* 6.3 (2017), pp. 134–136. ISSN: 21622256. DOI: 10.1109/MCE.2017.2685019.

[151] Rajeev Alur et al. "Systems Computing Challenges in the Internet of Things". In: (Apr. 2016). URL: http://arxiv.org/abs/1604.02980.

[152] Jason R. C. Nurse, Ahmad Atamli, and Andrew Martin. "Towards a Usable Framework for Modelling Security and Privacy Risks in the Smart Home". In: 2016, pp. 255–267. DOI: 10.1007/978-3-319-39381-0{_}23. URL: http://link.springer.com/10.1007/978-3-319-39381-0_23.

[153] Saurabh Dutta, Stuart Madnick, and Ger Joyce. "SecureUse: Balancing security and usability within system design". In: *Communications in Computer and Information Science*. Vol. 617. 2016, pp. 471–475. ISBN: 9783319405476. DOI: 10.1007/978-3-319-40548-3{_}78. URL: http://link.springer.com/10.1007/978-3-319-40548-3_78.

[154] Ka Ping Yee. "Aligning security and usability". In: *IEEE Security and Privacy* 2.5 (2004), pp. 48–55. ISSN: 15407993. DOI: 10.1109/MSP.2004.64.

[155] Deanna D. Caputo et al. "Barriers to Usable Security? Three Organizational Case Studies". In: *IEEE Security & Privacy* 14.5 (Sept. 2016), pp. 22–32. ISSN: 1540-7993. DOI: 10.1109/MSP.2016.95. URL: http://ieeexplore.ieee.org/document/7676139/.

[156] D. Balfanz et al. "In search of usable security: five lessons from the field". In: *IEEE Security & Privacy Magazine* 2.5 (Sept. 2004), pp. 19–24. ISSN: 1540-7993. DOI: 10.1109/MSP.2004.71. URL: http://ieeexplore.ieee.org/document/1341405/.

[157] Andrea Atzeni, Shamal Faily, and Ruggero Galloni. "Usable Security". In: *Encyclopedia of Information Science and Technology, Fourth Edition*. Ed. by Mehdi Khosrow-Pour D.B.A. IGI Global, 2018, pp. 5004–5013. ISBN: 9781522522553. DOI: 10.4018/978-1-5225-2255-3.ch433. URL: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-5225-2255-3%20http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-5225-2255-3.ch433.

[158] Nina Gerber, Benjamin Reinheimer, and Melanie Volkamer. "Investigating People's Privacy Risk Perception". In: *Proceedings on Privacy Enhancing Technologies* 2019.3 (July 2019), pp. 267–288. ISSN: 2299-0984. DOI: 10.2478/popets-2019-0047. URL: https://www.sciendo.com/article/10.2478/popets-2019-0047.

[159] Reyhan Duezguen et al. "How to Increase Smart Home Security and Privacy Risk Perception". In: *2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, Oct. 2021, pp. 997–1004. ISBN: 978-1-6654-1658-0. DOI: 10.1109/TrustCom53373.2021.00138. URL: https://ieeexplore.ieee.org/document/9724299/.

[160] Natã M Barbosa, Zhuohao Zhang, and Yang Wang. "Do Privacy and Security Matter to Everyone ? Quantifying and Clustering User-Centric Considerations About Smart Home Device Adoption This paper is included in the Proceedings of the Sixteenth Symposium on Usable Privacy and Security ." In: *Proceedings of the Sixteenth Symposium on Usable Privacy and Security* (2020), pp. 417–435.

[161] Pardis Emami-Naeini et al. "Ask the experts: What should be on an IoT privacy and security label?" In: *Proceedings - IEEE Symposium on Security and Privacy* 2020-May (2020), pp. 447–464. ISSN: 10816011. DOI: 10.1109/SP40000.2020.00043.

[162] Madiha Tabassum et al. "" I don ' t own the data ": End User Perceptions of Smart Home Device Data Practices and Risks This paper is included in the Proceedings of the". In: *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security* (2019), pp. 435–450. URL: https://dl.acm.org/citation.cfm?id=3361509.

[163] Pierre Antoine Vervier and Yun Shen. "Before toasters rise up: A view into the emerging IoT threat landscape". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11050 LNCS (2018), pp. 556–576. ISSN: 16113349. DOI: 10.1007/978-3-030-00470-5{_}26.

[164] Julie M. Haney, Susanne M. Furman, and Yasemin Acar. "Smart home security and privacy mitigations: Consumer perceptions, practices, and challenges". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12210 LNCS (2020), pp. 393–411. ISSN: 16113349. DOI: 10.1007/978-3-030-50309-3{_}26.

[165] Paul Dunphy et al. “Understanding the Experience-Centeredness of Privacy and Security Technologies”. In: *Proceedings of the 2014 workshop on New Security Paradigms Workshop - NSPW '14*. Vol. 15-18-Sept. New York, New York, USA: ACM Press, 2014, pp. 83–94. ISBN: 9781450330626. DOI: 10.1145/2683467.2683475. URL: http://dl.acm.org/citation.cfm?doid=2683467.2683475.

[166] Helen Collard and Jo Briggs. “Creative Toolkits for TIPS”. In: vol. 51. September. 2020, pp. 39–55. ISBN: 9781137456250. DOI: 10.1007/978-3-030-66504-3{_}3. URL: http://link.springer.com/10.1007/978-3-030-66504-3_3.

[167] Stefan Victora. “IoT Guard : Usable Transparency and Control Over Smart Home IoT Devices”. PhD thesis. Institut für Information Systems Engineering, 2020. DOI: 10.34726/hss.2020.53900.

[168] Maria Bada, Angela M. Sasse, and Jason R. C. Nurse. “Cyber Security Awareness Campaigns: Why do they fail to change behaviour?” In: July (Jan. 2019), p. 38. URL: http://discovery.ucl.ac.uk/1468954/1/Awareness%20CampaignsDraftWorkingPaper.pdf%20http://arxiv.org/abs/1901.02672.

[169] George Chalhoub et al. “Factoring user experience into the security and privacy design of smart home devices: A case study”. In: *Conference on Human Factors in Computing Systems - Proceedings* (2020), pp. 1–9. DOI: 10.1145/3334480.3382850.

[170] Denis Feth, Andreas Maier, and Svenja Polst. “A user-centered model for usable security and privacy”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10292 LNCS (2017), pp. 74–89. ISSN: 16113349. DOI: 10.1007/978-3-319-58460-7{_}6.

[171] Marthie Grobler, Raj Gaire, and Surya Nepal. “User, Usage and Usability: Redefining Human Centric Cyber Security”. In: *Frontiers in Big Data* 4.March (2021), pp. 1–18. ISSN: 2624909X. DOI: 10.3389/fdata.2021.583723.

[172] Yanlin Li et al. “Minibox: A two-way sandbox for x86 native code”. In: *Proceedings of the 2014 USENIX Annual Technical Conference, USENIX ATC 2014*. Philadelphia, PA: USENIX Association, June 2014, pp. 409–420. ISBN: 9781931971102. URL: https://www.usenix.org/conference/atc14/technical-sessions/presentation/li_yanlin.

[173] James R. Lewis. “The System Usability Scale: Past, Present, and Future”. In: *International Journal of Human–Computer Interaction* 34.7 (July 2018), pp. 577–590. ISSN: 1044-7318. DOI: 10.1080/10447318.2018.1455307. URL: https://doi.org/10.1080/10447318.2018.1455307%20https://www.tandfonline.com/doi/full/10.1080/10447318.2018.1455307.