Security and privacy in 5G

by

Emanuele Catania

Submitted to the Department of Electrical, Electronic and Computer Engineering
in partial fulfillment of the requirements for the degree of

Ph.D. in Systems, Energy, Computer Science and Telecommunication Engineering

UNIVERSITÀ DEGLI STUDI DI CATANIA

SICILIAE STVDIVM GENERALE · 1434 ·

Chiar.mo Prof. Aurelio La Corte
Supervisor

Chiar.mo Prof. Paolo Arena
Coordinator

Security and privacy in 5G

by

Emanuele Catania

Submitted to the Department of Electrical, Electronic and Computer Engineering
on September 1, 2018, in partial fulfillment of the
requirements for the degree of
Ph.D. in Systems, Energy, Computer Science and Telecommunication Engineering

Abstract

The current advancements of communication systems and their applications have changed our lives and will influence them further in the future. Next generation 5G networks will represent a salient technological breakthrough that combines old and new technologies and involves, among all, new models of service provisioning and resource sharing. In particular, they will lead to the emergence of mechanisms and architectures towards the on-demand multi-tenant philosophy. In this new ecosystem, it will be necessary to address the trust question among stakeholders as well as their security.

The 5G revolution brings new pitfalls due to novel forms of human-to-device interactions and the even higher pervasiveness of the technology in human life. As an example, in the Internet of Things (IoT), devices equipped with sensing, processing, storage and decision-making capabilities, can actively interact with one another and with humans. Although their design could strictly adhere to the principles of privacy and security, several factors, such as weak implementations of communication protocols, metadata information exchange, and architectural flaws, could jeopardise the security and privacy of their owners. Moreover, the augmented complexity and heterogeneity deriving from the ultra-densification of communication infrastructures, although it can improve data rate, reduce delay, and coverage of cellular networks, might raise new threats to the privacy of network subscribers.

In the first part of this thesis, we provide an overview of 5G networks and analyse the security, trust, and privacy problems in it. Then, we discuss the mutual impact of security and privacy of stakeholders and the use of semantic reasoning systems for the trust evaluation. In this vein, we studied the features of security ontologies that can influence the automated threat identification process and laid out a road towards ontologies simplification. In the second part of this thesis, we give a brief introduction to the privacy issues in the IoT. Then, we propose a methodology of analysis for identification of privacy threats in the IoT which can explore the privacy issue space from different perspectives and at various levels of abstraction. In the third part of this thesis, we explore the effect of both user equipment and access points densification on the location privacy.

We characterised the relationship between density of users and the success of attacks aiming at disclosing the location of subscribers. Hence, we propose a mitigation strategy founded on the concept of virtual cells.

Keywords: 5G, trust, privacy, security, semantic accuracy, differential semantic variance, network centrality, Internet of Things, Ultra-Dense Networks

Thesis Supervisor: Chiar.mo Prof. Aurelio La Corte

Acknowledgments

Entering an international doctoral course after roughly nine years from the end of my master degree has been very challenging. It qualified me and provided me with expertise that I hardly ever acquired. However, I never could have done it without the support of some important persons to me.

First and foremost I wish to warmly thank my PhD thesis advisor, Prof. Aurelio La Corte, for having walked me into the interesting worlds of telecommunication networks and security, and for supporting me throughout the PhD process. He has been a mentor, guide, and reference.

Special thanks go to thank Dr Ahmed Elmokashfi for all our productive conversations of research in the area of 5G networks and supporting my investigation from February 2018 now on and during my visit at Simula Research Laboratory in Oslo. I would like to thank the reviewers for having evaluated my thesis, taking the time to read it through, and for all their estimable comments and insightful recommendations. I am also thankful to all the members of Openlab at the University of Catania for providing such friendly environments.

I want to include in this expression of gratitude my best friends Giovanni, Gianfranco, Carlo, Mario, Antonio, Francesco, Laura, Tiziana, Gabriella, Guido, Giuseppe, Andrea, Biagio, Filippo, Luca, Riccardo for spending so much great time together, for our talks about many intriguing subjects on politics, history, music, literature, and cinema. I am thankful to Paulo Valenordica and Jangshui Hong for our pleasant talks on culture diversities, family, entrepreneurship, and positive thinking.

Above all, I sincerely thank my wife Giorgia, who upheld and encouraged me with patience during these years of study and research, my son Sergio, who has given my and my wife‘s effort meaning, and my loving parents and sisters.

Contents

List of Figures

List of Tables

Abbreviations

5G PPP 5G Infrastructure Public Private Partnership

API application program interfaces

AS access stratum

CAPEX capital expences for hardware

CDMA code-division multiple access

CN core network

CP Cyclic Prefix

DARPA Defence Advance Research Project Agency

DDOS Distributed Denial of Service

DOS Denial of Service

eMBB Enhanced Mobile Broadband

eMTC Enhanced Machine Type Communication

eNodeB E-UTRAN Node B

GUTI global unique temporary identifier

HSS Home Subscriber Server

IMSI International Mobile Subscriber Identity

IoT Internet of Things

KPI key performance indicator

LOS Line-of-Sight

LPWAN low-power wireless area network

LTE Long Term Evolution

LTE-A LTE Advanced

LTE-U LTE for Unlicensed spectrum

M2M machine-to-machine

MAP Macrocell Access Point

MIB Master Informaition Block

MME Mobility Management Entity

mMTC Massive Machine Type Communication

MNO Mobile network operator

NADF Network Attack and Defence Framework

NB-IoT Narrow Band IoT

NFV Network Function Virtualisation

OF OpenFlow

OFDMA orthogonal frequency-division multiple access

OPEX capital expences for operations

PCFICH Physical Control Format Indicator Channel

PCI Physical Cell Identity

PDCCH Physical Downlink Control Channel

PDSCH Physical Downlink Shared Channel

PGW Packet Data Network Gateway

PLMN Public Land Mobile Network

PSS Primary Synchronization Signal

QoE Quality of Experience

RAN radio Access Network

RLF radio link failure

RRC Radio Resource Control

RSRP reference signal received power

RSSI received signal strength indicator

S-TMSI serving temporary mobile subscriber identity

SND Software-Defined Network

SINR signal-to-interference-plus-noise ratio

SLA Service Level Agreement

SNR signal-to-noise ratio

SSS Secondary Synchronization Signal

TA tracking area

UDN Ultra-Dense Network

UE mobile device

URLLC Ultra-Reliable and Low Latency Communications

WLAN Wireless Local Area Network

ZF Zachman Framework

Chapter 1

Introduction

1.1 Motivation

The evolution of nowadays communication systems brings with her new opportunities and will determine a profound modification of our lifestyle and human-to-systems interaction. The 5G is a novel network paradigm that combines old and new technologies and answer the calls for growing data demand due to an increasing number of entities accessing data communication services (see Figure 1-1) and novel, resource-consuming applications (e.g., 4k ultra-HD video streaming, virtual and augmented reality). In the vision of 5G Infrastructure Public Private Partnership (5G PPP), in future nations an enormous quantity of heterogeneous data and knowledge will be everywhere accessible to everyone and in real-time.

Driven by new business use cases for the development and implementation of 5G, 5G PPP identified new actors/stakeholders that interact with each other to foster novel network services and applications. However, new challenging security problems might emerge in this potentially untrustful environment. As an example, if a provider of network services and applications suffers from security vulnerabilities, served stakeholders might be exposed to attacks to their security to court [85]. Therefore, accurately determining security vulnerabilities by considering interdependences among 5G actors might be of great interest.

It is universally recognised that 5G should fill the gap in term of capacity, data

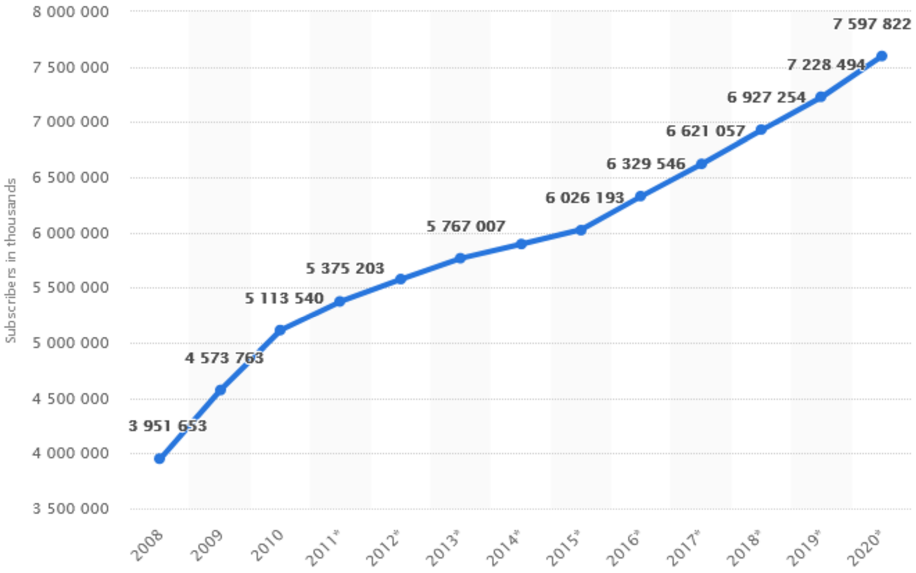


Figure (1-1) Global mobile subscribers to 2020.

rate, massive device connectivity, end-to-end latency, and Quality of Experience (QoE) between what 4G offers and what the novel services and applications need. For this reason, a dramatic transformation in the design and architectures of wireless networks is required. Furthermore, new insights might originate from the research of human-to-network modes of interaction and projections of data traffic model in the network of the future.

In years to come, the most data traffic of communication networks is expected to come from smart devices [18]. Hence, it is crucial that the current cellular network develops to foster the prospected deployment of IoT systems and applications. In this view, the cellular communication infrastructure could cooperate with other wireless network technologies (e.g. Wireless Local Area Network (WLAN), relay-assisted and device-to-device communications, wireless personal area networks, LTE for Unlicensed spectrum (LTE-U)).

As to provide more and more efficient services and anticipate needs, financial, social, and health information can day-by-day be stored, manipulated, and analysed by service providers. However, although of the benefits that it can offer, information processing and transmission through networked systems could affect the security and privacy of their owners and subjects. Hence, the integrated, service-oriented network of the future, while ensuring both connectivity and privacy to end-users, should address the security on the network as a whole [5]. However, studying and identifying privacy issues in the 5G is not straightforward. Indeed, 5G is a complex eco-system in which well-known vulnerabilities adds up to threats proper of the internet and software systems.

1.1.1 Goals

In this work, we are involved in studying the data and the location privacy of end-users in 5G. Given its tremendous complexity, we envisage that a thorough understanding and analysis of security threats in the 5G is both very relevant and challenging to achieve. Hence, first, we study the automatic identification of security threats in 5G, discussing security ontologies, their accuracy and complexity against threats eliciting accuracy and time of analysis. Then, we deepen the consequence of the massive spread of connected IoT devices on the data and location privacy of people. Thus, we analyse the effect of the modification of the current cellular network towards the ultra-dense paradigm on the location privacy of end-users.

1.2 Plan of the thesis and contribution

In this section, we present a list and a brief summary of the chapters included in this thesis.

In Chapter 2 we review the overall architecture of 5G and focus on the network slicing concept, the IoT, and Ultra-Dense Networks (UDNs). Therefore, we provide an overview of the security, privacy and trust in 5G and we recall some important privacy metrics. Next, we review the past-to-present literature on the analysis of ontologies in general, and of security ontologies in particular. We conclude the chapter overviewing the related work on the security and privacy in the IoT and UDNs.

In Chapter 3 we discuss deeper into the automatic threat identification. We evaluate the accuracy and complexity of ontologies to feeding automatic reasoning. Then, we discuss the trade-off between the quality and the time needed for eliciting the security threats in 5G systems.

In Chapter 4 we study the privacy issue in the context of the IoT. We survey the literature on security and privacy frameworks for communication networks. Further, we propose a privacy framework for the privacy assessment for the IoT.

In Chapter 5 we address the problem of the location privacy in ultra-dense networks. We provide some discussion on the role of both access point and subscriber numerosity on end-user location privacy and propose a mitigation strategy based on the concept of virtual cell for mobility purpose. Then, we weigh the conjoint effect of portable device spread and implementation of ultra-dense networks to the privacy of devices‘ owners.

Finally, in Chapter 6 we provide the concluding remarks of this thesis.

1.3 Publications

- E. Catania, A. La Corte, Privacy evaluation of IoT devices in Ultra-Dense Networks, to be submitted

- E. Catania, A. Di Stefano, A. La Corte, M Scatà, Study on the Semantic Accuracy of Ontologies Emerging from Folksonomies, Expert Systems (2018), accepted with major revisions

- E. Catania and A. La Corte, IoT Privacy in 5G Networks, IoTBDS 2018 Conference, 19- 21 March 2018, Madeira, Portugal

- E. Catania and A. La Corte, Location Privacy in Virtual Cell-Equipped Ultra-Dense Networks, NTMS 2018 Conference, 26-28 February, Paris, France

- M. Scatà, A. Di Stefano, A. La Corte, P. Liò, E. Catania, E. Guardo, S. Pagano, Combining evolutionary game theory and network theory to analyze human cooperation patterns. Chaos, Solitons & Fractals (2016), 91, 17-24.

- A. Di Stefano, M. Scatà, A. La Corte, P. Liò, E. Catania, E. Guardo, S. Pagano. Quantifying the Role of Homophily in Human Cooperation Using Multiplex Evolutionary Game Theory, PloS one 10.10 (2015): e0140646.

- E. Catania, A. Di Stefano, E. Guardo, A. La Corte, S. Pagano, M. Scatà. Energy Awareness and the Role of "Critical Mass" in Smart Cities, International Refereed Journal of Engineering and Science (IRJES), Volume 4, Issue 7, July 2015, pp. 38-43. ISSN: 2319-183X (online); 22319-1821 (print).

- P. Motta, E. Catania, E. Guardo, A. La Corte, S. Pagano. Benefits of nanosatellite network for smart metering technological infrastructure in wide areas, Tartu Conference on Space Science and Technology Tartu, Estonia, 22-24 September 2014, poster & invited talk

Chapter 2

Literature Review

In this Chapter, we provide some background that will be recalled later in this thesis. In particular, we describe architectures, main enabling technologies, key features, and known security vulnerabilities concerning next-generation 5G networks. Among all the key technologies, we deepen ultra-dense networks and the IoT. In line with the aim of this thesis, we provide some background on privacy and related metrics at the end of this Chapter.

2.1 The 5G eco-system

The 5G network represents the answer of wireless communication infrastructure to the proliferation of new businesses for industry and vertical markets. Present society is evolving towards a new model of human-to-human and human-to-device interaction, in which hyper-connected autonomic entities continuously exchange, store, and elaborate information in real-time. This communication network revolution leads the way towards design and implementation of new services and application, driven by novel emerging use cases, vertical markets and industries (e.g., e-health, smart city, connected cars, haptic communication, virtual and augmented reality).

5G is envisioned to provide high-speed connectivity (1Gbps), low-latency (at maximum 1 ms, to be successful for mission-critical application and systems) and support for low power devices (e.g., sensors) with lifespans up to several years without the

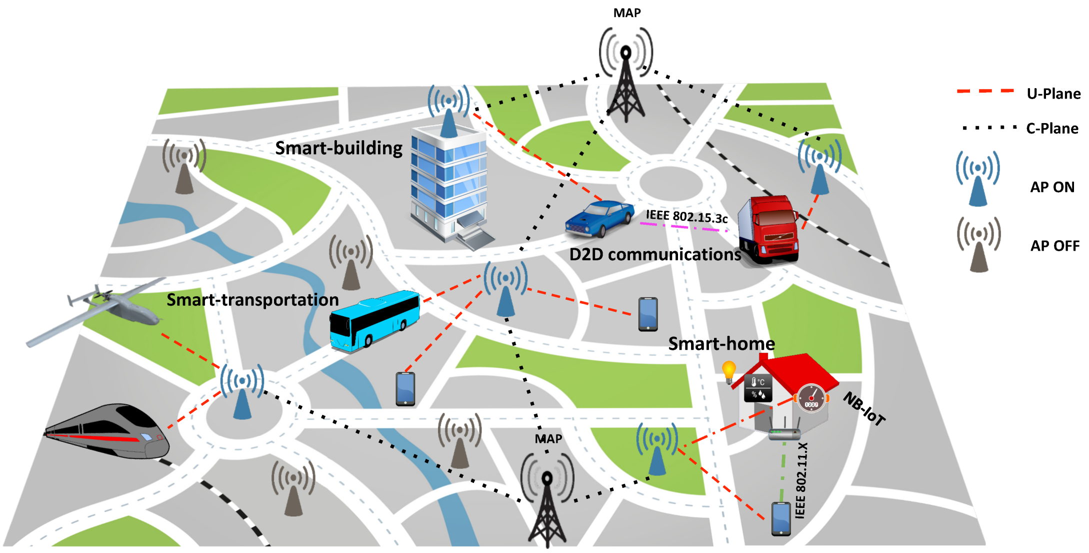


Figure (2-1) This picture shows a general 5G network architecture and actors involved in communication processes.

need for human assistance [53]. However, such a performance enhancement might require a more efficient spectrum utilisation in addition to an improvement of network flexibility. Ubiquity is one of the most important key features expected for 5G. It will support the aforementioned application and services, also fostering the communication and cooperation among autonomous, connected devices. Small cells and UDNs, may provide the network with the adaptability and plasticity for implementing ubiquitous connectivity [5]. A general 5G network scenario describing the interconnectivity among different technologies in a multi-tier, ultra dense network is shown in Figure 2-1. Activation status of the multitude access points is centrally managed by a Software-Defined Network (SND) controller. The main responsibility of access points is providing devices with the user-plane connectivity whilst Macrocell Access Point (MAP) provides devices with the control-plane. The resulting data traffic will be significantly different with respect to those generated in current cellular networks. Then, coexistence of both human and device guises of interaction with the network will impose a profound modification of the manner key performance indicators (KPIs) are defined [90, 42, 88].

Among all, the SND paradigm is one of the principal keys to designing of flexible network architectures. By separating the control plane from the data plane, it provides new network control functionalities and abstractions [5]. In particular, network control functions, such as the network operating system, the network manager, and related application program interfacess (APIs) are programmable. In SNDs, the network architecture is composed of three parts, namely the aforementioned controller, the northbound interface (i.e., APIs between the controller and SDN applications), and the southbound interface (i.e., the bridge between the controller and the SDN-enabled infrastructure). OpenFlow (OF) is the most important protocol for SDNs. It enables communication between the controller and the SDN-enabled infrastructure, allowing applications running on OF switches to manage packet forwarding and lookup among switches and routers. Complementarily to SDN, Network Function Virtualisations (NFVs) are responsible for implementing network functions (e.g., the routing, network address translation, firewalls, intrusion detection,) in software. In 5G, it is expected that NFVs will offer scalability, resilience, and flexibility by virtualising the core network systems (such as MME, PGW, HSS, etc.) over pooling computing resources. Since both NFVs and SND controllers can run on commodity servers, they may offer scalability, flexibility, and reduction of capital expences for hardware (CAPEX) and capital expences for operations (OPEX).

The virtualisation of network functions, the shift towards the software-defined networking, as well as an efficient management and orchestration of resources are the foudation for network slicing. According to the 5G PPP Architecture Working Group, "the network slice is a composition of adequately configured network functions, network applications, and the underlying cloud infrastructure (physical, virtual or even emulated resources, RAN resources etc.), that are bundled together to meet the requirements of a specific use case, e.g., bandwidth, latency, processing, and resiliency, coupled with a business purpose". Since it can cope with the demand for dedicated and on-demand services, slicing the physical network into multiple isolated logical networks has emerged as a key to satisfy the demand for a wide range of vertical sectors. The network slices will cover the entire protocol stack. They will span from the underlying virtualised hardware resources up to network functions and services running on top of them, thus answering the demands of extremely various use cases, to all the network segments including core, transport, radio, wired, and edge networks.

The scientific community and the industry are agreed that 5G logical network slices will enable the creation of multi-domain, multi-technology, and tenant-specific networks [72, 39]. Network slicing will revolutionise the networking panorama. By abstracting, isolating, orchestrating, and softwarising, it separates logical network segments from the underlying physical network resources, then enhancing the network capabilities and flexibility. Mobile network operators (MNOs) will provide the technological ground for the "as-a-service" paradigm, meeting the use cases spanning from Enhanced Mobile Broadband (eMBB), to Ultra-Reliable and Low Latency Communications (URLLC) and Massive Machine Type Communications (mMTCs). The end-to-end vision of network slicing will start from the mobile network edge, up until the core network (CN). One of the most important point with regard to network slicing features is to build dedicated logical networks specifically designed to exhibit customised functionality. Differently from legacy systems, telco services (e.g., SMS and voice) will not be hosted on the operator‘ systems, but implemented as interconnection of (virtualised) network functions. Hence, 5G networks will pave the way towards an importat shift from the monolitic network design to a more flexible and dynamic composition approach. Sharing of commercial off-the-shelf physical devices will be done through multiplexing and multitasking.

Thanks to NFV and SND different tenant can buid their services upon the same physical infrastructure. Since multiplexing happen at the infrastructure level, subscribers will experience better QoE. Moreover, that approach can foster interoperability among operators.

Although, the infrastructure resources could be shared among different network slice instances, every provider may use a specific control and cloud management system. Advanced orchestration and automation are required to release the configuration burden from users and to enable an integrated end-to-end solution. Softwarisation of functions and systems and predictive analytics make it possible to effectively scale, change and elastically manage the network service by mean of recursive structures. However, measurement of their performance in compliance with Service Level Agree-

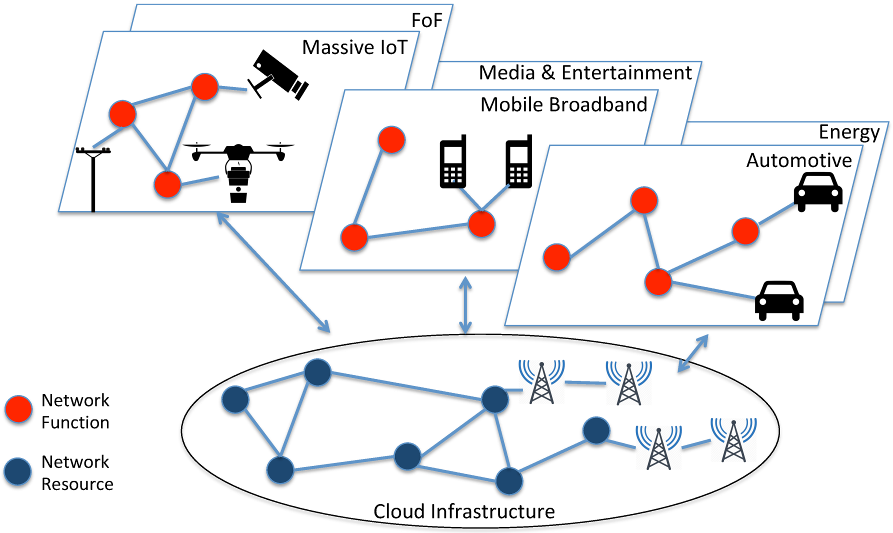


Figure (2-2) The network slice is a composition of network functions and applications together with the underlying cloud infrastructure, that are bundled together to meet, among all, the requirements of resilience , bandwidth, latency, and processing [69].

ments (SLAs) will be necessary for the dynamic slice instantiation and activation. The 5G network will be coping with the outstanding increasing information exchange throughout the whole system. Therefore, the security challenge in 5G will not be confined to guaranteeing reliable and trustworthy connectivity to end users.

2.2 Security, Privacy and Trust in 5G

The 5G system is expected to realise precise actions to cope with the increasing information traffic and provide reliable connectivity and security to the entire network. Security challenges will involve confidentiality of communications, authorisation and access control (of UEs), accounting, integrity, availability, and authentication. The 5G system is required to cope with a significant number of potential classes of threats that can be originated by failures (both internal and external), maliciuous actions of stakeholders, and external entities. Authentication will be particularly challenging. Due to the time required for operations at remote servers (in the order of hundreds of milliseconds), current approaches to authentication will not be suitable anymore in 5G, which is inconsistent with the specification of latency (ideally equal to zero). In general, current security and privacy mechanisms are at odds with latency requirements as stronger approaches are both more computationally expensive and time-demanding. Moreover, the authentication issue will be worsened by the ultra-densification trend. Then, as to cope with the resulting rise in the number of handovers and continuous authentication of user equipments over small cells, the ultra-dense setting will require novel mechanisms for transferring the security context. This problem is discussed in the Chapter 5. Although it might seem a good solution to detect intruders and secure the network, monitoring a large number of UEs served by 5G systems is not a trivial task. Since it does not require the execution of complex operations, the physical layer security [94] might represent a viable solution to secure the 5G. Furthermore, it offers a valid alternative to the adoption of third-party security providers and cryptographic approach for authentication and information privacy.

New business models and applications highlight the role of cooperation and strong interdependence among several actors of 5G and the paramount value of the trust among them. Delivery of services will require a synergistic action of multiple actors, that will try to ensure a given quality of service in agreement with the service level agreements (SLAs). Therefore, the trust among involved parties is the glue to build novel and evolved services. It and its model can track the reputation of actors as a function of their performance, thus promoting the birth of a chain of responsibilities for granting a proper level of service for the end-users and an optimal state for the network system.

In the next generation network it is expected to be a growing awereness of responsibilities, driven by the need of having a clear picture of risks. A metric to measure actors reputation and trust might be founded on security measures they actually put in place, their compliance to policies, and availability. Monitored resources might be both the physical and virtual ones required for the delivery of services. The 5G-ENSURE project [85] exploited a comprehensive method of risks identification built on machine understandable models enforced with expert knowledge on potential risks for the system to be analysed. By using a machine-understandable model, this analysis direction can automatically identify and carefully every known way in which the network may be attacked or accidentally compromised, also reducing the probability that any vulnerabilities could be overlooked. In Chapter 3 we discuss the automated threat identification and provide some considerations in this vein.

2.2.1 Privacy metrics

In this section we recall some important works and definitions of privacy. In general, we can distinguish two broad approaches to individual‘s privacy. The first is focused on protecting identities, and the second trying to preserve their data. When the users‘ identity are protected, linking users‘ data must be unfeasible or very hard to obtain. Most techniques that aim to protect user‘s data are usually based on disclosing of modified versions of original dataset. In this manner, aversaries can access only to inaccurate information. In this direction, Samarati et .al [71] introduced one of the most prominent privacy definition, namely the k-anonymity. Using the notion of quasi-identifiers (that is, a set of information and attributes that cannot disclose real identity of individuals when taken singularly, but can do it if used in combination with other data sources), authors proved that hiding or reducing the granularity of the set, can make an individual indistinguishable from other k-1 subjects. That is, any combination of quasi-identifiers (attributes of individuals) could not disclose real identities of a target, but make it undistinguishable from other k-1 ones. In [23], authors defined the concept of differential privacy, which provides a condition on the public release of datasets. It states that given two public release of dataset differing for only one record are differentially private if outputs from their querying are equally likely.

Although k-anonymity and differential privacy were considered first in the context of statistical databases, they have been successfully applied in the field of location-based systems. The scientific community proposed other variants of k-anonymity over the years, such as l-diversity [50] or t-closeness [59]. The former variant reaquires that for each combination of non-senitive information of individuals made public (e.g., present into the released set) there must be at least l different and well-represented values realeased of each sensitive data. The latter principle imposes that the difference of the distribution of values of a sentitive attribute in an equivalence class must be no more than a threshold t with respect to the released set. Anyway, protecting the users‘ identity might not be enough to protect their privacy. In fact, adversaries might disclose real identity of individuals by combining anonymised information with other publicly available dataset or data of their own.

Location, together with financial and medical information, belongs to a long list of personal data that can be exploited as a starting point to derive new insightful and valuable information on individuals (e.g., data on personal life, religion and political beliefs, financial and professional information). Protecting location privacy, might mean quantify the error of the adversary trying to infer the real location of victims. Privacy frameworks founded on definition of k-anonymity, differential privacy, or mix-zones [10] can, although with limitations [82], preserve location privacy of individuals.

2.2.2 Ontologies and security threats classification

Many studies have been published on the using and developing of ontologies in the security area [68, 67, 92, 11, 84]. In [11], authors claimed the importance of ontologies for classifying security threats. They provided a review of the most relevant papers on the topic of security ontologies and their applications and found that, although they offer a useful contribution to the community knowledge, ontologies proposed in the literature do not provide a comprehensive knowledge of the subject. An important contribution to the literature in the field was given in [84]. Authors suggested a classification of ontologies in eight different security classes, namely web-oriented, taxonomies, requirement, risk-based, initial, general, and modelling. In particular, we are interested in the latter classification.

Due to their extensibility, security ontological models can flexibly define concepts of a knowledge domain, spanning from a generic to detailed representation of facts. Moreover, their sharing enable the collaboration among system managers/network administrators and cybersecurity experts. As a benefit, they can improve in effectiveness and speed to respond to security threats. However, there are still some issues and challenges to be addressed in order to assess the semantic quality of security ontologies. Authors of [26] explored the problem of quality assessment of ontologies and provided a statistical characterisation of "good ontologies". In [34], the authors described how a shared thesaurus can be exploited for the development of ontologies in many stages of their lifecycle. Authors claimed that the use and maintenance of a shared thesaurus may facilitate the development and interoperability of ontologies. A taxonomy can be considered as a simple variant of a thesaurus. Hence, instead of a thesaurus, it could be possible to use a shared taxonomy, able to organise words in a hierarchical structure, according to a similarity of meaning.

In their analysis, the authors of [36] stated that semantic reasoning (i.e., deriving information not explicitly made available by an ontology) may be speeded-up by reducing the size of ontologies, limiting the number of independent paths and the degree of classes, and making the inheritance a tree-like graph. However, time constraints are not the only limiting factors for semantic reasoning systems. Indeed, semantic accuracy may influence the precision of the semantic inference tasks and, consequently, the output of reasoning systems. In [26] it has been claimed that taxonomic features, namely number of classes, depth and breadth variances, are the best predictors of ontologies' semantic accuracy, even though this measure of accuracy strongly depends on the ontology size. This challenging issue was coped and overcome by [87], by defining an empirical aggregated measure of ontologies' semantic accuracy based on the variance computation of semantic dispersion of their taxonomic structures.

2.2.3 Security threats in Ultra-dense networks

Motivated by the forecast of wireless traffic in the years to come [18], network densification represent a viable solution to cope with scarce spectrum resources. The idea behind the network densification trend is that network performance in term of bandwidht, spectrum resuse, network capacity, and energy consumption can improve through proper access points deployment. Densification is fulfilled by positioning small cell access points indoors in buildings, and outdoors in public areas. Both small and macro cells coexist in a multi-tier, software-defined network architecture. In UDNs, small cells can be classified into pico and femto cells (i.e., fully-functioning base stations, capable of perform all the functions of the protocol stack within a limited coverage area) and macro-extentions access systems (such as Relays and Remote Radio Heads, that extends the signal coverage and can perform some or all physical layers functions of the protocol stack). Follows a brief overview of the characteristics and challenges of ultra-dense networks [35]:

- Users are sorrounded by many small cells with a low power and small footprint. The inter-site distance is of the order at maximum of tens meters.

- For the sake of interference and energy consumption reduction, small cells are not all active at the same time instant. In particular, small cells can be turned on and off depending on connectivity demand by users and in consideration of the aforementioned optimisations.

- Drastic interference between neighboring cells limits network densification. To this aim, strict interference management schemes are required to mitigate the inter-cell interferences .

- The backhaul of a small cell in UDNs environments might limit its capacity. indeed, as the network evolves towards the ultra dense paradigm, it might be very difficult to guarantee an ideal high-speed and low-delay backhaul for each small cell.

- In canonical cellular networks, both the spectrum reuse and reuse pattern are at the level of a cluster of cells. In UDN context, there would be a need for a paradigm shift in the frequency reuse concept. Indeed, in code-division multiple access (CDMA) and orthogonal frequency-division multiple access (OFDMA), this reuse scheme converges to one when the spectrum is reused in each cell.

- Due to the high probability of Line-of-Sight (LOS) transmissions (i.e., dominant LOS component in the received signal), the propagation modelling in UDN should consider both Rician and Raylight models for multi-path fading. Indeed, the distance between BSs and users is small enough to have a high probability of LOS transmissions stressing the need for considering different propagation models.

Densification trend of communication networks could seriously jeopardize location privacy of mobile nodes [24]. Previous work related to UDNs has mainly focused on studying the effect of densification concerning handover [70], signal-to-interference-plus-noise ratio (SINR), and network cost (i.e., energy consumption, hardware, and cabling) [45]. Although the problems of privacy and security in UDNs have been coped in [24, 22, 89], further consideration on location privacy would be required. Leakage of information such as requesters‘ identifier data, usage information, the time of information request [81], and their combination [63] might undermine mobile users‘ location privacy.

It is likely that, during a UDN communication session, moving users perform many authentications to more than one access point. Identities, context information, and pairwise keys are exchanged among parties. Thus, as described in [22], as to reduce the risk of impersonation and man-in-the-middle attack, an authentication handover module could be introduced. According to authors, physical layer attributes can be a viable authentication solution to lessen the burden on both access points and users of using the widely adopted cryptographic mechanisms. Anyway, such an approach does not prevent a passive adversary from acquiring or inferring sensitive information on mobile users (e.g., their location).

Location privacy could be preserved through pseudonymization, anonymization and path perturbation. Anyway, during access point-to-user associations, exchanged meta-data could allow an adversary to disclose useful information on nodes' location. In this direction, Farhang et al. [24] observed that algorithms to associate mobile nodes to access points could reveal private information on the whereabouts of mobile nodes. Furthermore, since eavesdropped information could be sufficient to detect the presence or the absence of a user in a specific area with enough precision, the smaller the size of cells is, the better the location identification of users is. In addition, by combining the aforementioned information with additional knowledge, the adversary could infer the boundaries within which mobile nodes are likely to be located [29]. When such boundaries contain more than one mobile node, individual's location privacy could be safeguarded. k-anonymity [71] has been significantly adopted overtime to quantify users' privacy in location-based services. Unfortunately, one of the main limits of k-anonymity-based approaches is that k-anonymity cannot be proved to be satisfied without considering adversary's auxiliary information. Recently, because of its independence from the prior knowledge of the adversary, differential privacy has gained momentum. Anyway, both differential privacy and its derived definitions (e.g., geo-indistinguishability) still appear impractical for protecting of spatiotemporal information. Gramaglia et al. [30] introduced the notion of $k^{\tau,\epsilon}$-anonymity which can be seen as a variation of k^m-anonymity. It has been conceived to attain the uninformative principle, thus guaranteeing that an adversary cannot infer from eavesdropped information longer fragment of users' trajectories. However, since k-anonymity is a special case of $k^{\tau,\epsilon}$-anonymity and for the sake of simplicity, in this thesis we make use of k-anonymity as a metric for measuring location privacy

2.2.4 Security and Privacy in the IoT

The concept of the IoT was introduced as long ago as 1999 with the diffusion of Wireless Sensor Network technologies and the spread of the Radio Frequency Identification techniques. The IoT is a network of physical objects, smart and personal devices (such as smart-watches, smartphones), health monitors, vehicles, buildings, and appliances (and many other entities) that are revolutionising in many ways our daily lives. Devices can sense and send information to remote servers, communicate and cooperate with each other, and take decisions autonomously on our behalf. By elaborating and properly combining the acquired information, new smart entities such as smart homes, smart cities, healthcare and intelligent transport systems can spring into life [5]. The overwhelming growing rate of the number of interconnected devices was already discernible in early 2012, since yet then there were more than nine billion IoT devices active worldwide. These impressive numbers were and will be propelled by the financial market and in various domains, such as healthcare, public services and transportation. In the near future, the number of wirelessly connected devices will dramatically increase [18] and hugely influence the design and requirements of next-generation cellular networks. However, the notable IoT diffusion and potential interaction among a huge number of entities might lead to worrying security issues. This is further supported by the statements of the Defence Advance Research Project Agency (DARPA) on the difficulty to establish a general purpose security strategy model for the IoT. In addition, several challenges in term of scalability, latency, the reliability of messages delivery, management of intermittent transmission behaviour and support of multiple wireless technologies need to be addressed. Because of design trade-offs in term of cost, complexity, and energy consumption, many devices in the IoT are usually resource-limited.

As to cope with unauthorised access, data theft, and eavesdroppings, devices should be provided with authentication, authorisation mechanisms, and data preservation capabilities, ensuring freshness, authenticity, confidentiality, and integrity of information. Privacy (i.e., unlinkability, data secrecy, and anonymity) should be accurately preserved since personal and sensitive information could be stolen and abused by an adversary. In particular, sensitive data should be preserved by encryption before of being sent, since it prevents that transmitted data can be intercepted and easily read by passive adversaries. Nevertheless, encrypting the information might require using computationally expensive cryptographic primitives (e.g. pairing- based cryptography), which could not be executed by every IoT device.

In order to identify suitable cryptographic approaches for the IoT, Malina et al. [51] measured the performance of the most used primitives (such as RSA, secure hashing algorithms and AES) on some of the most common micro-controllers (ARM, MSP430f X) equipping IoT devices. They found that while operations of hashing and symmetric ciphering take few milliseconds and can also run on very limited microcontrollers, stronger approaches, such as RSA asymmetric signing (by a 2048-bit private key), can cause delays into hundreds of milliseconds, which are intolerable in real-time IoT applications.

Computationally complex operations could be carried out remotely or on communication gateways. Although they determine a reduction of both energy consumption and computation for devices [80], these approaches require trustful gateways and protected communications. Pseudonymization can hide the real identity of both devices and users. Hence, it could be a viable technique to protect entities from being traced. Anyway, as suggested in [7] and in [79], when they act within a sufficiently wide time window of observation, eavesdroppers might disclose real victims‘ identifier. For example, as described later in this thesis, when they connect to an LTE-based network, IoT devices can decode messages broadcasted by E-UTRAN Node Bs (eNodeBs) to search for their (temporary or unique) identifier. A passive adversary could exploit decoded information to retrieve associations among temporary and unique identifier. Furthermore, colluding IoT users positioned in proximity of locations occasionally visited by the IoT target (i.e. the victim), even though protected by a pseudonym, might reveal to the attacker the target's real identity and its private activities [98].

In our knowledge of the IoT, confidentiality, trust, and privacy are main key issues to the development of IoT services and applications. Then, identifying and adopting a privacy framework for the IoT is of paramount importance [83, 56]. Furthermore, since they are provided with internet connectivity, the IoT devices can be exposed to a long list of threats such as, to name a few, the Denial of Service (DOS), Distributed Denial of Service (DDOS), SYN Flood, and Smurfing attack. As to mitigate advanced persistent threats against the IoT, the authors of [91] proposed the Network Attack and Defence Framework (NADF) and provided a quantitative evaluation of the effectiveness of their approach. Since NADF is built upon the ZF, it can manage the security problem from both microscopic and macroscopic point of view, involving management, technology, and strategies for security.

The IoT is a very complicated ecosystem, whose protection would require both a holistic and specialised approach and that should span from the software to the network side. Then, we envisage that the research direction described in [91] is viable to assess security risk for the IoT. Anyway, this work does not study the IoT privacy problem, which is closely related to security, but most focuses on data protection and the right on data. The privacy issue in the IoT is discussed in [64]. In their work, authors provide a comprehensive overview of the IoT and on its privacy challenges, and draw our attention to the most important characteristics a privacy framework for the IoT should have, that is identity, temporal, location and query privacy in addition to interoperability, data minimisation, and accountability.

In [61], Perera et. al call into question some past assumption about privacy-by-design strategies. In particular, authors claim that minimisation of data acquisition, storage, retention time, number of data source, and granularity in conjunction with encryption of data, communications, processing, and hiding of routing and location information, could lessen the privacy risk in term of unauthorised access and misuse of data. Authors claimed that exploiting cryptographic approaches can guarantee the privacy in the IoT. Moreover, they stated that anonymous routing systems (e.g., TOR) is a viable direction to hide the routing traffic in the IoT. Even if assumptions in [61] seems to be well-founded, this study they overlook the technological limitations that may affect IoT devices [51, 41].

In their analysis of privacy requirements for the IoT, Kung et. al [41] question the need for embedding privacy engineering methods into IoT systems and applications. They underlined that apparently non-personal information, when linked together, may disclose personal and sensitive information on IoT users and their activities. They maked a distinction between a privacy framework and a privacy engineering framework. In particular, the latter extends the former providing reference to well-known privacy engineering fundamentals or concepts (i.e., privacy engineering and privacy-by-design, privacy objectives, and privacy properties to be protected), identifying actors (i.e., the stakeholders) and their roles within a system. Moreover, it includes protection of privacy engineering and common privacy engineering terminologies. In addition, the authors of [41] made a distinction between privacy engineering for IoT subsystem and privacy engineering for IoT system. In particular, they defined a subsystem of the IoT as an independent entity upon which the IoT system is built. By integrating sub-systems together, it is possible to provide the IoT system with functionalities it needs. As an example, the LTE cellular network when providing wireless connectivity to devices is a subsystem for the IoT. In general, suppliers of IoT subsystems are not aware of IoT-specific systems and applications requirements. Then, keen attention should be paid from the IoT system engineering perspective to prevent and manage security and privacy issues that can derive from heterogeneous system integration.

Chapter 3

Semantic quality of threat ontologies for assessing the trust in 5G networks

The 5G-ENSURE project [85] highlighted the effect of the security vulnerabilities of the stakeholders on the trust in 5G. By modelling the 5G system and providing a semantic reasoning system with a core ontology, the 5G-ENSURE‘s method elicited the expected security risks in term of availability, confidentiality, integrity, overload, and unreliability for stakeholders and their devices. Then, by identifying the effect of both security threats and misbehaviour of a stakeholder over other actors, it was possible to describe the trust relationships within the 5G system. Since the outcomes of analysis were a function of the system model and the core ontology considered, it might be interesting to estimate how the ontology‘ features could affect the quality and the temporal performance of semantic reasonings.

In this Chapter, we discuss the impact of security ontologies and their characteristics on the capacities of semantic reasoning systems. In particular, we analyse the semantic accuracy and complexity of ontologies. Then, we introduce a function called differential semantic variance based on the concepts of eigenvector centrality and semantic variance. It allows grouping taxonomy terms of an ontology according to their contribution to the whole ontology‘s accuracy. This work gives key insights into the assessment of the relevance of taxonomy terms in the evaluation of semantic accuracy, shedding light on the crucial role played by network theory and its structural properties, such as eigenvector centrality.

3.1 Introduction

In Chapter 2, we considered correspondences between privacy and security and between trust and security. In this section, we shed light on the correlation between privacy and trust. The risk of revealing private and sensitive information is tightly coupled with the trust a trustee has on trustors. The use cases described in [69] revealed that during their functioning, systems and stakeholders are both involved in numerous tasks in which information is exchanged. Hence, mechanisms in 5G are conceived with the assumption of trust relationships among parties. In [85], it has been finalised with a method of analysis aimed at security threat identification. It was based on the 5G network modelling and a semantic reasoning system fed with a security threat ontology (hereinafter called core ontology).

Among all, several security threats are specific privacy concerns. Indeed, they involve the information of both stakeholders and service providers and the right on it. Privacy (and security, as well) assessment methods depend on system models and ontologies taken into consideration. Therefore, in this chapter, we investigate the role of ontologies and their accuracy in the automated security and privacy threats identification.

Ontologies find an application field in security areas, such as risk management and quality of service analysis in next-generation networks [44, 42, 43]. By exploiting classes, instances, relations, and properties to formally represent concepts, ontologies provide a well-structured and machine-readable representation of information, thus enabling knowledge-based applications [47, 25] to manage and construe data from a semantic perspective. In particular, the output of these applications depends strictly on the ontologies‘ features. Consisting of either modifying (i.e., extending, specialising, assembling) [28, 27] or merging together ontologies [38], the ontology re-use diminishes the costs for ontology development and provides the systems with a commonly agreed knowledge of the domain of interest. Nevertheless, when such systems make use of very large and complex ontologies, semantic reasoning and then threats eliciting may be extremely time-consuming. In this Chapter, we provide a method for trading-off between the complexity of core ontologies and the speed of threat identification tasks.

3.2 On the accuracy of threat ontologies

Below are some definitions and explanations about terminologies and the list of symbols and notations (see Table 3.1) that will be used throughout this chapter.

An ontology is a tuple characterised by a set of concepts C, a taxonomy induced on concepts, a set of taxonomic relations, a set of terms related to concepts, and both relations and mapping among terms, concepts and relations [58]. In this section, we take into account only the taxonomy H induced on concepts. Taxonomies may be defined as an organization of terms in hierarchical structures. Terms groups, belonging to a taxonomy and selected according to a specific selection function, will henceforth be referred to as taxonomy‘slices. The semantic accuracy of an ontology is a measure of how much coherent and suitable the description of ontological components (e.g., classes and relationships)and their definition are.

This section provides some important insights into building semantically accurate ontologies. To this aim we introduce and describe the measure of differential semantic variance. Let H be a taxonomy. In particular, we suppose that H can be represented

Symbol	Description
C	set of concepts
$\|C\|$	measure of a set of concepts
H	taxonomy induced on concepts of an ontology
c_i	$i-th$ term of a taxonomy
$ROOT(\cdot)$	root element
$\sigma(\cdot)$	semantic variance
$\Delta\sigma(\cdot)$	differential semantic variance
α	depth level in a taxonomy
$dist(\cdot,\cdot)$	distance between two elements of a graph
$evcent(\cdot)$	eigenvector centrality

Table (3.1) The list of symbols and notations used in this chapter.

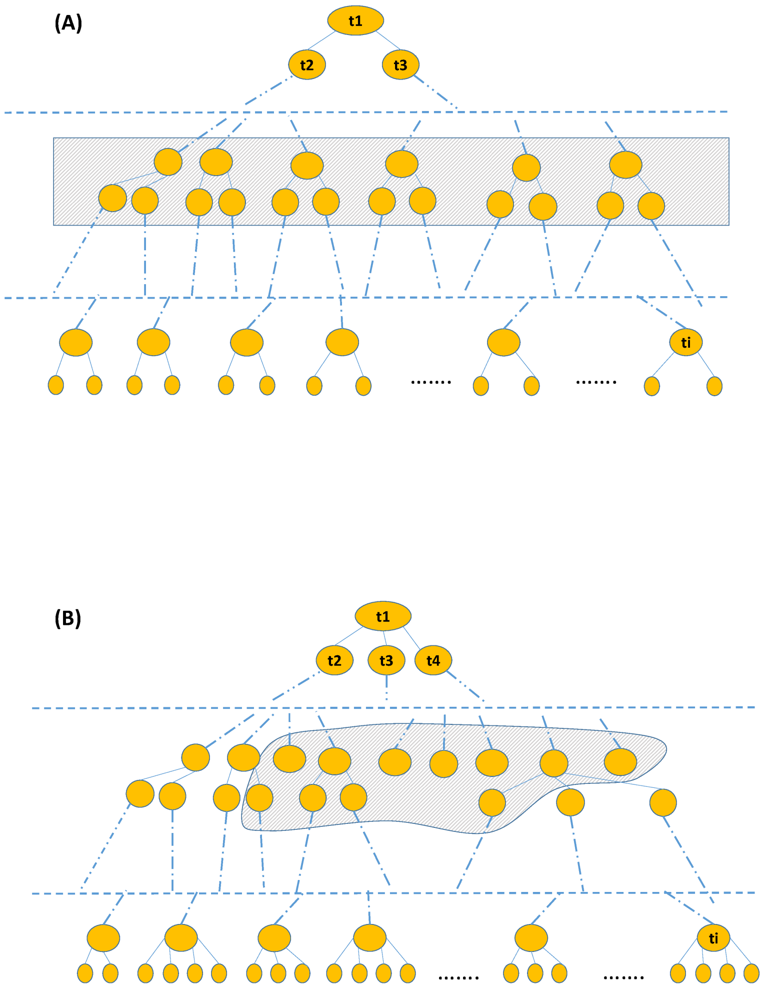


Figure (3-1) (A) Perfect k-ary tree. (B) Real taxonomy

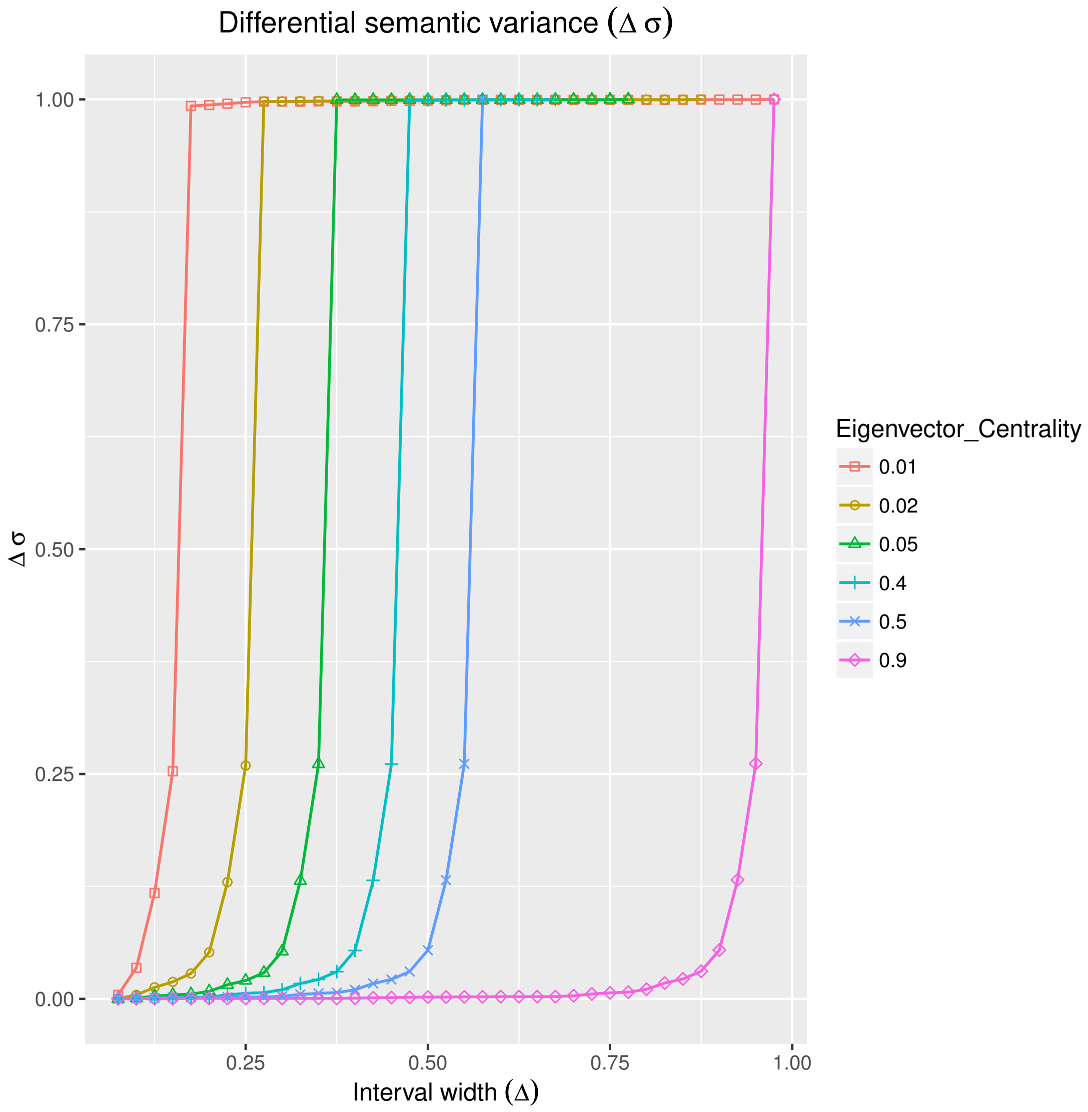


Figure (3-2) In this figure, we show the evaluation of $\Delta\sigma$ as a function of the interval width Δ and for different values of nodes‘ eigenvector centrality values, e_c.

as a perfect k-ary tree. Let N be the number of nodes attached to each branch and L the maximum depth of H. It is easy to demonstrate that the cardinality C of H is equal to:

$$|C|= \frac{N^L - 1}{N - 1}$$

We exploit the definition of semantic distance d between two nodes as in [8]. Thus, given a term c_i at depth k in H, the semantic distance between it and the root of H, denoted by $ROOT(H)$, is equal to:

$$dist(c_i, ROOT(H)) = \log{(1 + \frac{k-1}{k})} \tag{3.1}$$

Recalling the definition of semantic variance σ as in [87] and using equation (3.1)

$$\sigma(H) = \frac{\sum_{c_i \in C} dist(c_i, ROOT(H))^2}{|C|} = \frac{\sum_{k=1}^{L} N^k \log^2{(1 + \frac{k-1}{k})}}{|C|} \tag{3.2}$$

Let $\alpha \in \mathbb{N}$ be a depth level in H and $\Delta \in \mathbb{N}$ an integer number such that $\alpha + \Delta \in [1, L]$, equation (3.2) may be re-written as:

$$\sigma(H) = \frac{\sum_{k=01, k \notin [\alpha, \alpha+\Delta]}^{L} N^k \log^2{(1 + \frac{k-1}{k})}}{|C|} + \frac{\sum_{k=\alpha}^{\alpha+\Delta} N^k \log^2{(1 + \frac{k-1}{k})}}{|C|} \tag{3.3}$$

The second term on the right side of equation 3.3 represents the contribution to the semantic variance of a slice of H whose terms are located at depths between α and $\alpha + \Delta$. In Figure 3-1 A, a taxonomy as a perfect k-ary tree is presented. Since items may be described through a composition of tags mapped within a proper taxonomy H, it could be interesting studying how such tags can influence the semantic quality of item's representation. Dotted area encloses those tags whose depth in H can be represented within the interval [α,$\alpha + \Delta$]. By dividing both members of the equation

3.3 by σ, we can define the function $\Delta\sigma$, named as ‘Differential Semantic Variance’:

$$\Delta\sigma = \frac{\sum_{k=\alpha}^{\alpha+\Delta} N^k \log^2 (1 + \frac{k-1}{k})}{|C|\sigma(H)}, \ \Delta\sigma \in [0, 1] \tag{3.4}$$

Generalising, let us consider a more complex and robust organisation of terms with respect to a perfect tree. A more general approach to taxonomy slicing can exploit the concept of network centrality, where centrality measures offer an indication of the importance of network nodes. In Figure 3-1 B, an example of real taxonomy is shown. Dotted area encloses only those nodes whose eigenvector centrality values are in the range $[e_c - \frac{\Delta}{2}; e_c + \frac{\Delta}{2}]$. The classification of nodes is important to understand in many contest, such as social networks, how it could change radically the dynamics of phenomena inside the networks which influences the behaviours and choices of users [76, 74]. Many measures of centrality have been defined in the literature (e.g., degree, betweenness, Katz)[12]. In our model, we consider the eigenvector centrality e_c as a centrality measure. Below, we will clarify the reasons behind our choice. Let us consider a perfect k-ary tree (see Figure 3-1 A.). For a given value of e_c coherently with the tree structure (i.e., at least one node of H has e_c as eigenvector centrality measure), all the terms of H having centrality equal to e_c are located at the same depth level. Hence, considering only the terms within the following interval of eigenvector centralities $I = [e_c - \frac{\Delta}{2}; e_c + \frac{\Delta}{2}]$, is equivalent to slice H. Then, the measure of eigenvector centrality, indicated by using the function $evcent(\dot{:})$ can be exploited to generalise the equation 3.4 as follows:

$$\Delta\sigma(H, I) = \frac{\sum_{R} dist^2(c_i, ROOT(H))}{|C|\sigma(H)} \tag{3.5}$$

with $R = \{c_i \in H | evcent(c_i) \in I\}$. Both equations 3.4 and 3.5 state that terms (or nodes) showing a very low eigenvector centrality provide key contributions to the ontology’s semantic accuracy. Indeed, the lower the term‘s centrality is, the greater the value of $dist(c_i, ROOT(H))$ is.

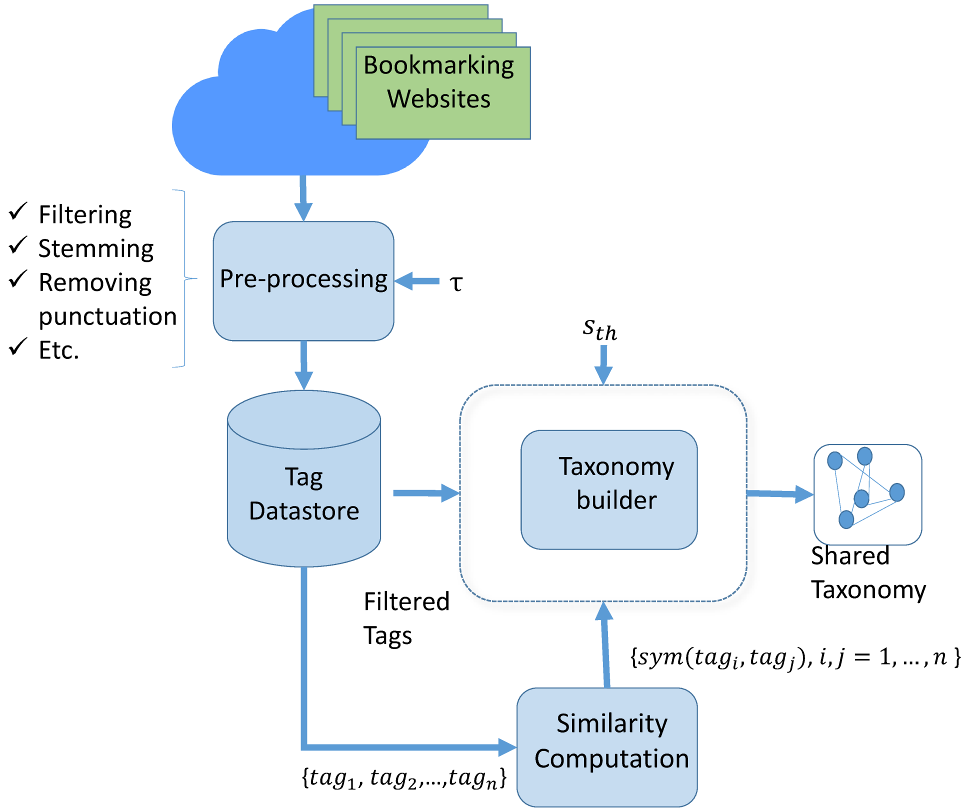


Figure (3-3) Extraction of a taxonomy from the Movielens Database. The output of the tag processing generates a taxonomy that can be used as a shared thesaurus.

3.3 Results

We used the Movielens database [31] as a source of information to estimate the goodness of the proposed method of analysis. Usually exploited to evaluate the performance of semantic reasoning systems, this dataset contains 20, 000, 263 ratings on a 5-star scale applied to 27, 278 movies by 138, 493 users. Moreover, it contains 465, 564 tags inserted by users to categorise movies. Hence, we designed a methodology to build a taxonomy from the aforementioned user-contributed term list, which is also called folksonomy. Our methodology consists of three stages: first, building a taxonomy from a folksonomy; then, analysing its semantic accuracy through the definition of semantic variance; finally, determining the contribution of taxonomy‘s slices to the taxonomy‘s semantic accuracy by introducing the notion of differential semantic variance.

Algorithm 1 Building of Taxonomy from a list of terms

```
Input:
$T = \{tag_i, i = 1, .., n\}$; is the tag set
$TI = \{(tag_i, item_j), i = 1, ...n_R; j = 1, ..., m\}$; is the set of pairs tag-to-item; $n_R \geq n$
$paramenters$ : $\tau$ is the lower bound frequency threshold;
$s_{th}$ is the similarity threshold.
Output: $taxonomy\ H$
  $T_{elb} \leftarrow removePunctuation(T)$
  $T_{elb} \leftarrow removeStopWords(T_{elb})$
  $T_{elb} \leftarrow removeNumbers(T_{elb})$
  $T_{elb} \leftarrow stemming(T_{elb})$
  $T_{elb} \leftarrow stripeWhiteSpace(T_{elb})$
  $T_{filt} \leftarrow filterTerms(T_{elb}, TI, \tau)$ {remove those terms from $T_{elb}$ whose frequency in $TI$ is smaller than $\tau$; this function is a list of terms sorted by descendent generality;}

  $M_{sym} \leftarrow computeSimilarity(T_{filt}, s_{th})$ {computeSimilarity() determines similarity values by couple of terms in $T_{filt}$, and return as an output a symmetric matrix $M_{sym}$. Row names of $M_{sym}$ are labelled with terms of $T_{filt}$}
  $dim \leftarrow nrows(M_{sym})$ {number of rows in $M_{sym}$}
  $parent \leftarrow rowName(M_{sym}[1, 1])$
  $H \leftarrow append(parent)$
  for ($row\ in\ 2\ :\ dim$) do
    for ($col\ in\ row\ :\ dim$) do
      if $M_{sym}[row, col] >= s_{th}$ then
        $parent \leftarrow append(rowName(M_{sym}[row, col]))$
      else
        $residualTerms \leftarrow rowName(M_{sym}[row, col])$
      end if
    end for
    $parent \leftarrow rowName(M_{sym}[row, 1])$
  end for
  for ($term_c\ in\ residualTerms$) do
    if ($term_c\ not\ in\ H$) then
      $H \leftarrow append(term_c)$
    end if
  end for
```

Our steps for building a taxonomy (see Algorithm 1 and Figure 3-3) are similar to those indicated in [9] and allow elaborating folksonomies extracted from one of many social bookmarking websites (e.g., BibSonomy, Delicious, Reddit, Pinterest, Digg). To remove punctuation, stop-words, common words endings, and white-spaces, the tags should be pre-processed first, sorted by generality and then filtered by frequency.

Terms are filtered by frequency through the introduction of the parameter τ, which acts as a lower bound frequency. Then, a tag-to-tag co-occurrence network is built. We evaluated a tag as the most generic one, when its centrality was the greatest in the co-occurrence network. Differently from [9], instead of degree centrality as a measure of centrality, we have considered the eigenvector centrality [13], which extends the concept of degree centrality, by quantifying not only the number of links of each node in the network, but also the quality of such connections [20, 75]. A hierarchical scheme (i.e., a taxonomy) has been assembled by taking into account the most generic term t_1 (from now on called 'root node') of the above mentioned sorted list, and and thus combining the latter with less generic but more similar terms (i.e., root‘s children). Similarities among terms have been obtained by exploiting the “Swoogle” semantic search engine [21].Then, the procedure has been iteratively repeated for all root‘s descendant terms.

The data set includes information on the tag-genome, namely a set of tags that can be used to encode movies‘ properties. After we derived the number of occurrences for each tag, we computed pair-wise similarities for all those tags whose frequency was greater than a threshold τ. Our method represents an alternative way to predict tag relevance for a specific domain. Overall, a solution for this problem requires evaluations by domain experts or users communities.

In this work, tags‘ relevance has been evaluated by measuring their eigenvector centrality in H. According to our analysis, relevant tags are all those localised in a slice of H marked by a small value of eigenvector centrality. In order to measure the contribution of the semantic accuracy of tag-genome‘s tags to the emerging ontology, we built a taxonomy H and mapped the tag-genome‘s tags onto H. Then, we measured their eigenvector centrality.

Table 3.2 reports the percentage of tag-genome‘s nodes having eigenvector centrality within specific ranges. Our results are consistent with previous achievements: indeed, the most semantically relevant tags (i.e., having small values of eigenvector centrality) of the 'Movielens' data set, obtained by using the differential semantic variance, contain most of the tag-genome. As we can see from Table 3.2, when $e_c \leqslant 0.2$,

Eigenvector centrality range	Percentage of tag-genomes nodes as function of eigenvector centrality ranges
> 0.9	0.1
$0.5 - 0.9$	0.4%
$0.4 - 0.5$	0.4%
$0.3 - 0.4$	0.5%
$0.2 - 0.3$	2.3%
$0.1 - 0.2$	8.8%
$0.05 - 0.1$	18.5%
$0.02 - 0.05$	32.5%
$0.01 - 0.02$	36.5%

Table (3.2) Eigenvector centrality distribution of all tag-genome‘s nodes

our results contain more than 0.96% of the tag-genome. In general, the set of tags identified through our approach depends on the choice of Δ and e_c. Thus, by tuning these parameters, the proportion of tags contained into the tag-genome can be varied. Moreover, the majority of the tag-genome‘s nodes has very low eigenvector centrality values, and only a few tag-genome‘s tag (little more than 1%) show higher eigenvector centrality values. Thus, most of the tag-genome‘s tag are confined within a well-defined region R of the taxonomy H, in which all nodes exhibit low eigenvector centrality values.

As mentioned in the literature review (see Chapter 2), ontologies can be generated by properly elaborating folksonomies [14, 32]. Previous studies evaluating such folksonomy-sourced ontologies [9] highlighted how the user community knowledge includes semantic information and that their quality is comparable to that of manually-built ontologies. A numerical evaluation of their accuracy (or quality) has been proposed in [87] by computing the variance of semantic dispersion of ontologies‘ taxonomic structures.

The most remarkable result emerging from our analysis is that the identification of semantically relevant tags from a tag space could be carried out through network theory. Furthermore, our approach (see equation 3.5) allows us to obtain results similar to those produced by evaluations from a selected group of users of a tagging system [77]. It should be noted that the tag set we identified is larger than the tag-genome, so that it includes and extends the tag-genome. A possible reason behind these differences might be that while tag-genome‘s tags relevance depends on both explicit and implicit users‘evaluation (consequently, it may be influenced by personal inclinations, indecisions, and opinions), in our work tags are evaluated only on the basis of implicit and aggregated user‘s behaviours (i.e., the number of users of a specific tag by a certain number of users).

3.4 Conclusion

As to secure next generation communication systems, reasoning systems have gained momentum more than ever. The complexity of the current and future communication networks makes threats mitigation very challenging.

In this chapter, we propose a method, based on the network theory and a novel definition of differential semantic variance, for reducing the complexity of security ontologies. Findings suggest that the semantic relevance of portion of ontologies could be determined through network theory and starting from the measure of eigenvector centrality, thereby also reducing the employment of domain experts or user communities.

As to assess the contribution of taxonomy terms to the semantic accuracy of ontologies, we defined the concept of differential semantic variance. It is a novel measure of the contribution to the semantic dispersion of slices of taxonomic structures. Moreover, we have introduced a methodology which can represent an important step forward towards the simplification of the automated process for security analysis.

Chapter 4

Evaluation of IoT Privacy in Ultra-Dense Networks

In the IoT, devices, equipped with sensing, processing, storage and decision-making capabilities, actively interact with one another and with humans. Although their design could strictly adhere to the principles of privacy and security, several factors, such as weak implementations of communication protocols, metadata information exchange, and architectural flaws, could jeopardise the security and privacy their owners. Moreover, the ultra-densification trend of the current communication infrastructure combined with its complexity and variability rises new threats to the privacy. In this chapter, we provide a brief introduction to privacy issues in the IoT. Afterwards, we describe how the evolution of the current wireless communication infrastructure might worsen the privacy problem in the IoT. Then, we propose a methodology that analyses and identifies privacy threats from different perspectives and at various levels of abstraction.

4.1 Introduction

By definition, the IoT is a composition of physical entities capable of sensing, computing and acting in response to the information they can acquire and manage [78]. Thanks to this paradigm, “people and things can be connected anytime, anyplace, with anything and anyone, ideally using any path/network and any service" [37].

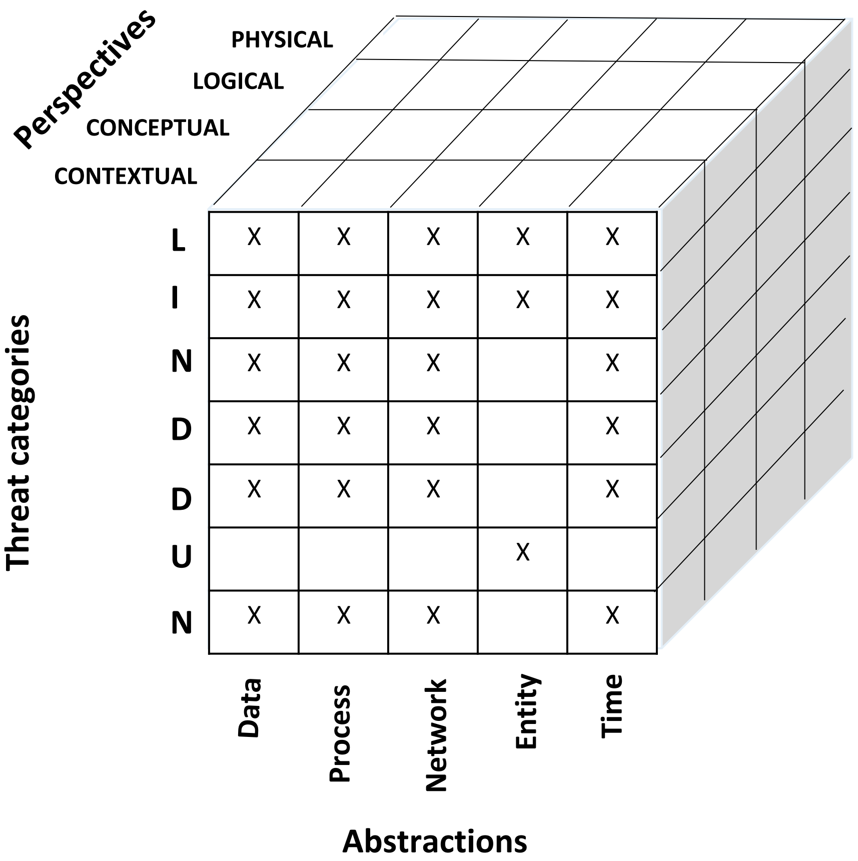
Mobility, scalability, interoperability, and resource constraints characterise the million interconnected both wireless and wired devices of which the IoT is composed [64]. Ubiquity is one of the most important key features expected for the underlying communication support. Undoubtedly, cellular networks, due to their diffusion, enable IoT implementation and also provide stable transmissions and acceptable delays. However, they cannot support machine-to-machine (M2M) communication. Indeed, the intermittent behaviour and small-sized data packet characterising M2M transmissions might easily exceed their uplink capacity.

The most data traffic of communication networks is expected to come from smart devices in the future [18]. Hence, it is crucial that the current cellular network will develop to foster the broad deployment of IoT systems and applications. In this direction, the cellular communication infrastructure could cooperate with other wireless network technologies (e.g. WLAN, relay-assisted and device-to-device communications, wireless personal area networks, LTE-U). Furthermore, since UDNs can allow very high connectivity and data rate, the evolution of networks towards the ultra-dense paradigm could meet the future systems communication requirements.

4.1.1 Motivation

Although the benefits that it may produce, the IoT might cause severe security implications. Inability or unwillingness of devices owner to update and fix devices' security flaws, limited capability of devices, and the lack of, or incompatibility among communication standards make hard addressing the security challenges in the IoT [52].

Leakage of sensitive information is one of the most severe menaces to the privacy. In fact, since they are often equipped with resource-limited microcontrollers, devices in the IoT shall not have strong security and cryptographic functions [51]. For this reason, a growing body of literature has evaluated and proposed lightweight encryption algorithms and privacy-by-design methodologies. Moreover, since it is an evolving, heterogeneous, and broad technological environment, it could be very diffi-



	Data	Process	Network	Entity	Time
L	X	X	X	X	X
I	X	X	X	X	X
N	X	X	X		X
D	X	X	X		X
D	X	X	X		X
U				X	
N	X	X	X		X

Figure (4-1) Illustration of coupling among privacy threat categories, elements of system models (named as *Abstractions*), and perspectives of elements descriptions.

cult staking the privacy to the whole IoT. Furthermore, the paradigmatic revolution that is already overwhelming current communication networks raises new security challenges. Then, a method of investigation for identifying privacy weaknesses fitting well with the complexity of the IoT would be beneficial.

4.1.2 Contribution

The aim of this chapter is twofold. First, it is to provide privacy engineering of a privacy assessment tool. Second, it is to address the privacy of the complex and heterogeneous IoT. By taking inspiration from the popular Zachman and the LINDDUN frameworks, this study provides a systematic approach to exploring the privacy domain of the IoT. It is a bottom-up methodology of analysis that exploits different standpoint of system models to disclose new privacy threats and to reason around

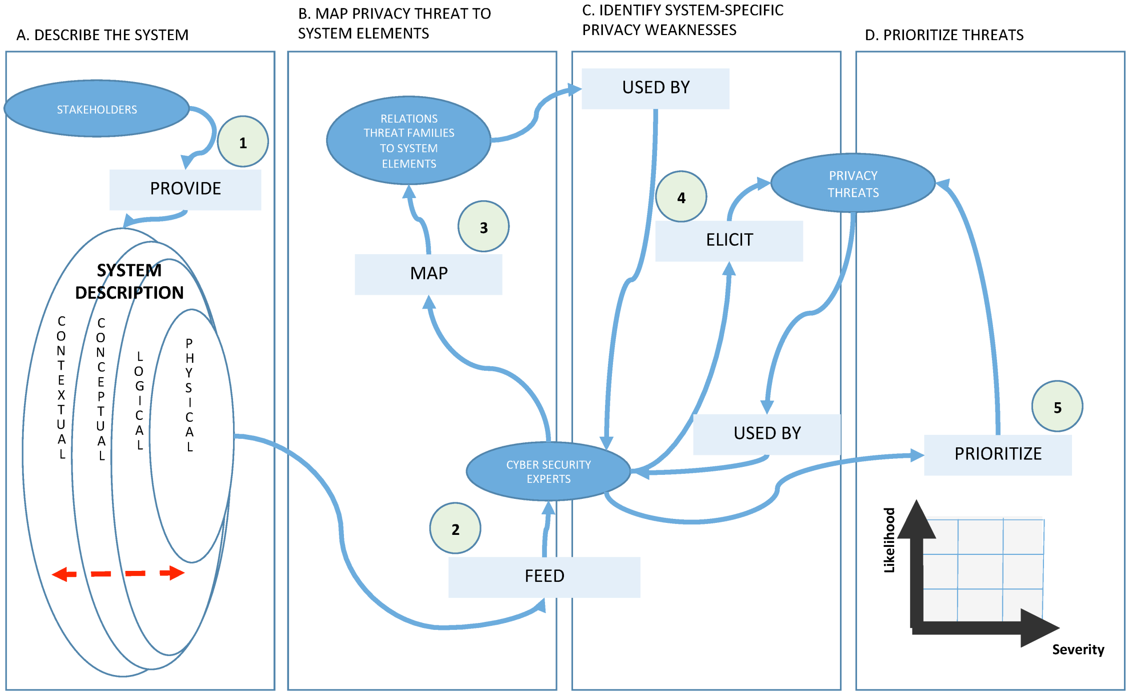


Figure (4-2) This figure shows the proposed privacy threat modelling. The method consists of four steps, namely "describe the system", "map privacy threats to system elements", "identify system-specific weaknesses", and "prioritise threats". Actors and related actions they perform during the threat identification are represented respectively as ellipses and rectangles.

the causes of well-note ones.

Here we argue that although one can exploit specific protection approaches in the design and implementation stages, privacy objectives in the IoT could not be achieved because of its complexity. Then, only a comprehensive understanding of the motivations behind privacy weaknesses might lead to the identification of proper mitigation actions. Therefore, we propose an organic, multi-faceted methodology of analysis for the wide and diverse IoT.

4.2 UDN and IoT Privacy

Security and privacy in the IoT are attracting widespread interest due to pervasiveness and diffusion of new IoT applications. Moreover, recent cyber-attacks (e.g., Mirai) have shown how the IoT ecosystem could be insecure and prone to be exploited for malicious activities [62].

Limited capabilities of devices, due to assembling components and energy constraints, together with massive wireless communications enhance the likelihood, effectiveness and impact of privacy attacks against the IoT ecosystem. In fact, devices cannot implement powerful security functions because of their limitedness. It all adds up to the wireless communication exposure to eavesdropping and other security attacks (i.e. jamming attacks). Furthermore, new and challenging dangers may gain strength as the underlying communication infrastructure evolves towards the ultra-dense model [17]. In this section, we analyse the effect of the deployment of resource-limited devices under UDNs coverage. UDNs are defined as networks in which the density of access nodes is at least a magnitude greater than the those of users. They can effectively cope with the future networks data requirements, also providing energy and spectrum efficiency. Composed of heterogeneous nodes with different radio access technologies (e.g. LTE, Wi-Max, IEEE 802.15.x), transmit powers, and coverage area, UDNs are characterised by a multi-tier architecture (see Figure 2-1). In detail, high-power nodes and low-power nodes, with large and small radio coverage, are placed respectively in macro-cell tiers and in small-cell tiers. Cellular communication infrastructure, if from the one hand make it possible to offer ubiquitous connectivity to the most devices, from the other hand is inefficient for transmitting small, infrequent data of M2M communications. Moreover, communications under cellular network coverage could make it possible to track events and entities (i.e., access points and subscribers) involved in data transmissions [7, 15], thus affecting their location privacy.

The spatial distribution of low-power nodes might influence the whole network security, as asserted in [17]. Specifically, the probability of positive secrecy rate, that is the capacity deviation of the operating channel from the eavesdropper channel, increases as the density of low-power nodes growths (until a critical point, after which is not observed any enhancement in term if secrecy performance). Moreover, the higher the density of transmitting entities is, the higher the risk of information eavesdropping [94]. Undeniably, while moving under a UDN coverage, entities are likely to be subject to more handovers than they do in conventional networks, making it possible for untrusted subjects to observe the just mentioned processes and acquire precious information.

Albeit finding trusted security organisations responsible for credential distribution could solve the above-mentioned problem [86], undesired network delays due to a large number of involved devices, in addition to high costs, make their adoption infeasible in practice. Physical layer security is a valid alternative to the adoption of third-party security providers and to the cryptographic approach. Indeed, in addition to having high scalability, it does not require the execution of complex operations. Even computationally powerful adversaries, in fact, cannot compromise the network security [94].

In [96] Yu et. al underlined that systematically mitigating the security issue in UDN requires defining a framework able to explore effectively the UDN's attack space. They claimed that developing better mitigation strategies against security attacks requires a meticulous comprehension of systems weaknesses. Only then, it is possible to determine the objectives, the targets and the impact of attacks. Anyway, the proposed framework does not address the complexity of the IoT.

In [17] we found an early attempt to offer some point of consideration on security and privacy of devices communication in UDN. Chen et. al investigated the wireless network security in view of the cellular network densification trend. Then, they identified weak links in the security and privacy chain in both network and device domains. However, their work does not provide any address to systematically mitigate security and privacy issues.

4.3 Methods

To tackle the problem of privacy in the IoT, we propose an assessment methodology which combines the popular ZF [97] with the LINDDUN framework [93]. The proposed approach aims at providing a tool for acquiring awareness about, and then react to, privacy weaknesses that might affect the system from both microscopic and macroscopic perspectives. LINDDUN is mainly a methodological approach, which uses data flow diagrams to list entities, processes, data flows, and data stores. Then, by mean of further successive steps, it maps, elicits and prioritises the threats, guiding towards the identification of mitigation strategies and privacy enhancing technologies.

The ZF allows logically organising and classifying artefacts involved in the design and development of information systems. Different perspectives match with different aspects of the system, allowing decomposing the verification of privacy properties in small, though sometimes interdependent, modules. Privacy assessment on IoT applications is a large complex task that requires a systematic verification approach on both software and hardware.

We claim that the LINDUNN framework is not well-suited to address the privacy problem in the IoT domain. Indeed, it does not consider the effect of physical location in which event happens (e.g., authentications, data exchange) neither correlates locations with time information. That information together with knowledge of entities, data flows, and processes might help to understand the motivations behind privacy issues and to identify better privacy enhancing solutions.

Here we give some explanation about the "Perspective" dimensions of our proposal.

- Contextual (i.e., what the system should do): refers to the description of information, processes, locations, entities, events, and motivations. It gives an overall, also non-detailed, view of purposes, extents, and relationships among elements of the IoT ecosystem or its subsystems.

- Conceptual (i.e., how the system should operate): gives an overview of models, semantic relationships, and processes.

- Logical: analyses the processing structure, how applications are architected, rules, and information models.

- Physical: aims at analysing the IoT from the technical points of view (technology constrained models) providing information on physical quantities and parameters.

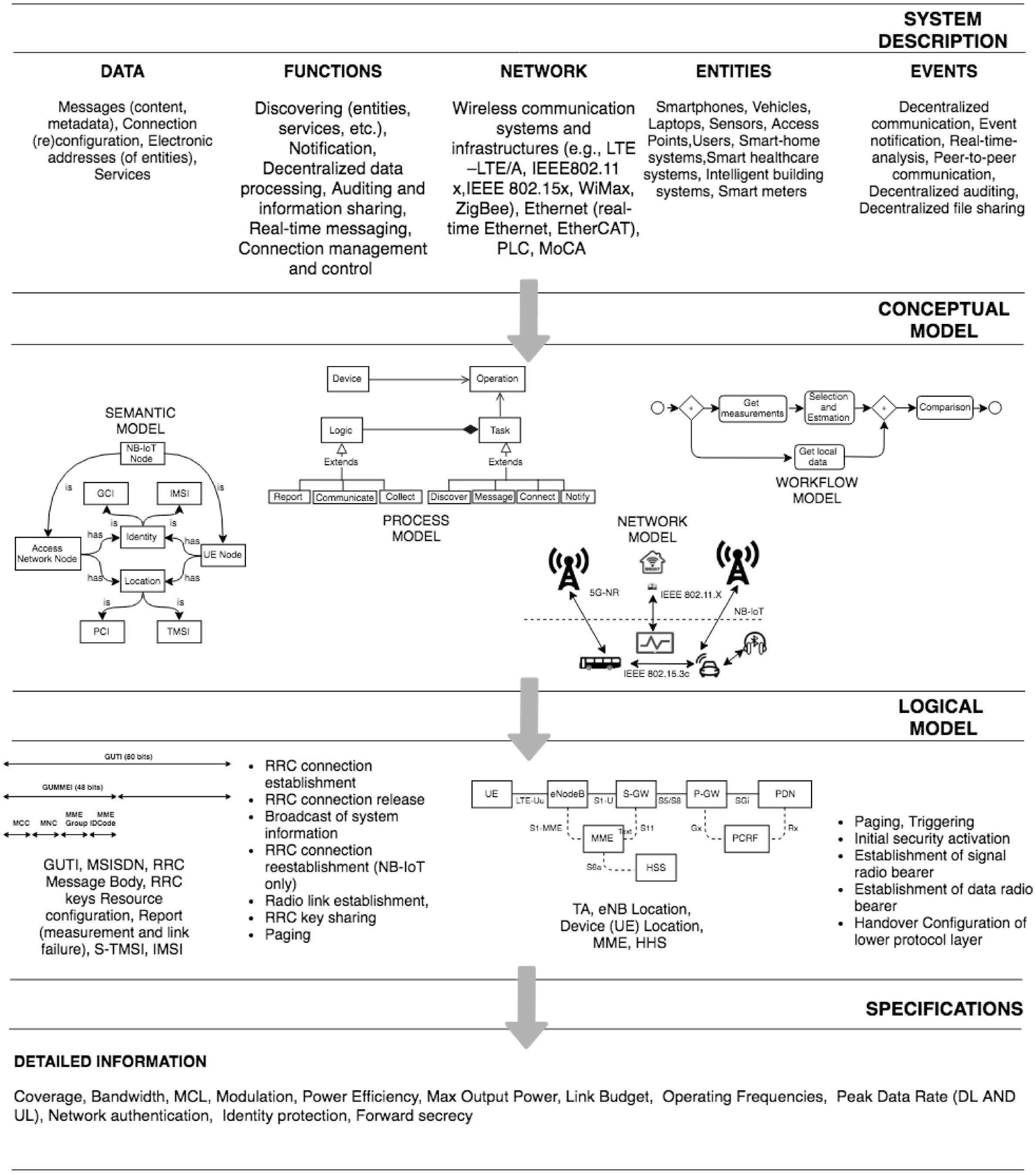

SYSTEM DESCRIPTION
DATA
Messages (content, metadata), Connection (re)configuration, Electronic addresses (of entities), Services
FUNCTIONS
Discovering (entities, services, etc.), Notification, Decentralized data processing, Auditing and information sharing, Real-time messaging, Connection management and control
NETWORK
Wireless communication systems and infrastructures (e.g., LTE –LTE/A, IEEE802.11 x,IEEE 802.15x, WiMax, ZigBee), Ethernet (real-time Ethernet, EtherCAT), PLC, MoCA
ENTITIES
Smartphones, Vehicles, Laptops, Sensors, Access Points,Users, Smart-home systems,Smart healthcare systems, Intelligent building systems, Smart meters
EVENTS
Decentralized communication, Event notification, Real-time-analysis, Peer-to-peer communication, Decentralized auditing, Decentralized file sharing
CONCEPTUAL MODEL
SEMANTIC MODEL
NB-IoT Node
is
is
GCI
IMSI
is
is
has
Identity
has
Access Network Node
UE Node
has
Location
has
is
is
PCI
TMSI
Device
Operation
Logic
Task
Extends
Extends
Report
Communicate
Collect
Discover
Message
Connect
Notify
PROCESS MODEL
Get measurements
Selection and Estmation
Comparison
Get local data
WORKFLOW MODEL
NETWORK MODEL
5G-NR
IEEE 802.11.X
NB-IoT
IEEE 802.15.3c
LOGICAL MODEL
GUTI (80 bits)
GUMMEI (48 bits)
MCC
MNC
MME Group
MME IDCode
GUTI, MSISDN, RRC Message Body, RRC keys Resource configuration, Report (measurement and link failure), S-TMSI, IMSI
• RRC connection establishment
• RRC connection release
• Broadcast of system information
• RRC connection reestablishment (NB-IoT only)
• Radio link establishment,
• RRC key sharing
• Paging
UE
LTE-Uu
eNodeB
S1-U
S-GW
S5/S8
P-GW
SGi
PDN
S1-MME
MME
S11
Text
Gx
PCRF
Rx
S6a
HSS
TA, eNB Location, Device (UE) Location, MME, HHS
• Paging, Triggering
• Initial security activation
• Establishment of signal radio bearer
• Establishment of data radio bearer
• Handover Configuration of lower protocol layer
SPECIFICATIONS
DETAILED INFORMATION
Coverage, Bandwidth, MCL, Modulation, Power Efficiency, Max Output Power, Link Budget, Operating Frequencies, Peak Data Rate (DL AND UL), Network authentication, Identity protection, Forward secrecy

Figure (4-3) Top-down system analysis based on the ZF. ZF uses a two-dimensional matrix for classifying and describing the artefacts created during the design and development of information systems. Technical information reported in this figure was retrieved using a taxonomy on the IoT [95] and a smartphone data taxonomy [57]

Each perspective aims at identifying privacy threats by analysing systems from different viewpoints and may be ground for investigations into the threats causes. Notably, the description of privacy vulnerabilities can be carried out from various angles. Furthermore, data, processes, locations, entities, and event (see Figure 4-3) can be related to each other. For example, the connection (re)configuration information may be related to the connection management and control processes (see Figure 4-3).

Privacy threats can be grouped into seven families, that is linkability, identifiability, non-repudiation, detectability, distinguishability, unawareness (of information content), and non-compliance to policy. For the sake of completeness, we report the definition of threat categories, as indicated in [93]. Linkability occurs when two entities can be related to each other. Identifiability refers to a capability of an adversary to infer the identity of an entity. Non-repudiation stands for the inability of a subject to demonstrate that he could not carry out a specific action. Detectability implies that it possible detect whether an entity exists or not. Disclosure of information happens when individuals information can be accessed by unauthorised entities. Unawareness is related to unconsciousness about supplied information to the system. To conclude, non-compliance refers to the inability of the system to be compliant with regulations, policies, and agreements with users.

4.4 Privacy Threat Analysis

IoT applications require both data and communications security, in addition to ubiquitous connectivity. In this section, we list and analyse the IoT abstractions at each perspective (see Figure 4-1).

4.4.1 Contextual and Conceptual Perspectives

The Contextual view reviews and tackle the privacy problem from a very non-concrete perspective. Representing systems architectures at a high-level of abstraction may allow identifying critical elements involved in communication processes. Figure 2-1 provides a simplified view of the IoT in UDN. Smart homes and their appliances, vehicles, and user equipment are some of the interconnected entities within the network. Access points may be deployed both in public and in private areas. Both application fields and protection objectives affect privacy specifications. For example, in smart home systems, a privacy objective could be concealing presence or absence of persons, consumption habits, and apparatus installed inside houses. To give just a few examples, in pay-as-you-drive insurance, black-box car insurance, and car-sharing services objectives could be protecting routes and information of guide style of drivers. Such a measure guarantees protection against linkability, identifiability and disclosure of information threats. For the listed cases, avoiding fine-grained information communication and using encrypted channels might reduce the risk of private information disclosure. When a device communicates sensed data to a remote service, linkability, identifiability, detectability, and disclosure of information threats might violate the system. Issues might derive from identifiability of remote services to which they connect. Further, problems might become more threatening when devices settings and services are set only by the manufacturers and cannot be modified by end users. As an instance, traffic analysis might be sufficient to identify smart appliances installed within a home. Hence, adversaries, by exploiting known vulnerabilities of devices, could steal or infer private users‘ information. By mean of modelling techniques it is straightforward to report the just discussed problems to the more specific conceptual perspective (see Figure 4-3). Although both wired and wireless communications can be studied, we deepened only the latter throughout this thesis. Radio communications are by nature exposed to eavesdropping. Then, the probability of privacy violations in wireless communications is higher than in wired interconnections. We considered as a communication infrastructure, a multi-tier, ultra-dense and heterogeneous network. Several wireless communication technologies and their related protocols (such as IEEE 802.11x, IEEE 802.15.x, WiMax, ZigBee, and LTE/LTE-A/LTE-U) could be analysed. Anyway, we focused on LTE-based technologies.

4.4.2 Logical and Physical Perspectives: A Protocol Layer-Wise Privacy Issues Identification

In this section, we use the proposed method to identify the privacy issues at different Network Protocol Stacks. Logical analysis is wider-ranging than that of the Contextual and Conceptual. In fact, many technologies should be analysed (see Table 4.1). Anyway, in this thesis, we studied only one LTE-based communication technology, namely the Narrow Band IoT (NB-IoT).

Table (4.1) Comparison of IoT technologies [3, 1, 2, 4]

	Licensed		**Unlicensed**	
	NB-IoT	eMTC	LoRaWAN	Sigfox
Coverage	<15 Km	<11 Km	<11 Km	<13 Km
Bandwidth	180 KHz	1.4 MHz	125 KHz	200 KHz
MCL	164 dB	164 dB	157 dB	153 dB
Modulation	OFDMA	OFDMA	SS Chip	UNB / GFSK / BPSK
Battery Life	>10 years	>10 years	>10 years	>10 years
Power Efficiency	Medium High	Medium	Very High	Very High
Max Message per Day	Unlimited	Unlimited	Unlimited	UL : 140 Msgs /day
Max Output Power	20 dBm	23/30 dBm	20 dBm	20 dBm
Link Budget	146 dB	150 dB	154 dB	151 dB
Operating Frequencies	700–900 MHZ	700–900 MHZ, 1.4 GHz	400–900 MHz	800 MHz
DL Peak Data Rate	234.7 kbps	800 Kbps	50 kbps	600 bps
UL Peak Data Rate	204.8 kbps	1 Mbps	50 kbps	100 bps
Cost	Low	Medium	Low	Very low
Mobility	Limited	Yes	Yes	Limited
Localization	No	No	Yes	No
Globally unique Identifiers	IMSI	IMSI	Optional (DevEUI)	Yes (32 bits)
Network authentication	LTE AKA	LTE AKA	Optional	No
Identity protection	TMSI	TMSI	Partial (DevAddr)	No
Forward secrecy	No	No	No	No

Table (4.2) PHY-layer privacy vulnerabilities [65, 46, 60]

Data	Process	Network	Entity	Event	Threat Category
PSS-SSS	Downlink synchronization	UE and eNB Location	UE-eNB	Initial Access	Linkability Identifiability Non-repudiation
PBCH-PHICH	Downlink Broadcast information	UE and eNB Location	UE-eNB	Initial access	Linkability Non-repudiation
PCFICH -CFI-PDDCH	Control information	UE and eNB Location	UE-eNB	Resource allocation	Non-repudiation Detectability
PDSCH–SIB1	Downlink transport information	UE and eNB Location	UE-eNB	Downlink data transport	Non-repudiation Identifiability Linkability

In low-power wireless area networks (LPWANs), low-cost technologies and implementations, and device positioning might affect the network coverage. Anyway, thanks to solutions such as retransmission and low-frequency modulation, LPWAN can overcome the coverage problem. These techniques make requirements in term of signal strength and signal-to-noise ratio (SNR) less stringent than in conventional technologies.

As an example, NB-IoT specifications consider acceptable a maximum coupling loss 20 dB greater with respect LTE (see Table 4.1). In this case, coverage enhancement is obtained through signal retransmissions (until 128 times for the uplink and 2048 times for the downlink). Then, the receivers combine the multiple copies of the same received signal until the resulting SNR becomes acceptable. It is obvious that multiple retransmissions of the same information make signal senders seriously exposed to eavesdropping.

Exploiting the LTE network as a part of the IoT communication infrastructure produces economic benefits, provides pervasive connectivity and offers a certain level of security of communications, as well. Indeed, it integrates various authentication

Table (4.3) Entities, protocols, and procedures involved during the UE-to-Core network communications (in LTE)

#	Symbol	Description
Paging		Refers to the process in which the MME needs to locate an UE in a particular area and to deliver network services, such as incoming calls.
Radio Resource Control	(RRC)	Includes a set of functions to manage connectivity between UE and eNodeB, that is broadcasted information (sent by eNBs over a broadcast channel) and UE measurement reports or radio link failure (RLF) sent by UEs
Access Stratum (AS)		Is a functional layer within LTE protocol stack, responsible for radio resource management and data transportation over the wireless channel
Access Stratum Security Context		The purpose of AS security context is to deliver RRC messages between an UE and an access point (eNodeB) through the control plane, and IP packets through the user plane using AS security keys.
Radio Link Failure report	(RLF)	It allow detecting connection failures caused by intra-LTE mobility and intersystem handovers between LTE, GSM, and 3G networks.
Measurement report		It includes throughput measurements, latency, reference signal received power (RSRP), received signal strength indicator (RSSI), as well as information about dropped calls and, sometimes, latitude and longitude.

and encryption algorithms (e.g., EPS AKA, SNOW 3G, MILENAGE). Therefore, as asserted in [80], embedding a SIM card into devices provides security arrangements to the IoT. Anyway, as explained below, security and privacy of devices might still be under risk. As to explain the motivations behind the afore statement, it is useful to provide an overview of the LTE Radio Resource Control (RRC) protocol.

Let us briefly introduce the concept of access stratum (AS) security. The AS security keys are generated every time a new radio link is established (that is when a mobile device moves from the IDLE state to CONNECTED state). When the AS security setup is completed, the mobile device (UE) and the eNodeB share an RRC integrity key, an RRC encryption key, and a user plane encryption key. Here we report the procedure described in LTE specification to locate UEs. The MME generates a paging message (see Table 4.3 for further details) and forwards it to the eNodeBs within a tracking area (TA). Thus, eNodeBs broadcast a radio RRC paging message[49]. Paging messages contain identities of UEs such as serving temporary mobile subscriber identitys (S-TMSIs). S-TMSI is a temporary identifier and it is part of a global unique temporary identifier (GUTI). When they are in the IDLE state, UEs decode RRC paging messages and search for their International Mobile Subscriber Identitys (IMSIs) in it. If their IMSIs matches, UEs initiate a new Attach procedure to receive a GUTI. RRC messages specify UEs which information it should be returned in response (e.g., Measurement report or RLF report). The reported behaviour of the RRC protocol, although with some modification, is still observable in two downstream licensed technologies, namely Enhanced Machine Type Communication (eMTC) and NB-IoT [49].

As to cope with small data transmissions, NB-IoT is provided with two optimizations, that is RRC connection suspend/resume procedure, and data transmission over control plane signalling [66]. Moreover, it is not provided with measurement reports and handover management [48]. Until the serving eNodeB does not release the connection or a link failure happens, NB-IoT devices stay in the connected mode. When the connection is interrupted, they go to the idle state. When necessary, it can trigger the RRC connection reestablishment procedure.

Usually, cellular networks protect the identity of subscribers by providing them with temporary identifiers. Unfortunately, in some circumstances, unique identifiers (IMSIs) of connected devices can be accessible to malicious entities. Indeed, if triggered when users are in IDLE state, RRC paging messages could be exploited to correlate IMSIs and GUTIs to TAs [40, 48]. In fact, RRC paging lacks encryption protection in its first phase [79]. Thus, when an NB-IoT device crosses a cell boundary (in case of moving or transported device) and the MME generates a paging message to locate it, the IMSI could be sent in clear. Hence, an adversary could steal such an information and identify the device's location. These problems add up to the known vulnerabilities and shortcomings affecting the NB-IoT (see Figure 4-4). For the sake of completeness, in Table 4.2 we reported some results of the privacy analysis on the physical layer of the LTE protocol stack. Follows the description of identified vulnerabilities:

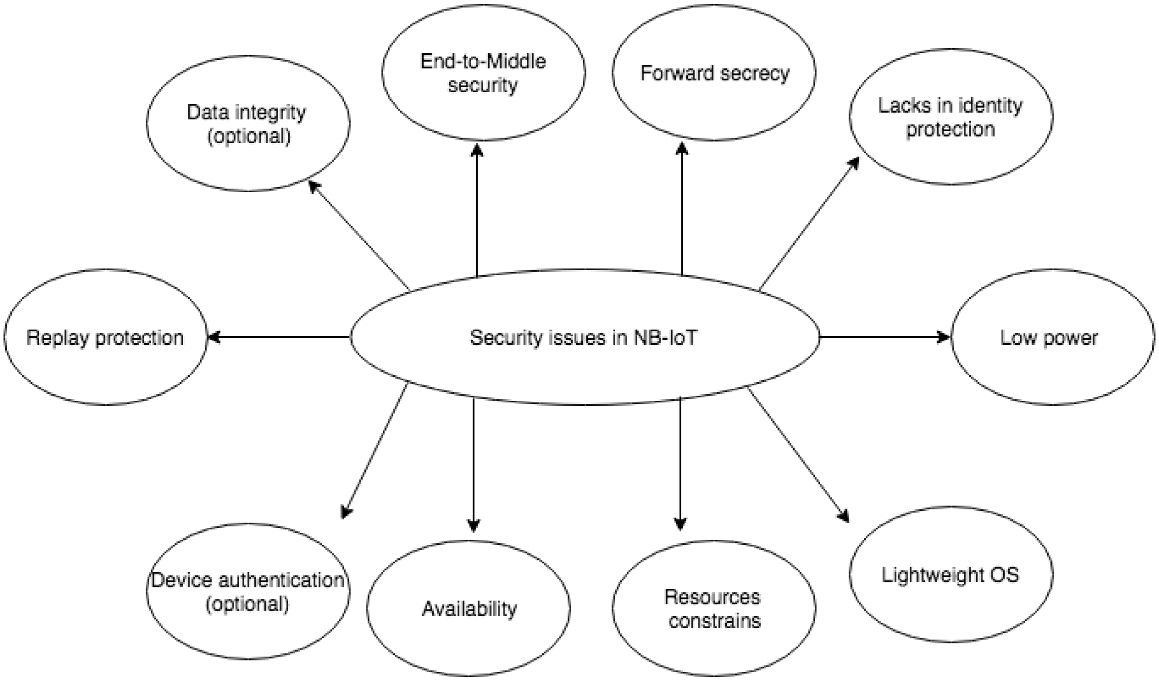

End-to-Middle security
Forward secrecy
Data integrity (optional)
Lacks in identity protection
Replay protection
Security issues in NB-IoT
Low power
Device authentication (optional)
Availability
Resources constrains
Lightweight OS

Figure (4-4) This figure reports the main security drawbacks and shortcomings affecting NB-IoT [16]

PSS - SSS By detecting the Primary Synchronization Signal (PSS), the UE determines the cell's physical layer identity and acquires time and frequency synchronization. The Secondary Synchronization Signal (SSS) provides the UE with the physical cell identity group. The physical cell identity group together with the physical layer identity provides the full Physical Cell Identity (PCI). Through the SSS, the UE also learns about the Cyclic Prefix (CP) type and the duplexing mode used by the cell.

PBCH - PHICH Broadcast channel carries Master Informaition Block (MIB) information, which is required by any UE to get initial access to the cell. This allows an adversary to sniff this traffic and extract all details about cell and network configurations.

PCFICH – CFI - PDDCH UEs decode the CFI value from Physical Control Format Indicator Channel (PCFICH). PCFICH gives the information of the frame structure carrying Physical Downlink Control Channel (PDCCH). Hence, PCFICH is the key to decode the control information and becomes highly vulnerable.

PDSCH – SIB1 Physical Downlink Shared Channel (PDSCH) carries SIB messages, which are not encrypted. SIB1 message carries vital information like Public Land Mobile Network (PLMN) identity.

4.5 Discussion

This chapter aims to systematically study the privacy problem of the IoT in ultra-dense networks. An effective privacy threat mitigation strategy targeted at the complexity of the IoT ought to analyse the security problem from multiple points of view. In this context, the scientific literature [91, 41, 93, 64] has stressed the need for a privacy-aware methodological approach. With this in mind, we provided a methodology that combines the capability of the LINDDUN framework to disclose privacy issues, with the multi-view descriptive capability of the Zachman framework. We would like to underline that our approach fits well with [41]. Indeed, it leads the threat discovery process from a high-level system description towards detailed system representations, thus addressing privacy threats originating from IoT subsystems. Taking the IoT model as an input, it provides cues on privacy vulnerability to and from each IoT network component and process.

As to better observe the complexity of phenomena and interactions in the IoT, our method of studying extends the LINDDUN framework, including new dimensions, namely the “Network” and the “Time”. By accessing such pieces of information even in the early stages of analysis, one might improve chances to identify privacy weaknesses and to react to them. Indeed, even small data leakage on time, location and their combination, might reveal ties among persons [19], but also between persons and devices. The privacy traits on which our method focuses on, namely identity privacy, data minimisation, temporal and location privacy, agree reasonably well with the guidelines reported in [64]. Moreover, in line with [41], our approach pays particular attention to correlation among non-personal information. Indeed, linking them together may disclose personal and sensitive information on IoT users and their activities. Our method analyses and highlights privacy problems and their sources. Anyway, it does not drive towards the identification of mitigation strategies. Moreover, it does not provide a mathematical framework for quantitative risk evaluations. Anyway, our method is generic and agnostic for the risk quantification.

Here follows a brief word on the method effectivess with regard to the complexity of system models to be analysed. The analysis reported in Figure 4-3 is consistent with the first step of the methodology depicted in Figure 4-2. The more complete and detailed the system description is, the higher the probability of identifying and describing known and still unknown privacy threats. On the contrary, the more detailed analysis of systems, the more the privacy analysis might be intractable. Indeed correlations among system elements and interactions among devices of the IoT system and IoT sub-system may result in a very complicated system picture. As a possible solution, privacy analyses might be done with the help of computerised systems. Anyway, it requires translating information in a machine-understandable form. Defining a core ontology for this purpose could solve the problem.

To sum up, the proposed method provides support to identify and analyse privacy threats within the IoT. Network ultra-densification is one of the most leading technology solutions to the 5G implementation. Anyway, that transformation may seriously undermine the privacy of devices and their owners. The proposed approach extends the LINDDUN frameworks by introducing temporal and location information to the threats identification process. Moreover, taking a cue from the popular Zachman framework, it also addresses the privacy weaknesses identification by investigating the entangled IoT from four different points of view, namely contextual, conceptual, logical, and physical. Given its capability to make both a sweeping and detailed analysis, the proposed method could answer the call to mitigate the privacy problem in the complex 5G.

Chapter 5

Location Privacy in Ultra-Dense Networks

UDNs are attracting significant interest due to their ability to provide the next generation 5G cellular networks with the high data rate, low delay, and seamless coverage. Many factors, such as drastic interferences, energy constraints, and backhaul bottlenecks limit wireless networks densification. In this Chapter, we investigate the effect of mobile node densification, access node densification, and their aggregation into virtual entities, referred to as virtual cells, on location privacy. By simulations, we observed that implementing virtual cells might reduce the probability of mobile nodes tracking up to ten per cent. Moreover, experiments highlight that the success of tracking attacks has an inverse relationship to the number of moving nodes.

5.1 Introduction

Wireless networks are currently experiencing an ever higher data demand due to increasing number of entities accessing data communication services and novel, resource-consuming applications (e.g., 4k ultra-HD video streaming, virtual and augmented reality). Network densification, providing the operators with more flexibility and the users with seamless connectivity, is considered as one of the most interesting paradigm shift towards future 5G networks. UDNs may be defined as those networks in which the density of access points is (at minimum) a magnitude greater than those of users. Differences between traditional and UDNs (e.g., idle mode capabilities, the probability of line-of-sight communication, and severe interferences) interestingly affect the choice of modelling techniques and performance assessment metrics [35]. Several challenges, such as energy efficiency, and interference and handover management, should be carefully addressed. Certainly, UDNs security cannot be underestimated. Indeed, UDNs while providing reliable connectivity to users, should also cope with concerns related to information hiding, accounting, authentication, and authorisation, thus requiring novel, light-weight cryptographic protocols, and algorithms that fit the ultra-dense scenario [5]. Because of the frequent handovers and authentications, users of UDNs are more exposed to security threats (e.g., such as man-in-the-middle, denial of service, eavesdropping, impersonation, or identity matching) than in conventional cellular networks. In addition, since they adapt to the spatial distribution of mobile nodes and data load, UDNs may provide malicious entities with updated and valuable information about the location of mobile nodes (see Figure 5-2).

Therefore, in this chapter we analyse and propose an approach to cope with the location privacy issue.

5.2 Model

5.2.1 System Model

We considered a multi-tier network composed of M low-coverage access points (APs) underlying B high-coverage base stations deployed within an area R in which N nodes move. We supposed that both APs and nodes are distributed in R according to the Gauss distribution and that $M >> N$. A network controller C manage the APs activation and deactivation as a function of mobile nodes‘ location and data load condition. Moreover, C can logically aggregate APs producing the so-called virtual cells for mobility enhancement [54]. Mobile nodes send and receive messages using their pseudonyms, namely temporary identifiers that are updated and operated by

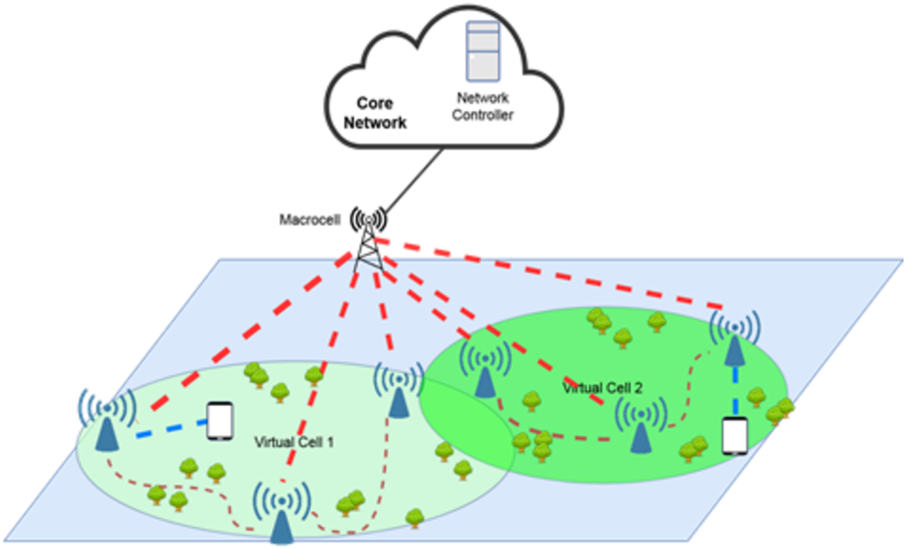
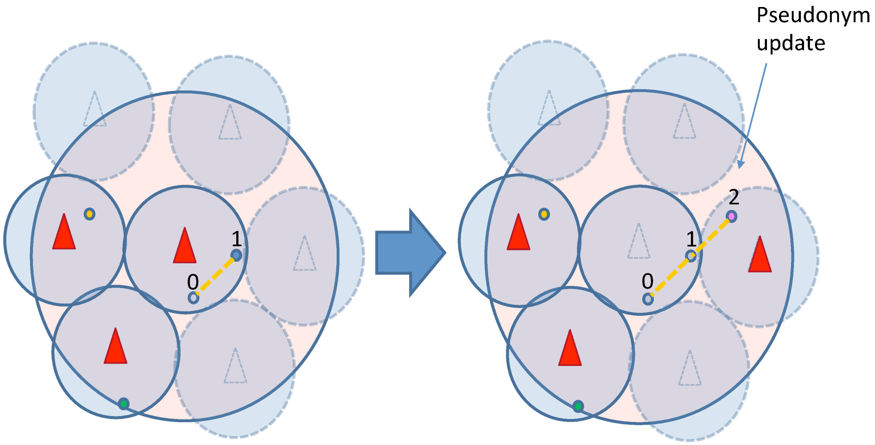


Figure (5-1) Representation of a heterogeneous two-layer cellular network. UEs receive high-data rate from the small-coverage APs in the U-plane while keeping the Radio Resource Control (RRC) connection with the macro cell from C-plane.



The goal of protection against linking attack is to make unlinkable subsequent pseudonyms associated with the same user

Figure (5-2) In this figure, red and shaded triangles represent respectively switched on and off access points. Objective of the adversary is to track the victim by acquiring associations pseudonym-to-AP and the pseudonym update

the network controller. Particularly, the controller governs the associations between nodes and APs. Besides, available resources are efficiently managed to provide a higher data rate in the User Plane. Each AP periodically transmits in broadcast its own physical ID and each mobile node, while monitoring the communication channel, calculates and reports to the network the received power, P_{rx}, for every AP detected. $P_{rx,i}$, namely the received signal strength by a node from AP_i, depends on the transmit power, on the antenna characteristics and average channel attenuation, and on the distance between the transmitter and receiver.

5.2.2 Attack Model

We supposed that the adversary could eavesdrop exchanged information between mobile nodes and APs. He also knows the position of every access point in R. An AP, identified with ap_j, is associated with a pseudonym x_i if such an AP serves the host adopting x_i as an identifier. The objective of the adversary is to keep track of pseudonym updates for each mobile node. To this aim, he/she stores and analyses information broadcasted by APs over time. A pseudonym registered at two different positions in R and at subsequent time instant may refer to either a displacement of the associated mobile node or to the association of it to another mobile node. The adversary can infer such an information by guessing the speed $v(\vec{t})$ at which nodes move and exploiting one or more mobility models. In this thesis, we supposed that the adversary infers nodes' position by using the random mobility model. In particular, given $\vec{v}_{est,i}(t)$ the speed at which u_i moves, and Δt_{trip} a likely duration of displacement for u_i, the area P_{u_i} of possible positions in which u_i might be located at successive time instants is:

$$P_{u_i}(t) = C_{u,i}(t) \cap \bigcup_{j=1, j \neq i}^{N} C_{u,j}(t) \tag{5.1}$$

in which $C_{u,i}(t) = \{p \in R | dist(p, p_i(t)) \leq \vec{v}_{est,i} * \Delta t_{trip}\}, i = 1 \ldots N$ p is a point of R, and $p_i(t)$ is the location (in R) of the AP associated with u_i at time t. The term $\bigcup_{j=1, j\neq i}^{N} C_{u,j}(t)$ is introduced to take into account the group of the closest active access points to $p_i(t)$ (in particular, their coverage might partially overlap those of the AP in $p_i(t)$) and that are associated with other pseudonyms, thus considering the case of pseudonyms update during the position identification process. When an AP serves k mobile nodes at the same time, those nodes might appear undistinguishable one to another to the adversary. In such a case, the effectiveness of the location tracking attack could be reduced.

5.3 Exploiting Virtual Cell to Reduce Location Attack Precision

In this section, we assess the impact of an attack on the location privacy of UEs and describe how virtual cell formation can mitigate the location-tracking threat. Attacks against privacy may success when direct identifiers or quasi-identifiers are available to adversaries. The goal of protection against linking attack is to make unlinkable subsequent pseudonyms associated with the same user. We consider an adversary whose strategy is to eavesdrop cellular network data broadcasted through wireless interfaces. Then, by analysing the acquired information and mixing them with background data of their own, adversaries could infer the location data of targeted victims.

As an attacker model, we considered a global adversary in a local environment, that can monitor every data communications wirelessly exchanged within limited areas. Devices could be either static or in motion. As an example, personal devices, namely systems that individuals usually bring with them, belong to the latter class of devices. Then, given the above preamble and since we are interested to discover privacy threats to persons, we consider more interesting deepen the location privacy of moving objects. A brief comment on static devices worth mentioning. Indeed, once their location is discovered, their privacy cannot be reversed. Hence, the impact of an attack upon static devices is, in principle, higher than those against devices in motion.

Through simulation, we observed the effect of cellular network densification on the location privacy of devices. As quasi - identifiers, we considered latitudes, longitudes and pseudonyms. Even though network operators timely update pseudonyms (e.g., T-IMSI) associated with USIM equipping the devices, adversaries could track pseudonym updates and, by matching and analysing the eavesdropped information, find relationships old-to-new temporary identifiers of victims. Thus, they could have chances to univocally map devices to their unique identifier [98, 79]. Let k be the number of mobile users served by the network. We suppose that the adversary at time instant $t = 0$ has already acquired the associations node-position-to-pseudonym for all the nodes. At the next time instant, nodes‘ pseudonyms change. Adversary tries to link old and new pseudonyms by exploiting eq. 5.1. Mobile nodes are always served by the closest switched-on access points. Moreover, the number of users served by each of access points is a function of its capacity. In addition, we suppose that capacity is the same for all the access points and it is known to the adversary. Let the capacity of access points be α Mbps and each active mobile node needs of β Mbps (with $\alpha > \beta$) bandwidth. Then, each access point can manage at most the integer part of the ratio $\frac{\beta}{\alpha}$ mobile nodes. The greater the number of nodes they can serve and the bigger their coverage area are, the lower the risk for nodes of being identified and traced is. This notwithstanding, the condition on coverage of cells is in stark contrast to network densification trend. However, as described below, exploiting the concept of virtual cell defined in [55] could make subsequent pseudonyms unlikable. The scientific literature has reported various definitions of the virtual cell. Samdanis et. al [73] introduced the concept of the virtual cell in Time Division LTE. Their proposal enabled users residing within an area covered by overlapping cells to access resources from more base stations at once. Meng et. al [55] proposed the concept of the virtual cell for mobility purposes. Their purpose was supplying moving users with a better connectivity experience. To this aim, they equipped the network with the capability to infer the mobility patterns of users. This feature limited user authentication during handovers. In both definitions, a virtual cell is a logical entity, obtained by aggregating physical cells.

Since it takes into account multiple constraints (e.g., energy consumption limits or signalling overheads), virtual cell formation is a multi-targeted task. Moreover, the variability of network characteristics (e.g., transmitting power at each access point, activation and de-activation of access points, and so forth) may render the aforementioned composition complex. For the sake of simplicity, we set the size of virtual cell to a fixed value of five as in [55].

Recalling the definition in [55], a virtual cell v_u is formed by estimating the direction and orientation of a node u:

$$v_u = \{AP_j | \phi_{u,ap_j} < \beta\}, 1 \leq j \leq M \tag{5.2}$$

in which ϕ_{u,ap_j} is the angle between the direction of u and the line joining his point position with the serving access point ap_j; the angle β is the threshold of prediction; $|v_u|$ represents the virtual cell size. We remark that the coverage of virtual cells might partially overlap and that thay can have APs in common. Thanks to the adoption of

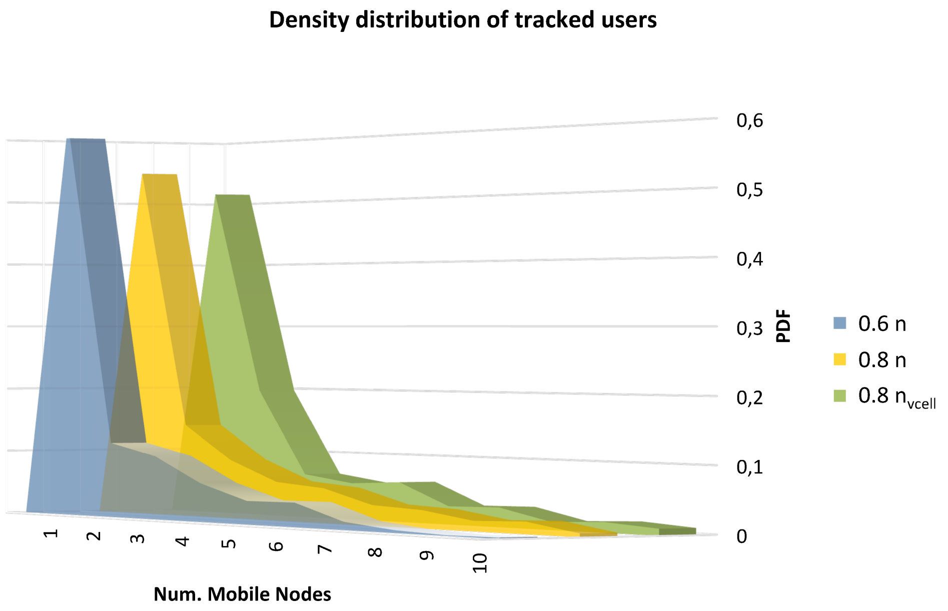


Figure (5-3) Virtual cell composition and the density of users per area unit n influence the probability density functions (PDF) of tracked users. This picture highlights the effect of a decrement of n of 20% and 40% on tracked users. As expected, the lower n is, the higher the likelihood that they can be tracked. Moreover, the success of attacks decreases when the network implements the virtual cells.

virtual cells, some level of uncertainty in term of position of nodes may be introduced, thus diminishing the effectiveness of location-tracking attack.

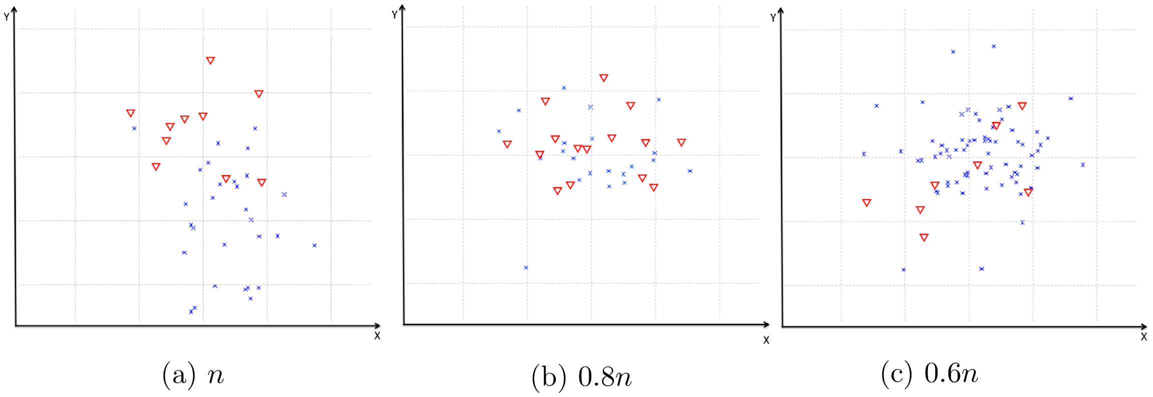
We simulated the movings of several users bringing with them their personal devices connected to an LTE network. Moreover, we supposed that users moved according to a random mobility model with constant speed of $5Km/h$. In particular, we considered a set of 2400 users moving within $A \subset \mathbb{R}^2, |A| = 6Km^2$. We also supposed that A was covered with $6{*}10^3$ small cells. Latitudes and longitudes of access points were modelled as normal distributions with standard deviation $\sigma = 6 * 10^-3$ and mean equal to $\mu = 0$ and $\mu = 39$ respectively. Each node was associated to one access point at time, as a function of the node proximity with respect to access points. In particular, a moble node n was associated to an access point ap_j iff the euclidean dinstance $d(n, ap_j) = min\{dist(a, ap_k), \forall k \in I \subset \mathbb{N}\}$. The topology of the network was controlled by a SDN controller and varied over time, depending on the relative position of nodes with respect the access points, interference level, and load condition. For the sake of simulation, we supposed that each acces point was able to serve at maximum three mobile nodes. At each time instant $t \in T$, nodes were associated to pseudonyms $p \in \mathbb{P}$ according to the following rule:

$$g : A \quad \times \quad T \to \mathbb{P} \quad | \quad n_1, n_2 \in A \wedge t_0 \in T \Rightarrow$$
$$g(n_1, t_0) \neq g(n_2, t_0) \wedge \exists t_k \neq t_0 \quad | \quad g(n_1, t_k) = g(n_2, t_0)$$

Pseudonym updates were triggered at regular time intervals Δt.

Factors that influence the probability of successful attacks against location privacy of mobile nodes are their velocity and direction of displacement, their trajectories, and the proximity among mobile terminals. Indeed, if mobile devices do not change their location over time, connect to adjactent cells, or meet, it is easy for the attacker solve the pseudonym association puzzle. On the contrary, when more than one mobile node camps within the same small cell their anonymity might be protected [15].

We heuristically inspected the sensitivity of the location identification attack to the reduction of moving users (see Figure 5-4(a), 5-4(b), and 5-4(c)). We observed



(a) n (b) $0.8n$ (c) $0.6n$

Figure (5-4) From left to right, experiments report the effect onto the identification of mobile nodes of reduction of the density of connected devices, n, from 20%(b) to 40%(c) without and with virtual cell support. Each cross within the maps represents a point of successful identification of a mobile device, whilst triangles represent points of identification of devices when networks implement virtual cells.

that the greater the number of connected devices, the lower the likelihood of location identification. Experiments show that the scale of the attack is a function of the number of connected and moving devices within the covered area. A significative improvement to location privacy derives from the employment of virtual cells.

Implementing virtual cells for mobility purposes introduces uncertainty to the inferred nodes position, thus reducing the effectiveness of location-tracking and real identity disclosure. Then, we repeated the same experiments equipping the network with virtual cells (see Figure 5-4(a), 5-4(b), and 5-4(c)). Comparison of experiment outcomes shows that virtual cells implementation can improve the network immunity to location identification attacks.

5.4 Discussion

Section 5.3 discussed the role and impact of network densification on the location privacy of their mobile users and present a use case of location attack. As a mitigation strategy, we reported the implementation of virtual cells. Location privacy of mobile nodes in UDNs depends on several factors. Our experiments show that the lower the number of mobile nodes with respect a fixed number of APs, the higher the protection from location disclosure attacks carried out by an adversary is (see Fig. 5-3). Furthermore, forming virtual cells may cut down the effectiveness of tracking attacks.

It is widely accepted that human mobility can affect the performance of wireless networks. Hence, it is likely that our choice of the random mobility model could have affected the simulation outcomes. As discussed in [6, 33], to provide networked systems with security and privacy at each layer of protocol stacks should be addressed. Thus, although implementing virtual cells can reduce the impact of attacks against location privacy of users (see Figure 5-4), further studies should be carried out in this direction since it addresses only the RRC layer of LTE and does not cope with the other protocol layers.

To sum up, despite the limitations of our approach, our findings suggest that, in addition to enhance the network's performance, virtual cell may improve the location privacy of mobile subscribers.

Chapter 6

Conclusion

In this thesis, we discussed the rising security and privacy issues in 5G networks fed by the complex interactions among stakeholders and the adoption of new technological paradigms.

The aim of this work was to provide a contribution to identifying security threats and mitigating menaces to the data and location privacy of individuals. In particular, by studying the characteristics of security ontologies through the network theory and introducing the novel concept of differential semantic variance, we provided a method for reducing the complexity of security ontologies but guaranteeing good performance of reasoning systems. Thanks to this result, the automatic identification of security threats of stakeholders can be streamlined and speeded up. In the light of this, security analysis of complex architectures might be performed effectively and within a reasonable timeframe. Among all the advantages, the 5G will promote the diffusion of novel services and real-time application only figured so far. A large amount of information will transit through it. Therefore, the next generation mobile network should be able, a fortiori when sensitive, to preserve it.

It is universally recognised that the transition to 5G will determine an outstanding revolution in the fashion in which persons will interact with technology. The IoT is an integral part of this technological breakthrough. Anyway, due to their features are driven mainly by economic reasons, the most spread devices in the IoT cannot guarantee their owners with the expected and required level of security and privacy.

As to assist in remedying security failure in the IoT, we proposed a methodology of analysis for the privacy issues identification in the IoT. By fusing together the LINDDUN and the Zachman frameworks, and extending the former by considering the temporal and location information, the proposed approach can address privacy weaknesses in the IoT from different points of view, operating either in the holistic and specialised way. Therefore, it can answer the call to mitigate the privacy problem in the complex and heterogeneous IoT environment.

Since it can easily be accessed by unintended actors, location information of mobile devices is one of the most valuable information that is required to be protected in 5G. However, the ultra-densification trend of wireless networks might seriously threaten it. We analysed the correlation between mobile device density and access points density on the location privacy of mobile devices. Then, we found that the implementation of virtual cells, namely logical aggregation of access points, could overcome the location privacy problem.

Bibliography

[1] 3GPP. Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE. Technical Report (TR) 36.888, 3rd Generation Partnership Project (3GPP), 06 2013. Version 12.0.0.

[2] 3GPP. Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT). Technical Specification (TS) 45.820, 3rd Generation Partnership Project (3GPP), 12 2015. Version 13.1.0.

[3] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); NB-IOT; Technical Report for BS and UE radio transmission and reception. Technical Report (TR) 36.802, 3rd Generation Partnership Project (3GPP), 06 2016. Version 13.0.0.

[4] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode. Technical Specification (TS) 36.304, 3rd Generation Partnership Project (3GPP), 01 2018. Version 14.5.0.

[5] Ian F Akyildiz, Shuai Nie, Shih-Chun Lin, and Manoj Chandrasekaran. 5g roadmap: 10 key enabling technologies. *Computer Networks*, 106:17–48, 2016.

[6] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393 – 422, 2002.

[7] D. A. Bailey. Moving 2 mishap: M2m's impact on privacy and safety. *IEEE Security Privacy*, 10(1):84–87, Jan 2012.

[8] Montserrat Batet, David Sánchez, and Aida Valls. An ontology-based measure to compute semantic similarity in biomedicine. *Journal of Biomedical Informatics*, 44(1):118 – 125, 2011. Ontologies for Clinical and Translational Research.

[9] Dominik Benz, Andreas Hotho, Gerd Stumme, and Stefan Stützer. Semantics made by you and me: Self-emerging ontologies can capture the diversity of shared knowledge. In *in Proc. of the 2nd Web Science Conference (WebSci10*, 2010.

[10] A.R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[11] C. Blanco, J. Lasheras, E. Fernández-Medina, R. Valencia-García, and A. Toval. Basis for an integrated security ontology according to a systematic review of existing proposals. *Computer Standards and Interfaces*, 33(4):372–388, 2011.

[12] Phillip Bonacich. Power and centrality: A family of measures. *American journal of sociology*, 92(5):1170–1182, 1987.

[13] Phillip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3):191–201, 2001.

[14] Iván Cantador, Martin Szomszor, Harith Alani, Miriam Fernández Sánchez, and Pablo Castells. Enriching ontological user profiles with tagging history for multi-domain recommendations. In *CEUR Workshop Proceedings*. Yannis Avrithis, 2008.

[15] E. Catania and A. La Corte. Location privacy in virtual cell-equipped ultra-dense networks. In *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–4, Feb 2018.

[16] M. Chen, Y. Miao, Y. Hao, and K. Hwang. Narrow band internet of things. *IEEE Access*, 5:20557–20577, 2017.

[17] S. Chen, R. Ma, H. H. Chen, H. Zhang, W. Meng, and J. Liu. Machine-to-machine communications in ultra-dense networks - a survey. *IEEE Communications Surveys Tutorials*, 19(3):1478–1503, thirdquarter 2017.

[18] Cisco Visual Networking Index Cisco. Global mobile data traffic forecast update, 2015-2020 white paper, 2016. *Online: https://www. cisco. com/c/dam/m/en_ in/innovation/enterprise/assets/mobile-white-paper-c11-520862. pdf.*

[19] D.J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences of the United States of America*, 107(52):22436–22441, 2010.

[20] Alessandro Di Stefano, Marialisa Scatà, Aurelio La Corte, Pietro Liò, Emanuele Catania, Ermanno Guardo, and Salvatore Pagano. Quantifying the role of homophily in human cooperation using multiplex evolutionary game theory. *PloS one*, 10(10):e0140646, 2015.

[21] Li Ding, Tim Finin, Anupam Joshi, Rong Pan, R Scott Cost, Yun Peng, Pavan Reddivari, Vishal Doshi, and Joel Sachs. Swoogle: a search and metadata engine for the semantic web. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 652–659. ACM, 2004.

[22] X. Duan and X. Wang. Authentication handover and privacy protection in 5g hetnets using software-defined networking. *IEEE Communications Magazine*, 53(4):28–35, April 2015.

[23] C. Dwork. Differential privacy. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4052 LNCS:1–12, 2006.

[24] S. Farhang, Y. Hayel, and Q. Zhu. Phy-layer location privacy-preserving access point selection mechanism in next-generation wireless networks. In *2015 IEEE Conference on Communications and Network Security (CNS)*, pages 263–271, Sept 2015.

[25] Aissa Fellah, Mimoun Malki, and Atilla Elci. A similarity measure across ontologies for web services discovery. *International Journal of Information Technology and Web Engineering*, 11(1):22–43, jan 2016.

[26] Miriam Fernández, Chwhynny Overbeeke, Marta Sabou, and Enrico Motta. What makes a good ontology? a case-study in fine-grained knowledge reuse". In *The Semantic Web*, pages 61–75, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[27] Andrew Flahive, David Taniar, and Wenny Rahayu. Ontology as a service (OaaS): extending sub-ontologies on the cloud. *Concurrency and Computation: Practice and Experience*, 27(8):2028–2040, oct 2014.

[28] Bart Gajderowicz, Alireza Sadeghian, and Mikhail Soutchanski. Ontology enhancement through inductive decision trees. In Fernando Bobillo, Paulo C. G. Costa, Claudia d'Amato, Nicola Fanizzi, Kathryn B. Laskey, Kenneth J. Laskey, Thomas Lukasiewicz, Matthias Nickles, and Michael Pool, editors, *Uncertainty Reasoning for the Semantic Web II*, pages 262–281, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[29] Gabriel Ghinita, Maria Luisa Damiani, Claudio Silvestri, and Elisa Bertino. Preventing velocity-based linkage attacks in location-aware applications. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, pages 246–255, New York, NY, USA, 2009. ACM.

[30] M. Gramaglia, M. Fiore, A. Tarable, and A. Banchs. Preserving mobile subscriber privacy in open datasets of spatiotemporal trajectories. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, May 2017.

[31] F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(4):19, 2016.

[32] I-Ching Hsu. Integrating ontology technology with folksonomies for personalized social tag recommendation. *Applied Soft Computing*, 13(8):3745 – 3750, 2013.

[33] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. Kwak. The internet of things for health care: A comprehensive survey. *IEEE Access*, 3:678–708, 2015.

[34] Antonio Jimeno-Yepes, Ernesto Jiménez-Ruiz, Rafael Berlanga-Llavori, and Dietrich Rebholz-Schuhmann. Reuse of terminological resources for efficient ontological engineering in life sciences. *BMC bioinformatics*, 10(10):S4, 2009.

[35] Mahmoud Kamel, Walaa Hamouda, and Amr Youssef. Ultra-dense networks: A survey. *IEEE Communications Surveys & Tutorials*, 18(4):2522–2545, 2016.

[36] Yong-Bin Kang, Yuan-Fang Li, and Shonali Krishnaswamy. Predicting reasoning performance using ontology metrics. In Philippe Cudré-Mauroux, Jeff Heflin, Evren Sirin, Tania Tudorache, Jérôme Euzenat, Manfred Hauswirth, Josiane Xavier Parreira, Jim Hendler, Guus Schreiber, Abraham Bernstein, and Eva Blomqvist, editors, *The Semantic Web – ISWC 2012*, pages 198–214, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[37] M Kende. Internet society global internet report 2014. *Internet Society*, 146:8332–8336, 2014.

[38] Jungmin Kim, Pankoo Kim, and Hyunsook Chung. Ontology construction using online ontologies based on selection, mapping and merging. *International Journal of Web and Grid Services*, 7(2):170, 2011.

[39] Teemu Koponen, Keith Amidon, Peter Balland, Martín Casado, Anupam Chanda, Bryan Fulton, Igor Ganichev, Jesse Gross, Paul Ingram, Ethan J Jackson, et al. Network virtualization in multi-tenant datacenters. In *NSDI*, volume 14, pages 203–216, 2014.

[40] Denis Foo Kune, John Koelndorfer, Nicholas Hopper, and Yongdae Kim. Location leaks on the gsm air interface. In *Proceedings of the 19th Annual Network & Distributed System Security Symposium*, Feb 2012.

[41] Antonio Kung, Frank Kargl, Santiago Suppan, Jorge Cuellar, Henrich C. Pöhls, Adam Kapovits, Nicolás Notario McDonnell, and Yod Samuel Martin. *A Privacy Engineering Framework for the Internet of Things*, pages 163–202. Springer International Publishing, Cham, 2017.

[42] Aurelio La Corte and Marialisa Scatá. Security and qos analysis for next generation networks. In *Information Society (i-Society), 2011 International Conference on*, pages 248–253. IEEE, 2011.

[43] Aurelio La Corte and Marialisa Scatá. Convergence, security and quality in the ngn. *International Journal of Internet Technology and Secured Transactions*, 4(4):327–343, 2012.

[44] Aurelio La Corte, Marialisa Scatá, and Evelina Giacchi. A bio-inspired approach for risk analysis of ict systems. In *International Conference on Computational Science and Its Applications*, pages 652–666. Springer, 2011.

[45] S. Lee and K. Huang. Coverage and economy of cellular networks with many base stations. *IEEE Communications Letters*, 16(7):1038–1040, July 2012.

[46] M. Lichtman, R. P. Jover, M. Labib, R. Rao, V. Marojevic, and J. H. Reed. Lte/lte-a jamming, spoofing, and sniffing: threat assessment and mitigation. *IEEE Communications Magazine*, 54(4):54–61, April 2016.

[47] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. *Content-based Recommender Systems: State of the Art and Trends*, pages 73–105. Springer US, Boston, MA, 2011.

[48] ETSI LTE. Evolved universal terrestrial radio access (e-utra); radio resource control (rrc); protocol specification (3gpp ts 36.331, version 15.21. 0 release 15), june 2018.

[49] ETSI LTE. Evolved universal terrestrial radio access (e-utra) and evolved universal terrestrial radio access network (e-utran)(3gpp ts 36.300, version 8.11. 0 release 8), december 2009. *ETSI TS*, 136(300):V8, 2015.

[50] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*, 1(1), 2007.

[51] Lukas Malina, Jan Hajny, Radek Fujdiak, and Jiri Hosek. On perspective of security and privacy-preserving solutions in the internet of things. *Computer Networks*, 102:83 – 95, 2016.

[52] Nayana Mannilthodi and Jinesh M Kannimoola. Secure iot: An improbable reality. 2017.

[53] M Maternia, S Eddine El Ayoubi, M Fallgren, P Spapis, Y Qi, D Martín-Sacristán, Óscar Carrasco, M Fresia, M Payaró, M Schubert, et al. 5g ppp use cases and performance evaluation models. *see https://5g-ppp. eu/wp-content/uploads/2014/02/5G-PPP-use-cases-and-performance-evaluation-modeling_ v1. 0. pdf*, 2016.

[54] N. Meng, H. Zhang, and H. Lu. Virtual cell-based mobility enhancement and performance evaluation in ultra-dense networks. In *2016 IEEE Wireless Communications and Networking Conference*, pages 1–6, April 2016.

[55] Na Meng, Hongtao Zhang, and Haitao Lu. Virtual cell-based mobility enhancement and performance evaluation in ultra-dense networks. In *Wireless Communications and Networking Conference (WCNC), 2016 IEEE*, pages 1–6. IEEE, 2016.

[56] Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, and Imrich Chlamtac. Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, 10(7):1497 – 1516, 2012.

[57] Alexios Mylonas, Vasilis Meletiadis, Bill Tsoumas, Lilian Mitrou, and Dimitris Gritzalis. Smartphone forensics: A proactive investigation scheme for evidence acquisition. In Dimitris Gritzalis, Steven Furnell, and Marianthi Theoharidou, editors, *Information Security and Privacy Research*, pages 249–260, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[58] Mathias Niepert, Cameron Buckner, and Colin Allen. Answer set programming on expert feedback to populate and extend dynamic ontologies. In *FLAIRS Conference*, pages 500–505, 2008.

[59] L. Ninghui, L. Tiancheng, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. pages 106–115, 2007.

[60] NIST. NIST Special Publication 800-187 Guide to LTE Security. Technical Specification (TS) 800-187, NIST, 12 2017.

[61] Charith Perera, Ciaran McCormick, Arosha K. Bandara, Blaine A. Price, and Bashar Nuseibeh. Privacy-by-design framework for assessing internet of things applications and platforms. In *Proceedings of the 6th International Conference on the Internet of Things*, IoT'16, pages 83–92, New York, NY, USA, 2016. ACM.

[62] Giovanni Perrone, Massimo Vecchio, Riccardo Pecori, and Raffaele Giaffreda. The day after mirai: A survey on mqtt security solutions after the largest cyber-attack carried out through an army of iot devices. In *Proceedings of the 2nd International Conference on Internet of Things, Big Data and Security - Volume 1: IoTBDS,*, pages 246–253. INSTICC, SciTePress, 2017.

[63] Nayot Poolsappasit and Indrakshi Ray. Towards achieving personalized privacy for location-based services. *Trans. Data Privacy*, 2(1):77–99, April 2009.

[64] P. Porambage, M. Ylianttila, C. Schmitt, P. Kumar, A. Gurtov, and A. V. Vasilakos. The quest for privacy in the internet of things. *IEEE Cloud Computing*, 3(2):36–45, Mar.-Apr. 2016.

[65] T. Pushpalata and S. Y. Chaudhari. Need of physical layer security in lte: Analysis of vulnerabilities in lte physical layer. In *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pages 1722–1727, March 2017.

[66] R. Ratasuk, N. Mangalvedhe, Y. Zhang, M. Robert, and J. P. Koskinen. Overview of narrowband iot in lte rel-13. In *2016 IEEE Conference on Standards for Communications and Networking (CSCN)*, pages 1–7, Oct 2016.

[67] A. Razzaq, Z. Anwar, H.F. Ahmad, K. Latif, and F. Munir. Ontology for attack detection: An intelligent approach to web application security. *Computers and Security*, 45:124–146, 2014.

[68] A. Razzaq, K. Latif, H. Farooq Ahmad, A. Hur, Z. Anwar, and P.C. Bloodsworth. Semantic security against web application attacks. *Information Sciences*, 254:19–38, 2014.

[69] Simone Redana, A Kaloxylos, A Galis, P Rost, and V Jungnickel. View on 5g architecture. *White paper of the 5G-PPP architecture WG*, 2016.

[70] S. Sadr and R. S. Adve. Handoff rate and coverage analysis in multi-tier heterogeneous networks. *IEEE Transactions on Wireless Communications*, 14(5):2626–2638, May 2015.

[71] P. Samarati. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010–1027, 2001.

[72] K. Samdanis, X. Costa-Perez, and V. Sciancalepore. From network sharing to multi-tenancy: The 5g network slice broker. *IEEE Communications Magazine*, 54(7):32–39, 2016.

[73] K. Samdanis, R. Shrivastava, A. Prasad, P. Rost, and D. Grace. Virtual cells: Enhancing the resource allocation efficiency for td-lte. In *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, pages 1–5, Sept 2014.

[74] Marialisa Scatà, Alessandro Di Stefano, Aurelio La Corte, and Pietro Liò. Quantifying the propagation of distress and mental disorders in social networks. *Scientific reports*, 8(1):5005, 2018.

[75] Marialisa Scatà, Alessandro Di Stefano, Aurelio La Corte, Pietro Liò, Emanuele Catania, Ermanno Guardo, and Salvatore Pagano. Combining evolutionary game theory and network theory to analyze human cooperation patterns. *Chaos, Solitons & Fractals*, 91:17–24, 2016.

[76] Marialisa Scatà, Alessandro Di Stefano, Pietro Liò, and Aurelio La Corte. The impact of heterogeneity and awareness in modeling epidemic spreading on multiplex networks. *Scientific reports*, 6:37105, 2016.

[77] Shilad Sen, F. Maxwell Harper, Adam LaPitz, and John Riedl. The quest for quality tags. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, GROUP '07, pages 361–370, New York, NY, USA, 2007. ACM.

[78] Arbia Riahi Sfar, Enrico Natalizio, Yacine Challal, and Zied Chtourou. A roadmap for security challenges in the internet of things. *Digital Communications and Networks*, 4(2):118 – 137, 2018.

[79] Altaf Shaik, Ravishankar Borgaonkar, N. Asokan, Valtteri Niemi, and Jean-Pierre Seifert. Practical attacks against privacy and availability in 4g/lte mobile communication systems. *CoRR*, abs/1510.07563, 2015.

[80] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh. Machine-type communications: current status and future perspectives toward 5g systems. *IEEE Communications Magazine*, 53(9):10–17, September 2015.

[81] Hang Shen, Guangwei Bai, Mei Yang, and Zhonghui Wang. Protecting trajectory privacy: A user-centric analysis. *Journal of Network and Computer Applications*, 82:128 – 139, 2017.

[82] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux. Quantifying location privacy. pages 247–262, 2011.

[83] S. Sicari, A. Rizzardi, L.A. Grieco, and A. Coen-Porisini. Security, privacy and trust in internet of things: The road ahead. *Computer Networks*, 76:146 – 164, 2015.

[84] A. Souag, C. Salinesi, and I. Comyn-Wattiau. Ontologies for security requirements: A literature survey and classification. *Lecture Notes in Business Information Processing*, 112 LNBIP:61–69, 2012.

[85] Mike Surridge, Toby Wilkinson, Peter Maynard, Stephen C. Phillips, Gianluca Correndo, and Stefanie Wiegand. Deliverable d2. 5 trust model (final).

[86] J. Swetina, G. Lu, P. Jacobs, F. Ennesser, and J. Song. Toward a standardized common m2m service layer platform: Introduction to onem2m. *IEEE Wireless Communications*, 21(3):20–26, June 2014.

[87] David Sánchez, Montserrat Batet, Sergio Martínez, and Josep Domingo-Ferrer. Semantic variance: An intuitive measure for ontology accuracy evaluation. *Engineering Applications of Artificial Intelligence*, 39:89 – 99, 2015.

[88] V.O. Tikhvinskiy and G. Bochechka. Prospects and qos requirements in 5g networks. *Journal of Telecommunications and Information Technology*, 2015(1):23–26, 2015.

[89] Vassilios Vassilakis, Emmanouil Panaousis, and Haralambos Mouratidis. Security challenges of small cell as a service in virtualized mobile edge computing environments. In Sara Foresti and Javier Lopez, editors, *Information Security Theory and Practice*, pages 70–84, Cham, 2016. Springer International Publishing.

[90] Gary White, Vivek Nallur, and Siobhán Clarke. Quality of service approaches in iot: A systematic mapping. *Journal of Systems and Software*, 132:186 – 203, 2017.

[91] Chensi Wu, Yuqing Zhang, and Ying Dong. Application research on network attacks and defenses with zachman framework. In Zheng Yan, Refik Molva, Wojciech Mazurczyk, and Raimo Kantola, editors, *Network and System Security*, pages 439–449, Cham, 2017. Springer International Publishing.

[92] S. Wu, Y. Zhang, and W. Cao. Network security assessment using a semantic reasoning and graph based approach. *Computers and Electrical Engineering*, 64:96–109, 2017.

[93] Kim Wuyts. Privacy threats in software architectures, 2015.

[94] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. D. Renzo. Safeguarding 5g wireless communication networks using physical layer security. *IEEE Communications Magazine*, 53(4):20–27, April 2015.

[95] I. Yaqoob, E. Ahmed, I. A. T. Hashem, A. I. A. Ahmed, A. Gani, M. Imran, and M. Guizani. Internet of things architecture: Recent advances, taxonomy, requirements, and open challenges. *IEEE Wireless Communications*, 24(3):10–16, June 2017.

[96] W. Yu, H. Xu, H. Zhang, D. Griffith, and N. Golmie. Ultra-dense networks: Survey of state of the art and future directions. In *2016 25th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–10, Aug 2016.

[97] J. A. Zachman. A framework for information systems architecture. *IBM Systems Journal*, 26(3):276–292, 1987.

[98] Jun Zhou, Zhenfu Cao, Xiaolei Dong, and Athanasios V Vasilakos. Security and privacy for cloud-based iot: challenges. *IEEE Communications Magazine*, 55(1):26–33, 2017.