THÈSE DE DOCTORAT
ORANGE LABS ET TÉLÉCOM SUDPARIS
EN CO-ACCREDITATION
AVEC L'UNIVERSITÉ PIERRE ET MARIE CURIE - PARIS 6

ÉCOLE DOCTORALE INFORMATIQUE, TÉLÉCOMMUNICATION ET ÉLECTRONIQUE DE PARIS

Spécialité

INFORMATIQUE ET RESEAUX

Présentée par

CAO HỮU Quyết

Policy-based Usage Control for Trustworthy Data Sharing in Smart Cities

Soutenue le 8/Juin/2017 devant le jury composé de:

Rapporteurs:

M. Yacine Gahrmi-Doudane	Professeur à Université de La Rochelle, France
M. Alberto Leone-Garcia	Professeur à Université de Toronto, Canada

Examinateurs:

M. Giovanni Pau	Professeur à UPMC - Paris 6, France
Mme. Lila Boukhatem	Maître de Conférences à Université Paris-Sud, France

Co-Encadrant de Thèse:

M. Giyyarpuram Madhusudan	Ingénieur de Recherche à Orange Labs, France

Directeur de Thèse:

M. Noël Crespi	Professeur à Télécom SudParis, France

Thèse numéro : 2017TELE0010

— *To my late Grandfather,*

Acknowledgements

First of all, I would like to express my deepest gratitude to my advisors. Being my industrial advisor, Mr. Giyyarpuram Madhusudan was really close to me. He guided and helped me greatly with my research. His motivation, patience, valuable suggestions, and his moral support is highly appreciable. Prof. Noël Crespi helped me with his excellent guidance, care, patience and precious remarks throughout the thesis. I feel lucky to have found such wonderful researchers as my thesis advisors.

I'd like to thank the reviewers of my thesis, Prof. Alberto Leon-Garcia (University of Toronto, Canada) and Prof. Yacine Gahrmi-Doudane (Université de La Rochelle, France) for their valuable feedback to improve the quality of my work.

My thanks to Orange Labs and the team BIZZ/MIS/CITY for providing an exceptional and friendly environment for the research and development. The interactions with different helpful people in Orange has led to the development of my dissertation. My sincere thanks to everybody involved in these fruitful discussions.

I am also grateful to my co-authors, Dr. Reza Farahbakhsh, Dr. Son N. Han, Dr. Imran Khan, Dr. Gyu Myoung Lee, Dr. Fano Ramparany, Mr. Nguyen B. Truong for their useful comments and suggestions. I need a special word of thanks them all.

Lastly a big thanks to my family. They have provided me great support and motivation during my thesis.

My apologies to those that I may have missed. Thanks to all of you. Without their help and support I would never have finished my dissertation.

CAO HỮU Quyết
Grenoble
06 February 2017

Résumé

Dans le domaine de “smart cities” ou “villes connectées”, les technologies de l'information et de la communication sont intégrées aux services traditionnels de la ville (eau, électricité, gaz, transports collectifs, équipements publics, bâtiments, etc.) pour améliorer la qualité des services urbains ou encore pour réduire les coûts.

Les données dans la ville connectée sont généralement produites par une grande variété d'acteurs. Ces données devraient être partagées entre diverses applications ou services. Or, il y a un problème, comment les acteurs peuvent-ils exercer un contrôle sur la façon dont leurs données vont être utilisées ?

C'est important car pour encourager le partage des données, nous devons établir des relations de confiance entre acteurs. Les acteurs ont confiance s'ils ont la capacité à contrôler l'utilisation de leurs données.

Nous prendrons en compte les obligations définies par les acteurs pour leurs données : (*i*) Abstraction de certaines informations, (*ii*) Granularité spatio-temporelle, (*iii*) Classification des acteurs et des objectifs, et (*iv*) Monétisation des données.

Mes contributions sont: (*i*) un modèle de contrôle d'utilisation des données. Ce modèle répond aux obligations définies par les acteurs pour leur données. (*ii*) une plateforme en tant que service. La plateforme a rajouté des composants nécessaire pour permettre la transparence et la traçabilité d'utilisation des données basée sur le modèle. (*iii*) un outil de visualisation. C'est l'implémentation d'un prototype pour que les acteurs puissent exercer un contrôle sur la façon dont leurs données vont être utilisées. (*iv*) une évaluation de la performance et l'impact de notre solution.

Ces solutions permettent l'établissement des relations de confiance pour le partage des données basée sur le modèle de contrôle d'utilisation des données.

Ma thèse se déroule dans le cadre d'un contrat CIFRE, en partenariat d'Orange Labs avec l'Université Pierre et Marie Curie et Télécom SudParis. Mon directeur de thèse s'appelle Noël Crespi, leader de l'équipe de recherche Service Architecture Lab, qui se centre sur les services futurs. Je travaille

dans l'équipe de ORANGE/IMT/OLPS/BIZZ/MIS/CITY. Cette équipe est gérée par MADILLO Pierre et mon encadrant s'appelle MADHUSUDAN Giyyarpuram. Ma thèse apporte sa contribution au projet de Smart Cities Trials pour la mission principale d'établissement des relations de confiance pour le partage des données de Smart Cities. J'ai commencé ma thèse à Meylan en Décembre 2013. Les résultats de ma thèse peuvent être appliqués à la plateforme IoT Datavenue d'Orange.

Abstract

In smart cities, Information and Communication Technologies, in particular Internet of Things (IoT) Technologies, are integrated into traditional services of our city, for example waste management, air pollution monitoring, and parking to improve quality while reducing costs of these services.

IoT data in this context are generated by different actors, such as service providers, developers, and municipal authorities. These data should be shared among applications or services. However, in traditional scenario, there is no sharing of IoT data between them. Each actor consumes data from sensors deployed on behalf of that actor, and network infrastructure maybe shared.

In order to encourage IoT data sharing, we need to establish the confidence between the actors. Exercising control over the usage of data by other actors is critical in building trust. Thus, the actors should have an ability to exercise control on how their data are going to be used. This major issue have not been treated in IoT namely Usage Control.

In this thesis, we take into account obligations defined by the actors for their data (*i*) Abstraction of certain information, (*ii*) Spatial and temporal granularity, (*iii*) Classification of actors and purposes, and (*iv*) Monetization of data. For example, requirements of data usage in Intelligent parking applications are (*i*) Data owners have full access to all the details, (*ii*) Municipal authorities can access the average occupancy of parking place per street on an hourly basis, (*iii*) Commercial service providers can access only statistical data over a zone and a weekly basis, and (*iv*) Monetization of data can be based on subscription types or users roles.

Thesis contributions include: (*i*) Policy-based Data Usage Control Model (DUPO) responds to the obligations defined by actors to their data. (*ii*) Trustworthy Data Sharing Platform as a Service allows transparency and traceability of data usage with open APIs based on the DUPO and Semantic technologies. (*iii*) Visualization Tool Prototype enables actors to exercise control on how their data will be used. (*iv*) Evaluation of the performance and the impact of our solution. The results show that the performance of

the added trust is not affecting of the system.

Mistrust might hamper public acceptance of IoT data sharing in smart cities. Our solution is key which will establish the trust between data owners and consumers by taking into account the obligations of the data owners. It is useful for data operators who would like to provide an open data platform with efficient enablers to partners, data-based services to clients, and ability to attract partners to share data on their platforms.

Keywords: Internet of Things (IoT); Smart Cities; Trust Model; Data Sharing, Data Usage Control, Data Usage Policy

List of Publications

I. N.B. Truong, Quyet H. Cao, T.W.Um, and G.M.Lee, "A Holistic Trust Computation Framework for Social Internet of Things", *submitted to IEEE International Conference on Communications (ICC)*, Paris, France, 2017.

II. Quyet H. Cao, G. Madhusudan, R. Farahbakhsh, and N. Crespi, "Policy-based Usage Control for A Trustworthy Data Sharing Platform in Smart Cities", *Journal of Future Generation Computer Systems*, 2017.

III. N.B. Truong, Quyet H. Cao, T.W.Um, and G.M.Lee, "Leverage a Trust Service Platform for Data Usage Control in Smart Cities", *In IEEE Global Communications Conference (GLOBECOM)*, Washington DC, USA, 2016.

IV. Quyet H. Cao, I. Khan, R. Farahbakhsh, G. Madhusudan, G.M. Lee, and N. Crespi, "A Trust Model for Data Sharing in Smart Cities", *In IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, 2016.

V. F. Ramparany and Quyet H. Cao, "A Semantic Approach to IoT Data Aggregation and Interpretation applied to Home Automation", *In IEEE International Conference on Internet of Things and Applications (IoTA)*, Pune, India, 2016.

VI. Quyet H. Cao, G. Madhusudan, R. Farahbakhsh, and N. Crespi, "Usage Control for Data Handling in Smart Cities", *In IEEE Global Communications Conference (GLOBECOM)*, San Diego, CA, USA, 2015.

VII. G.M. Lee, Y. Kim, W.S. Rhee, and Quyet H. Cao, "Introducing for common mapping principles between the Base Ontology and external ontologies", *MAS-2015-0627-oneM2M*, Sophia Antipolis, France, 2015.

VIII. G.M. Lee, Y. Kim, W.S. Rhee, and Quyet H. Cao, "Revisions for common mapping principles between the Base Ontology and external ontologies", *MAS-2015-0675-oneM2M*, Beijing, China, 2015.

IX. S.N. Han, Quyet H. Cao, B. Alinia, and N. Crespi, "Design, Implementation, and Evaluation of 6LoWPAN for Home and Building Automation", *In ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, Marrakech, Morocco, 2015.

Table of contents

List of Figures

List of Tables

"If we knew what it was we were doing,
it would not be called research, would it?"

Quote by Albert Einstein

Chapter 1

Introduction

Contents

1.1 Context and Problem

In smart cities, the Information and Communication Technologies are generally integrated into traditional services of our city to improve quality while reducing costs of these services (Townsend 2013). Nowadays, the new communication paradigm goes beyond traditional inter-personal interactions, as it involves interactions between devices under the umbrella of the Internet of Things (IoT)(Atzori et al. 2010, Gubbi et al. 2013) technologies which are among the main vehicles for realizing this vision. IoT data can be collected from huge amount of interactions across a large number of devices, and in the near future, large scale IoT applications in smart cities will become a reality. It could enhance a city's innovation capacity as well as provide significant

socioeconomic value for the cities(Zanella et al. 2014). In deploying such applications, the participation of citizens and other players in both data collection and in the emergence of new services is needed.

Currently, applications for smart cities are mostly developed in a vertical manner, with no sharing of data or resources between different players (Sanchez et al. 2014). Many of these vertical applications would benefit from using information sources of different origins to enhance their own services. The landscape consists of a diversity of actors, both public and private, who provide a large variety of services. These applications include energy management for public buildings, waste management, public lighting, mobility management, intelligent parking solutions and a whole range of new services that are being conceived for smart cities (Khatoun & Zeadally 2016). The actors involved in these applications tend to vary with the specific domain, as each comes with its own ecosystem. However we can identify several broad categories of actors: institutional actors (such as districts, municipalities), equipment manufacturers, network operators, infrastructure providers and service providers. With the development of the IoT, the range of actors involved will be enlarged to include micro companies, value-added service providers (such as aggregations, compositions and mashups) and end users. The need for a horizontal platform, which federates information from these disparate sources is particularly important. This intermediation platform, for actors with different and sometimes contradictory requirements brings its own set of challenges. For this horizontal approach to succeed, the platform needs to ensure that the business interests of the different participants are fully honored.

The main requirement to have a successful IoT data sharing in this context is that participants contribute and share their data. One example is when people are able to share their data related to different events by leveraging the sensing capabilities of their smartphones. This crowd-sensing is a recent trend (Christin 2016) and may soon outperform traditional data collection methods such as using pre-installed sensors. However, crowd-sensing may involve privacy issues for device owners. For example, some of the data collected by smartphones may contain sensitive information such as the location of the owners. In addition, the data in smart cities may come

from a variety of sources and potentially undergo several transformations, such as aggregation and composition, before reaching their final destination. The IoT data may also be shared for common usage through linked data sets such as Linked Open Data (Berners-Lee 2006). Therefore, to achieve trustworthy data sharing in smart cities, the shared platform should be able to: (*i*) establish the trust between different players to share their data, (*ii*) solve a potential conflict of interest between actors, (*iii*) achieve competitive advantages, and (*iv*) hide or abstract some information with usage control. A lack of this would inevitably lead to mistrust and might hamper public acceptance of IoT data sharing.

Trust has many facets, but one critical element in the IoT is the ability for each participant to exercise control on how their data is going to be used. Although this is an important research topic, but still it has not yet been treated in a proper manner in the context of smart cities. In this thesis, we claim that we would have trustworthy data sharing in smart cities, if we provide mechanisms for transparency and traceability of data usage by using a policy-based data usage control model. It will take into account of following issues

- There is still no specific data usage control model to express the constraints and obligations on the use of IoT data among participants.
- Transparency and traceability of data usage are also essential in the context of smart cities.
- A visualization tool is needed to help users customize their policies in an interactive format, which allows them to explore and monitor the consequences of certain changes to how their data is allowed to be used.

1.2 Motivating Scenario and Research Questions

To illustrate better the current issues of trust and control for data sharing cases in smart cities, we first present a general motivating scenario with a use case for intelligent parking, and then raise some research questions that will be addressed through this thesis.

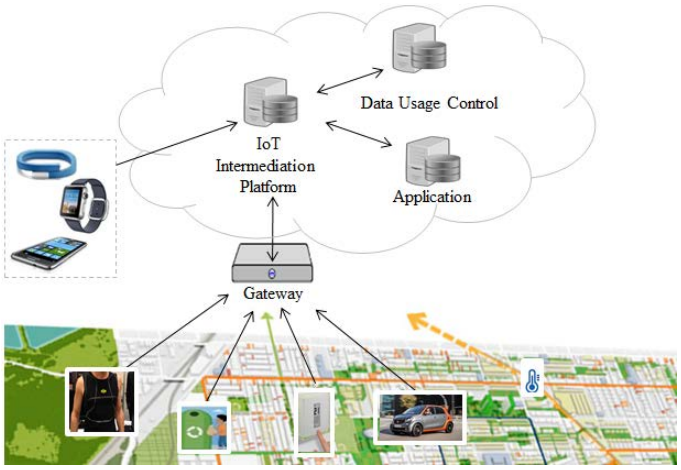

Figure 1.1: An overall schema of a Data Sharing Scenario in Smart Cities.

1.2.1 Smart Cities Data Sharing Scenario

Figure 1.1 shows our overall smart cities motivating scenario. Various sensors are deployed for sensing data in cities by service providers or citizens. We have different applications or services which may share their data or resources between them. Examples of such services include intelligent parking solutions, waste management, public lighting, air quality monitoring, and crowd-participatory sensing applications. A shared platform, may be provided by a data operator, will be used by the diverse applications. In this platform, a data usage control module is needed to deal with issues of trust and control. This module allows data providers to exercise some control over the generated data by their sensors and ensure that the policies put in place by the data producers are respected by data consumers.

We use a context of an Intelligent Parking Application (IPA) to demonstrate our motivation. This application has three main use cases: (*i*) monitoring data parking places, (*ii*) unexpected uses by data consumers, and (*iii*) observations of data usage. The generated sensor data are used not only by this application but also by other applications. Data owners therefore must define data usage policies to control the usage of their data. We have different data consumers such as municipal authorities, application developers and commercial operators. They can request to access data at the granularity and scope that the data owners has specified. We present examples of the

data usage policies in our scenario as follows

- The data owner (the company that deploys and is the owner of the parking sensors) will have full access to all the details generated by all the individual parking sensors.
- The data owner is willing to make available the average occupancy of parking places per street on an hourly basis to municipal authorities.
- However, the data owner will only offer commercial service providers statistical data, only per zone and only a on a weekly basis.
- The monetization of data is allowed, based on subscription type or on a user's role, for example.

1.2.2 Requirements reflected as our Research Questions

The main requirement in our scenario is about data usage control: How data is used after access to it has been granted? It is related to two main research questions.

- How do data owners define their data usage policies? ($Q1$)
- How do platforms ensure responsible data usage? ($Q2$)

For the first question, we focus on following aspects: (i) What are the main criteria to define the policies? (ii) How do we deal with potential conflict between dependent policies? and (iii) How do data owners exercise some control the usage of their data?

The second one deal with perspectives: (i) How do the platform process the data consumers' request and offer an explanation when the request is refused? (ii) How do the platform trace data usage? and (iii) How do data owners customize their policies and explore the consequences of certain change?

1.3 Contributions of the Thesis

In this thesis we propose a comprehensive trustworthy data sharing approach to deal with the above-mentioned issues of trust and control in the

context of an intermediation platform for smart cities. The main contributions are three-fold:

- First we propose a policy-based data usage control model, called DUPO, to capture the diversity of obligations and constraints that data owners impose on the use of data. In particular, this model covers the major data usage requirements such as spatio-temporal granularity, abstraction/masking of certain information, conditions depending upon the class of actor/purpose, and the monetization of data. The conceptual model, its formal theory based on defeasible logic (DL), and illustrative scenario are presented.
- Based on the DUPO and semantic technologies we define a trustworthy data sharing platform which enhances data usage transparency and traceability in the context of smart cities. It includes core components for data usage control in perspectives of data providers, data consumers, and IoT intermediation platform. We also illustrate procedures for trustworthy data sharing in the platform.
- Finally, a proof-of-concept is developed, its implementation choices and a visualization tool prototype which help users to control and monitor easily how their data is shared. We then do a preliminary performance analysis for the proposed solution.

1.4 Thesis Organization

The thesis is organized as follows:

- Chapter I presents the introduction including thesis context, problem, motivating scenario, and research questions.
- Chapter II provides the basic concepts and background about smart cities, data models and intermediation platforms.
- Chapter III covers the state-of-the-art of trust and control enhancing technologies.
- Chapter IV presents the novel policy-based data usage control model (DUPO).

- Chapter V introduces the trustworthy data sharing framework based on the DUPO and semantic technologies.
- Chapter VI illustrates the prototype implementation and experiment's results.
- Chapter VII concludes the thesis and provides some ideas for the future direction of this research.
- Finally the main research papers are attached with this thesis in the following order.

"The doors will be opened to those who are bold enough to knock."

Quote by Louise Haye

Chapter 2

Basic Concepts and Background

Contents

This chapter discusses the basic concepts and background of our thesis domain. At the beginning, we started with the overview about concepts and applications of smart cities in section 2.1. In section 2.2, data models are identified in aspect of context information management and semantic modeling, to show their contributions for data exchange, and integration from different sources in the IoT environments. Finally the technical background about IoT intermediation platforms are presented in section 2.3. It is focused on high level architecture, key functionalities for semantics, and existing core platforms.

2.1 Smart Cities

In the early 1990s the phrase "smart cities" was coined to signify how urban development was turning towards technology, innovation and globalisation (Gibson et al. 1992). However, it is more recent interest in because of the strong concern for sustainability, and the rise of new technologies, such as mobile devices, the semantic web, cloud computing, and the Internet of Things (IoT) promoting real world user interfaces (Schaffers et al. 2011). In this section, we describe more about smart cities with its concepts and applications.

2.1.1 Concepts

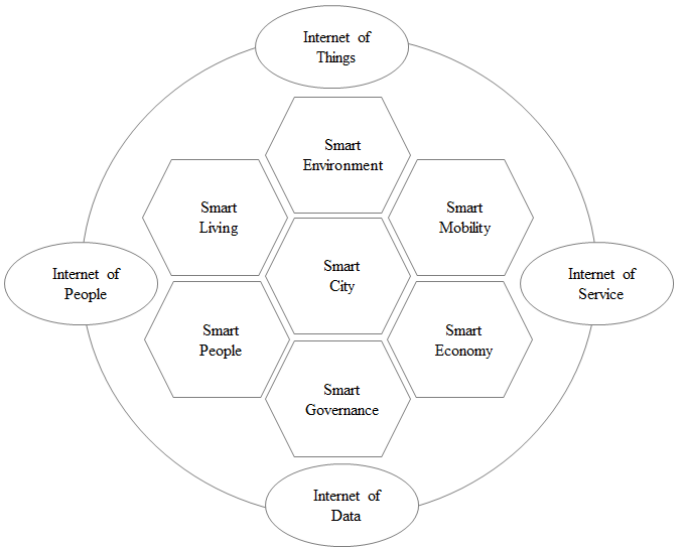

Figure 2.1: A Smart City Model

A city is "smart" if it provides better efficiency for urban planning through a variety of technologies (Miorandi et al. 2012). According to Townsend (2013), smart cities is defined as "places where information technology is combined with infrastructure, architecture, everyday objects, and our bodies to address social, economic, and environmental problems". The European Parliament [1] proposed a smart cities definition: "It is a city seeking to

[1] Mapping smart cities in EU: http://www.smartcities.at/assets/Publikationen/Weitere-

address public issues via information and communication technology (ICT)-based solutions on the basis of a multi-stakeholder, municipality based partnership". The smart cities concept is also defined by Schaffers et al. (2011) which particular interest in the application of the IoT paradigm (Zanella et al. 2014) to an urban context, as it responds to the strong push of many national governments to adopt ICT solutions in the management of public affairs. Although there is not yet a formal and widely accepted definition of "smart cities," the final aim is to make a better use of the public resources, increasing the quality of the services offered to the citizens, while reducing the costs. The technological advancement such as the IoT (Atzori et al. 2010, Roberti 2010, Harmon et al. 2015) are among the main vehicles for realizing this vision in developing a city that can adapt to the needs of its citizens.

2.1.2 Applications

Several smart cities initiatives have started to better serve citizens and to improve their quality of life. Barki et al. (2015) mentions concepts of city automation with green applications that allow the saving of energy. Khatoun & Zeadally (2016) introduces more general smart cities model in Figure 2.1 which consists of six components of applications such as Smart Governance, Smart Mobility, Smart People, Smart Living, Smart Environment, and Smart Economy. These components have also been considered in the European Smart Cities project (http://www.smart-cities.eu) to define a ranking criterion that can be used to assess the level of "smartness" of European cities. In addition, the smart cities indications(Lazaroiu & Roscia 2012) are presented in Table 2.1.

As a cornerstone of smart cities, smart governance means various stakeholders' engagement in decision making and public services (Nam & Pardo 2011). The initiatives in smart governance has introduced by Schaffers et al. (2011), such as government services to citizens, decision making, participation, direct democracy, monitoring and measurements. Electronic governance (e-governance) is key application which focuses on a government's performance through the electronic medium to facilitate an efficient, speedy, transparent process for disseminating information to the public(Paskaleva

Publikationen-zum-Thema/mappingsmartcities.pdf

Table 2.1: Smart Cities Indicators

Smart Cities	**Indicators**
Smart Economy	Innovative spirit
	Entrepreneurship
	Economic image & trademarks
	Productivity
	Flexibility of labor market
	International embeddedness
Smart Mobility	Local accessibility
	(Inter-)national accessibility
	Availability of ICT-infrastructure
	Sustainable, innovative and safe transport systems
Smart Environment	Attractiveness of natural conditions
	Pollution
	Environmental protection
	Sustainable resource management
Smart People	Level of qualification
	Affinity to lifelong learning
	Social and ethnic plurality
	Flexibility
	Creativity
	Cosmopolitanism/open-mindedness
	Participation in public life
Smart Living	Cultural facilities
	Health conditions
	Individual safety
	Housing quality
	Education facilities
	Touristic attractiveness
	Social cohesion
Smart Governance	Participation in decision-making
	Public and social services
	Transparent governance
	Political strategies & perspectives

2009). For example, the citizens are allowed to fulfill their civic and social responsibilities through a Web portal (Khatoun & Zeadally 2016).

Smart mobility is related to accessibility within the city as well as outside the city and availability of ICT-infrastructure, sustainable, innovative and safe transportation systems (Buhalis & Amaranggana 2013). According to Benevolo et al. (2016), Smart Mobility objectives are in the following six categories: (*i*) reducing pollution; (*ii*) reducing traffic congestion; increasing people safety; (*iii*) reducing noise pollution; (*iv*) improving transfer speed; (*v*) reducing transfer costs. Intelligent Transport Systems(ITS) are key applications which could collect, storage and process data, information and knowledge aiming at planning, implementing and evaluating integrated initiatives of Smart Mobility.

The concept smart people is used with regard to social and human capital. It comprises various factors like the level of qualification, affinity to life long learning, social and ethnic plurality, flexibility, creativity, cosmopolitanism or open-mindedness, and participation in public life (Nam & Pardo 2011). Smart people is an important component in smart cities (Caragliu et al. 2011) and their initiatives need to be able to provide advanced applications and services to citizens, such as initiatives supporting distance learning and online courses as a way to reach this result (Letaifa 2015).

The initiatives under Smart Living are mainly targeting new technology adoptions for improving high quality of life in terms of services, enhancing attractiveness for tourists, and promoting social cohesion and safety (Letaifa 2015). According to Lazaroiu & Roscia (2012), smart living comprises various aspects of cultural facilities, health conditions, individual safety, housing quality, education facilities, touristic attractiveness.

Smart Environment is described by attractiveness of natural conditions (climate, green space etc.), pollution, and sustainable resource management and also by efforts towards environmental protection (Balakrishna 2012).

Smart Economy includes factors around economic competitiveness such as innovative spirit entrepreneurship, economic image & trademarks, productivity, and flexibility of the labour market as well as the international embeddedness (Balakrishna 2012).

2.2 Data Models

A variety of data models have been proposed in the IoT environments influencing data exchange, and integration from different sources. In the following, we present a information model for context management and a semantic data model. We aim to introduce related background about data model which can be used in the context of smart cities.

2.2.1 Context Information

The standard NGSI (OMA 2010) provides specification for Context Information management. The central aspect of the NGSI 10 is the concept of entities, that are the virtual representation of all kinds of physical objects in the real world and any available information about entities is expressed in the form of attributes which are composed of attribute name and attribute type. Figure 2.2 gives some examples on entities that can be used as Context Entities.

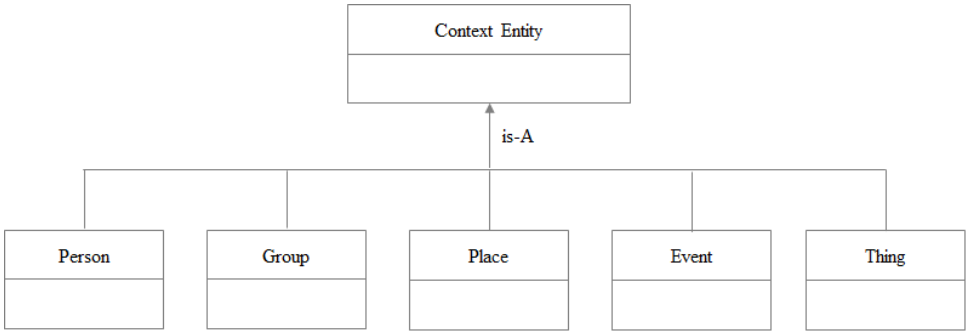

Figure 2.2: Examples of Context Entities

Context Entities are described by the Context Information Model. The Context Information Model details how Context Information is structured and associated to Context Entities in order to describe their situation. In this model, Context Information is organized as Context Elements, which contain set of Context Attributes and associated metadata. Details on this model in Figure 2.3 are provided in the following subsections.

- **Context Entity ids and types**

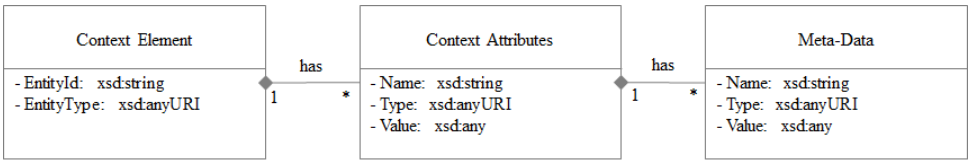

Figure 2.3: Context Information Model

Context Entities are identified using an entity identifier (EntityId). The optional entity type may be needed when the EntityId doesn't contain type information or when the EntityId is only unique per entity type. The entity type is defined as a URI, and thus can be for example an ontology reference (as URL) or a namespace (as URN).

- **Context attributes and attribute domains**

  A context attribute represents atomic Context Information. An attribute is defined as a set of information, namely a name, a type, a value and a set of associated metadata (e.g. timestamp, expires, source). The attribute value is expressed as any content, including strings or opaque objects represented using standard formats. An attribute domain represents the grouping of multiple attributes. Attribute domains allow requestors to specify a set of attributes of interest using a single string as attribute domain name.

2.2.2 Semantic Modeling

A semantic approach for data model is being pushed forward within several standard such as the ETSI (2011), OneM2M (2015), and the W3C Web of Things[2]. The semantic-supported data model can either be data model based on ontologies or existing data models extensible with semantic annotations. The value of semantic technologies has been recognized for data integration, modeling, and processing.

Data integration research has been focused in database schema integration approaches and the use of ontologies and related semantic technologies to provide data consistency among heterogeneous data set schemas. Ontology is a formal specification of a conceptualization that is defining concepts as

[2]http://www.w3.org/WoT/

objects with their properties and relationships versus other concepts. Therefore, Ontology can be defined as a linguistic artifact that defines a shared vocabulary of basic concepts for discourse about a piece of reality (subject domain) and specifies those concepts including operations. In the context of IoT, ontologies are expected to define resources (e.g. sensors), observation and measurement data (e.g. temperature), domain concepts (e.g. location), services (e.g. device functions) and other data sources. By use of ontologies, semantics are added to data models to create common understanding of data among people and system, and also facilitate data sharing and reuse from different sources.

One major benefit of expressing data representation with semantic language relates to its ability to provide high level and expressive abstractions. For instance, in the IoT, data abstraction is concerned with the ways that the physical world is perceived and managed. Currently, there are numerous efforts to provide ontologies for various domains. For example, for sensors we have an SSN ontology(Compton et al. 2012) that was developed and proposed at the W3C for standardization. Other ontologies include the Smart Appliance REFerence(SAREF) ontology developed by TNO[3], which covers popular sensors and actuators. OneM2M ontology provides a basic ontology reference to model devices, services, functions, and operations in IoT. As an ontology being developed, the base ontology replies on interworking with other IoT ontologies to model IoT devices. Recently, Linked Open Vocabularies for the Internet of Things (LOV4IoT)[4] referenced to more than 300 existing ontology-based projects relevant for the IoT.

The Web since its origins has been a vehicle of data interchange. However, automatic discovery and integration of Web data has been impractical until the availability of the RDF framework and RDF data sources. The flagship initiative on this area, Linked Data (Berners-Lee 2006) has fostered both the size of the structured Web data and its exploitation (Bizer et al. 2009). Also, the Web currently explores other approaches based on embedded JSON information or microformats using the tag facilities for HTML. In particular, a specific syntax for using JSON called JSON-LD has been

[3]https://www.tno.nl

[4]http://www.sensormeasurement.appspot.com/?p=ontologies

recently introduced to serialize LinkedData with the motivation to reduce the size of RDF documents compared to the size yielded by XML serialization.The IoT data and resources are published as Linked DataBerners-Lee (2006). A specific syntax for using JSON called JSON-LD has been recently introduced to serialize LinkedData with the motivation to reduce the size of RDF documents compared to the size yielded by XML serialization. One of the pillars of this idea is the possibility of retrieve specific data in the web of data; this task is performed by SPARQL (Hartig et al. 2009), a SQL-like language that enables querying a RDF store.

2.3 Intermediation Platforms

Accoring to Sanchez et al. (2014), IoT deployments in smart cities are often limited to vertical applications and use cases' specific goals, rather on city-wide transformation goals. We have seen successfully deployed and operated IoT systems in a wide range of application areas such as smart water management, urban freight management and logistics systems, urban mobility systems, as well as smart energy systems. These systems have helped the cities to improve certain aspects of citizens' quality of life, but have not been exploited towards establishing wider innovation ecosystems, which would enable them to maximize their innovation potential through access to a larger pool of innovation resources beyond public funding(Manyika et al. 2015). The main challenge in particular is lack of interoperability and scalability. Urban IoT deployments tend to form disaggregated silos (Schiele et al. 2014), which result in information and applications fragmentation. Thus, there is a need for IoT interoperability across different deployments, which could enable repurposing and reuse of costly IoT infrastructures. Recent technological advances provide the means for addressing the challenge. We aim to provide a technical background for an intermediation platform in context of smart cities.

2.3.1 High Level Architecture

In an effort to improve interoperability between different M2M solutions, standardisation bodies have conducted research into the creation of compre-

hensive frameworks that support various enabling technologies in order to make it of practical value. A standard of particular interest is the European Telecommunications Standards Institute (ETSI) SmartM2M. As it provides for a common, distributed middleware between different applications and devices (ETSI 2011). Figure 2.4 shows the standard ETSI SmartM2M high-level architecture.

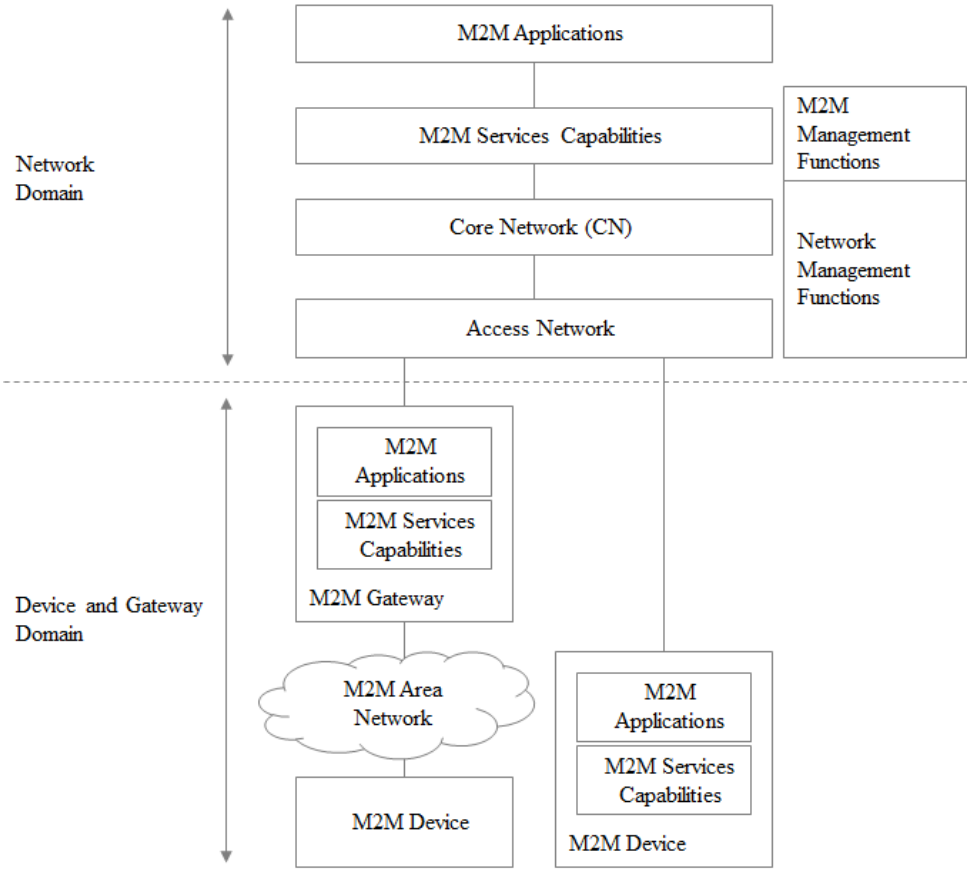

Figure 2.4: ETSI M2M High Level Architecture

The ETSI high level architecture for M2M includes a Device and Gateway Domain and a Network domain. M2M Devices are able to run M2M Applications using M2M Service Capabilities. It connects directly to the Network Domain via the Access network. M2M Devices can also connect to the Network Domain via an M2M Gateway using the M2M Area Network. It may provide service to other devices (e.g. legacy) connected to it that

are hidden from the Network Domain. M2M Gateway are able to act as a proxy for the Network Domain towards the M2M Devices that are connected to it. The M2M Gateway also runs M2M Applications using M2M Service Capabilities. Access Network allows the M2M Device and Gateway Domain to communicate with the Core Network. M2M Service Capabilities Layer provides M2M functions that are to be shared by different Applications. It exposes functions through a set of open interfaces. M2M applications then run the service logic and use M2M Service Capabilities accessible via the open interfaces.

We also particularly consider the previous work (Khan, Belqasmi, Glitho, Crespi, Morrow & Polakos 2015) which introduces a multilayer architecture for supporting multiple applications and service to be provisioned over a deployed Wireless Sensor Network (WSN). The architecture is also enhanced with new layers, entities and functionalities to allow interactions between WSN IaaS and PaaS to develop and deploy WSN applications and services. In addition to this, they deal with the issue of provisioning semantic applications and services over a virtualized WSN (Khan, Jafrin, Errounda, Glitho, Crespi, Morrow & Polakos 2015). The work propose the concept of base ontology which is independent of any application domain and truly reflects the deployed WSN infrastructure. This opens up the possibility for WSN infrastructure owners to offer their network to a variety of users from different domains.

2.3.2 Existing Core Platforms

There are relevant existing core platforms which provide the interoperability, security & privacy, and citizen-centric functionalities for building an intermediation platform.

The choice of oneM2M platform guarantees the semantic interoperability of IoT ecosystems. Figure 2.5 shows a generic functional model to support semantics for various M2M applications OneM2M (2015).The functionalities are logically composed of three main parts for service access, abstraction & semantics, and data access.

- For the service access, it provides an interface with various M2M applications. In the semantic analysis and query, the requests from an

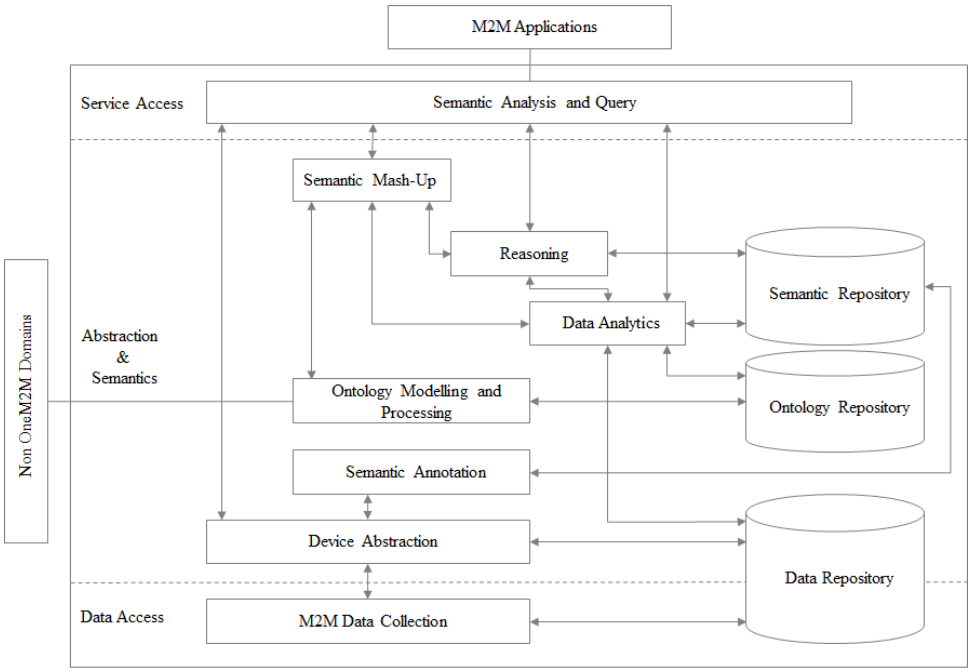

Figure 2.5: OneM2M Generic functional model for supporting semantics

M2M application are analyzed semantically. Based on the analysis, it creates semantic query messages and sends the messages to functional components in abstraction and semantics for requesting semantic information. After obtaining the requested information, it responds to the M2M application.

- For the abstraction & semantics, it performs main functionalities for semantics to M2M data and resources. Reasoning is a mechanism to derive a new implicit knowledge from semantically annotated data and to answer complex user query. Ontology repository provides a way for storing, retrieving and maintaining of ontology which is described as OWL or RDF. Ontology modeling and Processing is for building an ontology which is used to model a domain and support reasoning about concepts, classifying, storing and providing discovery function of published/modeled ontologies from external and internal of the M2M domain. The ontologies are converted and stored in Ontology repository. Semantic mash-up provides functionalities to support

new services through the creation of new virtual devices, which do not exist in physical world, by obtaining semantic information through semantic descriptions from existing M2M resources in the M2M System. Semantic annotation of M2M resources is a method for adding semantic information to M2M resources so that it provides consistent data translation and data interoperability to heterogeneous M2M applications. Semantics repository stores the semantics annotations of resources. Device abstraction is a process of mapping between a set of Device Application Information Models and an Abstract Application Information Model according to a specified set of rules. It allows to communicate with multiple, different but semantically similar devices through a virtual device that offers the functionality of the abstracted Application Information Model. Data repository basically stores new data and also provides functions to support the search, modification and deletion of the stored data.

- The data access functionality provides connections with a device and/or a gateway for accessing M2M data. It has M2M Data Collection which raw data from devices with sensors and/or gateways are collected and stored in the data repository.

FIWARE (2016) has proposed many generic enablers for the IoT which could act as back-end elements of core platform which will provide cloud-based functionalities for BigData storage, processing and analytics, with emphasis on the processing of IoT data streams. The generic enablers prescribed as part of the conceptual architecture include: (*i*) Context Broker enabling the retrieval of aggregated information from Internet-of-Things deployments, including multiple devices and gateways; (*ii*) BigData analytics engines, such as the COSCOS GE, yet the deployment of other analytics engines are possible; (*iii*) GEs for data storage and management, such as the context streams GE for generation, storage and analysis of data streams. (*iv*) GEs for the semantic annotation of IoT streams. FIWARE provide strong security at the level of services interactions and at the level of data encryption. Other IoT Platforms such as IERC-AIOTI[5] platform deployments can be deployed at

[5]http://www.aioti.org/

the back-end cloud infrastructure in order to provide data storage, semantic annotation and analytics, and OM2M[6] is an open source implementation of the OM2M platform which is already deployed and validated in the scope of smart buildings. Next, OpenIoT(Soldatos et al. 2015) is also an open source platform providing front-end and back-end functionalities, including interoperability across diverse IoT streams. Recently, LWM2M (Lightweight M2M) of OMA has been applied to various IoT standard platforms to provides standards-based security, subscriptions, and notification functionalities. It also provides a COAP interface for integration with other IoT systems.

2.4 Summary

In this chapter, we summary background of smart cities, including its concepts and applications. The data models are explained in aspect of context information management and semantic modeling, to show their contributions for data exchange, and integration from different sources in the IoT environments. The technical background about IoT intermediation platforms are presented in detail of high level architecture, key functionalities for semantics, and existing core platforms.

[6]http://www.eclipse.org/om2m/

"You can't stop the waves, but you can learn to surf."

Quote by Jon Kabat-Zinn

Chapter 3

State of the Art: Trust and Control

Contents

This chapter aims to provide a state of the art about trust and control enhancing technologies. In the following, we cover a great deal of the research which has focused on earning trust and gaining respect and confidence related to data sharing in different situations, such as Web, Social Networks, Ubiquitous Computing, Cloud Computing, Wireless Sensor Network, and IoT. The relevant studies can be categorized as following, privacy preservation in section 3.1, data licensing in section 3.2, access control in section 3.3, usage control in section 3.4, and trust computation in section 3.5. We then identify the gaps about data usage control in context of an intermediation platform for smart cities in section 3.6.

3.1 Privacy Preservation

In recent years a number of concepts and principles have been proposed in research domain called privacy-preserving. According to Le Métayer (2016), technologies designed to enhance privacy can be classified into two main categories: (*i*) Technologies for avoiding or reducing as much as possible the disclosure of personal data, hence enforcing the data minimisation principle; and (*ii*) Technologies for enforcing the rights of the subject if personal data is disclosed or processed. The former focuses on data minimization while the latter focuses on enforcing policies in data processing. In particular, Gürses et al. (2011) states that data minimization should be the foundational principle. Kung et al. (2011) defines three principles, minimisation, enforcement and transparency. Kung (2014) integrates accountability with transparency and adds a fourth principle, modifiability. Hoepman (2014) focuses on four data oriented strategies (minimize, hide, separate, aggregate), and four process oriented strategies (inform, control, enforce, demonstrate). Moreover, The International Working Group on Data Protection in Telecommunications [1] has studied on privacy and big data. It recommends the following

- The use of a clear legal basis: the need for valid consent (free, specific, unambiguous); or the data controller's legitimate processing interest as long as these interest are not overridden by the interests of the individual
- Anonymization: the need for properly engineered anonymization schemes
- Transparency: the need for each individual to have access to his or her profile, including information on which algorithms have been used, with information provided in a clear and understandable format.
- Privacy-by-design[2] and Accountability: the need for Privacy Impact Assessments as tools, the need for data controllers to demonstrate that they are being accountable.

[1] ISO/IEC 29190, Information technology – Security techniques – Privacy capability assessment model International Working Group on Data Protection in Telecommunications. Working Paper on Big Data and Privacy, Privacy principles under pressure in the age of Big Data analytics

[2] Privacy-by-Design. http://www.ipc.on.ca/english/Privacy/Introduction-to-PbD/

Privacy is a major issue when it comes to data sharing and the challenge is to provide techniques allowing data publishers to publish data in such a way that he does not breach the privacy of the data subjects yet still retains sufficient utility for the data recipients (Fung et al. 2010). However, we do not mainly focus on privacy. What does go along with data usage control is the notion of the levels of abstraction that the data producer wishes to provide. It means that which level of information should be shared. These abstractions could be studied to provide mechanisms that can be used by a privacy module. Speiser et al. (2011) specified the conceptual policy model for privacy in the figure 3.1 to deal with this issue of abstraction of information. However, there is still no specific data usage control model to express the constraints and obligations on the use of IoT data among participants. In particular, this model have to response to the obligations defined by the actors for their data such as (*i*) Abstraction of certain information, (*ii*) Spatial and temporal granularity, (*iii*) Classification of actors and purposes, and (*iv*) Monetization of data.

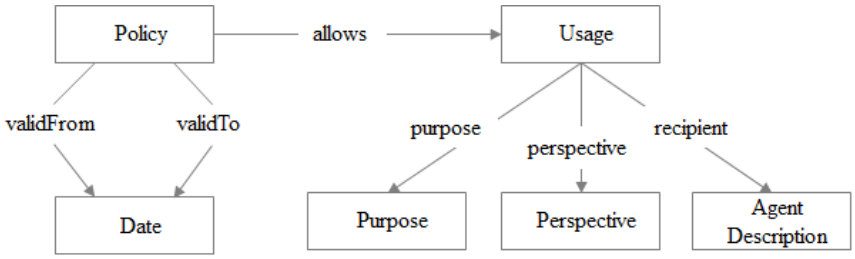

Figure 3.1: Conceptual Policy Model (Speiser et al. 2011)

Moreover, information accountability is complementary to privacy. Shifting to accountability as the basis for considering information sharing and disclosure is more tractable than abstract notions of privacy. In the context of a social network, Pato et al. (2011) proposed the solution which encourage responsible use of information by combining clearly expressed usage policies with systems for detecting misuse. We also consider that transparency and traceability are essential in the context of smart cities. However, there are still lack of mechanisms to allow for automated data usage control and traceability of data usage in this context.

3.2 Data Licensing

Data licensing is an active research domain which enables self-description of data consisting in licensing terms. A licensing vocabulary example is introduced by (Rotolo et al. 2013) in Figure 3.2. The licensing terms aim to specify the admitted use and re-use of the data by third parties. Heath & Bizer (2011) identified that the absence of clarity for data consumers about the terms under which they can reuse a particular dataset, and the absence of common guidelines for data licensing, are likely to hinder use and reuse of data.

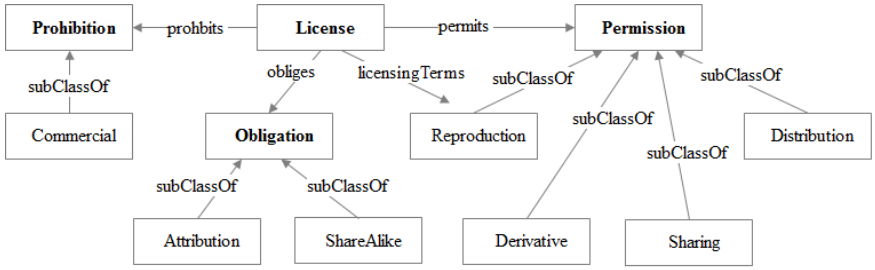

Figure 3.2: The l4lod lightweight vocabulary (Rotolo et al. 2013)

In order to support for generating licenses, Cabrio et al. (2014) address the research question: *How to support users in defining RDF licenses from natural language ones?* In particular, they study about RDF representation of licenses using CCRel and ODRL vocabularies, and classification problem in supervised learning with support vector machines. A synopsis of the overall framework are also provided to generate RDF licenses specification from natural language license.

Krötzsch & Speiser (2011) develop a general policy modelling language, then instantiated with OWL DL and Datalog, for supporting self-referential policies as expressed by CC. It aims to model licenses as part of the data to enable easy exchange and automated processing. A simple provenance model illustrated informally in Figure 3.3 to specify the conditions of the policy. This model in particular represent the origin of the artefact, and the context in which it has been published. A semantic framework is also introduced for evaluating ShareAlike recursive statements.

In the paper by Governatori et al. (2013), they also address the research

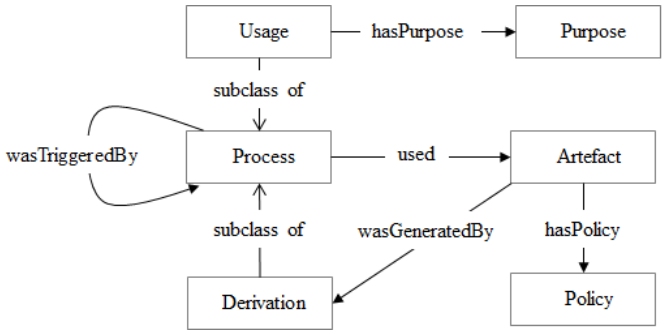

Figure 3.3: A provenance information model (Krötzsch & Speiser 2011)

question: *How to express the licensing terms associated to data coming from heterogeneous distributed sources?* In fact, they focus on licenses compatibility and composition and propose a framework to compose in a compliant and automated way the licensing terms associated to a set of heterogeneous data to produce a single composite license. The rationale behind this framework is to build a composite license starting from the single licensing terms associated to heterogeneous data. They adopt deontic logic to ensure the compliance of the composite license with respect to the single licenses composing it.

Morevover, Pucella & Weissman (2002) propose a logic to check whether the user's actions follow the licenses' specifications. Nadah et al. (2007) propose to assist licensors' work by providing them a generic way to instantiate licenses, independent from specific formats. Gordon (2011) present a legal prototype for analyzing open source licenses compatibility using the Carneades argumentation system.

This is an important research topic that potential deals with issues of trust and control of data usage, but existing solutions have not yet been focused on improving the data usage transparency and traceability and there is no way to express the constraints and obligations on the use of IoT data among participants. In our work, we may adapt formal logical methods based on the work done in the area of defeasible logic (DL) by (Governatori et al. 2013, Rotolo et al. 2013) to express the obligations and conditions in usage control policies and to provide the enforcement of these policies. The data usage policies can reprsent as regular defeasible rules with obligations,

permissions and prohibitions operators. The advantage of DL is its efficiency as possibility to compute the set of consequences in linear time and solving issues of rule conflict as well. However, we need to work more on the challenge to demonstrate that these policies are correctly enforced. It is also essential to find a language with sufficient expressive power which capture all the obligations and conditions that actors impose on the use of the data.

3.3 Access Control

Access Control (AC) is a key issue to enable a secure and trustworthy data sharing as it regulates who can access protected information or services. Many mechanisms have already been specified to control the access toward software systems. The basic model is the access control matrix (ACM), which specifies for each combination of user and type of access, whether it is allowed or not (Lampson 1974). Access control frameworks rely on access control lists (ALCs) that define which users can access the data. For example, Hollenbach et al. (2009) present a system where providers control access to RDF documents using Web Access Control vocabulary (WAC) [3]. Similarly to ACLs, other approaches that specify allowed accesses not for user identities but for roles, called role-based access control (RBAC) (Ferraiolo & Kuhn 2009). Sandhu et al. (1996) briefly discuss possible ways of going beyond RBAC such as Attribute Based Access Control (ABAC), a model that grants access according to client attributes, instead of relying on AC lists. Access control models may consider not only the information about the consumer who is accessing the data, but also the context of the request, e.g., time, location. Covington et al. (2001) present an approach where the notion of role proposed in RBAC is used to capture the environment in which the access requests are made. Cuppens & Cuppens-Boulahia (2008) propose an Organization Based Access Control model (OrBAC) that contains contextual conditions. Toninelli et al. (2006) use context-awareness to control access to resources, and semantic technologies for policy specification.

While the security aspects in access control have been dealt with extensively, issues to address transparency and traceability of data usage are still

[3] http://www.w3.org/wiki/WebAccessControl

subjects of research. Also, access control cannot deal with situations where information is published on purpose but should still have restricted usages (Speiser & Harth 2012). The data-purpose algebra by Hanson et al. (2007) mentions the modelling of usage restrictions of data and the transformation of the restrictions when data is processed. In their approach, a data item is associated with its content, the agent who produced it, the set of purposes for which usage is allowed and a set of categories. Depending on the process performed on a data item a function is defined that transforms the allowed usages. However, a mechanism needed to response the general idea of modeling the constraints and obligations about data usage requirements defined by data owners.This model has to extend to treat the issues of data usage transparency and traceability in the IoT.

3.4 Usage Control Mechanisms

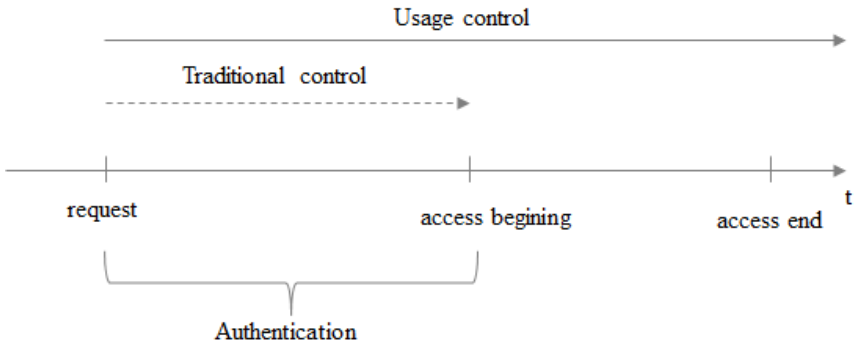

Figure 3.4: Usage control definition(Wu et al. 2015)

Usage control goes further than access control by regulating usage of information after initial access was granted (Pretschner & Walter 2008). The usage control definition is shown in Figure 3.4 by Wu et al. (2015). UCON model is a theoretical foundation for usage control and initially propose by Park & Sandhu (2002) with a purpose of being addressed to emerging digital environments. It enables two advanced features to cope with dynamic networking environment: (*i*) mutability of attributes, and (*ii*) continuity of an access decision. Basically, UCON keeps track of changes of attributes

and policies when access is in progress, resulting in being able to change permission decision. Then an authorization system revokes granted rights or terminates resource usages accordingly. The permission decision is determined based on three factors called Authorizations, Obligations and Conditions. Authorizations are predicates over subjects (data consumers) and/or objects (stakeholders, data) attributes and put constrains on them to judge and grant the subjects a certain right on the objects. Obligations is a novel component in UCON model that examines the accomplishment of compulsory tasks that subjects have being done to objects before, during and after access period. Conditions are constrains from environment attributes, not related to both subjects and objects but affect the usage decision process (Park & Sandhu 2004). A notable advantage of UCON is the expressiveness of policies and obligations applied in various access scenarios. UCON not only conveys capability of existing access control models but also goes beyond them. It also has been well studied and continue to be improved by authors already mentioned by Lazouski et al. (2010). Pretschner, Hilty, Florian, Schaefer & Walter (2008) give an overview of enforcement of usage control and also present a corresponding language (Hilty et al. 2007) and enforcement mechanisms (Pretschner, Hilty, Basin, Schaefer & Walter 2008).

Data in IoT environments is generated by a large variety of participants including end users and potentially undergoes several transformations such as aggregation and/or composition before finally being consumed. Usage control may deal with policies and mechanisms to ensure that consumers fulfill the obligations and conditions that data owners desires to impose on its utilization. While usage control has been used recently in different domain such as social networks and semantic web, to the author's knowledge it has not yet been applied in the IoT. The main focus of our study are on issues that we consider have not been treated in IoT. In fact, what does go along with usage control is the notion of the levels of abstraction that the producer wishes to provide, for instance mean data over a day and over a geographical zone rather than individual elements from each sensor and for each time period. The main technical challenges are to express the obligations and conditions in usage control policies and to ensure the transparency

and traceability of the policy enforcement rules. Actors also need to have an easily interpretable tool to demonstrate in a clear fashion the reasoning behind the rules. Such a visualization tool helps the actor understand the implications of the different choices that are made while defining the policies as well as understanding at runtime how conflicts have been resolved. The latter aspect is critical for accountability.

3.5 Trust Computation

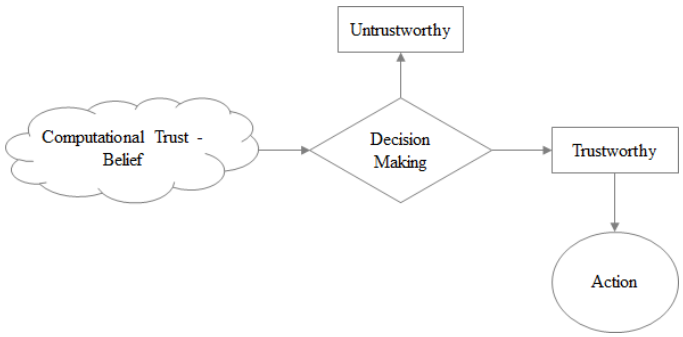

Figure 3.5: The relationship between computational trust and trustworthiness(Lu 2011)

A comprehensive summary on trust has been investigated in (Sicari et al. 2015, Yan et al. 2014). Technically, trust is a measurable belief of a trustor in a trustee that the trustee will provide or perform a given task in an expected manner within a specific trust context, for a specific period, measured by a set of trust metrics (TMs) using trust computation (Um et al. 2016). Basically, measuring trust requires us to deal with three questions (Truong et al. 2016): (*i*) What are TMs needed to realize trust?, (*ii*) how are the TMs extracted from various types of data? and (*iii*) how are the TMs aggregated to assemble an overall trust value?

Figure 3.5 presents the relationship between computational trust and trustworthiness(Lu 2011). Several approaches have been proposed to compute trustworthiness based on direct information (direct trust). In this regard, transactions between trustors and trustees are established; and during these transactions, several credentials are generated for evaluating trust

value. Others have measured trust based on third-party opinions (indirect trust) by accumulating feedback after interactions. Following this, a reputation value is then calculated by using heuristic algorithms and used to indicate trust.

Regarding the trust aspect, there is a concern for IoT data sharing, as there are various risks, emphasizing the need for trustworthiness. Trust has many facets, but one critical element in the IoT is the ability for each participant to exercise control on how their data is going to be used. This concrete trust model is still needed to deal with the issue of trust and control in the IoT.

3.6 Summary

Note that several solutions have been proposed for different research activities which have investigated supporting for confidence related to data sharing in different domains. We has categorized the different axes of confidence as follows: (*i*) Privacy Preservation; (*ii*) Data Licensing; (*iii*) Access Control; (*iv*) Usage Control; and (*v*) Trust Computation. While the security aspects such as confidentiality, privacy and access control have been dealt with extensively, the issue of data usage control and data usage transparency and traceability have not been treated in IoT. In fact, usage control is about how data is used after access to it has been granted. There is still no specific data usage control model to express the constraints and obligations on the use of IoT data among participants. It is still lack of data usage transparency and traceability which is essential in this context of smart cities. A visualization tool is also needed to help users customize their policies in an interactive format, which allows them to explore and monitor the consequences of certain changes to how their data is allowed to be used.

"If you don't have any shadows, you're not standing in the light."

Quote by Lady Gaga

Chapter 4

Data Usage Control Model

Contents

Usage control is concerned with how data is used after access to it has been granted. It has potential to deal with the issue of trust and control in context of smart cities. However, there is still no specific data usage

control model to express the constraints and obligations on the use of IoT data among participants. This chapter thus aims to address the research question: *How do the actors define the constraints and obligations?*. In particular, we focus on following aspects: (*i*) *What are the main criteria to define the data usage policies?* (*ii*) *How do we deal with potential conflict between dependent policies?* and (*iii*) *How do data owners exercise some control the usage of their data?*.

In fact, we take into account following obligations defined by the actors for their data: (*i*) Abstraction of certain information, (*ii*) Spatial and temporal granularity, (*iii*) Classification of actors and purposes, and (*iv*) Monetization of data.

We propose a policy-based data usage control model, called DUPO, including its conceptual model in section 4.1, formal theory based on defeasible logic (DL) in section 4.2, and practical expression are explained in section 4.3.

4.1 Conceptual Model

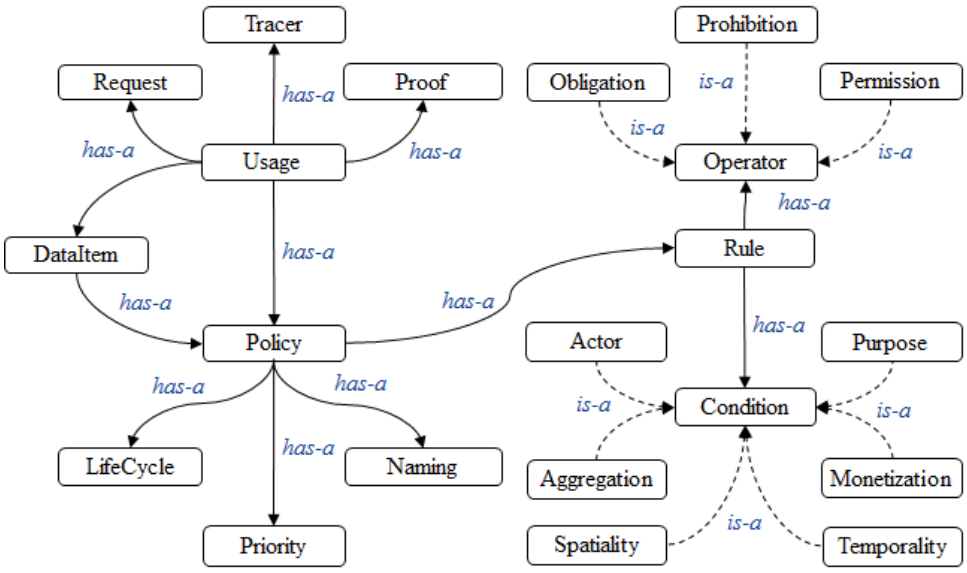

Figure 4.1: Conceptual view of the DUPO model

Figure 4.1 presents the conceptual view of the *DUPO* model. For each

data item, the first step is to define a data usage policy which will be attached to a set of data items. Next, the policy is created by using modal operators (Obligation, Prohibition, and Permission) and data usage conditions: (*i*) class of actors, (*ii*) granularity (Spatiality, Temporality, and Aggregation), (*iii*) class of purposes, and (*iv*) monetization constraints. We also manage the naming, life cycle, and priority of defined policies. Next we focus on the aspects of data usage transparency and traceability.

We explain the concepts of the DUPO model in more detail.

4.1.1 Data Items

A *Data Item* is an individual of the Context Element based on the Information Model OMA (2010) standard specification which is used in the European Project FI-WARE FIWARE (2016) for Context Management. An *Entity Element* is a container used to exchange information about an entity. It contains the following information: (*i*) an entity ID including the name and the type, (*ii*) a list of the entity attributes, (*iii*) (optionally) the name of an attribute domain that logically groups together a set of entity's attributes, and (*iv*) (optionally) a list of metadata that applies to all the attribute values of the given domain. We formally define a *Data Item* by using XML DTD, as mentioned in Listing 4.1.

```
1  <!DOCTYPE DUPO[
2  <!ELEMENT DataItem(EntityElement)>
3  <!ELEMENT EntityElement(EntityID, AttributeDomainName?,
       EntityAttributeList, DomainMetadata?)>
4  <!ELEMENT EntityID(Id, Type)>
5  <!ELEMENT EntityAttributeList(EntityAttribute*)>
6  <!ELEMENT EntityAttribute(Name, Type, EntitytValue, EntityMetadata+)>
7  <!ELEMENT DomainMetadata(EntityMetadata*)>
8  <!ELEMENT EntityMetadata(Name, Type, Value)>
9  ...
10 ]>
```

Listing 4.1: XML DTD Definition of Data Item.

4.1.2 Conditions

The Condition contains (optionally) the following expressions: (*i*) Spatio-Temporal Granularities, (*ii*) Aggregation Granularities, (*iii*) Conditions by

Actors, (*iv*) Conditions by Purposes, and (*v*) Conditions of Monetization. We formally define conditions by using XML DTD, as shown in Listing 4.2.

```
<!DOCTYPE DUPO[
<!ELEMENT Condition(Temporality*, Spatiality*, Aggregation*, Actor*,
    Purpose*, Monetization*)>
<!ELEMENT Spatiality(SpatialScope*)>
<!ELEMENT Temporality(TemporalScope*)>
<!ELEMENT Aggregation(AggregateScope*)>
<!ELEMENT Actor(ActorScope*)>
<!ELEMENT Purpose(PurposeScope*)>
<!ELEMENT Monetization(MonetizationScope*)>
<!ELEMENT TemporalScope(Secondly?, Minutly?, Hourly?, Daily?, Weekly?,
    Monthly?, Yearly?, Any?)>
<!ELEMENT SpatialScope(Street?, Zone?, Any?)>
<!ELEMENT ActorScope(DataOwner?, MulnicipalAuthority?,
    ComercicalServiceProvider?)>
<!ELEMENT AggregateScope(Detail?, Average?, Statistic?, Any?)>
<!ELEMENT PurposeScope(CommercialUse?, Any?)>
<!ELEMENT MonetizationScope(Level?, Any?)>
...
]>
```

Listing 4.2: XML DTD Definition of Condition.

4.1.3 Operators

An Operator contains (optionally) model operators: (*i*) Obligation (*ii*) Prohibition, and (*iii*) Permission. The formal definitions are created using XML DTD as presented in Listing 4.3.

```
<!DOCTYPE DUPO[
<!ELEMENT Operator(Obligation?, Prohibition?, Permission?)>
...
]>
```

Listing 4.3: XML DTD Definition of Operator.

4.1.4 Policies

A *Policy* has its name, lifecycle, priority, and a collection of rules which is created by defining the *Operator* on the individual *Condition*. Listing 4.4 formally defines the policy using XML DTD.

```
<!DOCTYPE DUPO[
```

```
<!ELEMENT Policy(Name, LifeCycle, Priority?, Rule*)>
<!ELEMENT Name(URI?)>
<!ELEMENT LifeCycle(Duration?, Datetime?)>
<!ELEMENT Rule(Operator?, Condition?)>
...
]>
```

Listing 4.4: XML DTD Definition of Policy.

4.1.5 Usage

An *Usage* is created by a consumer's request, related policies, and response data. The data could be a proof justification, a tracked data usage, or a list of returned data items. This component has a purpose for transparency and traceability of data usage. We formally define the Usage using XML DTD in Listing 4.5.

```
<!DOCTYPE DUPO[
<!ELEMENT Usage(Request, Policy*, Data*)>
<!ELEMENT Request(Rule?)>
<!ELEMENT Data(Tracker?, Proof?, DataItem*)>
...
]>
```

Listing 4.5: XML DTD Definition of Usage.

4.2 Formal Theory

Formal theory of the *DUPO* is based on the general concept of DL, which is a non-monotonic formalism that deals with incomplete and conflicting information, originally proposed by Nute (Nute 1994). In particular, we build on earlier works extending DL with modal and deontic operators, as presented in Governatori (Governatori & Rotolo 2008, Governatori et al. 2013) and Antoniou (Antoniou et al. 2001, 2009). Deontic logic is concerned with concepts of obligations, permissions and prohibitions, allowing such relationships to be captured with each entity. There are some proposed formalisms for dealing with reasoning, handling and solving the normative conflicts that arise between rules and exceptions. However, DL is one of the best solutions which can manage all these aspects in an efficient and computationally tractable way (Governatori et al. 2013). Moreover, DL offers enhanced representational capabilities and low computational complexity (Kontopoulos

et al. 2008). According to (Governatori & Rotolo 2008), when DL is enriched with modal deontic operators, the complexity does not increase in most cases. We define the DUPO theory and its proof as follows.

4.2.1 DUPO Theory

Let $PROP$ be a set of propositional atom. A set of literals $Lit = PROP \cup \{\neg p | p \in PROP\}$. Let $MOD = \{O, P, F\}$ be the set of basic deontic modalities (Obligation, Permission, and Forbiddance/Prohibition). A set of modal literals $ModLit = \{[X]l, \neg[X]l | l \in Lit, X \in MOD\}$.

Let Lbl be a set of arbitrary labels. R is a set of base and deontic rules. A base rule is expressed $r : A(r) \hookrightarrow C(r)$, while a deontic rule is $r : A(r) \hookrightarrow_X C(r)$, where (i) A unique label $r \in Lbl$, (ii) The antecedent (or body) $A(r) = a_1, ..., a_n$, $a_i \in Lit \cup ModLit, 1 \leq i \leq n$; (iii) An arrow $\hookrightarrow \in \{\rightarrow, \Rightarrow, \rightsquigarrow\}$, denotes the type of rules: strict rules, defeasible rules and defeaters, respectively, (iv) $X \in MOD$, and (v) The consequent (or head) $C(r) = b, b \in Lit$.

The different rules have the following meaning. Strict rules can never be defeated, while defeasible rules can be defeated by contrary evidence. Defeater rules are only used to prevent certain conclusions.

Definition 1. *A theory* $DUPO = (F^{DUPO}, R^{DUPO}, >)$, *where* $i) F^{DUPO} \subseteq Lit \cup ModLit$ *is a finite set of facts,* $ii) R^{DUPO} \subseteq R$ *is a finite set of rules and* $iii)$ $>$ *is a superiority relation for priorities among the non-strict rules in* R^{DUPO}.

4.2.2 Theory Proof and Conclusions

A conclusion derived from $DUPO$ is a tagged literal and it is classified as follows: $+\Delta q$ means that literal q is definitely provable in $DUPO$; $-\Delta q$ means that literal q is definitely rejected in $DUPO$; $+\partial q$ means that literal q is defeasibly provable in $DUPO$; and $-\partial q$ means that literal q is defeasibly rejected in $DUPO$.

A proof $P = (P(1), ..., P(n))$ in D is a finite sequence of tagged literals of type $+\Delta q$, $-\Delta q$, $+\partial q$ and $-\partial q$.

We denote the set of all strict rules in R by R_s, R_{sd} for the set of strict and defeasible rules, and $R[q]$ for the set of rules whose head is q. $P[1..i]$ denotes the initial part of the sequence of length i. The proof conditions (Antoniou et al. 2001, 2009) for the conclusions are formally defined as follows

$+\Delta : If\ P(i+1) = +\Delta q\ then\ either$
$(1)\ q \in F\ or$
$(2)\ \exists r \in R_s[q] \forall a \in A(r) : +\Delta a \in P[1..i].$
$-\Delta : If\ P(i+1) = -\Delta q\ then$
$(1)\ q \notin F\ and$
$(2)\ \forall r \in R_s[q] \exists a \in A(r) : -\Delta a \in P[1..i].$
$+\partial :\ If\ P(i+1) = +\partial q\ then\ either$
$(1) +\Delta q \in P[1..i]\ or$
$(2)(2.1) \exists r \in R_{sd}[q] \forall a \in A(r) : +\partial a \in P[1..i]\ and$
$(2.2) -\Delta \neg q \in P[1..i]\ and$
$(2.3) \forall s \in R[\neg q]\ either$
$(2.3.1) \exists a \in A(s) : -\partial a \in P[1..i]\ or$
$(2.3.2) \exists t \in R_{sd}[q]\ such\ that$
$\forall a \in A(t) : +\partial a \in P[1..i]\ and\ t > s.$
$-\partial :\ If\ P(i+1) = -\partial q\ then$
$(1) -\Delta q \in P[1..i]\ and$
$(2)(2.1) \forall r \in R_{sd}[q] \exists a \in A(r) : -\partial a \in P[1..i]\ or$
$(2.2) +\Delta \neg q \in P[1..i]\ or$
$(2.3) \exists s \in R[\neg q]\ such\ that$
$(2.3.1) \forall a \in A(s) : +\partial a \in P[1..i]\ and$
$(2.3.2) \forall t \in R_{sd}[q]\ either$
$\exists a \in A(t) : -\partial a \in P[1..i]\ or\ t \not> s.$

4.2.3 Consumer's Requests and Policy Composition

The DUPO theory proof are used as an efficient method for reasoning consumer's requests. We formally define a consumer's request by using a defeasible rule

Definition 2. *A consumer's request is a deontic rule* $rq : actor(a), [P]condition(c) \Rightarrow_O request(r)$*, where i)*$rq \in Lbl$ *is a unique label for the request, ii)*$a \in Lit$ *is*

an actor a in the DUPO, iii)$c \in ModLit$ *is a list of permission conditions for the actor*(a)*, iv)* $\Rightarrow_O$ *denotes a defeasible rule with modality Obligatory, v)*$r \in Lit$ *is a consequent of the request.*

It means if actor a is the case and c are permission conditions, then the request r is obligation in the DUPO. We then base on the DUPO conclusions $+\Delta[O]request(r)$, $-\Delta[O]request(r)$, $+\partial[O]request(r)$, and $-\partial[O]request(r)$ to prove this consumer's request. In case the conclusions are $+\Delta[O]request$, or $+\partial[O]request(r)$, the consumer's request is provable in the DUPO. Otherwise, the consumer's request is rejected.

In fact, we also have to compose the different data usage policies from several participants to process the consumer's request. Thus, we define a complete theory DUPO including policy composition as follows

Definition 3. *A theory* $DUPO^C = (F^{DUPO}, P^C, \{R^r\}_{r \in P^C}, R^q, >)$*, where i)*$F^{DUPO}$ *is a finite set of facts in DUPO, ii)*P^C *is a finite set of data usage policies in DUPO, iii)*$\{R^r\}_{r \in P^C}$ *is a finite set of rules of the policies, iv)*R^q *is a rule of consumer's request, v)* $>$ *is a superiority relation for priorities among the non-strict rules.*

For policy composition, if at least one of the policy involved in the composition owns a rule, then also policy composition owns it *(OR Composition)*. In other cases, if all the policies involved in the composition own a clause, then also policy composition owns it *(AND Composition)*.

4.3 Practical Expression

We have shown the conceptual model, and formal theory of the DUPO. In this section, we present a practical expression in the illustrative scenario and including policy language requirements.

4.3.1 Illustrative Scenario

To explain more the DUPO, we consider an example that a commercial service provider requests all the parking data details of a street on an hourly basis. We already have a data usage policy that states commercial service providers are only permitted access to statistical data over a zone, and that

Table 4.1: Requirements for Data Usage Policies

Types	Requirements	DUPO
Conditions	Abstraction of certain information	x
	Spatial and Temporal granularity	x
	Classification of actors and purposes	x
	Monetization of data	x
Modality Operators	Permission	x
	Prohibition	x
	Obligation	x
Policies	Linked Rules	x
	Naming	x
	Monotonic vs Non-monotonic Rules	x
	Rules Priorities	x
	Policy Life Cycle	x
Data Usage Justification	Transparency	x
	Traceability	x

on a weekly basis in Section 1.2.1. Thus, this consumer's request is refused with a proof justification. Otherwise, the related data items will be returned and data usage is tracked. This example basically covers the usage control requirements and related concepts in the DUPO model:

$$
\begin{aligned}
Actor &= (CommercialServiceOperator),\\
Aggregation &= (Detail, StatisticalData),\\
Spatiality &= (StreetLevel, ZoneLevel),\\
Temporality &= (Hourly, Weekly),\\
Operator &= (Obligation, Prohibition, Permission),\\
Subscription &= (ConsumerRequest),\\
Proof &= (ProofJustification),\\
Track &= (TrackedDataUsage).
\end{aligned}
$$

4.3.2 Requirements

In Table 4.1, we introduce all requirements that play a central role in policy expression in our DUPO model.

4.3.3 Practical Example

We now use our formal language to express facts, rules for usage policies, and consumer's request related to the scenario.

We have a fact about a commercial operator (CO) that wish to request the data. It is presented as follows:

$$F^{DUPO} = \{CommercialOperator(CO)\}$$

We express all rules related to the usage policy for commercial service operators, only statistical data will be made available over a zone and on a weekly basis. They are represented with the use of defeasible rules, as follows:

$$\begin{aligned} R^{DUPO} = \{&r_{1,c} : CO \Rightarrow_P SpatialScope(zone),\\ &r_{2,c} : CO \Rightarrow_F \neg SpatialScope(zone),\\ &r_{3,c} : CO \Rightarrow_P TemporalScope(weekly),\\ &r_{4,c} : CO \Rightarrow_F \neg TemporalScope(weekly),\\ &r_{5,c} : CO \Rightarrow_P AbstractScope(statistic),\\ &r_{6,c} : CO \Rightarrow_F \neg AbstractScope(statistic)\} \end{aligned}$$

The commercial service operator requests all the detail of the parking data over a street on a hourly basis. It is represented with the use of defeasible rules, as follows:

$$\begin{aligned} r : CO, [P]SpatialScope(street),\\ [P]TemporalScope(hourly),\\ [P]AbstractionScope(detail)\\ \Rightarrow_O ConsumerRequest \end{aligned}$$

Based on the DUPO theory, we have the conlusions, $-\Delta[O]ConsumerRequest$, $-\partial[O]ConsumerRequest$. It means that $ConsumerRequest$is not defeasible provable, so the request is refused.

4.4 Summary

In this chapter we proposed a new policy-based data usage control model, called DUPO, which responds to the obligations defined by actors to their data. The conceptual model for usage control with its formalization using defeasible logic have been presented. We also explain its practical expression. In fact, we use the concept of usage control as a starting point from which to propose the DUPO by defining the data usage policies based on spatio-temporal granularity, the abstraction/masking of certain information, conditions depending on the class of actors or purposes, and allowing the monetization of data. The data usage policies have been built on regular defeasible rules. Next, we aim to focus on a trustworthy data sharing platform in context of smart cities.

"Nobody can go back and start a new beginning,
but anyone can start today and make a new ending"

Quote by Maria Robinson

Chapter 5

Trustworthy Data Sharing Platform

Contents

To deal with the issue of trust and control in the context of shared platforms in smart cities, we have proposed the data usage control model

(DUPO) to capture the diversity of obligations and constraints that data owners impose on the use of data in the previous chapter. However, the architectural support to provide data usage transparency and traceability is still lacking, motivating us to develop this type of architectural support for the shared platform.

In particular, this chapter aims to address the research question: *How does the platform ensure responsible data usage?*. We focus on following aspects:(*i*) *How do the platform process the data consumers' request and offer an explanation when the request is refused?* (*ii*) *How do the platform trace data usage? and* (*iii*) *How do data owners customize their policies and explore the consequences of certain change?*.

We based on the DUPO and semantic technologies to tackle this issue and propose a trustworthy data sharing platform. It includes a multi-layer system architecture, core components for data usage control in perspectives of data providers, data consumers, and IoT intermediation platforms, and a trustworthy data sharing mechanism is also illustrated in detail of sequence steps and procedures.

The rest of this chapter is organized as follows. Section 5.1 presents the overall system architecture. Section 5.2 discusses in detail of the platform and the core components are presented in Section 5.3. The trustworthy data sharing procedure is discussed in Section 5.4, and finally Section 5.5 concludes the chapter.

5.1 Overall System Architecture

We build on the previous works in (Khan, Jafrin, Errounda, Glitho, Crespi, Morrow & Polakos 2015) which deal withs the simultaneous acquisition of data by multiple applications and services from deployed sensors. These applications and services can be traditional as well as semantic-based Wireless Sensor Network (WSN) applications. When required, the sensor data can be annotated using sensor domain ontology, such as the Semantic Sensor Network (SSN)(Compton et al. 2012). However, all of this data is sent directly to the consumers (platform or end-user applications) without allowing the owners of the data to enforce certain policies concerning its us-

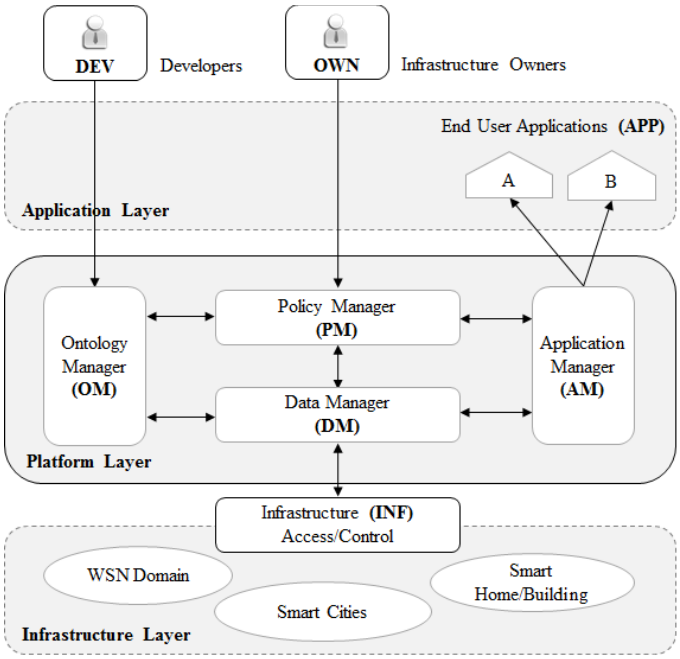

Figure 5.1: Overall System Architecture.

age. In other words, it is assumed that the data is always trusted, which may not be true. For example, issues such as how the same data can be shared among multiple end-users by using different policies based on their location, time or role (home users, city administration or law enforcement agency) are not addressed in the above-mentioned works. In addition, the two architectures mentioned above only consider WSNs as the source of data, whereas in the broader context of the IoT and smart cities, many types of devices, in addition to sensors, provide data to end-users.

Figure 5.1 shows the architecture designed for the proposed trustworthy data sharing platform. It contains the following three layers:

5.1.1 Infrastructure Layer

This bottom layer contains a variety of IoT objects that are deployed to send their data to different applications. Because of the IoT scenario, we consider that these IoT objects can belong to different domains, such as smart sensors from the WSN domain, smart street lights/traffic signal poles

from smart cities domain, or home alarm systems/intelligent HVAC systems from a smart home/building domain. We also consider that some kind of infrastructure access/control mechanism is used by each of these domains independently.

5.1.2 Platform Layer

The platform layer is the middle layer which are discussed detail in 5.2 which focuses on the following four functional entities, Ontology Manager (OM), Policy Manager (PM), Data Manager (DM), and Application Manager (AM). It contributes to the advancement of our previous architecture, in which the OM was used to work with the domain and trust ontologies. Here, the PM is used to work with trust policies, the DM is used to work with IoT data or resources from the infrastructure (INF), and the DM works with IoT applications.

5.1.3 Application Layer

The last layer, the application layer, contains end-user applications (APP) that receives the shared data from the infrastructure through the platform. We also consider that in most cases, the APP will receive and consume the sensor data (sent to it according to a pre-set policy) but also the data's owner (OWN) (probably) wants to know the data's usage.

5.2 Platform as a Service

Figure 5.2 shows an overall overview of the proposed platform. As it is shows in the figure, we have four groups connected to this platform as follow: (*i*) the connected objects, which can be special sensors or users' mobile phones. (*ii*) the data providers of historical records, additional data sets, etc., (*iii*) public data sources, which are open, e.g. calendars, directories, etc., and (*iv*) the array of business applications and developers accessing this platform, all using the shared data.

These are an ecosystem of developers that wish to exploit the data for commercial services or they can be government agencies charged with providing improved citizens services. Developers are able to ascertain data avail-

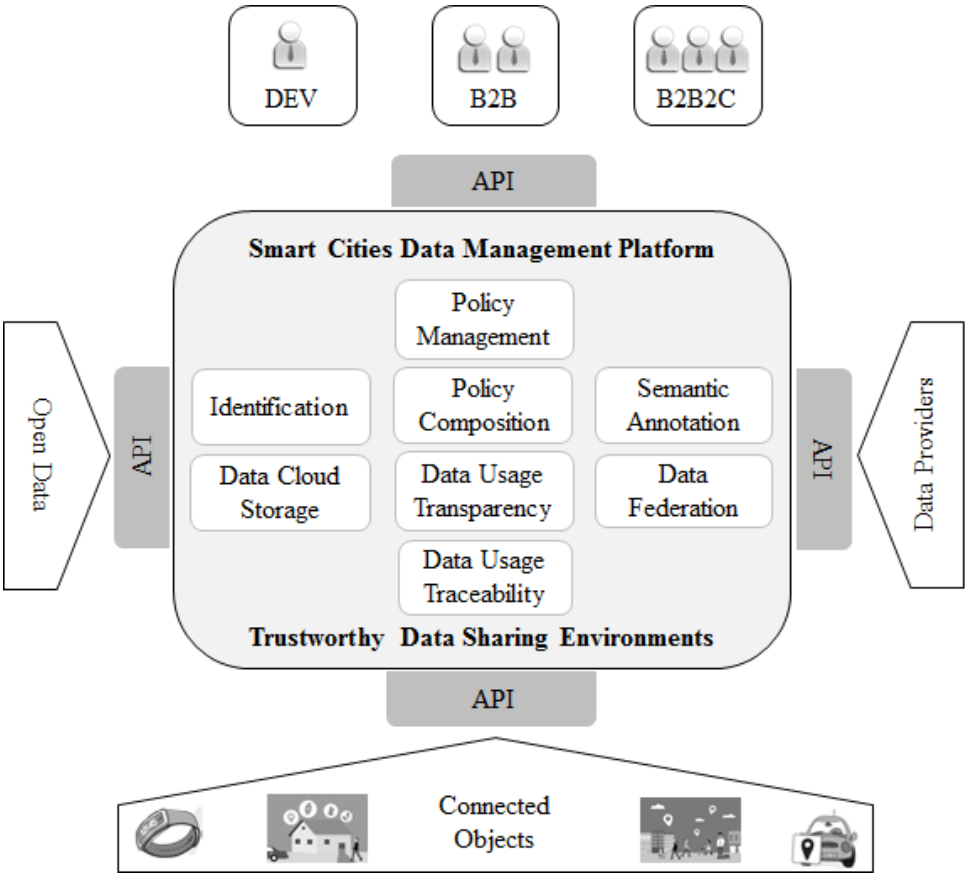

Figure 5.2: Overall Platform for Smart Cities Data Management

ability and the conditions of data usage, so they can quickly and reliably assess the feasibility of their intended development.

We focus only on the platform layer of the architecture and propose the platform as a service (PaaS). Other aspects of the architecture, such as the infrastructure layer (IaaS) and the application layer (SaaS) in the cloud computing paradigm, are out of the scope in this chapter and is a potential future direction. This platform is centralized computing and it includes main components and procedures that are developed based on the DUPO concepts and semantic technologies. In fact, we have added the core components APIs(Application Programming Interface) to allow the transparency and traceability of data usage, and support collaboration between the participants and interoperability of the services in the platform. The platform thus

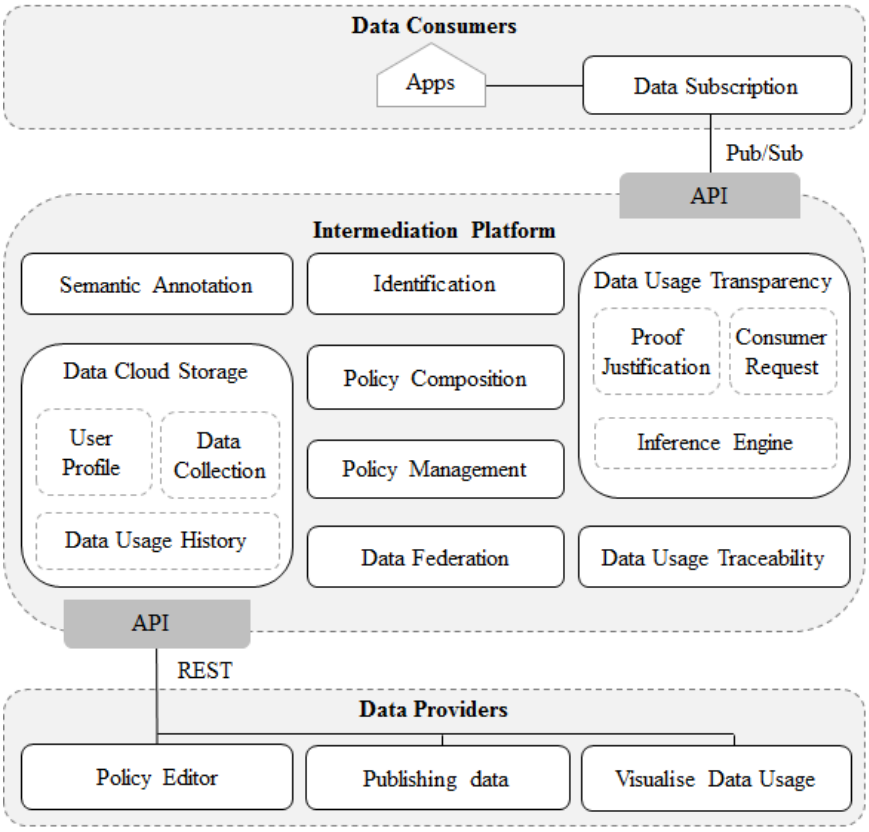

Figure 5.3: Data Usage Control Components

deals with issues of trust and control, and achieves competitive advantages to attract partners sharing their data using the open standard APIs

5.3 Data Usage Control Components

Figure 5.3 introduces data usage control components and relationships between them in three perspectives: data providers, data consumers, and the intermediation platform.

5.3.1 Data Providers

The data providers are able to publish their data to the intermediation platform. They are also provided an editor using which they can define the policy to exercise control on how the data is going to be used.

We develop RESTful web services to cover all needed functionalities for the data consumers. The service APIs are based on a subset of the principles of REpresentational State Transfer (REST) Fielding (2000), and are used by the data providers to manage their data items, policy/rules, and data usage history in the intermediation platform.

5.3.2 Data Consumers

Data consumers are allowed to request the data from the intermediation platform. They can visualize not only the responded data, but also the proof justification for trusting the results. Moreover, the stated obligations imposed by the data providers are reassured.

We develop Publish/Subscribe (Pub/Sub) services to cover all the needed functionalities for the data consumers. Pub/Sub is a highly-decoupled distribution model, where (generally) publishers produce information irrespective of consumers Bacon et al. (2008). In particular, the consumers are provided Pub/Sub APIs for data subscriptions, and proof justification from the intermediation platform.

5.3.3 Intermediation Platform

The platform aims to provide a trustworthy data sharing by enhancement of data usage transparency and traceability. We have several functionalities to ensure this goal as follow: (*i*) Identification of users' profile with reliable authentication, (*ii*) Policy Management for managing the defined policies for data usage, (*iii*) Policy Composition for defining the data usage policies and importing them at the platform level, (*iv*) Transparency for the fair processing of consumers requests, proof justification, and inference engine, (*v*) Traceability for tracing data usage history. It has other components that support (*vi*) Semantic Annotation, (*vii*) Data Cloud Storage for managing user profiles, data collection, and data usage history, and (*viii*) Data Federation for computing consumers' data response.

5.4 Trustworthy Data Sharing Procedures

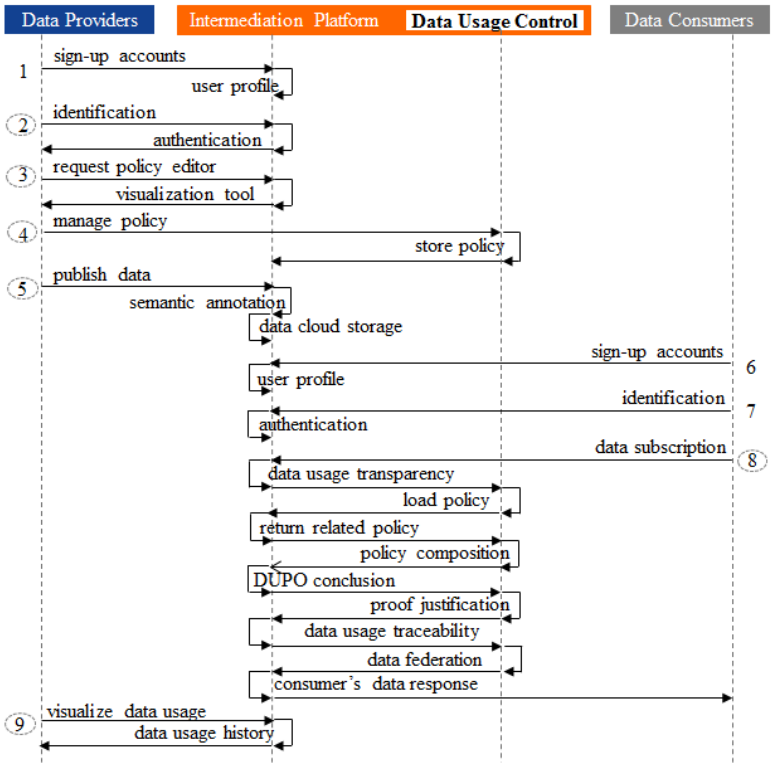

Figure 5.4: Trustworthy Data Sharing Procedures

Figure 5.4 presents the trustworthy data sharing procedures which shows the sequence of steps between data provider and consumer. Next we present the detail of the procedures with an illustrative scenario in the following parts:

5.4.1 Identification

As the first step, granting access to the platform is required. In the steps (1) and (6) of the figure 5.4, the data providers and consumers must create their accounts in the platform. After they are authenticated in steps (2) and (7), they have a secure access to the platform and use the APIs provided. These accounts are stored as user profiles on the Data Cloud Storage.

Mapping to the defined concepts of DUPO, the user profiles are facts about *actors*. In our scenario, we have known facts about commercial service operators (CO), data owners (DO), and municipal authorities (MA), which

are presented as follows:

$$F^{DUPO} = \{CO(X), DO(X), MA(X)\}$$

5.4.2 Policy Management

In order to manage the policy in the platform in step (4), we provide a visualization tool and the authenticated data providers need to request it in step (3).

Mapping to the concepts of DUPO, we illustrate all of the data usage policies in the scenario as are defined in (R^{DUPO}). In particular, the DO has a full access permission to all the details. This policy is represented with the use of defeasible rules, as follows:

$$r_{1,d} : DO(X) \Rightarrow_P TemporalScope(X, any),$$
$$r_{2,d} : DO(X) \Rightarrow_P SpatialScope(X, any),$$
$$r_{3,d} : DO(X) \Rightarrow_P AggregateScope(X, any),$$
$$r_{4,d} : DO(X) \Rightarrow_P PurposeScope(X, any)$$

The MA has permission to access the available average occupancy of parking places ($average$) per street on an hourly basis. This policy is represented with the use of defeasible rules, as follows:

$$r_{1,m} : MA(X) \Rightarrow_P SpatialScope(X, street),$$
$$r_{2,m} : MA(X) \Rightarrow_F \neg SpatialScope(X, street),$$
$$r_{3,m} : MA(X) \Rightarrow_P TemporalScope(X, hourly),$$
$$r_{4,m} : MA(X) \Rightarrow_F \neg TemporalScope(X, hourly),$$
$$r_{5,m} : MA(X) \Rightarrow_P AggregateScope(X, average),$$
$$r_{6,m} : MA(X) \Rightarrow_F \neg AggregateScope(X, average)$$

For CO, the consideration is that only statistical data will be available over a zone and on a weekly basis. This policy is represented with the use

of defeasible rules, as follows:

$$
\begin{aligned}
&r_{1,c} : CO(X) \Rightarrow_P SpatialScope(X, zone),\\
&r_{2,c} : CO(X) \Rightarrow_F \neg SpatialScope(X, zone),\\
&r_{3,c} : CO(X) \Rightarrow_P TemporalScope(X, weekly),\\
&r_{4,c} : CO(X) \Rightarrow_F \neg TemporalScope(X, weekly),\\
&r_{5,c} : CO(X) \Rightarrow_P AggregateScope(X, statistic),\\
&r_{6,c} : CO(X) \Rightarrow_F \neg AggregateScope(X, statistic)
\end{aligned}
$$

5.4.3 Publishing Data

The platform supports collection and securing storage of IoT data. In fact, data providers use REST APIs to publish their data in step (5) and the collected data will be stored in the Data Cloud Storage.

Mapping to concepts of DUPO, we present an example of data item using Context Element XML format in Listing 5.1. This data item contains the current state (*line* 9) of the parking sensor (*line* 3) in location (*line* 14) at timestamp (*line* 21).

```
1  <contextElement>
2    <entityId type="ParkingSensor" >
3      <id>ps1</id>
4    </entityId>
5    <contextAttributeList>
6      <contextAttribute>
7        <name>currentState</name>
8        <type>integer</type>
9        <contextValue>1</contextValue>
10     </contextAttribute>
11     <contextAttribute>
12       <name>location</name>
13       <type>string</type>
14       <contextValue>parkingspace1</contextValue>
15     </contextAttribute>
16   </contextAttributeList>
17   <domainMetadata>
18     <contextMetadata>
19       <name>timestamp</name>
```

```
20        <type>dateTime</type>
21        <value>2016-02-16T15:23:17.234+0200</value>
22      </contextMetadata>
23    </domainMetadata>
24 </contextElement>
```

Listing 5.1: Example of Data Item in XML format.

5.4.4 Data Subscription

For the data consumers, they could subscribe data usage in step (8). We implement the data usage transparency and traceability in the platform based on processing the consumer's request. Toward this end, the Data Usage Transparency component will load the related policies, perform a policy composition, deal with policy conflicts, and do policy enforcement based on defeasible reasoning to obtain the DUPO conclusions. In the case that the conclusion is defeasible provable, the Data Federation component will compute to return related data items. The data are filtered or aggregated following the request conditions and the rules extracted from the policy to return the results to the consumers. Every transaction of data usage will be stored as a new data items and later reported to the data owners. The Data Usage Traceability component ensures the traceability of the data usage. In other case, we provide proof justification to the consumer.

For mapping to the DUPO, we define the consumer's request in our scenario as a defeasible rule:

$$\begin{aligned} r : CO(X), [P]SpatialScope(X, street), \\ [P]TemporalScope(X, hourly), \\ [P]AggregateScope(X, detail) \\ \Rightarrow_O ConsumerRequest(X) \end{aligned}$$

This consumer's request is processed in the DUPO and the conclusions are
$-\Delta[O]ConsumerRequest(X)$, and $-\partial[O]ConsumerRequest$. Which means that $ConsumerRequest$ is defeasible rejected in $DUPO$, so the request is refused. We then apply (Kontopoulos et al. 2011) to provide a proof justification to the consumer.

5.4.5 Visualize Data Usage

The data providers could visualize their data usage in step (9), customize their policies, and explore the consequences of certain changes.

5.5 Summary

In this chapter, we focus on data usage transparency and traceability and propose the trustworthy data sharing platform in smart cities. It includes the system architecture, core components, and mechanism for transparency and traceability of data usage which has provided as a sequence diagram to the smart cities' stakeholders. In fact, we based on the DUPO and semantic technologies to develop this type of architectural support for the trustworthy data sharing platform. Next, we are going to provide a visualization tool to help users to customize their policies in an interactive format that allows them to explore the consequences of certain changes.

"One small candle may light a thousand"

Quote by William Bradford

Chapter 6

Visualization Tool and Evaluation

Contents

We have proposed the trustworthy data sharing platform in the previous chapter which based on the DUPO and semantic technologies. However, a visualization tool based on the platform is needed to help users customize their policies in an interactive format, which allows them to explore and monitor the consequences of certain changes to how their data is allowed to be used. This chapter thus focuses more on the research question: *How do data owners exercise some control the usage of their data?*. Moreover, we

provide an evaluation of the solution for the issue of trust and control in context of a shared platform in smart cities.

In section 6.1, we begin by prototype implementation with its overall proof-of-concept based on the platform in section 6.1.1, implementation choices in section 6.1.2, and prototype of visualization tool in section 6.1.3. We present next early performance analysis for the proposed solution for trust and control in section 6.2. Finally, we summary this chapter in section 6.3.

6.1 Prototype Implementation

6.1.1 Overall Proof-of-Concept

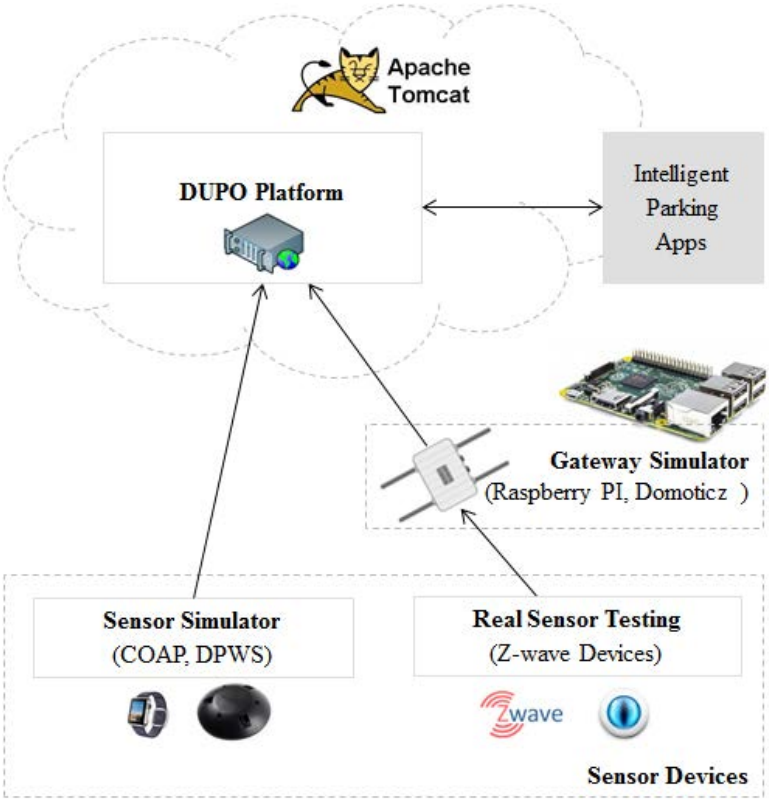

Figure 6.1: Overview of Proof-of-Concept

We define an overall implemented system for the proof-of-concept in Figure 6.1. The DUPO platform are developed to receive data from the sensors and process data subscription from the intelligent parking application (IPA). We used Apache Tomcat[1] as a web applications server to deploy our DUPO platform. The IPA is a RESTful service developed using Restlet[2], a framework for developing REST web services. The service requests the relevant data from the DUPO platform using the Pub/Sub APIs provided.

Sensor devices are simulated by using DPWS Simulator[3], and CoAP Simulator[4]. We also use the raspberry PI[5] to run the z-ware/ethernet gateway. All real Z-wave sensor devices emit z-wave messages that are caught by the gateway. This data can be processed locally by the raspberry. The simulated sensors and the gateway use the REST APIs provided to forward the data to our platform.

6.1.2 Implementation Choices

Figure 6.2 explains more about implementation choices for the proof-of-concept. We proposed essential technologies that are used to develop prototypes for the platform APIs.

We used Apache Jena Framework[6], an open source Java Framework for developing the functionalities of Data Annotation. In fact, the platform received the raw data from the sensors or the gateway, we aim to convert it to linked data(Berners-Lee 2006). A specific syntax called JSON-LD[7] is used to serialize Linked Data with the motivation to reduce the size of RDF documents compared to the size yielded by XML serialization. The linked data are stored in the Data Cloud Storage which use Virtuoso[8]. We also processed SPARQL query to implement the component of Data Federation.

We built on SPINdle(Lam & Governatori 2009) for functionalities of Data Usage Transparency, Traceability, Policy Composition, and Policy Manage-

[1] http://tomcat.apache.org/
[2] http://restlet.com/
[3] https://github.com/sonhan/dpwsim
[4] https://github.com/caohuuquyet/jhess/tree/master/jUCP
[5] http://www.materiel.net/barebone/raspberry-pi-type-b-106574.html
[6] https://jena.apache.org/
[7] https://www.w3.org/TR/json-ld/
[8] http://virtuoso.openlinksw.com/

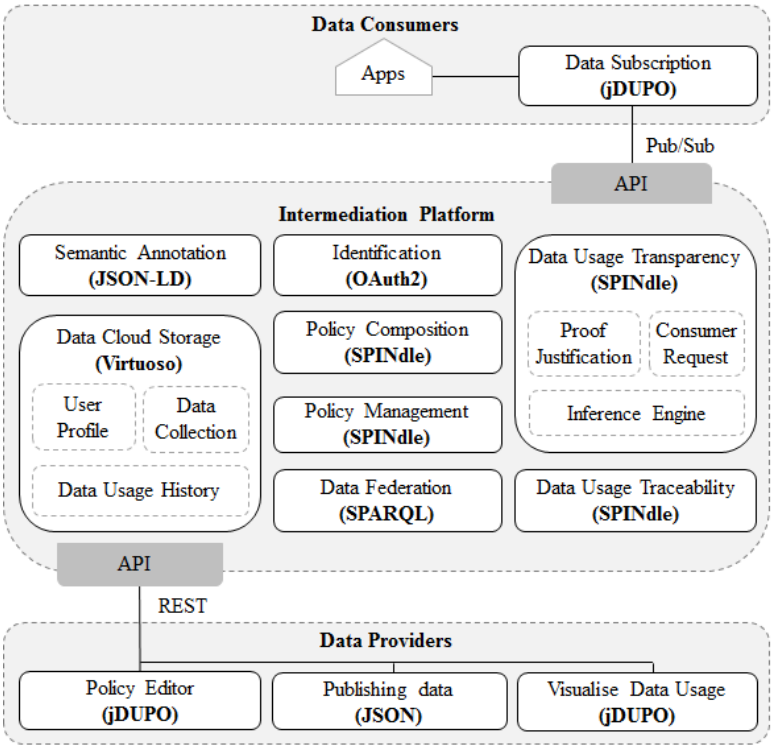

Figure 6.2: Implementation Choices of the Proof-of-concept

ment. It is a logic reasoner that can compute efficiently the consequences of DUPO theories (Lam & Governatori 2009).

The Identification component is used to granting access to the platform. In the prototype, we proposed to using OAuth[9] for this purpose.

In the next, we discuss about the jDUPO prototype that is used to edit policy, create consumer' request, and visualize data usage.

6.1.3 Visualization Tool Prototype

We developed a prototype version of our visualization tool namely *jDUPO* which aims to help users and data owners to customize their policies in a way that allows them to explore the consequences of each change and monitor how the data is going to be used after sharing it. We implemented an initial policy editor, including its functionalities for data usage control, and a short

[9]Auth: http://oauth.net/2/

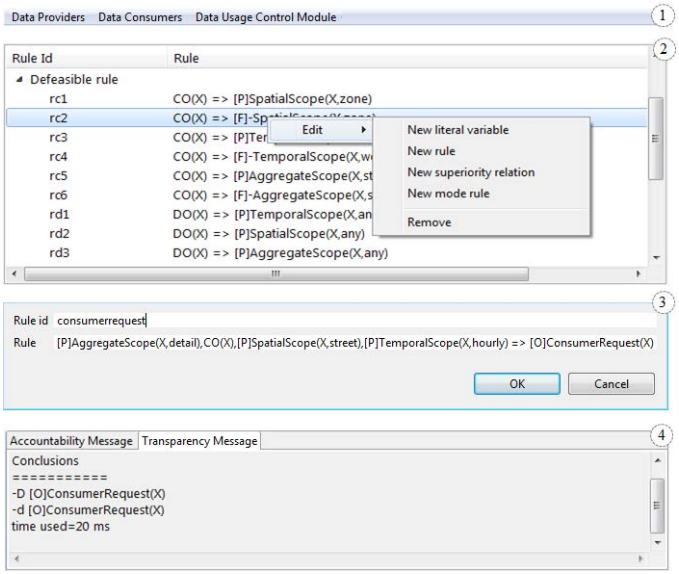

Figure 6.3: The main interface of the implemented Visualization tool Prototype (namely *jDUPO*)

demo illustrating the use case scenarios. Figure 6.3 shows a snapshot of a *jDUPO* interface which shows some of the prototype functionalities including (1) Main menu, (2) Policy editor, (3) Consumer's request, (4) Transparency and Traceability.

6.1.3.1 Policy Editor

In this prototype, we use SPINdle syntax to define facts, rules, and rule priorities for the DUPO. For example, Listing 6.1 shows the data usage policies for the *CO* in SPINdle syntax. End users, however, could also use *jDUPO* to edit their policies.

```
1 # Facts
2 >> CO(X)
3 # Defeasible rules
4 r1c: CO(X) =>[P] SpatialScope(X,zone)
5 r2c: CO(X) =>[F] -SpatialScope(X,zone)
6 r3c: CO(X) =>[P] TemporalScope(X,weekly)
7 r4c: CO(X) =>[F] -TemporalScope(X,weekly)
8 r5c: CO(X) =>[P] AggregateScope(X,statistic)
9 r6c: CO(X) =>[F] -AggregateScope(X,statistic)
```

```
10 ...
```

Listing 6.1: Data Usage Policies in SPINdle syntax.

6.1.3.2 Consumer's Request

We are able to use *jDUPO* to create a consumer's request as well. Listing 6.2 shows an example of a consumer's request in the SPINdle syntax.

```
1 # Consumer`s request
2 r: CO(X),[P]SpatialScope(X,street),[P]TemporalScope(X,hourly),[P]
      AggregateScope(X,detail) =>[O] ConsumerRequest(X)
```

Listing 6.2: Consumers' Request in SPINdle syntax

6.1.3.3 Transparency and Traceability

By using *jDUPO*, we are able to process the transparency and traceability of data usage. Listing 6.3 shows the conclusions of the consumer's request with an inference logger built on top of the SPINdle Reasoner.

```
1  # Conclusions
2  ==================
3  -D [O]ConsumerRequest(X)
4  -d [O]ConsumerRequest(X)
5  ...
6
7  === Inference Logger ===
8  Rule_00000
9  +-- [DEFEASIBLE] Discarded :- [-d [O]ConsumerRequest(X)]
10 ...
```

Listing 6.3: SPINdle-based Conclusions and Inference Logger.

6.1.3.4 Data Federation

For the prototype, we use SPARQL to query data. Listing 6.4 show an example to query the available parking data in the average occupancy of parking places per street on an hourly basis.

```
1 prefix tl: <http://purl.org/NET/c4dm/timeline.owl#>
2 prefix sao: <http://iot.ee.surrey.ac.uk/citypulse/resources/ontologies/
      sao.ttl#>
```

```
3  prefix ct: <http://www.insight-centre.org/citytraffic#>
4  prefix ns1: <http://purl.oclc.org/NET/ssnx/ssn#>
5  prefix xsd: <http://www.w3.org/2001/XMLSchema#>
6
7  SELECT ?latitude ?longitude ?time (AVG(xsd:integer(?m)) as ?sum)
8  WHERE {
9  ?observation sao:value ?m .
10 ?observation sao:time ?t .
11 ?t tl:at ?time1.
12 bind(CONCAT(STR(year(?time1)),
13 '-', STR(month(?time1)),  '-', STR(day(?time1)),
14 ':',  STR(hours(?time1))) as ?time).
15 ?observation ns1:featureOfInterest ?fi .
16 ?fi a sao:FeatureOfInterest .
17 ?fi ct:hasFirstNode ?v .
18 ?v ct:hasLatitude ?latitude .
19 ?v ct:hasLongitude ?longitude .
20 }
21 GROUP BY ?latitude ?longitude ?time
22 ORDER BY ?latitude ?time
```

Listing 6.4: SPARQL to query data.

6.2 Evaluation

6.2.1 Performance Analysis

In order to measure the performance of our solution, we conduct some experiments by using jDUPO and considering the intelligent parking use case. We run the prototype on a HP Elite Book 850 G3 computer with an Intel Core-i5-6300 2.4 GHz processor, 8 GB of RAM, and a 64-bit Windows 7 Enterprise operating system.

We also use the parking dataset in the European Project CityPulse CityPulse (2014). It includes a total of 8 parking lots providing information over a period of 6 months (55.264 data points in total) in the city of Aarhus.

In the following experiments, the query was that of a municipal authority asking for the average occupancy of parking places per street on an hourly basis.

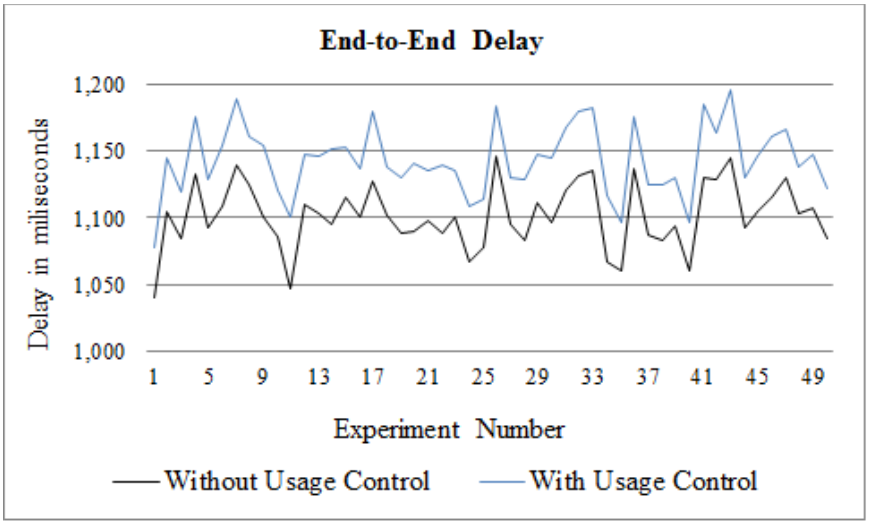

Figure 6.4: End-to-End Delay (E2ED).

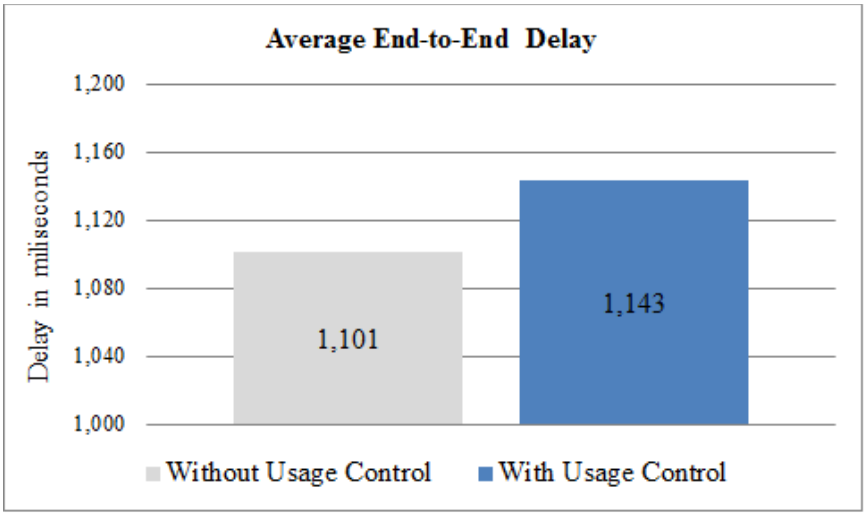

Figure 6.5: Average End-to-End Delay.

The performance was assessed in terms of the following metrics: End-to-End Delay (E2ED), Trust Computation Time (TCT), Impact on the Computational Time (ICT), and Memory Usage (IMU). E2ED is the time delay which takes to process the consumer request and get the data response. TCT is time used only for processing usage control. By increasing the number of rules, ICT and IMU were studied in terms of impact on computational time and memory usage.

The first experiment aims to compare performance result of the E2ED with and without usage control. The request is processed and repeated 50 times and the result of this experiment is shown in Figure 6.4. The highest value of E2ED with usage control is 1195ms, while the lowest one is 1078ms. In the other case, without usage control, the highest value of E2ED is 1146ms, and the lowest one is 1041ms. Figure 6.5 shows their average E2ED between two assumptions after 50 repetitions. As it can be seen, on the average the overhead of usage control in the first experiment is about 3.8%.

The second experiment aims to evaluate actual value of TCT with and without new instance cases. In the first case, we restarted jDUPO to create new instance for each request processing. In the second case, we used the same instance for subsequent consumer request. Based on that, we compared the TCT in delay milliseconds after 50 repetitions. Figure 6.6 shows the results of the experiment (with and without new instance respectively). The highest value of TCT with new instance is 56ms, while the lowest one is 35ms. The highest value of TCT without new instance is 37ms, and the lowest is around 4ms. Figure 6.7 also shows their average TCT between two cases. The overhead of trust computation without new instance is only 6ms and with new instance is 42ms.

The third experiment aims to evaluate the impact on the computation time (ICT) and the impact on the memory usage (IMU), we compare the time and memory usage which is needed for trust computation as number of rules increases from 1000 to 10000. Toward this end, 25 cases consisting of 10 runs of each were performed. The result of ICT consumed is shown in Figure 6.8. It shows that the computational time taken increased linearly ($y = 0.08x$, $R^2 = 0.99$) with increasing number of rules. In the case of IMU, Figure 6.9 shows the impact result on the memory usage. It demonstrate

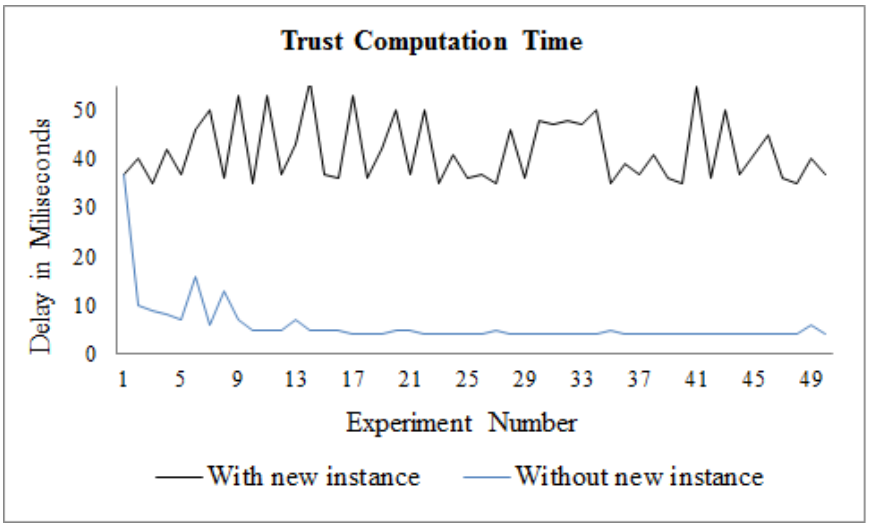

Figure 6.6: Trust Computation Time (TCT).

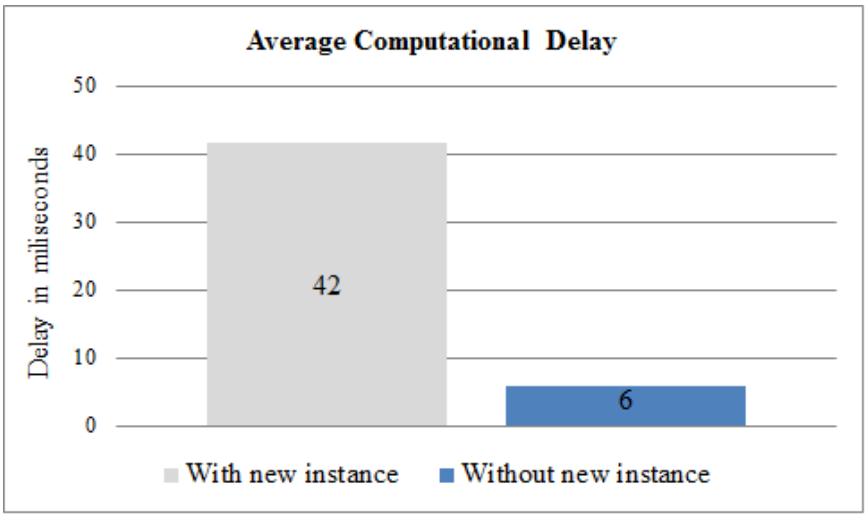

Figure 6.7: Average Computational Delay.

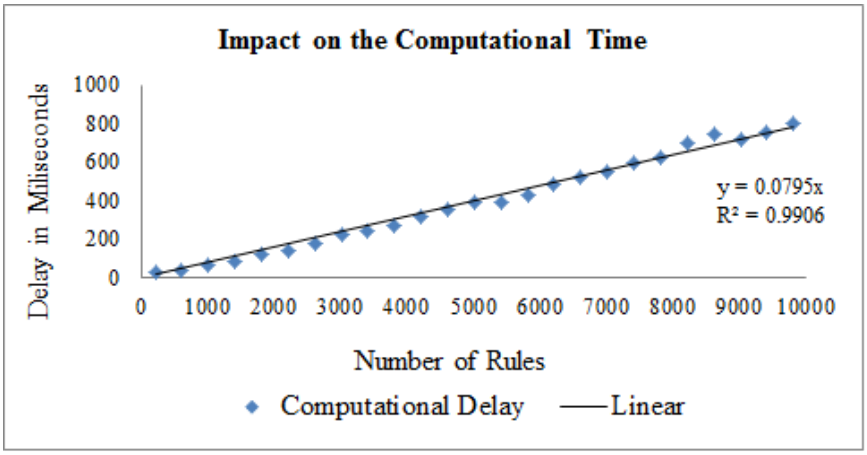

Figure 6.8: Impact on the Computational Time (ICT).

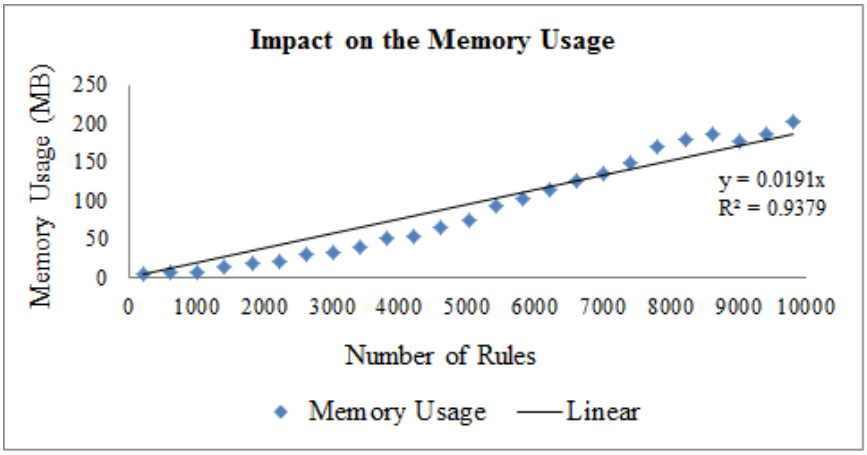

Figure 6.9: Impact on the Memory Usage (IMU).

that the memory also increased linearly with the increase of number of rules ($y = 0.02x$, $R^2 = 0.94$).

In conclusion the performance evaluation shows that the overhead of usage control stays reasonably in the range of 3.8% with new instance creation and about 0.5% without new instance. Also the growth in overhead of usage control stays linear even in very complex cases with thousands of rules.

6.2.2 Comparison with Related works

In this part, we aim to compare the general characteristic of the proposed framework in this study with other similar approaches. As it is mentioned earlier, we aim to tackle the issues of trust and control in the context of IoT smart cities use cases. In particular, we use the concept of usage control by (Pretschner & Walter 2008) as a starting point to develop the data usage control model that enables the expression and definition of obligations on data usage. It should be noted that usage control policies apply to an entire group of devices - for instance a particular class of sensors in a given geographical area and deployed by a specific actor. In particular spatial and temporal constraints that a data provider imposes on the usage of the data apply to the data generated by this group. We believe that is a novelty of our framework which has not been addressed by prior works. In addition the model not only decides whether to provide access to the data, but also provides an explanation for the decision.

To understand better the position of our framework in compare to other solutions, a comparative analysis of our proposed model *DUPO* with respect to others is provided in Table 6.1. In a relevant study, Speiser *et al.* Speiser et al. (2011) specified the conceptual policy model to deal with this issue of abstraction of information, but this model does not respond to the obligations defined by the actors for their data. In the context of a social network, Pato *et al.* Pato et al. (2011) proposed the solution which encourage responsible use of information by combining clearly expressed usage policies with systems for detecting misuse. However it does not address the issues in an IoT smart city use cases. In another study, Governatori *et al.* Governatori et al. (2013) focus on the data licensing using the composite license from the single licenses. Our trust model is policy-based usage control approach. We

Table 6.1: Comparative study previous approaches to our proposal based on different features.

	Speiser, et al. Speiser et al. (2011)	**Pato, et al. Pato et al. (2011)**	**Governatori, et al. Governatori et al. (2013)**	**DUPO**
Domain	Smart Grid	Web, Social Network	Web of Data	Smart Cities
Use cases Scenario	Energy Consumption	Health Insurance	Composite License	Intelligent Parking
Requirement	Usage Perspectives	Usage Restrictions	Set of Licenses	Data Usage Obligations
Policy Model	Yes	Yes	No	Yes
Policy Representation	RDF/N3 Syntax	AIR Language	Deontic Logic Semantics	Defeasible Rules
Deal with Rule conflict	No	No	Yes	Yes
Policy Composition	No	No	Yes	Yes
Trust Model	Abstraction Information	Information Accountability	Data Licensing	Policy-based Usage Control
Proposed Platform	No	Yes	No	DUPO Platform as a Service
Visualization Tool	No	Yes	No	SPINdle-based jDUPO
Evaluation	Policy Matching	No	No	With and without Usage Control

develop the formal theory and its proof based on DL, the data usage policies and each consumer requests are expressed as in regular DL rules. We also apply semantic technologies to IoT Data aggregation and interpretation. Lastly it worth to mention again that our contribution applies to a group of devices and in particular the constraints and obligation used in the policies apply to an aggregation of devices in spatial and temporal domains which is an novel part in *DUPO*. For IoT domain, we believe that this dimension is needed as millions of devices are involved and the appropriate level for usage control policies needs to be provide for higher level abstractions and not be restricted to individual device level. Considering all said so far, to the best of our knowledge, the ideas presented in this study are novel and different

from earlier efforts in the IoT domain

6.3 Summary

In summary, the proof-of-concept is developed based on the trustworthy data sharing platform and the visualization tool prototype is provided to help users easily control and monitor how their data are shared. We also are presented all experiments along with the results. Importantly, the evaluation results show that the performance of the added trust and control does not impact negatively on the system. We also do comparative study previous approaches to our proposal based on different features. It is confirmed that the ideas presented in this thesis are novel and provide new insight in the IoT domain for dealing with issues of trust and control.

"What is not started will never get finished"
Quote by Johann Wolfgang von Goethe

Chapter 7

Conclusion

Sharing data across multiple entities can be highly rewarding in terms of insights and usability but trust is the key point when stakeholders share data. One important aspect of building trust is for the data owner to be able to exercise control over the usage of the data by other actors. In this thesis, we concentrate on this issue namely Usage Control which have not been adequately addressed in the context of an intermediation platform for smart cities. In particular, a novel trust model is still needed to deal with this issue of trust and control in the IoT.

We do not mainly focus on security aspects such as privacy, data licensing, or access control. In fact, what does go along with data usage control is the notion of the levels of abstraction. It means that which level of information should be shared. These abstractions could be studied to provide mechanisms that can be used by a privacy module Speiser et al. (2011) or a data licensing model (Governatori et al. 2013). Also, Pato et al. (2011) introduce an information accountability model which is complementary to the security aspects. However, there is still no specific data usage control model to express the constraints and obligations on the use of IoT data among participants. In particular, this model have to response to the obligations defined by the actors for their data such as (*i*) Abstraction of certain information, (*ii*) Spatial and temporal granularity, (*iii*) Classification of actors and purposes, and (*iv*) Monetization of data. It is still lack of mechanisms to allow for automated data usage control and traceability of data usage in

the context of smart cities.

The main technical challenges are to express the obligations and conditions in usage control policies and to ensure the transparency and traceability of the policy enforcement rules. Actors also need to have an easily interpretable tool to demonstrate in a clear fashion the reasoning behind the rules. Such a visualization tool helps the actor understand the implications of the different choices that are made while defining the policies as well as understanding at runtime how conflicts have been resolved. The latter aspect is critical for accountability.

To deal with these issues, we start from the concept of usage control which is about how data is used after access to it has been granted (Pretschner & Walter 2008). We then focus on mechanisms for trustworthy data sharing in an IoT intermediation platform. The detail of thesis contributions is presented in section 7.1 along with some work items for the future work in section 7.2 that will aid in extending the work done in this thesis.

7.1 Summary of Contributions

Although data is a key part in smart cities, traditionally there has been no systematic effort to enable the sharing of data in a trustworthy manner among applications or services. In order to promote sharing of data, mechanisms need to be put into place to provide the different actors - data producers, data consumers, etc. means to control and visualize how their data or requests are being processed and used. In this thesis we deal with the key issue involved in trust which is usage control, i.e., how data is used once access to it has been granted. In total three contributions were made in this thesis.

Firstly, we proposed a model for policy-based data usage control (namely DUPO) with its conceptual model, formal theory, and illustrative scenario. This model responded to the diversity of obligations or data usage requirements that data owners impose on the use of their data. It also focused on the non-monotonic formalism which aims to handle the normative conflicts between rules, rules with deontic consequents, and exceptions, illustrated the logical reasoning applied when the policies are enforced in a computationally

tractable way. The illustrative made use of a smart city scenario aims to explain the model concepts.

Secondly, a trustworthy data sharing platform as service is then defined. It allowed transparency and traceability of data usage with the core components based on the DUPO and Semantic technologies. We also presented in detail the main procedures for the trustworthy data sharing in aspects of data owners, consumers, and an intermediation platform.

Thirdly, a proof-of-concept is developed and a visualization tool is provided to help users easily control and monitor how their data is shared. Finally, we investigated the performance of the system with the initial assumption about trust and control to compare the performance results with and without those assumptions. All experiments are presented along with the results and more importantly it showed that the performance of the added trust does not impact negatively on the system.

7.2 Research Directions

However, several other aspects that are not covered in this thesis, can be considered as future work.

The trust computing framework is firstly needed to enhance the performance on real-time responses in production systems. In particular, we will work more on employment of a specific trust computation model which is built on trust metrics, and attributes. We additionally aim to provide efficient query answering which could lead to the improvement of the reasoning mechanism for more complex use cases and for supporting real-time processing and scalability.

We also need to work more with the open standard APIs which attract partners to share data on the platform. For example, there are APIs which deliver the right data to partners, handle semantics variability, manage metadata along their usage and their value. Moreover, privacy is a major issue when it comes to data sharing. The main focus on this thesis is on issues of usage control. What does go along with usage control is the notion of the levels of abstraction that the producer wishes to provide. Thus, these abstractions could be extended to provide mechanisms that can be used by

a privacy module.

Another important evolution is to involve end-users in the evaluation of the proposed visualization tools in order to ensure their usability.

Bibliography

Antoniou, G., Billington, D., Governatori, G. & Maher, M. J. (2001), 'Representation Results for Defeasible Logic', *ACM Trans. Comput. Logic* **2**(2), 255–287.

Antoniou, G., Dimaresis, N. & Governatori, G. (2009), 'A modal and deontic defeasible reasoning system for modelling policies and multi-agent systems', *Expert Systems with Applications* **36**(2), 4125–4134.

Atzori, L., Iera, A. & Morabito, G. (2010), 'The Internet of Things: A Survey', *Computer Networks* **54**(15), 2787–2805.

Bacon, J., Eyers, D. M., Singh, J. & Pietzuch, P. R. (2008), Access control in publish/subscribe systems, *in* 'Proceedings of the second international conference on Distributed event-based systems', ACM, pp. 23–34.

Balakrishna, C. (2012), Enabling technologies for smart city services and applications, *in* '2012 Sixth International Conference on Next Generation Mobile Applications, Services and Technologies', IEEE, pp. 223–227.

Barki, A., Bouabdallah, A., Gharout, S. & Traoré, J. (2015), 'M2m security: Challenges and solutions', *IEEE Communications Surveys & Tutorials* **18**(2), 1241–1254.

Benevolo, C., Dameri, R. P. & D'Auria, B. (2016), Smart mobility in smart city, *in* 'Empowering Organizations', Springer, pp. 13–28.

Berners-Lee, T. (2006), Linked Data, *in* 'International Journal on Semantic Web and Information Systems', Vol. 4, W3C.

Bizer, C., T., H. & Berners-Lee, T. (2009), Linked Data - The Story So Far, *in* 'International Journal on Semantic Web and Information Systems', Vol. 5.

Buhalis, D. & Amaranggana, A. (2013), Smart tourism destinations, *in* 'Information and communication technologies in tourism 2014', Springer, pp. 553–564.

Cabrio, E., Aprosio, A. P. & Villata, S. (2014), These are your rights, *in* 'European Semantic Web Conference', Springer, pp. 255–269.

Caragliu, A., Del Bo, C. & Nijkamp, P. (2011), 'Smart cities in europe', *Journal of urban technology* **18**(2), 65–82.

Christin, D. (2016), 'Privacy in mobile participatory sensing: current trends and future challenges', *Journal of Systems and Software* **116**, 57–68.

CityPulse (2014), The parking dataset in the European Project CityPulse, Technical report, http://iot.ee.surrey.ac.uk:8080/datasets.html#parking.

Compton, M., Barnaghi, P., Bermudez, L., GarcíA-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A. et al. (2012), 'The SSN ontology of the W3C semantic sensor network incubator group', *Web Semantics: Science, Services and Agents on the World Wide Web* **17**, 25–32.

Covington, M. J., Long, W., Srinivasan, S., Dev, A. K., Ahamad, M. & Abowd, G. D. (2001), Securing context-aware applications using environment roles, *in* 'Proceedings of the sixth ACM symposium on Access control models and technologies', ACM, pp. 10–20.

Cuppens, F. & Cuppens-Boulahia, N. (2008), 'Modeling contextual security policies', *International Journal of Information Security* **7**(4), 285–305.

ETSI (2011), Machine-to-Machine communications (M2M); Functional architecture , Technical report, http://goo.gl/mv6qFZ.

Ferraiolo, D. F. & Kuhn, D. R. (2009), 'Role-based access controls', *arXiv preprint arXiv:0903.2171* .

Fielding, R. T. (2000), Architectural styles and the design of network-based software architectures, PhD thesis, University of California, Irvine.

FIWARE (2016), *FI-WARE Platform*, http://forge.fiware.eu/plugins/mediawiki/wiki/fiware/index.php.

Fung, B., Wang, K., Chen, R. & Yu, P. S. (2010), 'Privacy-preserving data publishing: A survey of recent developments', *ACM Computing Surveys (CSUR)* **42**(4), 14.

Gibson, D. V., Kozmetsky, G. & Smilor, R. W. (1992), *The technopolis phenomenon: Smart cities, fast systems, global networks*, Rowman & Littlefield.

Gordon, T. F. (2011), Analyzing open source license compatibility issues with carneades, *in* 'Proceedings of the 13th International Conference on Artificial Intelligence and Law', ACM, pp. 51–55.

Governatori, G. & Rotolo, A. (2008), 'BIO logical agents: Norms, beliefs, intentions in defeasible logic', *Autonomous Agents and Multi-Agent Systems* **17**(1), 36–69.

Governatori, G., Rotolo, A., Villata, S. & Gandon, F. (2013), One License to Compose Them All, *in* 'The Semantic Web–ISWC 2013', Springer, pp. 151–166.

Gubbi, J., Buyya, R., Marusic, S. & Palaniswami, M. (2013), 'Internet of things (iot): A vision, architectural elements, and future directions', *Future Generation Computer Systems* **29**(7), 1645–1660.

Gürses, S., Troncoso, C. & Diaz, C. (2011), 'Engineering privacy by design', *Computers, Privacy & Data Protection* **14**(3).

Hanson, C., Kagal, L., Berners-Lee, T., Sussman, G. J. & Weitzner, D. (2007), Data-purpose algebra: Modeling data usage policies, *in* 'Policies for Distributed Systems and Networks, 2007. POLICY'07. Eighth IEEE International Workshop on', IEEE, pp. 173–177.

Harmon, R. R., Castro-Leon, E. G. & Bhide, S. (2015), Smart cities and the internet of things, *in* '2015 Portland International Conference on Management of Engineering and Technology (PICMET)', pp. 485–494.

Hartig, O., Bizer, C. & Freytag, J. C. (2009), Executing SPARQL queries over the web of linked data, *in* 'The Semantic Web-ISWC', Springer Berlin Heidelberg, W3C Working Group, pp. 293–309.

Heath, T. & Bizer, C. (2011), 'Linked data: Evolving the web into a global data space', *Synthesis lectures on the semantic web: theory and technology* **1**(1), 1–136.

Hilty, M., Pretschner, A., Basin, D., Schaefer, C. & Walter, T. (2007), A policy language for distributed usage control, *in* 'European Symposium on Research in Computer Security', Springer, pp. 531–546.

Hoepman, J.-H. (2014), Privacy design strategies, *in* 'IFIP International Information Security Conference', Springer, pp. 446–459.

Hollenbach, J., Presbrey, J. & Berners-Lee, T. (2009), Using rdf metadata to enable access control on the social semantic web, *in* 'Proceedings of the Workshop on Collaborative Construction, Management and Linking of Structured Knowledge (CK2009)', Vol. 514.

Khan, I., Belqasmi, F., Glitho, R., Crespi, N., Morrow, M. & Polakos, P. (2015), 'Wireless Sensor Network virtualization: Early architecture and research perspectives', *IEEE Network* **29**(3), 104–112.

Khan, I., Jafrin, R., Errounda, F. Z., Glitho, R., Crespi, N., Morrow, M. & Polakos, P. (2015), A data annotation architecture for semantic applications in virtualized wireless sensor networks, *in* '2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)', IEEE, pp. 27–35.

Khatoun, R. & Zeadally, S. (2016), 'Smart Cities: Concepts, Architectures, Research Opportunities', *Commun. ACM* **59**(8), 46–57.

Kontopoulos, E., Bassiliades, N. & Antoniou, G. (2008), 'Deploying defeasible logic rule bases for the semantic web', *Data & Knowledge Engineering* **66**(1), 116 – 146.

Kontopoulos, E., Bassiliades, N. & Antoniou, G. (2011), 'Visualizing Semantic Web proofs of defeasible logic in the DR-DEVICE system', *Knowledge-Based Systems* **24**(3), 406–419.

Krötzsch, M. & Speiser, S. (2011), Sharealike your data: Self-referential usage policies for the semantic web, *in* 'International Semantic Web Conference', Springer, pp. 354–369.

Kung, A. (2014), Pears: privacy enhancing architectures, *in* 'Annual Privacy Forum', Springer, pp. 18–29.

Kung, A., Freytag, J.-C. & Kargl, F. (2011), Privacy-by-design in its applications, *in* 'World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a', IEEE, pp. 1–6.

Lam, H.-P. & Governatori, G. (2009), The making of SPINdle, *in* 'Rule Interchange and Applications', Springer, pp. 315–322.

Lampson, B. W. (1974), 'Protection', *ACM SIGOPS Operating Systems Review* **8**(1), 18–24.

Lazaroiu, G. C. & Roscia, M. (2012), 'Definition methodology for the smart cities model', *Energy* **47**(1), 326–332.

Lazouski, A., Martinelli, F. & Mori, P. (2010), 'Usage control in computer security: A survey', *Computer Science Review* **4**(2), 81–99.

Le Métayer, D. (2016), Whom to trust? using technology to enforce privacy, *in* 'Enforcing Privacy', Springer, pp. 395–437.

Letaifa, S. B. (2015), 'How to strategize smart cities: Revealing the smart model', *Journal of Business Research* **68**(7), 1414–1419.

Lu, G. (2011), Neural Trust Model for Multi-agent Systems, PhD thesis, University of Huddersfield.

Manyika, J., Chui, M., Bisson, P., Woetzel, J., Dobbs, R., Bughin, J. & Aharon, D. (2015), 'Unlocking the potential of the internet of things', *McKinsey Global Institute http://goo. gl/qzq5mV* .

Miorandi, D., Sicari, S., De Pellegrini, F. & Chlamtac, I. (2012), 'Internet of things: Vision, applications and research challenges', *Ad Hoc Networks* **10**(7), 1497–1516.

Nadah, N., De Rosnay, M. D. & Bachimont, B. (2007), Licensing digital content with a generic ontology: escaping from the jungle of rights expression languages, *in* 'Proceedings of the 11th international conference on Artificial intelligence and law', ACM, pp. 65–69.

Nam, T. & Pardo, T. A. (2011), Conceptualizing smart city with dimensions of technology, people, and institutions, *in* 'Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times', ACM, pp. 282–291.

Nute, D. (1994), Handbook of Logic in Artificial Intelligence and Logic Programming (Vol. 3), Oxford University Press, Inc., New York, NY, USA, chapter Defeasible Logic, pp. 353–395.

OMA (2010), NGSI Context Management, Technical report, http://goo.gl/mv6qFZ.

OneM2M (2015), Study of Abstraction and Semantics Enablements, Technical report.
URL: *http://goo.gl/2w98Y6*

Park, J. & Sandhu, R. (2002), Towards usage control models: beyond traditional access control, *in* 'Proceedings of the seventh ACM symposium on Access control models and technologies', ACM, pp. 57–64.

Park, J. & Sandhu, R. (2004), 'The UCON ABC usage control model', *ACM Transactions on Information and System Security (TISSEC)* **7**(1), 128–174.

Paskaleva, K. A. (2009), 'Enabling the smart city: The progress of city e-governance in europe', *International Journal of Innovation and Regional Development* **1**(4), 405–422.

Pato, J., Paradesi, S., Jacobi, I., Shih, F. & Wang, S. (2011), Aintno: Demonstration of Information Accountability on the Web, *in* 'Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on', IEEE, pp. 1072–1080.

Pretschner, A., Hilty, M., Basin, D., Schaefer, C. & Walter, T. (2008), Mechanisms for usage control, *in* 'Proceedings of the 2008 ACM symposium on Information, computer and communications security', ACM, pp. 240–244.

Pretschner, A., Hilty, M., Florian, S., Schaefer, C. & Walter, T. (2008), 'Usage Control Enforcement: Present and Future', *IEEE Security Privacy* **6**(4), 44–53.

Pretschner, A. & Walter, T. (2008), Negotiation of usage control policies-simply the best?, *in* 'Availability, Reliability and Security, 2008. ARES 08. Third International Conference on', IEEE, pp. 1135–1136.

Pucella, R. & Weissman, V. (2002), A logic for reasoning about digital rights, *in* 'Computer Security Foundations Workshop, 2002. Proceedings. 15th IEEE', IEEE, pp. 282–294.

Roberti, M. (2010), 'The Internet of Things Revisited', *RFID Journal* .

Rotolo, A., Villata, S. & Gandon, F. (2013), A deontic logic semantics for licenses composition in the web of data, *in* 'Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law', ACM, pp. 111–120.

Sanchez, L., Muñoz, L., Galache, J. A., Sotres, P., Santana, J. R., Gutierrez, V., Ramdhany, R., Gluhak, A., Krco, S., Theodoridis, E. et al. (2014), 'Smartsantander: Iot experimentation over a smart city testbed', *Computer Networks* **61**, 217–238.

Sandhu, R. S., Coynek, E. J., Feinsteink, H. L. & Youmank, C. E. (1996), 'Role-based access control models yz', *IEEE computer* **29**(2), 38–47.

Schaffers, H., Komninos, N., Pallot, M., Trousse, B., Nilsson, M. & Oliveira, A. (2011), Smart cities and the future internet Towards cooperation frameworks for open innovation, *in* 'The Future Internet Assembly', Springer, pp. 431–446.

Schiele, G., Soldatos, J. & Mitton, N. (2014), 'Moving towards interoperable internet-of-things deployments in smart cities', *Smart Cities* p. 16.

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

Soldatos, J., Kefalakis, N., Hauswirth, M., Serrano, M., Calbimonte, J.-P., Riahi, M., Aberer, K., Jayaraman, P. P., Zaslavsky, A., Žarko, I. P. et al. (2015), Openiot: Open source internet-of-things in the cloud, *in* 'Interoperability and Open-Source Solutions for the Internet of Things', Springer, pp. 13–25.

Speiser, S. & Harth, A. (2012), Data-centric privacy policies for smart grids, *in* '2012 AAAI Workshop— Semantic Cities. The AAAI Press, Palo Alto, California', pp. 31–36.

Speiser, S., Wagner, A., Raabe, O. & Harth, A. (2011), Web technologies and privacy policies for the smart grid, *in* 'Policies for Distributed Systems and Networks (POLICY), 2011 IEEE International Symposium on', IEEE, pp. 121–124.

Toninelli, A., Montanari, R., Kagal, L. & Lassila, O. (2006), A semantic context-aware access control framework for secure collaborations in pervasive computing environments, *in* 'International semantic web conference', Springer, pp. 473–486.

Townsend, A. M. (2013), *Smart cities: Big data, civic hackers, and the quest for a new utopia*, WW Norton & Company.

Truong, N. B., Um, T.-W. & Lee, G. M. (2016), 'A reputation and knowledge based trust service platform for trustworthy social internet of things', *Innovations in Clouds, Internet and Networks (ICIN), Paris, France* .

Um, T.-W., Lee, G. M. & Choi, J. K. (2016), 'Strengthening trust in the future social-cyber-physical infrastructure: an itu-t perspective', *IEEE Communications Magazine* **54**(9), 36–42.

Wu, J., Dong, M., Ota, K., Tariq, M. & Guo, L. (2015), 'Cross-domain fine-grained data usage control service for industrial wireless sensor networks', *IEEE Access* **3**, 2939–2949.

Yan, Z., Zhang, P. & Vasilakos, A. V. (2014), 'A Survey on Trust Management for Internet of Things', *Journal of Network and Computer Applications* **42**, 120 – 134.

Zanella, A., Bui, N., Castellani, A., Vangelista, L. & Zorzi, M. (2014), 'Internet of things for smart cities', *Internet of Things Journal, IEEE* **1**(1), 22–32.