PARCo: a Knowledge-Based Agent for Context-Sensitive Reasoning and Decision-Making Regarding Privacy

Author: Andrei Popescu (6246834)

Main Supervisor: Prof. dr. Pinar Yolum Birbil
Second Examiner: Prof. dr. mr. Henry Prakken

In fulfillment of the Master of Science -
Artificial Intelligence
Department of Information and Computing Sciences
Utrecht University
30 June 2019

Acknowledgements

I would like to express my gratitude to those people who have directly or indirectly supported me during my work on this research.

In first place, I want to thank my main supervisor Dr. Pinar Yolum, for diligently guiding me through the work involved in this research, and for her constant encouraging and positive attitude. Her support was of crucial value in the completion of this Thesis.

Secondly, I would like to thank Dr. Henry Prakken for his meticulous analysis of the early version of my work. His suggestions, helped me shaping the current version of the research as presented in this document.

Lastly, I would like to express my gratitude to my parents who have always, unconditionally supported me.

Abstract

Privacy infringements in dynamic environments, such as online social networks and the Internet of Things, are still poorly addressed by traditional privacy regulations. Contextual Integrity (CI) has been proposed as an alternative definition of privacy, which describes privacy preservation in terms of appropriate information exchange, with respect to contextual norms. CI has inspired a number of approaches towards regulating appropriate information sharing in OSNs and the IoT. Nevertheless, due to its scale and heterogeneity, the IoT environment in particular, has been addressed by a limited number of efficient methods. The literature suggests that available approaches would benefit from a definition of contexts which can capture contexts' relations, allowing contexts' inference from fragmentary information. Furthermore, these approaches should display decision-making capabilities in partially observable and incomplete-information environments. This study proposes PARCo, a knowledge-based agent which reasons on its internal context representation implemented by way of an OWL ontology and subsequently uses argumentation to take an information sharing decision in the IoT environment. PARCo and its components are evaluated with respect to a selection of IoT scenarios and compared to a previous approach.

Contents

List of Tables

List of Figures

1 Introduction

It is commonly acknowledged that the challenge of privacy protection has amplified in the digital era. Due to rapid technological developments, the collection and subsequent exchange of data have become increasingly more effortless to accomplish. One growing technology that favours mass generation and in general, movement of data, is the Internet of Things (IoT)[Weber, 2010, 2015; Sicari et al., 2015; Miorandi et al., 2012; Ziegeldorf et al., 2014; Atzori et al., 2010]. It is commonly recognized that the number of connected devices is ever increasing in the IoT. Part of these devices are being delegated the task of managing a growing amount of information on behalf of their owners. Such information might be intrinsically sensitive or become so in particular contexts. Clearly, the treatment of this information might put at risk and even violate the privacy of Internet of Things participants.

Legal frameworks, such as the U.S. Constitution, are currently being employed in solving privacy debates, as they protect individuals against governmental abuse of personal information, through limiting access to sensitive data and avoiding information breaches [Nissenbaum, 2004]. In Europe, the treatment of personal data is regulated by law as well. For instance, the European Union relies on recently updated regulations for the protection of personal data, commonly known as the General Data Protection Regulation (GDPR). The GDPR is an updated policy with respect to the Data Protection Directive introduced by the European Community decades ago, with the intent to regulate data treatment within Europe. The GDPR has the purpose of providing more homogeneous regulations among European member states, regarding the treatment of personal data from organizations and at the same time, to protect individuals against improper data collection and processing from organizations [Voigt and von dem Bussche, 2017]. In particular, the GDPR emphasizes some key aspects regarding the collection and processing of personal data, such as individuals' consent as a lawful basis for data collection and processing and among others, individuals' rights of deletion, objection and data portability. More concisely, the GDPR attempts to reduce abusive personal data collection and processing, and is intended to provide individuals with transparency and some degree of control over their sensitive data.

Even though the GDPR represents a step forward with respect to previously available privacy policies, by tackling abusive data collection and unclear data treatment from

organizations, such policy does not provide solid guidance on issues arising in scenarios such as public surveillance, online social networks or the Internet of Things. Different studies agree that in these environments, privacy is not necessarily endangered by abusive treatment of sensitive data, rather privacy violations may occur as a result of inappropriate information sharing. For example, in online social networks privacy violations might arise as the result of a multitude of users' interactions, which taken alone do not represent a violation [Kökciyan and Yolum, 2016]. In fact, different studies share a position of distrust with respect to the effectiveness of static privacy policies in scenarios like public surveillance, OSNs and IoT. These studies agree that the net distinction between sensitive and not sensitive data is not effective or sufficient to tackle implicit privacy violations arising in the above mentioned scenarios scenarios [Nissenbaum, 2004; Criado and Such, 2015; Weber, 2010, 2015; Ziegeldorf et al., 2014; Kökciyan and Yolum, 2016; Kökciyan et al., 2017].

Nissenbaum has addressed the scenario of public surveillance with respect to available privacy policies and has highlighted the three principles which according to her bound the core of all such policies. In specific, these regulations limit their guidance with respect to the following principles: i. protecting individuals against governments intrusions; ii. restricting access to sensitive information and iii. limiting access into private spheres of life [Nissenbaum, 2004]. In her work, Nissenbaum provides examples, showing that the application of the three principles is often not straightforward and thus, such principles not adequate to provide solution to privacy disputes. Additionally, she explains why public surveillance does not fall under the scope of any of the mentioned principles. Specifically, since the collected information consists of public records, it is not considered sensitive, personal or confidential. Additionally, the collection of data occurs in public, therefore there is no intrusion in personal spheres. Yet, as Nissenbaum states, we have the intuition that public surveillance represents a privacy infringement.

Consequently she provides an alternative theory for privacy, namely Contextual Integrity (CI). This concept abandons the two main features common to available privacy policies, namely: i. not being conditioned from time, location or any other attribute; ii. their net distinctions between sensitive and not sensitive data, and private and public domains [Nissenbaum, 2004]. According to the intuition of Contextual Integrity, all situations in life are subject to norms regulating the appropriateness of actions in each situation. With respect to information exchange, in every context in which information

might be shared, there are certain norms regulating the appropriateness of information distribution. Note that as Nissenbaum explains, contexts across which an individual's life progresses, are not always clearly distinguishable nor mutually exclusive. In these grey areas, according to Nissenbaum, traditional accounts for privacy show their shortcomings and may lead to privacy infringements. Contextual Integrity is therefore said to be broken, when appropriateness or distribution norms are not respected, hence privacy is not maintained.

Public surveillance is not the only instance to have put existent privacy policies under the question mark. In a similar way, normative privacy policies have had limited effects with respect to online social networks. While policies such as the GDPR reduce the amount of collection and aggregation of users' data, such policies do not provide guidance on information sharing occurring among the participants of an OSN, as evident from available studies on the matter [Kökciyan and Yolum, 2016; Kökciyan et al., 2017; Criado and Such, 2015].

As Criado and Such point out, one of the main concerns in online social networks in terms of privacy violation, is the exchange of inappropriate information and the dissemination of such [Criado and Such, 2015]. The authors explain that there are contexts in which sharing certain information might result inappropriate, even if the information itself is not intrinsically sensitive, for example sharing political views at work. Moreover, it is risky to disclose information in a context within which such information was previously unknown. These two issues are problematic in OSNs because, as Criado and Such describe, users in an OSN often deal with situations where the context has no definite borders over time. That is intuitively true, since content shared in OSNs can be understood to have different contexts depending on the interactions of users with the content and even on interactions among users themselves. These interactions, which alone might not violate users' privacy, may result in an inappropriate information exchange or dissemination of sensitive details. From another perspective, as Criado and Such claim, users are not able to share information according to context as they would do in real life, as they have no control over other users' actions [Criado and Such, 2015].

OSNs privacy violations have also been tackled by Kökciyan and Yolum, who seem to agree with Criado and Such on the little control on the information flow in OSNs, and thus on privacy risks. In particular Kökciyan and Yolum have addressed violations that might occur by posting in online social networks. In specific, posts may contain information

regarding other users, which even though not intrinsically sensible, may become sensible with respect to the privacy preferences of other users. The work conducted by the authors aimed at detecting such violations and allow action to be taken accordingly [Kökciyan and Yolum, 2016]. Similarly, Kökciyan et al. address the risk of privacy violations due to inappropriate posts in OSNs. With respect to Kökciyan and Yolum thoug, the idea is to provide users with a tool to reason beforehand on their privacy preferences, and to prevent the sharing of a post which might violate privacy [Kökciyan et al., 2017].

The third important scenario which introduces significant challenges to users' privacy is the Internet of Things (IoT). While there is no universal IoT definition agreed upon, the common ground for all IoT visions is the IoT as a network of a vast number of smart objects, which are identifiable and can interact with each other towards achieving certain goals. As covered in the literature, the penetration of the IoT paradigm in everyday's life has both introduced commodities and risks. Advantages are evident in IoT environments such as comfortable homes and offices, smart factories, e-Health and so on. Nevertheless, a variety of studies have discussed the privacy and security challenges that the IoT is facing due to its characteristics [Atzori et al., 2010; Miorandi et al., 2012; Ziegeldorf et al., 2014; Weber, 2015; Sicari et al., 2015].

Firstly, due to heterogeneity and scale causes, it is unfeasible at the moment to picture an agreement on a common privacy policy across all entities involved in the Internet of Things [Sicari et al., 2015; Weber, 2015]. Secondly, similarly to the OSNs environment, privacy risks predominantly arise from improper information sharing occurring within such vast, heterogeneous and distributed network of smart devices [Kökciyan and Yolum, 2017; Weber, 2015; Sicari et al., 2015]. For this reasons, it is currently unfeasible for privacy policies to tackle the IoT, and the task becomes even harder with respect to future developments of the IoT, as pointed out by Weber [Weber, 2015].

One interesting study covering regulations for the IoT has been conducted by Weber in 2010, who has further updated his work in 2015 [Weber, 2015]. Weber has discussed specific IoT privacy challenges and has provided an overview of what regulations for such an environment should take into account. Weber covers aspects such as data minimization, anonymity and transparency, which according to him, need to be taken into account when creating regulations for the IoT. Some of these requirements are currently covered in the GDPR, nevertheless, there are a few obstacles that still make normative policies such as the GDPR insufficient for privacy protection within the IoT. Early in his paper, Weber

explains four elements that must be taken into account when building regulations for the IoT: i. global technology standards, ii. broad scope (encompassing all spheres of life), iii. verticality and iv. technical IoT aspects [Weber, 2015]. It is evident that current regulations do not properly accomplish Weber's suggestions simply by considering point i., according to which technology standards should be globally consistent. As Weber suggests, the scale of the IoT is still increasing and paired with its heterogeneity, the satisfaction of point i. tends to become more difficult. Moreover, available laws come into play when information tagged as personal is exchanged, however, as suggested by Weber, the IoT raw data are not intrinsically personal. These challenges lead Weber to his conclusion according to which IoT's use and privacy implications are largely unaddressed [Weber, 2015].

These challenges are further supported by Kökciyan and Yolum, who agree that the IoT is dynamic, heterogeneous and extended in scale [Kökciyan and Yolum, 2017]. The authors describe the IoT as being dynamic because participating entities do not share a common access point, like users do in online social networks by means of authentication pages. Such an access point would already represents a degree of control, as the entities that could possibly interact would be known. Instead, in the IoT environment, users' devices simply establish contact dynamically, while users themselves are not aware of other entities present in the environment or their capabilities and privacy policies [Kökciyan and Yolum, 2017]. Another important aspect is heterogeneity. Kökciyan and Yolum add that entities communicating in the IoT will not share the same capabilities nor purposes. For this reason, privacy policies might have different values according to the entity that is employing it.

More concisely, normative legal tools attempting to regulate data generation and treatment, fail to resolve privacy debates in scenarios like public surveillance, online social networks and even more complex environments such as the Internet of Things. A number of studies have tackled such environments with alternative approaches, one example being Nissenbaum's Contextual Integrity as a benchmark for privacy with respect to public surveillance. Overall, a prevailing result in the available work is the convenience of having entities capable of autonomous reasoning on privacy, by taking into account the context in which they are operating and their own privacy preferences [Nissenbaum, 2004; Criado and Such, 2015; Kökciyan and Yolum, 2016; Kökciyan et al., 2017; Kökciyan and Yolum, 2017; Ziegeldorf et al., 2014; Sicari et al., 2015].

1.1 Existing Approaches

Various work has been dedicated towards providing an account for privacy in terms of appropriate information exchange, in online social networks and the Internet of Things [Criado and Such, 2015; Fong, 2011; Kökciyan and Yolum, 2016; Krupa and Vercouter, 2012; Kökciyan et al., 2017; Kökciyan and Yolum, 2017].

With respect to online social networks, Fong has proposed a relationship-based access control (ReBAC) type of social network, as a suitable model for information exchange in the healthcare domain [Fong, 2011]. The author has formalised access control to sensible information, based on the relationship of users exchanging the information and the contexts in which the exchange occurs [Fong, 2011]. Fong has used a hierarchy of contexts to formalise the fact that two users having a certain relationship, might have different access authorizations according to the context. However, even though Fong's context hierarchy could be adapted to another domain, this hierarchy is not compelling for a IoT environment, as it represents contexts in a specific arrangement, from general to specific. In Fong's work, access to an information is provided based on the relationships of the information requester and owner wit respect to a specified context and all its ancestor contexts. Nevertheless, in a IoT scenario the domain of an information exchange is often implicit, thus, such a hierarchy would rarely be available. Instead, decisions would have to be taken with respect to contexts which might have much more dynamic relations.

Another approach addressing privacy in online communities, based on the context within which information is exchanged, roles and relationships, was proposed by Krupa and Vercouter [Krupa and Vercouter, 2012]. The authors have taken inspiration from Nissenbaum's Contextual Integrity concept, to model a network in which agents share information and avoid violating privacy in the sending phase. Accordingly, agents can send punishments when they receive an information which was not appropriate. Even though in the proposed approach, appropriateness of information exchange is managed with respect to contexts, Krupa and Vercouter do not focus on the contexts themselves and assume that a hierarchy of contexts would be available from organizational entities [Krupa and Vercouter, 2012]. However, in many occasions in the IoT environment, it is likely that contexts are not precisely known a priori, and reasoning on contexts might be required. With respect to online social networks, more interesting work has been done by Criado and Such, Kökciyan et al. and Kökciyan and Yolum [Criado and Such, 2015; Kökciyan et al., 2017; Kökciyan and Yolum, 2016].

Kökciyan and Yolum and Kökciyan et al. have handled in details the scenario of online social networks. Both studies have proposed frameworks that take into account the context and roles in accounting for privacy in OSNs. For example, Kökciyan and Yolum have developed an approach that detects privacy violations based on the roles and contexts of posts. In their work, privacy violations are provided with semantics, however, there is no need for decisions to be taken, as no active information sharing occurs. Instead, Kökciyan et al have proposed a model that should detect privacy violations beforehand, based on the privacy preferences of users. The authors used an ontology in order to represent the domain knowledge, in other words the information available in the social network. This consists of the relations between users, what they share and their privacy constraints [Kökciyan et al., 2017]. Semantic Rules were consequently used in the proposed work, in order to infer new information and further, by means of assumption-based argumentation, to decide whether the post is safe to be published or it violates other agents' privacy. Contexts can be assigned to posts by using the *isInContext()* relation. In the second half of their work, Kökciyan et al. mention that their approach can benefit from the contexts being represented in the ontology in a hierarchical manner, with the subclassof property. That is because by abstracting a privacy rule with respect to the ancestor context, it is possible to change the rejection reason, and hence change argumentation's result. Nevertheless, the authors do not elaborate more on this, nor do they consider other possible relationships among contexts.

Another interesting model has been developed by Criado et al., who have proposed the first model for what the authors have named implicit Contextual Integrity, in the attempt of addressing the dynamic contexts of OSNs [Criado and Such, 2015]. Specifically, the agent implementing their model, first finds the community interacting with its user. Consequently, the agent will it uses appropriateness functions in the exchange of messages between the user and her inferred community, in order to avoid privacy violations.

Yet, the model proposed by Criado and Such uses the interactions of users in order to infer the appropriateness of information exchange in a given context. In other words, even if the proposed approach is able to infer contexts for new participants in the information exchange, the reasoning regarding such exchange is based on the relations of users, and does not take into consideration the relations among contexts themselves. While users' relations represent a good factor to rely on, towards proper information exchange in online social networks, in the IoT environment it is desirable to be able to take information

sharing decisions, even in absence of detailed information regarding users' relations. In particular, in the IoT information might become available from unrelated sources. In that case, reasoning on the content of the information and the context in which it has to bee exchanged might be more appropriate.

With respect to the IoT, one approach targeting the Internet of Things has been proposed by Kökciyan and Yolum [Kökciyan and Yolum, 2017]. This approach attempted to provide a IoT entity with a way of figuring out which context it is in, and based on that, decide whether to share a piece of information or not. The authors have used predefined contexts in combination with contextual norms, in order to reason on which context might be active. By means of argumentation, Kökciyan and Yolum have implemented a decision making algorithm that eventually decides whether to share or not the information. As the work of Kökciyan and Yolum is the study that best approximates a desirable approach for the IoT scenario, it shall be discussed in the next section, alongside its improvable aspects.

From a general perspective, there are a few intuitions shared by the above mentioned studies, which are worth considering in the attempt to tackle privacy in the Internet of Things. The first aspect is the abandonment of focus on centralized codes of conduct, thus the emphasis of single reasoning entities, which together must achieve the common, greater goal of safe information exchange. This is especially the case since Nisseunbaum introduced the concept of privacy as the maintenance of Contextual Integrity [Nissenbaum, 2004]. Such vision of privacy has inspired a number of approaches to shift the focus on single appropriate information exchange actions, with the purpose of preventing violations in first place. Focusing on the appropriateness of information exchange leads to a second recurrent element. That is the consideration of context as a main factor in assessing whether the information exchange is appropriate or not. For example, the studies presented above have all used some type of contextualisation of the information in order to assess whether it is sensible or not.

Nevertheless, the context representations that have been proposed, often lack depth. In other words, available computational frameworks concentrate on the aspect of roles within contexts, and relationships among users. The contexts themselves usually have no relations among each other and are merely an attribute. Two exceptions are the hierarchy of contexts of Fong and the use of the subclassof relation to relate contexts in the ontology used by Kökciyan et al. [Fong, 2011; Kökciyan et al., 2017]. Nevertheless, in both studies

relations among contexts are still primitive and thus, such aspect is not fully exploited.

A second aspect emerging from available frameworks, is their privacy-preserving centered design [Krupa and Vercouter, 2012; Kökciyan and Yolum, 2016; Barth et al., 2006; Criado and Such, 2015; Fong, 2011]. These frameworks usually do not exchange information, unless the privacy constraints established by the system are satisfied. In other words, they tend to neglect trade-offs between privacy and other objectives a person might have. Consider for example a supermarket bonus-card, which benefits can only be obtained by sharing personal data with the supermarket company. If an agent implementing a rigid system as previously described, would have to decide whether to activate such a card, it would only rely on the privacy preferences in order to decide whether it is appropriate or not to share personal information with the supermarket company. The benefits of a bonus card would not be considered at all. However, it is not realistic to assume that privacy is the unique goal of any person using the mentioned frameworks. For this reason, contexts that might lead to privacy violations according to the privacy preference of a user, shall still be given a fair weight in the decision making process. Back to the example of the supermarket, the context of having an active bonus card, and thus less privacy, might, be desired in a similar measure as the one of having completely untouched privacy. One decision-making approach that has been adopted by a few of the above mentioned studies is argumentation [Kökciyan et al., 2017; Kökciyan and Yolum, 2017; Fox et al., 2007; Dijkstra et al., 2005]. Argumentation and its advantages shall be discussed in Section 2.3.

While various studies have tackled computational accounts for privacy in online social networks, the Internet of Things has not received equal attention. In particular, there is a gap in the literature between studies that have discussed the challenges of such an environment and work providing computational solutions for privacy in the IoT. It is the goal of this research to provide an account to privacy in the environment of the Internet of Things, based on the common intuitions and interesting concepts covered in past research.

1.2 Research Question

Ideally, an entity operating in the IoT while ensuring its owner's privacy, would benefit from context-based reasoning, in order to take proper decisions with respect to exchanging information. The research question is thus:

How to design ***a semantic representation of contexts****, to allow* ***reasoning*** *and* ***decision making*** *for* ***privacy****?*

In order to tackle the research question, it is convenient to take into consideration what has been emphasised in the introduction. Specifically, it is necessary to study what would be an appropriate representation of context; whether an ontology of concepts would be beneficial for reasoning on contexts; finally, it is important to determine how to perform decision making, in order to take actions towards information exchange. It is therefore convenient, to first address the following sub-questions, which answers would allow to tackle the main research question:

1. *How to represent* ***contexts*** *for reasoning on privacy?*
2. *Does an* ***ontology of concepts*** *bring any advantage with respect to reasoning on privacy contexts? If yes, which one?*
3. *How to exploit the context representation for* ***decision-making*** *on contexts for privacy?*

The remainder of this document is structured as follows: Section 2 provides relevant technical background information. In Section 3, available work related to context definitions, contextual integrity computational models and decision making through argumentation will be covered. In the Section 4, the approach proposed in this research will be explained in details, specifically with respect to reasoning on contexts. In Section 5, additional emphasis will be put on the agent and the algorithm that processes the input for the argumentation engine. Section 6 is concerned with evaluating the proposed reasoning agent in the scope of a variety of context scenarios meant to illustrate agent's reasoning capabilities. In Section 7, the evaluation and results will be discussed with respect to the main research questions. The discussions shall also mention limitations and future directions of this study. Finally, some concluding remarks will be provided given in Section 8.

2 Technical Background

This section aims at familiarising the reader with the most important technical aspects of this thesis. In order to do so, the work of Kökciyan and Yolum is first introduced, and its improvable aspects are mentioned. On this basis, the contributions of this research are stated. Consequently, ontologies as a mean to conceptualise knowledge and argumentation as a decision-making tool are being covered.

2.1 Reasoning on Privacy in the IoT

As mentioned earlier, a number of studies have provided insights on the challenges of addressing privacy in the IoT [Weber, 2015], [Weber, 2010], [Sicari et al., 2015], [Ziegeldorf et al., 2014], [Atzori et al., 2010], [Miorandi et al., 2012],[Kökciyan and Yolum, 2017]. Nevertheless, little work has specifically addressed the IoT environment with approaches to guarantee appropriate information sharing.

The work of Kökciyan and Yolum in particular, has first discussed such challenges, and has further tackled them by introducing a reasoning approach designed to take a decision regarding user's information, while maintaining her privacy safe. The authors introduced their reasoning model as ideally embedded within agents participating in the IoT. These agents communicate, sense and act in the environment on behalf of their users. When information has to be exchanged, an agent reasons on the possible contexts of its owner; if owner's contexts are appropriate with respect to the information to be shared, then the agent shall exchange the information. Consider the following two scenarios, which have been used by Kökciyan and Yolum as a IoT instance [Kökciyan and Yolum, 2017]:

> **Example 1**. A surveillance camera at Alice's work place records a video 24/7. The footage is not shared with others except when there is an emergency. Alice's boss Bob would like to access Alice's footage taken on November 30, with the claim that Alice might be in trouble and that her recent footage might help solve the situation. The camera needs to make a decision autonomously. Should the camera share the footage?

Example 2. The camera decides to consult Alice's home device and it turns out that Alice was not at home that day, either. Next, the camera consults the police department and finds out that there is a missing report for Alice. Is this enough evidence to decide Alice is in an emergency?

In order to handle the given examples, Kökciyan and Yolum have considered two elements: the representation of contexts inspired from Contextual Integrity and a method for decision making.

A context is represented as a set of norms and a set of relations [Kökciyan and Yolum, 2017]. Norms specify when a certain context is active, given that the information in their body is available. Relations instead, tell the agent from which other agents to collect additional information in the decision making phase. In other words, the agent will aim at collecting all available information in order to apply the most rules it can. As Kökciyan and Yolum explain, in their approach it is possible for an user to find himself in more than one context at a time. For this reason, they use degrees of belief, in terms of values between (0,1]. Thus, each piece of information stating that a certain agent is in a given context, as well as each norm used to infer a context are associated with a degree of belief.

This quantification of active contexts, is exploited in the decision making phase of the approach. Such phase is handled by means of argumentation, executed in the ASPIC engine. Hence, each information available about Alice, in "Prolog-like" predicate form, will become a deafisible premise if it has an associated dob value, or a strict premise otherwise. With respect to norms, Kökciyan and Yolum have decided to give a dob value of 0.9 to contextual norms, and to make the norms that manage the dob value strict. This means that the norms inferring contexts are defeasible, which as described by the authors, are attacable. The authors propose an algorithm that illustrates how such decision making occurs. On an intuitive level, the information availalbe at time 0, is used against the contextual norms in order to infer the active contexts. If there are missing predicates necessary for applying a contextual norm, the agent will ask the other agents, based on the relationships, whether they have the necessary available information. When all possible extra information is collected, the new belief base is used with the contextual norms in order to infer the active contexts [Kökciyan and Yolum, 2017].

As a final result, the argumentation will provide the state of the arguments that have been built for the specified query, in the case of the work of Kökciyan and Yolum,

shareFootage(alice,30). According to the preferred semantics, which is chosen for the evaluation, an argument is preferred if it cannot be defeated. Therefore if there is one argument for shareFootage(alice,30), which has not defeater, then the argumentation will consider such argument in the preferred extension, hence the footage will be shared.

The model was illustrated to produce significant results with respect to the two IoT scenarios proposed above. Consequently, Kökciyan and Yolum have proved their results theoretically meaningful, by showing that the agent can always reach a sharing decision, and that such decision is sound according to the Best Effort Theorem stated in the study. The latter ensures that: i. the agent contacts all agents that might provide missing information and ii. since every piece of information is associated with a dob value and the arguments attacking relations are constructed accordingly, the winning argument according to ASPIC will be the correct decision to output.

The approach proposed by Kökciyan and Yolum has provided an interesting account to privacy in the Internet of Things. In specific, it has encapsulated the main elements which, as a suggested by the literature introduction, are desired in order to tackle the IoT environment. These consist of an approach to privacy inspired by CI and reasoning on available information by means of argumentation in order to reach the decision that is most supported by such knowledge.

Nevertheless, the work done by Kökciyan and Yolum leaves space for improvements, as the authors mention in the future work suggestions. In particular, one of the two mentioned directions is to capture relations between contexts in order to improve the reasoning by allowing new inference of contexts and norms, prior to the decision making process [Kökciyan and Yolum, 2017]. Additionally, note that the authors have proposed their framework as suitable for an agent. Still, an agent instantiating their approach has not been developed since the ASPIC argumentation engine could alone execute the entire reasoning. However, in order to achieve enhanced reasoning on contexts, a more detailed representation of context is required. This could not be properly achieved just by using the argumentation engine. As it shall be revealed in Section 4.1, a software agent represents a natural way to exploit the knowledge representation and decision-making process. thus an agent might be suitable in order to encapsulate all used tools. More concisely, the following contributions are the focus of this research:

1. **use of an ontology for capturing the relations between contexts**

2. **development of an agent that exploits the representation of contexts in the decision-making phase**

2.2 Knowledge-Representation and OWL

An agent required to take privacy decisions while respecting Contextual Integrity, does intuitively demand a world representation to rely its reasoning on.

An ontology provides a natural way of modelling the world, as an ontology is basically a formalisation of a domain in terms of concepts and relations. For example, the formalisation of a company environment might include concepts such as Employee, Boss, Coffe Room, Office and so on. Relations may link these concepts among each other, for example an Employee might have a Boss and a Boss may be in his Office.

One way of creating an ontology is by using the standard Web Ontology Language (OWL). An ontology defined in OWL, consists of classes, properties and individuals [Behhofer, 2009]. Moreover, OWL allows the use of Boolean operators such as , quantifiers and the specification of characteristics of object properties. Examples of such characteristics are transitivity, symmetry, reciprocity and so on. Specifically, in this study it is relevant to consider the meaning of a transitive, functional and inverse functional property, as these properties hold an important role in knowledge representation as proposed in this work. The web documentation by Smith et al. provides further details on the expressivity of OWL [Smith et al., 2004]. With respect to the features of object properties the authors explain the following:

Transitive property - if A is related to B and B is related to C by a certain property, if the the property is transitive, then A is related to C by that property.

Functional property - A can be related to at most one individual B in the range of a functional property. If A is related to more than one individual in such a range, then those individuals refer to the same object, hence they are not distinct entities.

Inverse Functional property - intuitively an inverse functional property allows only one individual A to be put into relation with B. If more individuals related to B by the inverse functional property, then those individuals refer to the same object.

With respect to this study, other interesting OWL aspects covered by Smith et al.,

are the equivalence and subclass-of properties of classes [Smith et al., 2004]. The authors cover the meaning of each of them and further mention what is their difference.

SubClass Of property - this property intuitively relates a specific class to a more generic class. If X is a subclass of Y, then every instance of X is also an instance of Y. As the authors explain, the subclass-of relation is transitive, thus if X is a sublcass of Y and Y a subclass of Z then X is a sublcass of Z. For example, a pizza margherita is a subclass of the class Pizza. If a pizza margherita without tomato sauce is a sublcass of PizzaMargherita, then it is also a subclass of Pizza.

Equivalent To property - this property is used to indicate that two classes have precisely the same instances. As Smith et al. write, this property provides a powerful way to define a class based on the satisfaction of a property. For example, a home office context could be defined as equivalent to being at home some day. This would imply that all things that are at home some day are in a home office context. On the other hand, if we would have defined the home office context as a sublcass of all things being at home some day, there could still exist things that are at home some day, but are not in a home office context.

2.3 Argumentation

Given a problem and a set of possible solutions, decision making is the process of selecting the most appropriate solution from the set of alternatives, in order to tackle the problem at hand [Amgoud, 2009; Ouerdane et al., 2010; Russel and Norvig, 2009]. Even though decision making methods have been researched within a variety of fields, much focus has been put on formal ways of assessing the best solution in a set of alternatives, rather than focusing on what makes one solution the best [Amgoud, 2009; Ouerdane et al., 2010]. In particular, much work has been done by economists, who have developed techniques that can rank a set of alternative options by means of a utility function, which assesses how good an option is [Amgoud, 2009]. From this perspective, a rational decision is one that leads to an outcome which maximises such utility function [Russel and Norvig, 2009]. As Amgoud explains, this method expresses the whole decision-making process in analytical terms. For this reason, decision-making methods based on this principle, do not naturally match the way humans reach decisions [Ouerdane et al., 2010; Amgoud, 2009].

It is commonly accepted that in many domains, explainable decision-making is required, for example in the legal, military, security and medicine domains. Indeed, Both Amgoud and Ouerdane et al. agree that, in certain scenarios, the process of reaching a decision might be of equal importance with respect the decision itself. That is because domains like the above-mentioned ones, might require reasons motivating a certain decision. As a consequence, in many cases it is desirable to rely on decision-making methods that provide intuitive insights on how a decision was reached.

In this sense, a promising account to decision making is given by Argumentation. Intuitively, decision making modelled as argumentation, amounts to construct arguments for each possible solution, and to opt for the one with the strongest support [Ouerdane et al., 2010]. Computational accounts for argumentation emerged after a period when argumentation has been mainly employed, i. as a tool to treat non-monotonic reasoning formally and ii. as methodologies in the legal domain [Bench-Capon and Dunne, 2007].

Specifically in the Artificial Intelligence research field, due to its contributions, the work of Dung is considered to be a fundamental step in the study of argumentation [Dung, 1995]. In this early study, Dung has aimed at providing a general framework for argumentation, which exhibits important properties with respect to non-monotonic logics and functional programming. Dung's work's relevance lies as much in the proposed general argumentation framework, as in his results. One relevant result with respect to the work proposed in this document, is the demonstration that a variety of non-monotonic logics are an instance of Dung's argumentation theory.

Non-monotonic logics have been of interest in AI research, as a consequence of the limited account of classical logical reasoning with respect to "daily-life" reasoning [Ouerdane et al., 2010]. As Ouerdane et al. explain, the motivation was to formally account for defeasible inference, i.e., inference that can change when additional information is provided. Non-monotonic reasoning is covered in details by Reiter [Reiter, 2003]. The author provides the following typical example as a basic case of non-monotonic reasoning:

Example 1 [Reiter, 2003]

- By default, birds fly;
- Tweety is a bird;

- Therefore, Tweety flies.

It is known that birds fly and that Tweety is a bird. Thus, it seems reasonable to conclude that Tweety flies. As Reiter explains, in absence of additional information, it is justified to conclude that Tweety flies [Reiter, 2003]. On the other hand, if we knew that Tweety is an exceptional bird, e.g. a penguin, such conclusion shall change. This reasoning pattern is interesting in the scope of this research, since it naturally corresponds to the human way of reasoning and reaching decisions. Moreover, as Reiter explains, there are cases in which non-monotonic logics have the capability of reaching a conclusion in absence of information, in contrast to classical logics, which could not reach such conclusion [Reiter, 2003]. As Reiter describes, if flying things would be defined as those things that are birds, not emu, not dead and so on, by means of classical logics it would be impossible to infer that Tweety flies from the information that Tweety is a bird, because intuitively the other antecedents are not available.

As a consequence of argumentation's capabilities of representing non-monotonic reasoning and providing transparency w.r.t the reasoning process itself, such approach is a well suited approach to decision-making in the scope of this research. In particular, it is relevant to specify the argumentation instance employed in this work, as well as the computational tool that allows its implementation.

Since the requirements in terms of argumentation capabilities remain unchanged, the same implementation of the argumentation engine will be used in this study. The inference engine used in the decision making phase, was developed in the context of the European project ASPIC [Fox et al., 2007]. This argumentation engine has previously been employed by Kökciyan and Yolum in order to execute their proposed approach [Kökciyan and Yolum, 2017]. Fox et al. describe in more details the ASPIC inference engine which consists of four steps: i. arguments construction, ii. arguments valuation, iii. arguments interaction and iv. arguments status evaluation [Fox et al., 2007]. As Fox et al. describes, arguments are constructed from a knowledge base made of facts and strict and defeasible rules in the form of modus ponens. With respect to the argument valuation step, one relevant aspect is the possibility of assigning a weight, which in the scope of this study will be called a degree of belief, and the possibility of choosing the strategy in order to assign the support for the main argument. The weakest-link strategy compares all the sub arguments of the main argument, and assigns the lowest degree of belief. The last-link instead, assigns the

value of the highest defeasible inference in the support of an argument, and the lowest degree of belief of all inference paths, in case there were more than one defeasible rule on the last level of argument's subtree [Fox et al., 2007]. Another important aspect to be taken into consideration in the evaluation of arguments, is the semantics. In ASPIC it is possible to evaluate arguments according to the grounded or preferred semantics. Intuitively, the grounded semantics can be used to perform "skeptical" reasoning, since the grounded extension contains only those arguments who are labelled *In* in all status assignments. On the other hand, the preferred semantics is used for "credulous" inference. The preferred semantics, as in the work of Kökciyan and Yolum, is used in this research as a parameter in the ASPIC engine. This semantics allows to take a decision even if arguments are not universally justified, nevertheless they can defend themselves against attacking arguments. In the environment of the IoT tackled in this study, such an attitude is more reasonable than a skeptical attitute.

In conclusion of this subsection, it is also relevant to mention another theoretical remark of the ASPIC engine. As referred by the documentation provided with the argumentation engine, the ASPIC implementation relates to the work of Caminada and Amgoud, who have proposed the so called rationality postulates for argumentation systems [Caminada and Amgoud, 2005]. These postulates are principles which should be respected by argumentation frameworks in order to avoid counter-intuitive results. In the mentioned work, the authors have proposed two principles that argumentation systems should respect, namely closeness and consistency. Intuitively, closeness implies that the set of justified conclusions in the output of an argumentation system, should include all those arguments that can be built by means of the strict rules. As this set might be inconsistent, the second postulate for consistency is needed [Caminada and Amgoud, 2005]. According to the consistency postulate, two contradictory arguments should not be justified at the same time.

3 Related Work

3.1 Defining Contexts

A number of studies have addressed privacy in online social networks and the Internet of Things, by introducing the concept of context as a main element in defining appropriate information sharing [Kökciyan and Yolum, 2016; Criado and Such, 2015; Kökciyan et al., 2017]; some of these studies have taken direct inspiration from the Contextual Integrity concept [Barth et al., 2006; Krupa and Vercouter, 2012; Sicari et al., 2015; Kökciyan and Yolum, 2017]. Generally, all approaches exploit a concept of context in order to infer the appropriateness of the information being exchanged and therefore ensure privacy. Overall, only a few of the available studies have achieved a degree of flexibility with respect to reasoning on contexts. This was mostly possible by focusing on the relationships among these contexts.

For example, Barth et al. have aimed at formalising some aspects of Contextual Integrity, such as roles and contexts, in order to express privacy preferences and expectations, with a main focus on the healthcare domain [Barth et al., 2006]. The model proposed by Barth et al., refers to contexts as the elements of the partition set taken over the set of all roles. Hence, each context is a subset of roles. Each user can have different roles in different contexts, and therefore be active in multiple contexts. For example, as Barth et al. explain, a user Alice could be both active in a bank context by having the role of a customer and active in a healthcare context by being a medic. The way in which the authors achieve a degree of reasoning on contexts, is by structuring roles by means of a partial order. Therefore, if Alice would be active in the role of a psychiatrist, she would have to be also active in the role of a doctor according to the partial order[Barth et al., 2006]. Contexts and roles have a key function in the model proposed by Barth et al.. Roles within contexts are used to express that communication which is perfectly acceptable between a psychiatrist and patient might be completely unacceptable between a human resource specialist and a job applicant. Norms are used to make effective the constraints of a given context and by considering the roles, they limit what different figures can say in different contexts. Finally, with respect to contexts, the authors mention that multiple instances of a context may exist in their model. Intuitively this is possible by having different partitions over the set of roles. Nevertheless, Barth et al. do not elaborate on possible relations between contexts.

A different approach with respect to defining contexts as a subset of roles has been taken by Criado and Such, who have used a set of users as defining a certain context in online social networks [Criado and Such, 2015]. The intuition behind their work, is that a context is the set of users tagged in a certain message exchange, message which includes further information such as a set of relevant topics, a sender and a set of receivers. The model of Criado and Such allows users to be active in more than one contexts at the same time. In general, the work of Criado and Such focuses on achieving an implicit representation of contexts, in this case, through a set of users. For this reason, their work does not address possible relations among contexts, rather the perspective is hold on users interactions and the appropriateness of information exchange.

Another account to security and privacy is developed by Fong, who develops an access control model inspired by the Relationship-Based Access Control (ReBAC), which has been the emerging paradigm in social networks [Fong, 2011]. Fong argues that this paradigm is well suited in other environments, where the relationship between the information owner and accessor is fundamental for the information exchange. One contribution of Fong is that he defines the contextual aspect of relationships. The author uses a tree structure in order to represent and organise the different contexts, and to provide a mechanism to deduce the context scope within which to provide access based on the relationships. More specifically, contexts span from a specific treatment case to the general healthcare system context [Fong, 2011]. The tree structure used by Fong does relate contexts to each other, however, such relation is merely a subclass relation. With respect to one specific environment, such as the healthcare domain which was modelled by Fong, using exclusively the subclass relation might be possible. Nevertheless, in presence of heterogeneous contexts, more flexible relations might be demanded.

The contextual factor has also been taken into account by Kökciyan and Yolum in order to regulate information exchange. The authors have focused on a framework for detecting privacy violations in online social networks [Kökciyan and Yolum, 2016]. The framework manages the sharing of posts which are defined to contain, among others, certain types of content, to have an audience consisting of certain agents and to have a context. The notion of context is represented trough an attribute attached to a post, representing the type of context relevant to a post. Yet, the context attribute is not further exploited in the work proposed by Kökciyan and Yolum.

A similar use of contexts with respect to Kökciyan and Yolum has been proposed

by Krupa and Vercouter [Krupa and Vercouter, 2012]. In order to describe the context of a message shared in a virtual community, Krupa and Vercouter use a context tag. Each context has a defined set of roles, which regulates the appropriateness of message exchange based on the roles that a user has in a certain context. In this work, the authors assume that these contexts and associated roles are provided by external entities. Relations between possible context tags are not treated.

Similarly, privacy violations arising due to post sharing in OSNs, have been tackled by Kökciyan et al.. Their work shows a more elaborate use of contexts with respect to the work of Kökciyan and Yolum [Kökciyan et al., 2017]. Contexts are still an attribute that can be attached to posts by means of a relation *isInContext*. However, there is some degree of reasoning which might occur in the model of Kökciyan et al., since contexts are represented in an ontology by means of the *subclass* relation. In particular, this aspect is used in order to generalise a reason to reject a post, by checking the ancestors of the initial context. The subclass relation between contexts used by Kökciyan et al. is in some sense analogous to Fong's hierarchy of contexts, which similarly links contexts from the most general to the most specific. Even though Kökciyan et al. do not elaborate more on the use of the ontology of contexts, the use of such an ontology might provide further advantages, which shall be explored in the technical part of this thesis.

Contexts have been at the core of regulating information exchange in the work of Kökciyan and Yolum as well [Kökciyan and Yolum, 2017]. In their work, a slightly more sophisticated definition of contexts has been presented. Specifically, contexts are represented as a tuple containing a set of norms and a set of binary relations. Norms specify when a certain context is active and interestingly, a norm can be triggered by information regarding another context. Consider for example the following norm expressed as a First Order Logic rule:

```
inContext(A, emergency, T) <- info(atWork(A, T)), info(fire(T))
```

Such rule describes the emergency context in terms of being at work on a certain time T, and of a fire being in place at time T. Note that being at work is naturally used by the authors to define the work context, by means of the following rule:

```
inContext(A, work, T) <- info(atWork(A, T))
```

With respect to the relation set defined for each context, Kökciyan and Yolum use the set in order to gather information from relevant external agent. Nevertheless, relations among users are not in the focus of the authors. Overall, this way of representing contexts seem to provide an implicit way to represent relations between context, and such aspect might be enforced by means of a "structure" of contexts as used by Kökciyan and Yolum and Fong [Kökciyan and Yolum, 2016; Fong, 2011].

3.2 Computational Contextual Integrity

Contextual Integrity has been theoretically introduced by Nissenbaum as an alternative approach to privacy [Nissenbaum, 2004]. Her focus in order to explain such concept, has been on the environment of public surveillance, which is not optimally managed by privacy policies [Nissenbaum, 2004].

Ever since its proposal, different studies have explored the concept of Contextual Integrity within other scenarios which lack proper privacy accounts, such as social networks, online social networks and the IoT [Barth et al., 2006], [Criado and Such, 2015], [Krupa and Vercouter, 2012], [Kökciyan and Yolum, 2017].

One example is the work done by Barth et al., which attempts to formalise contextual integrity concepts in order to represent privacy guidelines, policies and expectations so that they can be processed by an information system [Barth et al., 2006]. Contextual Integrity is the desired status of the system, hence informational norms, as called by the authors, must be respected and there should be no information breach. In other words, informational norms are meant to ensure the transmission principles which would ensure Contextual Integrity and therefore privacy. Some of these principles are confidentiality and reciprocity and they differentiate communication between friends from communication between a medic and her patient. More practically, Barth et al. propose a model of communicating agents that have roles within contexts and base their communication on the exchange of messages containing attributes about other agents. The authors impose a structure on the roles, meaning that some roles can be a more specific than others and therefore have different privileges. Intuitively, users may have simultaneously different roles within different contexts. This exchange of messages between agents form a trace that must comply with the informational norms of a given context [Barth et al., 2006]. Barth et al. represent norms by means of temporal logic which was allowed by the authors

in order to formalise the transmission principles mentioned above.

The proposed framework is evaluated in terms of expressivity with respect to known privacy legislations such as the HIPAA Privacy Rule and the Children's Online Privacy Protection Act (COPPA). This expressivity however, is strongly related to the use of temporal logic and does not offer many options for context representation nor for reasoning on contexts in order to make a privacy decision. In other words, the work done by Barth et al. does not naturally extend to a IoT scenario, which is the main focus of this research.

Another computational model of Contextual Integrity is proposed by Criado and Such [Criado and Such, 2015]. The study first analyses the CI concept, including the meaning of CI violation which translates to the violation of the norms of a certain context within which such norms are valid. Criado and Such then try to provide a model that abstracts from the contextual integrity assumption according to which contexts are well defined beforehand and which is not relevant in situations such as OSNs and even less in the IoT. The authors provide therefore an implicit contextual integrity model where intelligent agents can learn contexts and their information sharing norms. The intent is that of avoiding exchange of inappropriate information and dissemination of sensitive information.

This is achieved by having agents that monitor the activity of their users and by computing appropriateness levels for the exchanged messages content within various contexts given different roles of users involved in the exchange. Technically, the intelligent assistant agent relies on an information model for implicit contextual integrity which models the transmission of messages among users by finding the community around a user, regulating sending and receiving of messages, computing appropriateness of messages and by means of a time dependent function. This later function decreases the appropriateness of messages regarding certain topics for given users and their likelihoods of knowing certain contexts [Criado and Such, 2015].

Criado and Such extensively analyse the behaviour of their proposed model in a scenario of a hundred communicating agents, on 36 topics and operating on behalf of users with different attitudes with respect to social norms. Their experiments show that Contextual Integrity is indeed favoured by their model in that less inappropriate messages are exchanged when the number of malicious users stays within reasonable values.

Even though the approach taken by Criado and Such is promising with respect to OSNs, there is no information regarding its applicability in a IoT scenario. In particular,

the context representation aspect and use of argumentation tools for decision making, which are desired in the framework proposed in this paper, are not emphasised by Criado and Such.

Based on the intuition that in contextual integrity any information transmission can trigger a privacy violation depending on the context of the transmission, Krupa and Vercouter use contextual integrity in order to deduce if the transmission of certain information will cause the violation of contextual integrity and therefore of privacy, based on an appropriateness level that they derive from Nissenbaum's Contextual Integrity Theory [Krupa and Vercouter, 2012]. Sanctions for violating agents are also introduced by the authors. The developed model is instantiated on a photo sharing application. The developed assistant agent must decide whether to share the picture given the information attached to it such as context, audience and content. Krupa and Vercouter develop a use case on which the agent first informs its user that sharing the information is inappropriate and then gives the final choice to the user. In case the user would still decide to go ahead and share the picture she would be socially excluded as she is not trustworthy. The approach taken by Krupa and Vercouter offers a tool for making decisions based on appropriateness of posts within social networks. The decision mechanism is heavily based on evaluating the appropriateness function based on the post information and the users involved in the information transmission. The authors do not deeply develop the context aspect of posts, nor consider argumentation as a tool for decision making.

3.3 Argumentation for Decision Making

An overview of argumentation, its emergence in the field of AI as a decision-making tool, and its capacity toward explainable reasoning, has been discussed in section Section 2.3. This subsection shall propose a few examples of studies which have used forms of argumentation, in order to address information sharing or information access control, for safeguarding privacy in dynamic or distributed environments [Fox et al., 2007; Kökciyan et al., 2017; Dijkstra et al., 2005, 2007; Kökciyan and Yolum, 2017].

As described in Section 2.3, the work of Dung is considered a crucial study in the field of Artificial Intelligence, that has inspired various work on argumentation systems. Dung's conclusions in particular, support the use of argumentation for decision-making. As described in 2.3, the proposed general framework provides a unified model, based on

which it is possible to compare a number of non-monotonic logics [Dung, 1995]. In support of his argument, Dung demonstrates that Reiter's default logic and Pollock's inductive defeasible logic are both instances of his general argumentation framework [Reiter, 2003; L. Pollock, 1987]. Furthermore, by tackling the n-person game and the stable marriage problem by means of argumentation, Dung concludes that argumentation can provide insights on many social and economic disputes. This opens a wide range of practical possibilities for possible uses of argumentation systems. Additional conclusions following from Dung's proofs, are that the proposed framework provides a common ground for knowledge representations to be compared and even communicate. Further, it is argued that logic programming is a form of argumentation as well, thus logic programming is an optimal tool to implement argumentation systems. Finally, Dung also concludes that negotiation is a form of argumentation, in fact, he states that negotiation is the process of finding the solution, whereas argumentation is the process of providing grounds to a proposed solution. Thus, argumentation is said to be part of negotiation. The conclusions of Dung's work have clearly shown the potential of argumentation systems for decision-making. As a result, the argumentation subject has widely developed in the AI research,and argumentation systems have been studied towards different goals.

One example is the work of Fox et al., who have analysed the use of argumentation in the healthcare domain [Fox et al., 2007]. In specific, the main focus of Fox et al. has been the way in which clinicians treat patients based on incomplete or massive information amounts. It is argued that clinicians reason on available information, rather than using analytical methods in order to take decisions. Fox et al. discuss the shortcomings of analytical approaches in medicine, in contrast to the benefits of argumentation. Finally, the authors discuss the ASPIC project with respect to the requirements of argumentation as developed within the medicine field.

The ASPIC project has also come in handy in the approach developed by Kökciyan and Yolum [Kökciyan and Yolum, 2017]. In this study, the ASPIC argumentation engine was conceptualised within an agent, in order to figure out the active contexts of agent's owner. Based on the active contexts and the privacy preference of the user, a decision would be reached on whether to share or not to share a piece of information containing data about the owner. As described in Section 3.1, Kökciyan and Yolum have represented contexts in terms of a set of norms and a set of relations. The set of norms in particular is important in the argumentation phase. These norms are specified in ASPIC in FOL

format; consequently, the argumentation engine builds arguments with the top statement equal to the head of the norm and assumptions equal to the predicates in the body of the norm. These arguments are effectively supported, when the assumptions of an argument are available in the domain knowledge.

Further work involving argumentation has been done by Kökciyan et al. [Kökciyan et al., 2017]. In this paper, Assumption-based argumentation (ABA) was used in order to simulate a dialogue for decision making between agents [Dung et al., 2009]. In specific, the users would perform decision-making based on their privacy preferences, in order to assess whether the publication of a certain post in an online social network, would violate or not somebody's privacy. According to the argumentation system, arguments are constructed from assumptions and rules. Agents can attack the assumptions of others according to the ABA definition of attack relation. Upon concluded argumentation, a user knows whether it is admissible to share a post, or else the sharing is rejected.

Another example of the employment of argumentation for regulating information exchange can be found in the work of Dikstra et al. [Dijkstra et al., 2005, 2007]. Dijkstra et al. provided a way to regulate the information flow among distributed entities, in this case Dutch police departments. In order to address the distributed structure of the Dutch police, the authors have relied on a multi-agent system (MAS) architecture. Entities participating in the MAS communicate under the rules of negotiation dialogues which can turn into persuasion dialogues in case two agents would not find a beneficial agreement [Dijkstra et al., 2005]. The authors provide relevant examples regarding interactions between police officers across departments, and consequently describe their multi-agent model. In specific, w.r.t to the communication between agents, Dijkstra et al. first define negotiation and persuasion dialogues individually. Consequently, they propose a hybrid scheme which starts as a negotiation dialogue, but can be triggered into a persuasion dialogue.

The outlined multi-agent system for regulating information exchange between Dutch police departments, has been consequently implemented by Dijkstra et al. [Dijkstra et al., 2007]. In other words, the main cycle of the agent is implemented, allowing the authors to test their approach. It is interesting to mention that similarly to Kökciyan and Yolum and Fox et al., the ASPIC inference engine has been used for the argumentation computation.

4 Methodology

Various accounts to privacy in challenging environments have been discussed in the literature review. The Methodology section shall uncover and analyse the approach proposed in this research. To begin with, a knowledge-based agent will be introduced as a promising autonomous reasoning entity with respect to the IoT environment. Next, agent's internal domain representation will be explained in details. Finally, agent's reasoning and decision making capabilities based on the proposed domain representation will be discussed.

4.1 An Agent Approach

Software Intelligent Agents is the branch of AI concerned with hardware, software or hybrid entities which act autonomously in an environment, towards achieving a certain goal. Intelligent Agents represent an interesting approach with respect to providing an account for privacy as outlined by Contextual Integrity.

Firstly, an agent is able to reason autonomously towards achieving a certain goal. Across an extended network of devices as the IoT, privacy goals might be extremely different based on the environments reached by the IoT. This calls for modular capabilities, which allow users' to tweak their privacy agent to perform according to specific environments and privacy preferences. When employing autonomously reasoning agents, it is possible for these to be tweaked in such a manner that they perform optimally in the desired environment. As analysed in the literature review, such an approach is likely to produce a better outcome with respect to what a universal privacy regulation would be capable of achieving.

The second potential benefit of using an intelligent agent is that agents are able to communicate with other agents. This is a strong advantage over using a software expert system, which would have to be provided with all information in some way. Instead, the agent is able to collect additional information by communicating with other agents. Intuitively, this is a huge advantage, since in a IoT scenario, where different systems are connected with each other, the communication between these systems represents the most time effective and reliable source of information when a decision has to be taken. It is not to be expected that in a vast interconnected world, humans would be aware of any available information to be fed into decision-making agents.

Russel and Norvig present a variety of architectures for Intelligent Agents (IAs),

such as reflex agents, model-based agents, goal-based agents, utility-based agents, learning agents and knowledge-based agent [Russel and Norvig, 2009]. Russel and Norvig introduce the reflex agent as the most basic architecture, immediately after the table-drive agent, which simply relies on a table directly mapping percepts to actions. The two authors increasingly add complexity to the agent's architecture by presenting all different mentioned structures. These architectures have advantages and disadvantages, and even with a defined environment, the decision on which is the most suitable for a given approach is not always straightforward. That is because it might be difficult to establish the boundaries that define the architecture of the chosen agent. Nevertheless, as previously announced in this study, reasoning towards a privacy decision requires a representation and manipulation of the domain of interest, in this case a portion of the Internet of Things. For such purpose, a knowledge-based agent is the most appropriate choice. As Russel and Norvig describe, such agents manipulate a formal representation of knowledge in order to infer additional knowledge and hence reach a decision regarding the action to be taken [Russel and Norvig, 2009]. As the domain modelling plays a key role in the whole reasoning and decision-making process, an analysis of the environment to model shall be provided from an agent point of view..

The environment of the Internet of Things has been discussed in the introductory section, and among its features it has been recognized as being extensive, dynamic and heterogeneous [Kökciyan and Yolum, 2017]. From the perspective of an Intelligent Agent, the IoT environment has the following characteristics:

- partially observable: it is not possible for an agent to sense all the surrounding environment; For instance, Alice's agent could not extensively sense all available contexts for information, but only its surroundings.

- dynamic: it is possible for the environment to change while the agent is taking a decision. For instance, new information might be available regarding Alice's situation, however the current agent would be unaware of it unless communicating again with the other agents.

- uncertain: the environment tackled by the current research does not change with respect to agent's actions. Generally, Alice can freely act in the environment, hence new information might become available from nearby agents, regardless the sharing decision of Alice's agent. For this reason the environment is not deterministic.

Nevertheless, the environment is not stochastic either. That is because in the scope of this study, it is not dealt with the probabilities of each possible environment state.

The analysis of the IoT environment from an Intelligent Agent standpoint, can be linked to the main challenges of the IoT that have been covered in the introductory section. The scale of the Internet of Things in particular, relate to the feature of agent's environment of being partially observable. An agent, will be able to sense, in this case communicate, with a limited number of agents. Reachable agents might not provide an exhaustive representation of the world.

The second characteristic of agent's environment, that of being dynamic, implies that changes in the environment might occur while the agent is taking a decision. This seems perfectly in line with how Kökciyan and Yolum have interpreted the dynamic feature of the IoT. According to the authors, being dynamic indicates the uncontrolled interaction of entities [Kökciyan and Yolum, 2017]. As these entities are not centrally controlled, moreover unaware of each other, the environment is continuously subject to change and thus, new information regarding Alice might become available at any time.

The use of a knowledge-based agent was well-considered, since the use of an ontology of concepts is desired for investigating improvements of reasoning on contexts. Furthermore, by considering environment's analysis from an Intelligent Agent point of view, it becomes more clear that such an architecture is well suited for the purpose of this research.

Having an agent reasoning on its internal representation of the domain allows to address a partially observable environment. The agent can communicate with other agents and add information to its knowledge base, in such a way that it gathers the most information possible, adding to a richer representation of the world.

The agent developed in the scope of this work, has the goal of taking a privacy decision on behalf of Alice, by reasoning on its internal representation of the world and inferring Alice's situation. The following subsections shall clarify the representation of the world at hand and its use towards a privacy decision.

4.2 Domain Representation

In this subsection, the the world model as used in this research is presented. In particular, an ontology is used to define the structure of contexts, their relations with one another and with the individuals in the world. For this end, a hierarchy of contexts shall be

developed. Consequently semantics will be added by means of OWL tools such as object properties, class properties and individuals. In parallel, context inference and decision making shall be touched, before being fully covered in the Reasoning sub section.

4.2.1 Context Definition

In order to represent the world of interest it is first necessary to provide a hierarchy of relevant concepts of the world. The hierarchy used by the agent proposed in this research is shown below, in Figure 1.

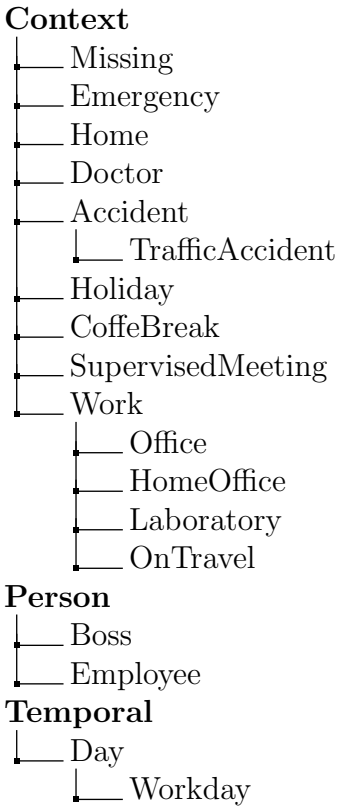

Figure 1: Concepts Hierarchy

As depicted in Figure 1, there are three essential concepts needed, the contexts themselves, people and days. As of now, such hierarchy has little to no semantics. This hierarchy on its own, only groups contexts, people and days, and additionally, it represents the fact that an employee is a type of Person, a traffic accident is a type of accident and so on.

By using the modelling tools available in OWL, it is possible to establish relations among different concepts of the hierarchy, and therefore to allow relevant inference. These tools are object properties, equivalence class expressions and sub-class expressions, just to name a few. Note that there are multiple ways in which such tools may provide semantics to the world. In this research, one specific way of establishing relationships between the concepts of the ontology is developed.

For instance, consider the following table grouping all individuals available in the ontology.

Class	Individual name
Workday	30
Employee	**Alice**
Employee	Tom
Boss	Bob
Boss	Jack

Table 1: Available Individuals

By looking at the concepts hierarchy in Figure 1 and the available individuals in Table 1, one can immediately notice that Table 1 does not contain any individual of a context Class. This is one peculiar aspect of the world modelling proposed in this study, and the reasons behind such choice shall become clear later in this subsection. With respect to Table 1, note that each individual can be put into relation with each other by means of object properties. For example, it is possible to assert that Alice is at work on the workday March 30st, that Bob is the superintendent on the workday March 30st, that Jack is not absent on the workday March 30st and that Tom has assigned as a supervisor Bob. This can be expressed by using the object properties atWork, isSuperIntendent, isAbsent and hasSupervisor. The following table gathers all defined object properties in the ontology at hand.

Property name	Feature	Domain- Range
emergency		Person - Day
missing		Person - Day
accident		Person - Day
trafficaccident		Person - Day
atDoctor		Person - Day
atHome		Person - Day
atWork		Person - Day
atHomeOffice		Person - Day
atNuclearLab		Person - Day
atOffice		Person - Day
fire		Person - Day
hasAllergy		Person - Day
hasBookedLab		Employee - Workday
hasSupervisor	TRANSITIVE	Person - Person
isAbsent		Person - Day
isSpring		Person - Day
isSuperIntendent	INVERSE FUNCTIONAL	Boss - Day
isSupervisedBy	FUNCTIONAL	Employee - Boss
isSupervisor		Boss - Day
isTalking		Person - Person
onCoffeBreak		Employee - Day
onHoliday		Person - Day
onleave		Person - Day
onTravel		Person - Day
shareFootage		SurveillanceCamera - Emergency
supervisedMeeting		Person - Day

Table 2: Defined Object Properties

Table 2 provides an overview of possible assertions in the modelled world. In other words, it shows what is possible to say about a certain person, be her an employee, a

boss or more generally, a person. Analysing the peculiarities of some of the available object properties,would only result clear after the reveal of the whole picture of the world representation created in this work.

While it is possible to assert a variety of facts about people, these assertions would merely put in relation a person and a day, and the only meaning left would be the name of the object property itself. In order to connect possible assertions to the context defined in Figure 1, equivalence class expressions are used.

Equivalence class expressions are properties of classes that allow the link between the class description of one class and the class description of another. For example, we might define the class Home to be equivalent to the set of all those individuals who are at home on some day. Hence, Home would be equivalent to *atHome some Day.* This allows to put into relationship individuals and contexts. More specifically, based on what information is known about a certain person, certain contexts' class descriptions are going to be matched, hence these contexts will inferred as a classes on the individual representing that person. Consider another example, that of the coffe break context. Being in a coffe break context is intuitively defined as that moment in which a person is talking to some person who is not a superintendent on that day. To represent such intuition in the ontology, the equivalence class expression *isTalking some (Person and not(isSuperIntendent value 30))* is set on the *CoffeBreak* context. In such a way, those individuals who are talking to some person who is not a superintendent on March the 30st, will be inferred as being in a *CoffeBreak* context. This way of representing contexts, respects the intuition of Contextual Integrity according to which it is possible for someone to be active in more than one contexts. If a person satisfies equivalence classes from different contexts, unless there is no contradictory information, these contexts will be inferred for that individual. For example, consider the individual Alice. Suppose that we know that Alice is at home (*alice: atHome value 30*), and that there has been a traffic accident on a road often take by Alice (*alice: trafficAccident value 30*). Both the *Home* and *TrafficAccident* classes will be deduced active on the individual *alice*. The following table gathers the definition of each context.

Context	Equivalence Class Expression
Missing	- (not (atHomeOffice value 30)) and (not (atOffice value 30)) and (not (onleave value 30)) - missing some Day
Emergency	- (atWork some (Day and (date$_d$*ayvalue*30)))*and* (fire some (Day and (dateday value 30))) - emergency some Day
Home	- atHome some Day
Accident	- accident some Day
TrafficAccident	- trafficAccident some Day
Holiday	- onHoliday some Day
CoffeBreak	- isTalking some (Person and (not (isSuperIntendent value 30))) - onCoffeBreak some day
SupervisedMeeting	- not (isSupervisedBy some (Boss and (isAbsent value 30))) - supervisedMeeting some Day
Work	- atWork some Day
Office	- atOffice some Day
HomeOffice	- (atHome some (Workday and (dateday value 30))) and (atHomeOffice some (Workday and (dateday value 30)))
Laboratory	- (hasBookedLab value 30) and (hasSupervisor value bob) - atNuclearLab some Day
OnTravel	- onTravel some Day

Table 3: Context and their equivalence class expressions

Note that this structure does not say anything on the extent to which these contexts are believed to be active. For instance, if we have the information that Alice is at home and the information that an accident occurred on a street often taken by Alice, even though both the *Home* and *TrafficAccident* contexts became active, there is no straightforward way to quantify the "activity level" of each context within the ontology. In this specific case, it would be intuitive to assume that Alice is most likely at home. Therefore the *Home* context should be believed active with a high degree of belief while the *TrafficAccident*

context should be considered active with a lower degree of belief. The way in which degrees of belief are used, and their effect on the decision making, will be covered in the agent section of this document.

The definition of contexts and their inference, represent one required aspect of current world model. As explained above, the taxonomy of contexts paired with equivalence class expressions and object properties, allows the the derivation of contexts on certain individuals of interest. For example, by means of the defined contexts and object properties, it would be possible to infer a variety of contexts on an individual representing Alice.

4.2.2 Share Mechanism

Another important aspect of the world model is the "sharing mechanism". Within the portion of IoT modelled in the current work, a share decision taken by Alice's agent, ideally is the consequence of having inferred information according to which Alice is in some sort of danger. In terms of contexts, there are contexts in which is intuitively more reasonable to trade privacy for safety, with respect to other contexts where such a trade-off would not be reasonable. In the scope of this study, the difference between contexts that encourage a share decision and contexts that encourage a not share decision, is implemented at the ontology level and it is further exploited by the agent in the argumentation phase. In particular, this implementation will be "translated" by the agent into positive or negative share norms, depending on whether a context represents a danger situation or not. Additionally, norms that link "dangerous" contexts to the *Emergency* context will be created by the agent. But how is the structure that allows the agent to create such norms represented in the ontology?

Referring to Table 2, it is easy to identify an exception w.r.t all possible assertions that lead to context inferences. The exception is the *shareFootage* object property. This object property does not play a crucial role in the reasoning phase with respect to Alice's context, since it will not lead to the inference of contexts. In fact, as evident from Table 3, the *shareFootage* property does not appear in any context's class definition. The *shareFootage* property has the goal of establishing, at the ontology level, what information would potentially have to be shared. As shown in Table 2, the range of the *shareFootage* property is the *Emergency* context, hence in the scope of this research, the agent will share Alice's footage if it has strong enough evidence of an emergency situation, with

respect to the rest of active contexts.

In the current model of the world, there are two contexts that can lead to an emergency situation, the *TrafficAccident* and *Missing* contexts. These two concepts are defined to be a sub-class of the expression *emergency some Day.*. Since the *Emergency* class is equivalent to *emergency some Day,* as shown in Table 3, the *Missing* and *TrafficAccident* classes are subclasses of the *Emergency* class. As covered in the background section dedicated to OWL, this implies that a the two contexts are also emergency contexts, but not the contrary. Therefore, an emergency context is not necessarily a missing context, in fact, it could be a traffic accident context as well. The subclass property of classes, allows to express that multiple situations could potentially represent an emergency for Alice. This would not be possible by means of equivalence class expressions, as these would force the two linked classes to have the same individuals, which might lead to inconsistencies in the ontology.

In summary, the interaction between the *shareFootage* property and the *Emergency* context is not strictly relevant for the inference of new contexts, but it is employed by the agent in order to create relevant norms. These norms will encode which contexts lead to an emergency contexts, in the case of this study, *TrafficAccident* and *Missing*. In other words, an Emergency context will not be inferred by the ontology for an individual *alice*. Rather, Alice could be inferred to be in a traffic accident or missing. Such information will further trigger an emergency context in the argumentation phase, according to the norms that have been generated by using the above described "share mechanism". For the sake of clarity, this section has only covered the ontological point of view, while the creation and triggering of norms will be explained in the section dedicated to the agent.

4.3 Reasoning

The reasoning performed by the knowledge-based agent proposed in this study can be seen as a two-step structured reasoning. First, the agent will consult its internal representation of the world with a series of information it has been provided as input, and second, it will attempt to take the best decision regarding the action to perform. The first step is highly dependent on agent's ontology of concepts and inference making, whereas the second step is achieved by means of argumentation.

4.3.1 Context Inference

What has been described in the previous subsection is the definition of contexts as used in the scope of this research. While part of the possible knowledge inference has been revealed, namely the inference of contexts for certain individuals, the main purpose of the previous subsection was to explain the underlying structure that allows such reasoning. In this part, an overview of the inference performed on such structure shall be provided.

Reasoning on concepts, relationships and individuals expressed in terms of OWL, is possible thanks to so called semantic reasoners. Reasoners have the purpose to derive knowledge which is not explicitly asserted in the ontology. Modern reasoners allow reasoning on concepts, relationships and individuals as expressed in the OWL language, and can even take into consideration rules expressed in First Order Logic (FOL). Nevertheless, in the scope of this research the use of FOL rules is not considered, as this would require additional knowledge about the domain. Moreover decidability, namely the capability of computing inferences in a finite time, would not be ensured [Smith et al., 2004]. The focus is put on exploiting the OWL DL specie of OWL as described in the OWL documentation, which provides both maximum expressivity and computational advantages such as computational completeness and decidability.

In specific, context inference benefit from OWL added semantics, as covered in the technical subsection on OWL. The features of object properties, such as transitive, functional and inverse functional are investigated as part of the core of the proposed research approach. Moreover, the subClassOf and equivalentTo class properties are also useful towards context inference as achieved in this work. In other words, by using an ontology modelled with OWL, it is possible for the reasoner to infer otherwise unknown information. The following example shall illustrate how such inference is achieved.

Consider the context *CoffeBreak* in Figure 1, and its class description definitions in Table 3. Consider the first one, *isTalking some (Person and (not (isSuperIntendent value 30))*. As previously mentioned, this expression intuitively means that an individual is on a coffe break context if he is talking to a person who is not the superintendent on that day. Suppose that it is known that Alice is talking to Jack, and that Bob is the superintendent on that day. Note that since the fact that Jack is not a superintendent on March the 30st, the equivalence class expression is not satisfied by Alice in a straightforward way. However, consider that the *isSuperIntendent* property is set to be inverse functional and that Bob and Jack are different people. Since Bob is the superintendent and *isSuperIntendent* is

inverse functional, no other person, different from Bob, can be a superintendent on the same day. The *CoffeBreak* context thus, is inferred for the individual *alice*.

While this research will not go into the details of the algorithms used by the reasoner, it is relevant to mentioned that HermiT is the semantic reasoner adopted in this research. HermiT has its own API available for use in Java.

4.3.2 Domain Representation for Decision Making

Previously in this section, it has been explained that by means of an ontology it is possible to perform part of the reasoning, specifically it is possible to infer which contexts may be active at the structural level of the ontology.

The representation of the world in terms of an ontology of concepts, introduces benefits with respect to the decision making phase as well. Such domain representation is exploited by the agent, in order to automatically create the domain knowledge as needed by the argumentation engine. In particular, the ontology holds almost all necessary information required in the decision making phase. Among others, it contains the collected information regarding Alice and other people in the world, in the form of object property assertions set on each individual; it contains the the contexts that have been inferred for Alice in the form of classes; finally, the ontology contains the definitions of contexts in terms of equivalence class expressions. This information shall be transposed into the correct syntax by the agent.

For instance, the contexts that have been inferred for Alice, need to become "Prolog-like" predicates in order to be suitable for the argumentation engine. For example, the information that Alice is in a *HomeOffice* context on day *30*, which in the ontology is express by means of the class *HomeOffice* as an inferred type for Alice, shall become a predicate of the form `atHomeOffice(alice, 30).`

Similarly, equivalence classes describing contexts, shall become "Prolog-like" norms, suitable for the argumentation engine. Take for example the equivalence class *atOffice some Day*, which defines the *Office* context. The equivalent "Prolog-like" norm would be `inContext(A, Office, T) <- atOffice(A,T).`

Finally, all information available about Alice and the other people in the domain, such as Tom, Bob and Jack, in the form of object property assertions, need to be parsed into "Prolog-like" predicates as well.

Note that in order to have a meaninful decision making with respect to active con-

texts, a quantification of the parsed information is required. Consider that Alice might find herself in more contexts at the same time, hence the need of differentiating between the level of "activity " of each context. Such quantification is done by the agent in the process of transposing all information available into an argumentation-syntax friendly specification, referred to as the domain knowledge.

As in the work of Kökciyan and Yolum, degrees of belief of values ranging in (1.0] are assigned to each information going into the argumentation engine [Kökciyan and Yolum, 2017]. For example, if the ontology has inferred that Alice is in a *HomeOffice* context, this information should be believed with a certain degree. By default, the agent will assign a value of 0.5.

More concisely, the representation of the domain by means of an ontology of concepts, offers a natural structure from which to extract the information required in the decision making phase. With the parsed domain representation, quantified by means of degrees of belief, it is possible to perform decision making by means of the argumentation engine. The achieved domain specification used for argumentation, shall be discussed in details in the Domain Knowledge subsection of the chapter dedicated to the agent.

5 PARCo

In the previous sections the agent has only been referred to, in order to provide context to the explanation of the relevant methods used in this research, such as an ontology of concepts and argumentation. This section shall cover in more details the internal design and functioning of the proposed agent, as well as to present its contribution with respect to the decision-making phase.

5.1 Agent Design

The current agent is required to group all the reasoning and decision making elements, namely the ontology of concepts and the ASPIC engine for decision making. In addition, it is expected to provide improvements possibilities with respect to the decision-making process. The knowledge-based agent architecture has been chosen as the most appropriate for the current system. This sub-section shall take an overview perspective on the agent, by analysing its internal structure. The purpose is to clarify where the contributions of this research find their place within the agent, as well as to present the contribution of the agent-approach itself.

As the agent at hand is a software agent, an intuitive way to concisely capture its structure is a flowchart. In Figure 1 the arrangement of the proposed agent is shown as a flowchart. As a remainder, flowcharts use rectangles to indicate processes, parallelograms to indicate input or output, and diamonds to indicate decisions.

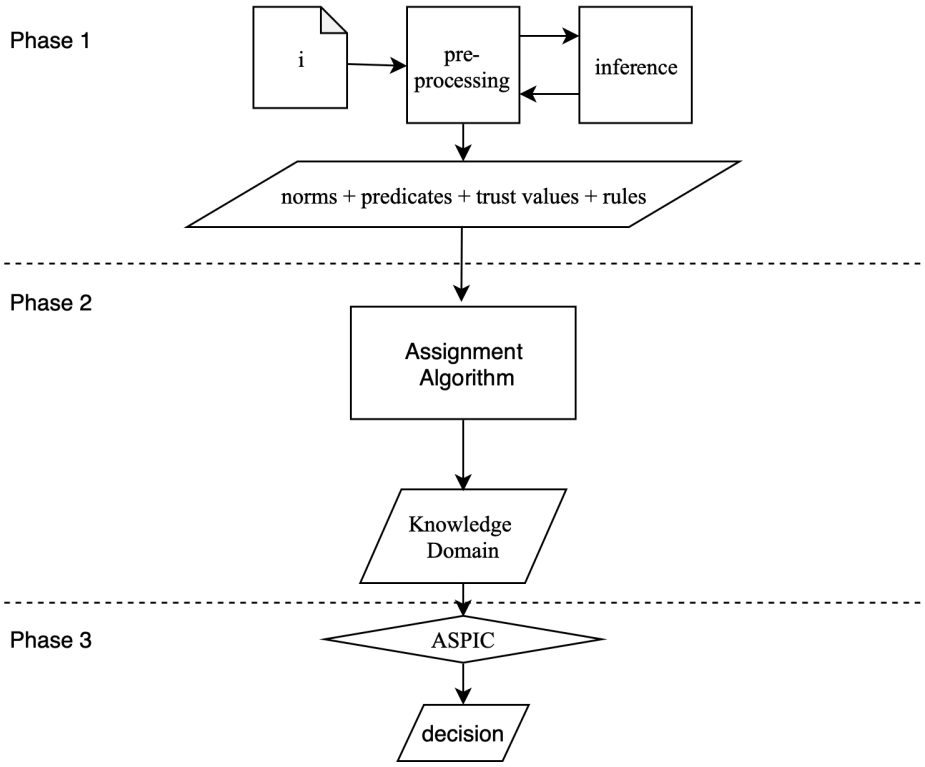

Figure 2: Agent's Architecture - Flowchart

For simplicity of explanation, it is convenient to divide the flowchart in three phases:

i. inference and information gathering

ii. improve knowledge domain for decision making

iii. decision making

The first phase is concerned with dealing with the input information and communicating with the ontology of concepts, in order to infer new knowledge and to create the context norms. As a result of the first phase, all necessary information required by the knowledge domain should be available.

The input file contains three pieces of information necessary for the agent: rules that regulate the degrees of belief of information predicates coming from other agents, trust values for other agents and input information predicates coming from consulted agents. The input file is a compact way to allow the execution of the proposed agent, nevertheless,

ideally the agent would receive information predicates directly by communicating with other agents.

Once the input has been read by the pre-processing block, this needs to communicate with the ontology for the inference step. First, the pre-processing process will pass information predicates to the ontology. Consequently, in the inference process, the reasoner will infer new knowledge which will be reconstructed as information predicates by the pre-processing block. The pre-processing part of the agent is also responsible for creating the context norms. For such a purpose, information available in the ontology is required. Specifically, equivalent class expression available on each contexts are used by the agent to create contextual norms. Once norms and inferred information predicates have been prepared, it is time for the agent to decide how such information should be manipulated before being pushed to the knowledge domain. This occurs in the second phase, where the Assignment Algorithm is shown to use a bias to alter available information before outputting it in the knowledge domain. The Assignment Algorithm is covered in details in section 5.

Finally, in Phase 3 the argumentation engine issues a decision based on the knowledge domain.

5.2 Domain Knowledge

The domain knowledge is the product of agent's reasoning of Phase 1, and it represents the input for the argumentation engine towards the final decision. It is therefore relevant to cover its content and syntax in more details.

As anticipated, the content of the domain knowledge conceptually split into four categories. The first category includes "Prolog-like" rules that manage the degree of belief of each information. The rules are provided in Table 4.

dob processing rule	dob
info(X) $\leftarrow says(A, X, B)$	0.9.
~d(A,X) $\leftarrow mydob(A, info(X), V3), <(V3, 0.7)$.	
info(X) $\leftarrow$ ~$d(A, X)$	0.2.
mydob(A,info(X),V3) $\leftarrow trust(A, V1), says(A, X, V2), mult(V1, V2, V3)$.	
mult(V1,V2,V3) $\leftarrow is(V3, *(V1, V2))$.	

Table 4: Domain Knowledge - dob processing rules

These rules are part of the model proposed by Kökciyan and Yolum, and in the scope of this research they are not modified or further elaborated [Kökciyan and Yolum, 2017]. Nevertheless, an intuitive explanation shall be provided for the sake of clarity.

Intuitively the purpose of these rules is to establish the way arguments are created within the argumentation framework, according to degrees of belief. For instance, the first rule `info(X) <- says(A, X, B)`, would tell the argumentation engine to create an argument of the form info(X) for X, whenever X has been "said" by agent A at time B. This newly created argument, might have counter arguments that can be created by the last 4 rules in Table 4. Before analysing what such rules do, it is relevant to consider Table 5 containing trust values for each external agent, as these values are used in the fourth rule of Table 4, and represent the second element of the domain knowledge.

Trust value
trust(sensor,0.9).
trust(boss,0.3).
trust(police,0.9).

Table 5: Domain Knowledge - Trust Values

These trust values are employed in the fourth rule together with the fifth rule of Table 5, in order to create an argument of type mydob(...). This argument will then trigger the second rule. Such rule however, has a condition "<(V3,0.7)". This condition

is basically stating that the head of that rule is true, if the degree of belief of the triggering predicate mydob(...) is less than 0.7. In such, a case the predicate ~d(A,X) will be true and it will trigger the third rule. This rule will create an argument for the information X with a lowe degree of belief, precisely 0.2.The last two elements useful for the domain knowledge are norms and information predicates specifying the scenario within which the final decision should be taken.

The current research deals with two types of norms. Contextual norms that allow the argumentation engine to create shareFootage types of arguments, and contextual norms that allow the engine to create inContext and info type of arguments. A glance at the first type is provided in Table 6.

shareFootage norm	dob
shareFootage(A, T) $\leftarrow inContext(A, emergency, T)$	0.9.
~shareFootage(A, T) $\leftarrow inContext(A, work, T)$	0.9.
~shareFootage(A, T) $\leftarrow inContext(A, office, T)$	0.9.
~shareFootage(A, T) $\leftarrow inContext(A, homeoffice, T)$	0.9.
~shareFootage(A, T) $\leftarrow inContext(A, ontravel, T)$	0.9.
. . .	...

Table 6: Domain Knowledge - (Example) shareFootage norms

As mentioned in the domain representation section, these norms are responsible for providing the ultimate arguments towards a privacy decision. That is the case because by their means, shareFootage arguments of the form shareFootage(A,T) are built. Since that is what the proposed agent is trying to provide a decision about, the shareFootage argument with its state will represent the answer. The table shows only a portion of the shareFootage norms generated according to the modelled domain, which show that a positive share argument is built when there are arguments sustaining that Alice is in an emergency context.

These norms are created by the agent by accessing the ontology, specifically the *shareFootage* object property, and based on the range of such property, the agent will decide whether to create a positive or negative shareFootage norm.

The second type of norms are the so called contextual norms. These norms allow the argumentation engine to create arguments related to the context of Alice. Table 6 groups a selection of the norms generated from the world model use in this research.

Contextual norm	dob
inContext(A, home, T) $\leftarrow info(atHome(A,T))$	0.9.
inContext(A, emergency, T) $\leftarrow info(emergency(A,T))$	0.9.
inContext(A, holiday, T) $\leftarrow info(onHoliday(A,T))$	0.9.
inContext(A, doctor, T) $\leftarrow info(atDoctor(A,T))$	0.9.
inContext(A, ontravel, T) $\leftarrow info(onTravel(A,T))$	0.9.
inContext(A, office, T) $\leftarrow info(atOffice(A,T))$	0.9.
. . .	...
info(missing(A, T))$\leftarrow info(\text{~}atHomeOffice(A,T)), info(\text{~}atOffice(A,T))$, info(~onleave(A, T)).	0.9
info(emergency(A, T)) $\leftarrow info(trafficaccident(A,T))$	0.9.
info(doctor(A, T)) $\leftarrow info(hasAllergy(A,T)), info(isSpring(A,T))$	0.9.
info(coffebreak(A, T)) $\leftarrow info(isTalking(A,B)), info(\text{~}isSuperIntendent(B,T))$	0.9.
. . .	..

Table 7: Domain knowledge - (Example) Contextual Norms

Table 6, divides the contextual norms according to the argument that they sustain. As you can see, the first group of contextual norms has as heads, predicates of the form inContext(A, context, T) and as bodies, a predicate of form info(inContext(A,T)). Suppose that there is available information that Alice is at home on March the 30st with a degree of belief of 0.8, expressed as info(atHome(alice,30)) 0.8. In such a case, the first contextual norm in Table 6 would trigger, creating the argument inContext(alice, home, 30) 0.8. Such argument will have as premises the contextual norm that has been triggered which has a dob of 0.9, and the information predicate, which has a dob of 0.8.

As you can notice, all contextual norms have a degree of belief of 0.9. Since the dob is smaller than 1, these rules are still defeasible, but offer a strong support to the created arguments. Having a 0.9 dob for all contextual norms, intuitively means that the

support given to an argument by the premise consisting of its creator norm, is strong. It is important to notice that, since every context argument is the result of a norm with dob 0.9, every context argument will have a strong premise with the high dob of 0.9. Nevertheless, the premise arguments which triggered the norms for different context, have different degrees of belief, hence, they provide different support to the argument created by the norm.

Table 8 gathers an example of possible information predicates necessary for the domain knowledge, which shall become arguments when processed by the argumentation engine.

Information Predicate	dob
info(atWork(alice, 30))	0.5.
info(~atHomeOffice(alice, 30))	0.8.
info(~onleave(alice, 30))	0.8.
info(workday(30)).	
info(~atHome(alice, 30))	0.6.
info(trafficaccident(alice, 30))	0.6.
. . .	...

Table 8: Domain Knowledge - (Example) Scenario specification

5.3 Agent Algorithm

As described in the previous subsection, the proposed agent covers the entire decision making process, from the input of gathered information, to the output of the final share decision. Agent's architecture described in Section 5.1 illustrates the different modules necessary in such process. While these components, namely the ontology and the argumentation engine, are fundamental building blocks used in this study, it is more compelling to consider the way in which they perform their function in the overall agent. This is regulated by the main algorithm of the agent.

Algorithm 1, as proposed below, illustrates the most relevant procedures and data structures underlying the functioning of the agent.

Algorithm 1: decide(q, i, o)

Input: q: initial query, o: ontology, i: input file

Output: d: decision without support, d2: decision with support

Data: dobRules: set of rules regulating dobs' computation

Data: tSet: set of trust predicates

Data: pSet: set of predicates specifying a scenario

Data: cIndividuals: set of contexts, each with the relative list of inferred individuals

Data: nSet: set of norms

```
1  dobRules ← readDobRules(i)
2  tSet ← readTrustValues(i)
3  pSet ← readPredicates(i)
4  updateOntology(pSet, o)
5  cIndividuals ← computeInferences(o)
6  pSet ← createPredicates(cIndividuals)
7  nSet ← createNorms(o)
8  w ← writeToWorldSpec(dobRules, nSet, tSet, pSet)
9  d ← argumentationEngine(q, w)
10 d2 ← decideWithSupport(q, pSet, cIndividuals)
11 return [d, d2]
```

The aforementioned algorithm broadly encompasses agent's behaviour. As a consequence, Algorithm 1 includes a variety of input, output and local data structures needed for evaluating the final return value, the privacy decision. For clarity, let us explain Algorithm 1 by means of an example. Consider the following scenario:

As long as the input is concerned, assume the input query requires the agent to decide whether the footage should be shared or not. As the query **q** must be passed to the argumentation engine, it must be expressed in a predicate form, thus **q** is equal to *shareFootage(alice,30)*. Assume there is an available ontology **o** conceptualising the domain. Finally assume the file **i** contains the rules according to which the degree of belief are managed in relation to trust, trust values of each agent and the specification of the scenario of interest in terms of predicates. As mentioned in the Methodology section,

trust management will not be further elaborated in this research, thus let us focus on the scenario specification. Consider the following basic input scenario, consisting of the following two information predicates:

Basic Scenario
`info(atOffice(alice, 30)) 0.5.`
`info(atDoctor(alice, 30)) 0.5.`

Consider the data structures used by Algorithm 1 to reach its decision. *tSet* is the set containing the trust values for each agent. *dobRules* is the set containing those rules which compute the dob of information predicates coming from other agents. These rules and the trust values will be kept general, as they are not the main focus. Another data structure is *pSet*. This the set of gathered information predicates. *nSet* is the set containing the norms generated by the agent by reading the ontology domain representation. Finally, *cIndividuals* is a map that relates each context to the individuals that are inferred to be in that context. These data structures are empty before the execution of Algorithm 1.

tSet = [ ]
dobRules = [ ]
pSet = [ ]
nSet = [ ]
cIndividuals = { }

The algorithm starts with three simple lines, which read the input file and store the information in the correct data structures.

tSet = [tValue1, tValue2, ...]
dobRules = [dobRule1, dobRule2, ...]
pSet = [info(atOffice(alice, 30)) 0.5, info(atDoctor(alice, 30)) 0.5]
nSet = [ ]
cIndividuals = { }

Once the input predicates have been retrieved from the input file, they need to be inserted in the ontology to allow correct inference. This is achieved by the *updateOntology* method on line 4. This method takes each predicate at a time and it checks what is the related object property in the ontology. Furthermore, it checks which are the individuals which it puts in relation. For example for the first information predicate, the method will look for an object property *atOffice* and for individuals *alice* and *30*. The *updateOntology* method finally inserts the predicate in the ontology as a positive or negative assertion axiom depending on the predicate itself.

Once relevant input information has been processed and has reached the ontology, it is the moment to exploit ontology's inference capabilities. In particular, on line 5 of Algorithm 1, the *computeInferences* method is illustrated to populate the *cIndividuals* data structure which as previously described, contains the set of contexts, each with the relative set of inferred individuals. For example, *cIndividuals* could contain:

tSet	= [tValue1, tValue2, ...]
dobRules	= [dobRule1, dobRule2, ...]
pSet	= [info(atOffice(alice, 30)) 0.5, info(atDoctor(alice, 30)) 0.5]
nSet	= [ ]
cIndividuals	= { (Office, [alice]), (Doctor, [alice]), (Work, [alice]), ...}

Note that in this example, the ontology has inferred that Alice is in a work context by using its representation of the domain, and the newly inserted information at line 4 according to which Alice is at her office.

The map *cIndividuals*, which now contains inferred information, is consequently employed by the *createPredicates* method in order to build the new predicates and hence update the predicate set *pSet*.

tSet	= [tValue1, tValue2, ...]
dobRules	= [dobRule1, dobRule2, ...]
pSet	= [info(atOffice(alice, 30)) 0.9, info(atDoctor(alice, 30)) 0.7, info(atWork(alice,30)) 0.5]
nSet	= [ ]
cIndividuals	= { (Office, [alice]), (Doctor, [alice]), (Work, [alice]), ...}

As you can notice, at this point there is no manipulation of the degrees of belief, thus the newly available predicate has a default dob of 0.5.

The last step towards having a complete *knowledge domain* suitable for the argumentation engine, is providing the contextual norms. The method *createNorms* on line 7 of Algorithm 1 is responsible for that. Such method is strictly tight to the syntax of the equivalence class expressions available on each context present in the ontology. The method reads such expressions and builds the appropriate norms. Some examples are:

- simple equivalence class expression: *atWork some Day*

  created **norm1**: `inContext(A, work, T) ← info(atWork(A, T)) 0.9.`

- aggregate equivalence class expression: *(not (atHomeOffice value 30)) and (not (atOffice value 30)) and (not (onleave value 30))*

  created **norm2**: `info(missing(A, T)) ← info( atHomeOffice(A, T)), info( atOffice(A, T)), info( onleave(A, T)) 0.9.`

  These norms are inserted in nSet immediately after creation.

tSet	= [tValue1, tValue2, ...]
dobRules	= [dobRule1, dobRule2, ...]
pSet	= [info(atOffice(alice, 30)) 0.9, info(atDoctor(alice, 30)) 0.7, info(atWork(alice,30)) 0.5]
nSet	= [norm1, norm2, ...]
cIndividuals	= { (Office, [alice]), (Doctor, [alice]), (Work, [alice]), ...}

Extending the domain knowledge ontology, but also representing different domains, might require different equivalence class expressions and would thus require the *createNorms* method to be adapted or extended to new syntax.

On line 8, Algorithm 1 groups all available information in a file representing the *knowledge domain* as required by the argumentation engine. As explained in the previous section, the argumentation engine has the duty of computing the strongest predicate and hence return the result of the initial query in *d*. This is represented on line 9 with the *argumentationEngine* method, which will return a share or not share decision in d.

Line 10 of Algorithm 1, shows an interesting possibility given by the agent approach proposed in this research. The decision reached in line 9 by calling the argumentation engine, has been evaluated by means of a *knowledge domain* directly built according to the inferences provided by the ontology. That means that, as mentioned previously, inferred predicates have been assigned the default dob value of 0.5. Nevertheless, there are multiple ways in which these inference knowledge may be processed before passed to the argumentation engine. One way is to manipulate the degree of belief of inferred predicates, based on certain metrics. This is the chosen path in this study, and it is represented on line 10 with the *decideWithSupport* algorithm call. Oberseve that *decideWithSupport*, differently from the rest of Algorithm 1's lines, is an algorithm on itself. In this way, the agent is modular with respect to the treatment of the *knowledge domain*. Hence, in presence of different domains, this algorithm may be very well swapped with a completely different algorithm for processing the inferences. The *decideWithSupport* implementation used in the current research, is provided in the Assignment Algorithm, shown bellow.

Algorithm 2: assignmentAlgorithm(q, pSet, cIndividuals)

Input: q: query, pSet: predicates, nSet: norms, cIndividuals: inferred individuals for each context, sSet: support values array

Output: d: decision

Data: pUSet: predicates to be updated

```
1 foreach context in cIndividuals do
2     foreach individual in context do
3         pUSet ← getPredicateToUpdate(pSet, individual, context)
4 pSet← updateSupportOfPredicates(pUSet, sSet)
5 w ← writeToWorldSpec(pSet)
6 d ← argumentationEngine(q, w)
7 return d
```

In a nutshell, the Assignment Algorithm considers the inferred contexts of individuals, updates the degrees of belief of predicates asserting such contexts, takes a decision based on the updated *knowledge domain* and returns the decision to Algorithm 1.

For its purpose, the Assignment Algorithm receives as input the query q, the set of predicates *pSet* and the *cIndividuals* map from Algorithm 1. In addition, an array of support values *sSet* is necessary. For simplicity assume that the support array contains static values. In a more advanced algorithm however, these could change based on many factors. For example, support values might change based on the subset of contexts where Alice has been inferred to be.

From line one to three, for each context and for each individual in the map of inferences *cIndividuals*, the predicates to be updated are loocked for in *pSet* and inserted in *pUSet*. As *cIndividuals* contains each context as an index and the relative inferred individuals as a value, the first loop iterates through each context, while the second through each individual inferred to be in that context. At each inner iteration, the method *getPredicateToUpdate* looks up the predicate set *pSet*, in order to find that predicate relating *individual* and *context*. If such predicate is found, by knowing that it relates *individual* and *context*, it is also known that is inferred by the ontology, and it is therefore desirable for update, hence inserted into *pUSet*.

Not all predicates are updated; only those who represent a direct context assertion.

For example, *info(atDoctor(alice,30))* would be updated, whereas *info(hasAllergy(alice,30))* would not, since the later does not alone trigger a context norm. That is because in the proposed implementation, the bias can be understood as the amount of support that is given to a certain context, when Alice is inferred to be in such context. Following Assignment Algorithm's clarification, one possible support instance, more specifically the one used in the ongoing study is provided.

The Assignment Algorithm takes care of the update itself on line 4, by calling the method *updateSupportOfPredicates*, which based on the predicate head picks the correct support from the *sSet* array. Finally, the newly updated list of predicates *pSet* is substituted with the old predicates in the the knowledge domain on line 5, and the argumentation engine produces a new decision on line 6, before returning it to Algorithm 1 on line 7.

Context	Support Value
Missing	0.8
Work	0.5
Traffic Accident	0.6
Doctor	0.7
Office	0.9
Laboratory	0.9
Travel	0.4
Holiday	0.7
Coffe Break	0.9
Supervised Meeting	0.9

Table 9: Assignment Algorithm - Bias values

Note that this table does not have the purpose of providing an universally applicable bias. On the contrary, its purpose is only auxiliary to showing the improvements with respect to decision making in an environment modelled as described in the previous section. Its values might therefore change from environment to environment. Even more abstractly, the application of a bias to manipulate degrees of belief might be altogether substituted with a more sophisticated approach towards the same purpose. That being

said, an intuitive interpretation of the currently used values is provided:

- Work = 0.5: the work assertion is quite weak in this domain and it merely means that Alice is supposed to be at work. That is because there are more specific work related contexts such as Office, Laboratory, Coffe Break and Supervised Meeting which deserve a much higher dob.
- Missing = 0.8:
- Traffic Accident = 0.6: Alice does not usually have accidents but is safe to keep medium high probability
- Doctor = 0.4: Alice does rarely go to the medic
- Office = 0.9: this assertions is precise and when asserted, it is expected that Alice is at the Office.
- Laboratory = 0.9: when asserted, there is high probability of being at the Laboratory; also because booking the Laboratory and a supervisor are required.
- Travel = 0.4: when asserted, it is poorly supported as there are multiple reasons not to be on travel.
- Holiday = 0.7: usually holidays are announced or known beforehand, nevertheless there reasons not to actually go on holiday.
- Coffe Break = 0.9: in the modelled domain, a Coffe Break is a context that is intuitively supported.
- Supervised Meeting = 0.9: in the work domain a supervised meeting is also well supported if information regarding such context is available.

6 Evaluation and Results

In this section, the behaviour of the proposed agent will be tested and analysed. The chosen evaluation method consists of executing the agent in a series of context scenarios, which are meant to expose its capabilities of reasoning on contexts and reaching a decision, and to show that such decision is in line with human intuition with respect to privacy. The selection of such scenarios is therefore central to a meaningful evaluation. The criteria followed in the selection are covered in the following subsection. Moreover, an assessment of the chosen scenarios, through a guided interview, is provided. Furthermore, prior to the context scenarios, this section also provides clarification regarding the meaning and use of the *missing* predicate in the scope of this research, since this predicate has decisive weight. Finally, the results achieved by the proposed agent are given in each context scenario at a time.

6.1 Scenarios Selection

In the scope of this research, context scenarios serve the purpose of evaluating the proposed agent. This subsection aims at clarifying the criteria followed in the selection of such scenarios.

From a high-level perspective, the overall goal of this research was to provide a more flexible computational account for privacy in challenging environments, with respect to what has been proposed to date. Such goal has been tackled by first, analysing available work conducted so far, and subsequently, by focusing on improvable aspects of the work done by Kökciyan and Yolum. The first contribution was to define the domain by means of an ontology in order to represent contexts and their definition and to capture their relations. A knowledge-based agent has been proposed as the main architecture of an agent that should exploit the representation of contexts, in order to achieve a more fine grained input for the decision making process. Additionally, the proposed agent aimed at improving decision making, by using an algorithm that manipulates the degrees of belief of available information.

There are thus three main aspects of the proposed agent, which need to be evaluated. The first aspect consists of the influence which the definition of contexts, has on the decision making process. Ideally, by reasoning on the ontology of contexts it would be possible to reach decisions otherwise unreachable. The second aspect is concerned with

the agent approach itself. In particular, it is desirable to assess whether an algorithm for manipulating the degree of believes has an impact on the decision making process. The third aspect to be evaluated, consists of the validity of the achieved decisions. One important benchmark that is considered in this research is the human intuition. Thus, agent's decisions should be in line with what humans would most likely decide in the same situation. This would show that the proposed agent behaves meaningully with respect to the real world.

In order for the three aspects to be evaluated by means of the scenarios, the following three criteria were taken into consideration at the time of scenario's creation:

- exhibit inference possibilities of the ontology of contexts
- exhibit the different influence of using or not using an algorithm for the manipulation of degrees of belief, on the decision making
- be decidable by human intuition, i.e, having a benchmark decision (see Section 6.2)

In other words, scenarios should in first place activate certain contexts by triggering contextual norms. Moreover, some of the scenarios should provide some specific information, in such a way that the features of object properties such as transitive, functional and inverse functional allow the ontology to infer otherwise unreachable knowledge. For example, suppose that Alice is considered to be in a Laboratory context if she is supervised by Bob, and that such information is not explicitly available. Also suppose that the *hasSupervisor* object property is transitive, and that the scenario has available information of Alice being supervised by Tom, and Tom being supervised by Bob. Such a scenario would satisfy the first criterion as it provides necessary information in order to show the capabilities of the proposed representation of contexts.

The second criterion is concerned with showing one of the advantages of the agent approach. As mentioned previously, scenarios should show the value of an approach to manipulating the domain knowledge, w.r.t. the decision making process. In the case of this study, the algorithm that manages degrees of belief of available information is the target of this criterion. In order show such benefits, it is assumed that the default information pieces provided by scenarios, have not the same degrees of belief as the ones assigned by the algorithm. Such coincidence would not have an influence on the decision making, rather it would only constitute a more straightforward scenario to manage. In

other words, the value of an agent would be evident if the decision-making process would be influenced by the algorithm that manipulates the degrees of belief.

While the first two criteria are concerned with exploiting the capabilities of the proposed model, the third one aims at providing a way to evaluate whether the decision reached by the model is meaningful. Since the chosen benchmark is human intuition, scenarios must be realistic enough to lead human intuition towards a certain decision. In other words, a trade-off between intuitively related information and unrelated information should be taken into account. In such a way, it would be possible to achieve a secure enough benchmark. i.e. golden result, against which to compare agent's decision. The assessment of the golden result shall be discussed in the next subsection.

6.2 Scenarios Assessment

Six scenarios were chosen for the evaluation phase, according to the criteria covered in the previous subsection. As explained above, scenarios need first, to exhibit agent's capabilities in terms of reasoning, and second, to depict situations which are addressable by human intuition. That is the case because by having the human intuition as a term of comparison, it is possible in the evaluation phase to quantify how well the agent performed. In particular, decisions taken by the agent would ideally be in line with the golden result represented by the human intuition. This would suggest both that the world has been modelled in a realistic manner and that the agent is able to take relevant decisions in the environment at hand. This subsection in particular, aims at clarifying how the golden result for each scenario was achieved and what the human intuition in such scenarios actually says.

A guided interview was conducted in order to asses, for each selected scenario, what humans would have had decided if they were to decide in place of Alice's agent. The majority of decisions for each scenario will determine the golden result for the relative scenario. The golden result will be compared to agent's decisions in the Results subsection, in order to validate whether agent's decisions can be considered relevant or not. The content of the guided interview is enclosed in the appendices, and it unfolded according to the following steps:

1. acquaintance with the IoT concept

2. introduction of the world at T0, when Bob is requesting the footage from the surveillance camera with the motive that Alice might be in danger

3. explanation of what a share or not share decision implies (privacy/safety trade-off)

4. introduction of the rules according to which knowledge is modelled in the domain and which should be used to deduce information in combination with the information of the scenarios

5. participants choice and questions

Six participants, three males and three females, took part in the interviews. All participants were in their twenties and their background included: Mechanical Engineering, Law, Philosophy, Artificial intelligence and Economics. Two out of six participants were employees while the rest were pursuing their higher-level studies. This sample is not exhaustive with respect to every intuition of privacy present across different generations, cultural contexts or fields. Nevertheless, the relatively young sample can give meaningful insights on the current human intuition of privacy in the proposed scenarios. That is because the chosen scenarios match sample's common knowledge and professional environment.

Participants were free to ask any kind of information that would allow a better understanding of the domain modelling, in other words, how the modelled world works. Once interviewees concluded the task of deciding whether to share or not to share for each of the six scenarios, they were asked if the domain rules were useful in assessing their answers. All participants were positive in this respect. While in the scope of this test it was not possible to measure to what extent the rules have been used, it is interesting to notice that almost all participants showed some doubt on whether to share or not to share, especially in scenarios which provided higher volume of information. This suggests that participants had to rely on their intuition while making the decision, as by simply applying the domain rules to the available information, a final decision could not distinctly be reached. This is especially relevant with respect to the last two scenarios, which provide extended enough set of information to challenge the human intuition on whether Alice might actually be in danger.

Once the interviews were finalised the results were gathered Table 10 below:

Person / Scenario	1 (M)	2(F)	3(M)	4(M)	5(F)	6(F)	**Majority**
1	s	s	s	s	s	ns	s
2	ns	ns	ns	ns	ns	s	ns
3	ns	ns	s	s	ns	ns	ns
4	s	s	s	ns	ns	ns	ns
5	ns	ns	ns	ns	ns	ns	ns
6	ns	ns	ns	ns	ns	ns	ns

Table 10: Decisions according to human intuition

As evident from Table 10, scenarios one, two, five and six achieved very clearly a majority decision. Even if not as striking, scenario three obtained a majority decision as well, whereas within scenario four participants have voted in equal number for both decisions. In such a situation it is tricky to provide an absolute best default decision. It could be argued that since participants were in doubt whether to share or not the information, a real danger was not intuitively present, thus the footage should not be shared. On the other hand, one could argue that is better for Alice to be safe, rather than ensuring her privacy. For the sake of this evaluation, it was decided that Alice's privacy has a strong value in case there is not clear danger, hence the footage should not be shared.

Finally, w.r.t. to the assessment of the proposed scenarios, it is intriguing to consider the most relevant questions asked by participants:

- Is the order of information important?
- **Do I have to explain why I take certain decisions?**
- What is the danger threshold?
- Does the surveillance camera have footage of Alice outside of the company?

The above displayed questions were noted down during the conducted interviews. The most asked questions was the second one, being asked in different forms by all six

participants. While it is not possible to conclude anything from such question, such a doubt indicates that the proposed scenarios have most likely triggered the human intuition to some degree, in the process of achieving a decision.

6.3 Context Scenarios

This subsection shall illustrate agent's behaviour in the scope of the selected scenarios. For each scenario, a short description of the scenario will be provided. This description should give the reader an idea of the importance of the scenario. Next, the information that Alice's agent has gathered from other agents is grouped together in a table which divides the information according to the time point when it got collected. Each piece of information is showed as a predicate followed by its degree of belief. As discussed in the methodology section, in this study input information has predefined degrees of belief, as the trust of agents is not used to manipulate the dobs. Finally, agent's behaviour will be analysed based on the table of available information and the results will be concisely illustrated in a table. The subsection concludes by presenting agent's achieved results in an overview table.

6.3.1 Scenario 1 (Missing Alice)

In the Methodology section, emphasis has been put on the missing concept. Scenario 1 shall illustrate the *missing* predicate being triggered and thus, leading to a share decision.

Suppose that on March the 30st, Alice is expected to be at work. Furthermore, it is not her home office day, yet she is not present at her office. Similarly to Example 2, Alice's agent gathers the information from her house sensor, according to which she is not at home. Even more, the surveillance camera at her work has not detected Alice being on leave. Additional information has become available from the news channel agent, which states that there has been an accident on the route that Alice usually takes to work.

The described scenario is expressed in the following information predicates:

Time	Information
T1	info(atWork(alice, 30)) 0.5. info(~atHomeOffice(alice, 30)) 0.8. info(~atOffice(alice, 30)) 0.8.
T2	info(workday(30)). info(~onleave(alice, 30)) 0.8. info(~atHome(alice, 30)) 0.6.
T3	info(trafficAccident(alice, 30)) 0.6.

Table 11: Scenario 1 - gathered information

In this scenario, since all necessary information is available, a missing predicate is triggered according to the norm showed below.

info(missing(A, T)) ← info(atHomeOffice(A, T)), info(atOffice(A, T)), info(onleave(A, T)) 0.9.

Nevertheless, since the trivial approach and the inference based approach do not regulate the degree of belief assignment, and default value is assumed to be 0.5, the missing argument is not powerful enough to result in a share decision in these two approaches. On the other hand, the more sophisticated approach uses support values to assign a new degree of belief to the inferred predicate *info(missing(alice, 30))*. According to the table provided in section 6.2, the support for the missing context is 0.8, which is strong enough to result in a share decision.

The following table, specifically in the third and fourth column, gathers the two different decisions taken by the agent, respectively based on only inferred contexts and with both inferred contexts and the support. The first and second column show what would be the intuitive decision within such scenario in the first column, and the decision that the argumentation engine will take without any inference in the second one.

Golden Result	No Inference	Inference	Inference + Support	object property feature
share	no share	no share	share	none

Table 12: Scenario 1 - agent's decisions

6.3.2 Scenario 2 (Office Day)

Let us slightly variate on Scenario 1. Suppose we differentiate between Alice's work location, in the sense that Alice has the possibility to work from home one day a week, whereas she is supposed to work from her office for the rest of the days. There are thus 2 types of ideal scenarios, those in which Alice is working from the office and those in which Alice is working from other locations, including her home, when having a home office day.

In Scenario 2, let us assert that Alice is not having her home office day, moreover she is at work, more specifically at her office. Additionally, consider the information of the traffic accident on the usual route taken by Alice still valid.

Time	Information

T1	`info(atWork(alice, 30)) 0.5.`
	`info(~atHomeOffice(alice, 30)) 0.8.`
	`info(atOffice(alice, 30)) 0.8.`

T2	`info(workday(30)).`
	`info(~onleave(alice, 30)) 0.8.`
	`info(~atHome(alice, 30)) 0.6.`

T3	`info(trafficAccident(alice, 30)) 0.6.`

Table 13: Scenario 2 - gathered information

As illustrated in the result table below, none of the approaches has decided to share

Alice's footage, which is intuitively correct. That is the case because we have the information of Alice being at her office. When asserted, this information has a high degree of belief by default. One argument has been built by the traffic accident information which has triggered the following norm into creating an *emergency* predicate:

```
info(emergency(A, T)) ← info(trafficaccident(A, T)) 0.9.
```

Nevertheless, the low degree of belief regarding the traffic accident did not allow such an argument to win against the strong belief of Alice being at her office.

Golden Result	No Inference	Inference	Inference + Support	object property feature
no share	no share	no share	no share	none

Table 14: Scenario 2 - agent's decisions

Scenario 2 has provided an intuitive situation to manage for all available approaches. As a result, all three methods behaved according to the golden result.

6.3.3 Scenario 3 (Home Office Day)

Suppose the situation becomes more ambiguous in the third context scenario. First, suppose that Alice is having her home office day and that we do not have additional information about her work position. This means that the following information predicate becomes positive:

info(homeOffice(alice, 30)) 0.8.

Note that in order for Alice to be inferred in a *HomeOffice* context, she should also be present at her home, which is not the case in the proposed context scenarios, as her house sensor claims the contrary. With respect to the remaining gathered information, assume that the traffic accident has still occurred on the route often taken by Alice, that Alice has an allergy, and that it is Spring.

Scenario 3 is represented as follows:

Time	Information
T1	`info(atWork(alice, 30)) 0.5.` `info(atHomeOffice(alice, 30)) 0.8.`

T2	`info(workday(30)).`
	`info(~onleave(alice, 30)) 0.8.`
	`info(~atHome(alice, 30)) 0.6.`

T3	`info(trafficAccident(alice, 30)) 0.6.`
	`info(hasAllergy(alice, 30)) 0.9.`
	`info(isSpring(alice, 30)) 0.4.`

Table 15: Scenario 3 - gathered information

Taking a sharing decision in such an ambiguous situation might be a little tricky. It is known that Alice has her home office day. Therefore, even though she is not at home according to her home sensor, she could be in a variety of other contexts, for example at the doctor due to a sudden allergic reaction, considered that she has pollen allergy and that it is Spring. It is interesting to notice that according to the golden result, which resembles the human intuition, her footage would better be shared. This intuition is followed by the inference only approach. This approach uses the default degrees of belief, and since the dob of the information regarding the traffic accident is higher that the dob of the doctor context, Alice is thought to be in danger, and thus her footage is shared.

From another perspective though, the support based algorithm decides not to share Alice's footage, as according to the bias used in the Assignment Algorithm, the idea that Alice might have an allergic reaction and therefore be at the doctor is supported more with respect to Alice having a traffic accident.

As mentioned in the Scenario Assessment subsection, the golden result sets an intuitive result for each scenario, however, there is no universal correct outcome. For this reason, it is interesting to notice the influence that a simple metric, such as a support, has on the final result. The flexibility offered by an algorithm such as the Assignment Algorithm, might thus be of strong value in ambiguous scenarios like the current one.

Golden Result	No Inference	Inference	Inference + Support	object property feature
no share	share	share	no share	none

Table 16: Scenario 3 - agent's decisions

6.3.4 Scenario 4 (Laboratory) - Transitive property

The following three scenarios shall illustrate ontology's potential in terms of inference, and its influence on the privacy agent. In particular, thanks to the transitive feature of object properties, the agent will be able to infer otherwise unreachable knowledge in the scope of Scenario 4.

For the sake of this scenario, consider that Alice works in a corporate. In such extended surroundings, knowing that Alice is expected to be at her office, but having the information about her being absent from the office, would intuitively still leave space for her being in other areas of the company and therefore, a missing or emergency context should not be immediately thought of.

Bob and Jack have not found Alice at her office and have decided to check the register of activity to find any hint on Alice's position. The two find out that Alice has booked a visit at the experimental laboratory, where experiments can be carried out by research employees. Operating in the laboratory however, requires both having scheduled an appointment, like Alice seems to have done, and having direct or indirect supervision from Bob who is the superintendent of the laboratory. Once the two conditions are satisfied, permission is granted to an employer whose presence is expected during the booked time. The following norm encapsulates this intuition according to the following norm:

```
info(nuclearlab(A, T)) ← info(hasBookedLab(A, T)),
info(hasSupervisor(A, bob)) 0.9.
```

In terms of available information thus, suppose that Alice should be at work, at her office specifically since it is not her home office day. Consider previously available information such as not being on leave, her absence from home, the traffic accident, Alice's allergy and the fact that it is Spring still holding. Furthermore, in trying to figure out Alice's situation, additional facts are learnt, such as Alice having booked the experimental laboratory for the 30st of March, her assigned supervisor being Tom and Tom's supervisor being Bob. Additionally, the whole team of Alice is on a trip to a research conference.

Time	Information
T1	`info(atWork(alice, 30)) 0.5.` `info(~atHomeOffice(alice, 30)) 0.8.`

T2	`info(workday(30)).`
	`info(~onleave(alice, 30)) 0.8.`
	`info(~atHome(alice, 30)) 0.6.`

T3	`info(trafficAccident(alice, 30)) 0.6.`
	`info(hasAllergy(alice, 30)) 0.9.`
	`info(isSpring(alice, 30)) 0.4.`
	`info(onTravel(alice, 30)) 0.4.`
	`info(hasBookedLab(alice, 30)) 0.9.`
	`info(hasSupervisor(alice, tom)) 0.9.`
	`info(hasSupervisor(tom, bob)) 0.9.`

Table 17: Scenario 4 - gathered information

Considering all available information in Scenario 4, Alice could be in a variety of contexts. She could be at the doctor, involved in a traffic accident or on travel with her team. Alice has also booked the use of the experimental laboratory, however, she has assigned as a supervisor Tom, which is not sufficient obtain permission. If Alice would have Bob's supervision, in the context of Scenario 4, that would mean that Alice is most likely at the laboratory as this context has a high degree of belief. Nevertheless, in order to reach the belief of Alice being at the laboratory, there is Bob's direct or indirect supervision which is missing. It is here that the inference capabilities of the ontology of

concepts come in handy.

As we have information regarding the booking done by Alice, and the supervision of Alice from Tom and of Tom from Bob, the ontology is able to the transitivity of the *hasSupervisor* object property, in order to infer that Alice is also supervised by Bob. At this point, the laboratory context can be inferred by the ontology. Notice that even though the inference-based approach inferred the laboratory context, the decision is still a share decision, as this context does not have a high enough degree of belief to compete with the degree of belief of the information regarding the traffic accident.

For this reason, the Assignment Algorithm is relevant within scenario 4. By using bias values, it is possible to personalize the decision with respect to Alice. In this case, it is desired to express that if Alice has both booked the laboratory and has Bob's supervision, she is likely to be at the laboratory performing experiments. This is done by the algorithm by assigning a fairly high degree of belief for the laboratory context. As evident from the following table, the inference+support approach does not share the footage, since Alice is most likely conducting experiments.

It is interesting to notice that the human intuition is inclined to share the footage in the context of this scenario. This could be attributed, among others, to the lack of intuition of the transitive aspect captured in this context.

Golden Result	No Inference	Inference	Inference + Support	object property feature
no share	share	share	no share	transitive

Table 18: Scenario 4 - agent's decisions

6.3.5 Scenario 5 (Supervised Meeting) - Functional Property

Scenario 4 has focused on the specific inference capability based on transitivity, introduced by the ontology. Yet another inference possibility offered by the ontology is related to

functional object properties. Similarly to Scenario 4, in Scenario 5 the final decision issued by the agent is influenced by knowledge inferred by the ontology, knowledge which would otherwise remain unknown.

Consider the context *SupervisedMeeting*, defined as follows: an individual is in a *SupervisedMeeting* context on the workday March the 30st, when she is not supervised by **any** boss who is absent on that workday. In order to represent such definition, the functional feature of object properties comes in handy and it is paired with the following equivalence class for the *SupervisedMeeting* context:

isSupervisedBy: Functional

not(isSupervisedBy some (Boss and isAbsent value 30)).

Now consider the following representation of Scenario 5, consisting of the same information predicates as Scenario 4 at T1 and T2, and the new information that Alice is supervised by Jack and Jack is not absent on March the 30st at T3.

Time	Information
T1	`info(atWork(alice, 30)) 0.5.` `info(~atHomeOffice(alice, 30)) 0.8.`

T2	`info(workday(30)).`
	`info(~onleave(alice, 30)) 0.8.`
	`info(~atHome(alice, 30)) 0.6.`

T3	`info(trafficAccident(alice, 30)) 0.6.`
	`info(hasAllergy(alice, 30)) 0.9.`
	`info(isSpring(alice, 30)) 0.4.`
	`info(onTravel(alice, 30)) 0.4.`
	`info(isSupervisedBy(alice, jack)) 0.9.`
	`info(~isAbsent(jack , 30)) 0.9.`

Table 19: Scenario 5 - gathered information

As presented in the table below, the only approach which has exploited the inferred information of Alice being at a supervised meeting, is the inference+support based approach. The trivial and the inference approach have both decided to share, regardless the fact that only the inference based one has inferred the *SupervisedMeeting* context. That is because both have acted on the higher degree of belief for the *TrafficAccident* context.

Golden Result	No Inference	Inference	Inference + Support	object property feature
no share	share	share	no share	functional

Table 20: Scenario 5 - agent's decisions

In this scenario, the difference in the final decision from the inference only based approach to the inference+bias approach has analogous explanation to the fourth scenario. In particular, the inference of the *SupervisedMeeting* context is achieved in both approaches, however, only the inference+bias approach is able to exploit it to keep Alice's privacy safe by not sharing the footage. Nevertheless, it is another point we shall direct our attention to.

One could argue that the trivial approach, which does not use an ontology for inference, could achieve a similar result by keeping the information predicates used above, and by using the following norm:

```
info(supervisedMeeting(A,T)) ← info(isSupervisedBy(A,B)),
info(~isAbsent(B,T)) 0.9.
```

While this approach might delude the writer to have achieved the same result, this is not the case, as without the a functional *isSupervisedBy* property, Alice could satisfy the norm above while still being supervised by bosses who are absent on the workday 30. This representation would therefore not respect the desired definition of the context *SupervisedMeeting*: an individual is in a *SupervisedMeeting* context when she is not supervised by **any** boss who is absent on the workday 30.

6.3.6 Scenario 6 (Coffe break) - Inverse Functional Property

One last inference possibility covered in the scope of this research is provided by the use of inverse functional properties, in the scope of Scenario 6. For the sake of this context scenario, consider a *CoffeBreak* context defined as follows: an individual is in a *CoffeBreak* context, if she is talking to some person who is not superintendent on the workday 30.

Suppose the *CoffeBreak* context is defined with the following equivalence expression in the ontology, and that the *isSuperIntendent* object property is set to be inverse functional.

isSuperIntendent: Inverse Functional

isTalking some (Person and (not (isSuperIntendent value 30)))

Now consider the following representation of Scenario 6, consisting of the same information as Scenario 5 and 4 at T1 and T2, and providing the new information that Alice is talking to Jack and Bob is superintendent on the workday 30 at T3.

Time	Information
T1	`info(atWork(alice, 30)) 0.5.` `info(~atHomeOffice(alice, 30)) 0.8.`

T2	`info(workday(30)).`
	`info(~onleave(alice, 30)) 0.8.`
	`info(~atHome(alice, 30)) 0.6.`

T3	`info(trafficAccident(alice, 30)) 0.6.`
	`info(hasAllergy(alice, 30)) 0.9.`
	`info(isSpring(alice, 30)) 0.4.`
	`info(onTravel(alice, 30)) 0.4.`
	`info(isTalking(alice, jack)) 0.9.`
	`info(isSuperIntendent(bob , 30)) 0.9.`

Table 21: Scenario 6 - gathered information

Since Bob and Jack are different people and since *isSuperIntendent* is an inverse functional property, Bob is the only superintendent on the March the 30st. Furthermore, since Alice is talking to Jack, the ontology is able to infer that Alice is not talking to any person that is superintendent on the workday 30, and it thus infers the *CoffeBreak* context.

Analogously to the effect of the *SupervisedMeeting* context on Scenario 5, the context *CoffeBreak* has been exploited by the inference+bias algorithm, to ensure Alice's privacy, and not to share her footage. That is the case because such algorithm has given a high degree of belief to the newly inferred context, *CoffeBreak*. On the other hand, the two simpler approaches have both considered that it would be better to share her footage taking into account the *TrafficAccident* active context.

Golden Result	No Inference	Inference	Inference + Support	object property feature
no share	share	share	no share	inverse functional

Table 22: Scenario 6 - agent's decisions

The following table gathers the achieved results. The Golden Result column, as covered in 6.2, represents the decision which humans achieve in the developed scenarios by using contexts' definitions and the information available. The third column displays the results achieved by manually constructing the domain knowledge described in 5.2. In this method, as in the approach of Kökciyan and Yolum, norms, scenario information and trust values are directly specified in the ASPIC argumentation engine. Thus, the reasoning on contexts occurs exclusively by triggering contextual norms and therefore having the consequent argumentation process.

The fourth column instead, displays the results provided by the agent proposed in this study, when using only the representation of contexts by means of an ontology, and not the algorithm for manipulating the domain knowledge prior to the decision-making phase. As it shall be discussed in the next section, even though the final result is not different from the most basic approach, new contexts are inferred and hence, new arguments are

built in the argumentation phase. Nevertheless, the agent needs to handle the degrees of belief of newly inferred information in a more specific way, rather than just assigning the default value of 0.5. This is achieved in the fifth column with the complete agent, PARCo. Finally, the sixth column provides information circa which object property feature has been used in the inference step.

Scenario	**Golden Result**	No Inference	Inference	**PARCo**	Object Property Feature
1	share	share	share	share	none
2	no share	no share	no share	no share	none
3	no share	share	share	no share	none
4	no share	share	share	no share	transitive
5	no share	share	share	no share	functional
6	no share	share	share	no share	inverse functional

Table 23: Different Approaches' decisions overview

7 Discussion

This section shall provide an interpretation of the final results displayed in the end of Section 6.3, and clarify what lies behind the eventual decisions that have been reached with PARCo. Most importantly, it shall formulate answers to the research questions addressed in this study. Finally, the main shortcomings of the current study and directions for future research will be discussed.

7.1 Results Discussion

The previous section has provided a method to evaluate the performance of the proposed model, with respect to the introduced contributions, namely: i. the representation of contexts withing an OWL ontology; ii. the integral implementation of the model as a knowledge-based agent, including an algorithm for processing the degrees of belief of the inferred information. Based on the gathered results, this section shall discuss the extent to which these contributions have affected the behaviour of the proposed model.

In conclusion of Section 6.3, Table 23 has displayed the behaviour of the proposed agent, PARCo, with respect to the golden result representing the human intuition, and with respect to the more elementary approaches: the basic approach, which does not use an ontology of concepts, and the inference-based approach, which uses an ontology of concepts but does not manipulate the generated inference.

One first aspect worth discussing is concerned with the results obtained by the most basic approach compared to the inference-based approach. In particular, Table 23 shows that both methods achieved the same final decisions in all the six scenarios proposed in this document. While at the level of the final decision there is no difference between the two techniques, it is interesting to provide further insights on the internal functioning of these methods. In particular, there is a distinction in terms of inference capabilities between the two approaches. The basic approach does not have any inference capability, in fact, all required information such as the norms and information predicates, ought to be inserted directly into the argumentation engine. Consequently, the ASPIC engine would perform the reasoning in its entirety. On the other hand, the inference based approach is performed by the agent, by using solely the ontology for inferring contexts and the ASPIC engine. Thus, the domain knowledge in the inference-based approach, might include additional information predicates resulting from ontology's inferred contexts. In

the scope of the six evaluation scenarios, distinct domain knowledge specifications w.r.t. the two approaches, occur in the last three scenarios: scenario four, five and six.

Essentially, within the first three scenarios, the equal final decision can be considered to present an accurate picture of the two approaches' capabilities. With respect to scenarios four, five and six instead, the equal final decisions displayed in Table 23 are not representative of the inferences capabilities of the two approaches. In other words, in the first three scenarios it is possible to manually create the domain knowledge, by inserting the input information predicates and the contextual norms defining each context, and to achieve the same arguments, in the argumentation phase, as would be achieved by using the domain knowledge produced by the inference-based approach. To understand this, it is helpful to consider both the way contexts are inferred by the reasoner in the ontology, and the construction of arguments supporting contexts, occurring in the argumentation engine. On one hand, the ontology uses the information predicates provided by a scenario, parsed into an object property assertion form, to match the equivalence class expressions defining contexts, and hence, infer contexts. On the other hand, the argumentation engine will use the information predicates of the scenario, to trigger certain contextual norms, which are nothing but the FOL representation of the equivalence class expressions defining contexts in the ontology. Considering that in the first three scenarios the inference achieved by the inference-based approach, mainly consists of matching equivalence class expressions, the resulting domain knowledge is equal to one that would have not used an ontology and a reasoner. Intuitively, there is no inference made for which the argumentation engine can not create an argument by using the scenario information and contextual norms. As a consequence, in the first three scenarios, both the basic approach and the inference-based approach lead to the same domain representation.

Instead, with respect to the last three scenarios, the two approaches have different representational power, even though the ultimate decision shown in Table 23 is equal. In the scope of these scenarios, the inference based approach is able to infer additional information due too the features of object properties, such as the transitive, functional and inverse functional properties, for scenarios four, five and six respectively. This information is not available in the basic approach to represent manually. In other words, the inference-based approach is able to infer additional contexts by using the proposed representation of contexts. Additionally, because of the subclass relation between the *Work* and *Office* context, and the available information in that Alice is at her office, in Scenario 2, it

is possible to infer the Office context for the user Alice, instead of both the *Work* and *Office* context. Inferring a specific context like the *Office* one, should intuitively provide more grounds to contrast the dismiss of Alice's footage. Nevertheless, without taking into account degrees of belief, inference alone is not enough to alter decision-making process. Even though the argumentation is able to construct arguments for more context by using the domain knowledge generated by the inference-based approach, generated context are assigned a default degree of belief of 0.5. This value does not allow the contexts inferred in the last three scenarios to be acceptable with respect to the arguments constructed towards a share decision.

7.2 Research Questions' Answers

The results discussion provided above, together with the representation of contexts developed in Section 4, are used in order to formulate the answer to the first research sub-question:

How to represent contexts for reasoning on privacy?

The current work aimed at developing a context representation that would enhance the inference of contexts. An improvement is not directly visible when comparing the final decision reached by the most trivial approach, that of manually creating the domain knowledge, w.r.t. the inference-based approach. However, as explained above, that is not related to the inference of contexts itself, rather to the assigned degrees of belief. In fact, an improvement of inference is achieved in the last three proposed scenarios, specifically by means of using transitive, functional and inverse functional object properties. Moreover, the improvement is clearly present in PARCo, which is able to achieve descisions close to the human intuition, among others, due to its context representation.

To answer the first sub-question, a suitable way of representing contexts is achievable by using an ontology of concepts in order to define a hierarchy of contexts. Within the hierarchy each context is defined in terms of equivalence class expressions which express when a user should be inferred in a given context, by means of object properties which can be asserted about users. The advantages for reasoning on privacy contexts are mostly allowed by the OWL semantic language for creating an ontology, and shall be discussed in the answer of the second research question:

Does an ontology of concepts bring any advantage with respect to reasoning on privacy contexts? If yes, which one?

Section 2.2 has provided an elementary introduction to the concept of an ontology and the OWL semantic language as needed for the purpose of this study. Specifically, some of the expressive capabilities of OWL have been listed. With respect to the achieved results, the most important are the transitive, functional and inverse functional object properties and the SubClassOf and EquivalentTo properties of classes. By representing the domain with such means, it was possible to improve knowledge inference. Specifically, direct advantages of an ontology w.r.t reasoning come from:

- *express contexts' relations by means of the EquivalentTo property*

It is possible to use multiple equivalence class expression on a context, making reference to object properties that might be related to other context. This is one way of representing a relation between contexts. An example is the *Missing* context which is defined as not being home, nor at the office or on leave.

- *express contexts' relations by means of the SubClassOf property*

Another possibility to put contexts into relation comes naturally from the SubClassOf relation. In the proposed approach one example is the *Work* context and its *HomeOffice* subclass. As shown in Table 3, a user is in a *HomeOffice* if he has a home office day and is at home on such day. Nevertheless, it might be the case that there is not enough information to activate a *HomeOffice* context. In that case it would be possible to use more general information, if available, in order to activate a more general context, in this case, *Work*. Note that when the ontology is inferring a specific context for a user, such as *HomeOffice*, it does not infer the general one, *Work*. Also note that this advantage, in big enough scenarios, might extend from being an inference advantage, to being a scalability and/or performance one, as the argumentation might be at ease by avoiding the construction of all the arguments for the most general contexts. Nevertheless, this aspect would require an analysis on its own and the tradeoff between leaving out general contexts arguments and using all information available is not yet clear.

- *taking into account the transitive, functional and inverse functional aspects of in-*

formation assertions

As emphasized in the evaluation section, by use of these properties it is possible to infer knowledge which would otherwise be not available. See scenarios 4,5 and 6 for examples.

The above mentioned advantages are evident with respect to the final decision, when used by PARCo, as shown in Table 23. This falls under the scope of the third subquestion:

How to exploit the context representation for decision-making on contexts for privacy?

Unlike the previous two, the answer to the third subquestion finds more evident support in the final decisions achieved in Table 23. In specific, PARCo has achieved the closest result to the Golden Result. That is the case because PARCo has exploited the context representation by using an algorithm to manipulate the degrees of belief of the achieved domain knowledge. This algorithm has been covered in the section dedicated to the agent 5.3, and as a brief remainder, it is a procedure that assigns degrees of belief to the information in a scenario, based on a bias established by the user of PARCo. More generally, the use of such an algorithm is a form of knowledge manipulation in the scope of the proposed knowledge-based agent. Many other techniques could be used by the agent to process its domain knowledge. One alternative example would involve contextual norms' degrees of belief being updated, for example, based on the hierarchy of contexts. For example, if there is available information for creating arguments for the *Work* and *Office* context, the algorithm could choose to enforce the more precise norm by assigning a higher degree of belief to it, w.r.t. to the norms activating the more general context, *Work*. Many other algorithms can potentially be set in place as an option to the proposed one. More concisely, by using an agent which uses an ontology for contexts inference, it is possible to create the domain knowledge and to further manipulate it prior to the argumentation phase. In this way, the eventual sharing decision better respects user's privacy preferences.

Each necessary element towards providing a computational approach to privacy in the Internet of Things has been analysed. It is now possible to formulate an answer to the main research question:

How to implement ***a semantic representation of contexts****, to allow* ***reasoning*** *and* ***decision making*** *for* ***privacy****?*

One promising approach to reason and make decisions while ensuring privacy in terms of information sharing in IoT scenarios, is to rely on a knowledge-based agent. Such an agent can use an ontology based knowledge representation, in order to derive knowledge from available, potentially incomplete, information. Subsequently, it can use an internal algorithm to influence the decision-making process according to its user's privacy preferences.

7.3 Limitations and Future Work

The current work has tackled appropriate information sharing in the Internet of Things by focusing on two main aspects. It is nevertheless the case that there is space for improvements from a variety of perspectives.

From a high-level perspective, the aspects to be taken care of would be to incorporate in the proposed model of trust management as proposed by Kökciyan and Yolum [Kökciyan and Yolum, 2017]. In particular, it would be necessary to differentiate between predicates of the form says(A, info(X), dob), and predicates of the form info(property(A,T)). The predicates containing the "says" keyword would be taken care by the norms described in Table 4, by using the trust values of external agents, available in Table 5.

There is another important aspect to develop, which has been touched by Kökciyan and Yolum in their framework. This is the use of relations withing contexts, towards more appropriate information sharing. Such aspect has been mentioned by Kökciyan and Yolum, as their agent uses a set of relations defined in a context, in order to decide from which agents additional information should be gathered. Nevertheless, this aspect has not been deepened enough in the approach proposed by the authors, nor in the scope of this study. Such aspect has theoretical value, as relations are an important factor to consider when treating privacy expressed as Contextual Integrity.

With respect to the context representation itself, a first aspect to be considered into more depth, is the possibility of inferring additional information by means of those OWL properties which have not been exploited in this research. For example, object properties may have more features than only transitive, functional and inverse functional. Additional

possibilities are symmetric, asymmetric, reflexive and irreflexive. These features might, in different scenarios than the one proposed, yield to additional knowledge.

Additionally, the context representation could be improved by addressing the hierarchy of contexts and their definition. Even though the EquivalentTo property is believed to be the most powerful available in OWL for the definition of contexts, the representation of contexts could be investigated more, by considering additional properties. One element to be considered, is that there is a trade-off between the the amount of equivalent class expressions used to define a context, and the freedom in defining such contexts. In other words, when defining a context to be equivalent to a number of equivalence class expressions, it is not the case that one expression can hold whilst another does not hold. Hence, it is not straightforward to use a variety of equivalence class expression to define one context while using object properties which are also involved in other contexts' definitions. A way to tackle this aspect is to further explore class properties available in OWL, such as the SubClassOf property that has been used on the *Work* context in this study and the DisjointWith property.

8 Conclusion

Substantial advancements of technology have allowed devices to penetrate different spheres of humans' lives and extensively manage the data generated in daily activities. This has increasingly raised concerns regarding users' privacy. Traditional regulations that should settle privacy disputes, such as the US Constitution and the GDPR, have obtained particularly little success with respect to dynamic scenarios like Online Social Networks and the Internet of Things.

Contextual Integrity has been proposed as an alternative account to privacy. This definition focuses on the appropriateness of information exchange with respect to contextual norms. In other words, privacy is maintained when information exchange respects the distribution norms in a certain context. Such concept has inspired a number of scholars to provide privacy-preserving approaches in the context of OSNs. Moreover, Kökciyan and Yolum have developed a model which, by taking inspiration from the CI concept, has regulated information exchange in a IoT scenario.

Considering the improvable aspects in the work of Kökciyan and Yolum as a starting point, this study has proposed a knowledge-based agent which uses an ontology of concepts, in order to represent contexts and argumentation as a mean to make a transparent share decision on top of available information. Specifically, the main contributions lie in: i. exploiting OWL expressivity for representing contexts and by means of a reasoner, infer new knowledge; ii. using a knowledge-based agent which manipulates such a domain knowledge in a convenient way, in this case by assigning degrees of belief, and uses it in the ASPIC argumentation engine in order to reach a share decision.

The work has been evaluated with respect to a selection of six IoT scenarios, for which the golden result has been established to be human intuition, assessed by means of guided interviews. Consequently, the inference capabilities have been discussed. Finally, PARCo has been shown to respect the human intuition in the selected scenarios, as a result of implementing the proposed representation of contexts, the algorithm for manipulating degrees of belief and the use of argumentation for decision making.

References

Amgoud, L. (2009). Argumentation for decision making. In *Argumentation in Artificial Intelligence*, pages 301–320. Springer Science+Business Media, LLC 2009.

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, 54(15):2787 – 2805.

Barth, A., Datta, A., Mitchell, J. C., and Nissenbaum, H. (2006). Privacy and contextual integrity: Framework and applications. In *2006 IEEE Symposium on Security and Privacy*, page 184–198.

Behhofer, S. (2009). Owl: Web ontology language. In *Encyclopedia of database systems*, pages 2008–2009. Springer US.

Bench-Capon, T. and Dunne, P. (2007). Argumentation in artificial intelligence. *Artificial Intelligence*, 171:619–641.

Caminada, M. and Amgoud, L. (2005). An axiomatic account of formal argumentation. In *Proceedings of the AAAI 2005*, pages 608–613.

Criado, N. and Such, J. (2015). Implicit contextual integrity in online social networks. *Information Sciences: An International Journal*, 325:48–69.

Dijkstra, P., Bex, F., Prakken, H., and de Vey Mestdagh, C. (2005). Towards a multi-agent system for regulated information exchange in crime investigations. *Artif. Intell. Law*, 13:133–151.

Dijkstra, P., Prakken, H., and de Vey Mestdagh, C. (2007). An implementation of norm-based agent negotiation. pages 167–175.

Dung, P. (1995). On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games* 1. *Artificial Intelligence*, 77:321–357.

Dung, P., Kowalski, R., and Toni, F. (2009). Assumption-based argumentation. In *Argumentation in Artificial Intelligence*, pages 199–218. Springer Science+Business Media, LLC 2009.

Fong, P. W. (2011). Relationship-based access control: Protection model and policy language. *Proceedings of the ACM Conference on Data and Application Security and Privacy (CODASPY)*, pages 191–202.

Fox, J., Glasspool, D., Grecu, D., Modgil, S., South, M., and Patkar, V. (2007). Argumentation-based inference and decision making–a medical perspective. *IEEE Intelligent Systems*, 22:34–41.

Krupa, Y. and Vercouter, L. (2012). Handling privacy as contextual integrity in decentralized virtual communities: The privacias framework. *Web Intelligence and Agent Systems*, 10:105–116.

Kökciyan, N., Yaglikci, N., and Yolum, P. (2017). An argumentation approach for resolving privacy disputes in online social networks. *ACM Transactions on Internet Technology*, 17:1–22.

Kökciyan, N. and Yolum, P. (2016). Priguard: A semantic approach to detect privacy violations in online social networks. *IEEE Transactions on Knowledge and Data Engineering*, 28:2724–2737.

Kökciyan, N. and Yolum, P. (2017). Context-based reasoning on privacy in internet of things. International Joint Conference on Artificial Intelligence (IJCAI-17), pages 4738–4744.

L. Pollock, J. (1987). Defeasible reasoning. *Cognitive Science*, 11:481–518.

Miorandi, D., Sicari, S., Pellegrini, F. D., and Chlamtac, I. (2012). Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, 10.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79:119.

Ouerdane, W., Maudet, N., and Tsoukiàs, A. (2010). Argumentation theory and decision aiding. *Trends in Multiple Criteria Decision Analysis*, 142.

Reiter, R. (2003). Nonmonotonic reasoning. *Annual Review of Computer Science*, 2:147–186.

Russel, S. J. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Malaysia; Pearson Education Limited 2016.

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

Smith, M. K., Welty, C., and McGuinness, D. L. (2004). Owl web ontology language guide. `https://www.w3.org/TR/2004/REC-owl-guide-20040210`. [Online; accessed 29-June-2019].

Voigt, P. and von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer International Publishing AG 2017.

Weber, R. H. (2010). Internet of things – new security and privacy challenges. *Computer Law Security Review*, 26:23–30.

Weber, R. H. (2015). Internet of things: Privacy issues revisited. *Computer Law and Security Review*, 31:618–627.

Ziegeldorf, J., Morchon, O. G., and Wehrle, K. (2014). Privacy in the internet of things: Threats and challenges. *Security and Communication Networks*, 7.

Appendices

Guided Interview

To share or not to share?

In the scope of this analysis, you will be asked to give your intuition regarding information sharing in a IoT scenario. Specifically, you will need to decide whether a surveillance camera should, or should not share a piece of footage with the requester of such footage. This footage is considered to hold relevant information regarding a person who might (not necessarily) be in danger.

Suppose that Bob, Alice's boss, does not know where Alice is. With the claim that Alice might be in danger, he wants to access the footage taken by the surveillance system, which might reveal additional information about Alice. Before taking a decision, the surveillance system communicates with other sources of information and gathers additional facts to provide a decision which best accounts for Alice's privacy. In particular, if the surveillance system thinks that Alice is not in danger, it would probably be reasonable not to share Alice's footage and therefore protecting her privacy. Viceversa, if Alice is in danger, it might be wiser to trade her privacy for additional information.

To take into account when deciding:

- a **share** decision means that Alice's privacy is weakened, but this can be in Alice's interest if she is in danger
- a **not share** decision means that Alice's privacy is enforced, but a small risk of Alice actually being in danger remains

Note: there is a cost associated with both over-sharing (always trading Alice's privacy for her safety) and under-sharing (always trading Alice's safety for her privacy), thus it is not encouraged.

Domain Rules

- Alice works at a company, in the physics department. Like most employees, she has the possibility of **working from home** once a week, while the rest of the days she is supposed to **work at her office**.
- Alice is considered **missing** when she is **not having a home office day**, yet she is **not at her office nor** has been seen **leaving** the company.
- Alice usually uses her car for transportation.
- Alice can carry out **laboratory** experiments. In order to do so, she **must book the laboratory.** Furthermore, **direct or indirect supervision** from her boss Bob is required.
- Alice is on a **break** when she is **talking to some person who is not a superintendent**.
- Usually, when her team is travelling, Alice joins.
- Alice is allergic to pollen. In Spring she **could** therefore suddenly need to go to a **Doctor**.
- Alice can be in a **supervised Meeting**. During such a meeting she **cannot be supervised by a Boss who is absent on the meeting day.**

For each of the following scenarios, consider the gathered information. Would you think reasonable of the surveillance system to share or not share the footage with Bob?

Scenario 1
It is a workday. (all information below applies during this day)
Alice is **not** having her home office day.
Alice is **not** at her office.
Alice has **not** been seen leaving the office.
There is an accident on the road usually taken by Alice.
Alice is **not** at home.
Share / Not Share

Scenario 2
It is a workday. (all information below applies during this day)
Alice is **not** having her home office day.
Alice is at her office.
Alice has **not** been seen leaving the office.
There is an accident on the road usually taken by Alice.
Alice is **not** at home.
Share / Not Share

Scenario 3
It is a workday. (all information below applies during this day)
Alice is having her home office day.
Alice has **not** been seen leaving the office.
There is an accident on the road usually taken by Alice.
Alice is **not** at home.
It is Spring.
Share / Not Share

Scenario 4
It is a workday. (all information below applies during this day)
Alice is **not** having her home office day.
Alice has **not** been seen leaving the office.
There is an accident on the road usually taken by Alice.
Alice is **not** at home.
It is Spring.
Alice's team is on travel to a conference.
The experimental laboratory has been booked by Alice for this day.
Alice has assigned Tom as supervisor for the laboratory.
Tom has assigned the boss Bob as a supervisor for the laboratory.
Share / Not Share

Scenario 5
It is a workday. (all information below applies during this day)
Alice is **not** having her home office day.
Alice has **not** been seen leaving the office.
There is an accident on the road usually taken by Alice.
Alice is **not** at home.
It is Spring.
Alice's team is on travel to a conference.
Alice has as assigned supervisor Jack (who is also a Boss).
Jack is not absent on this day.
Share / Not Share

Scenario 6
It is a workday. (all information below applies during this day)
Alice is **not** having her home office day.
Alice has **not** been seen leaving the office.
There is an accident on the road usually taken by Alice.
Alice is **not** at home.
It is Spring.
Alice's team is on travel to a conference.
Alice is talking to Jack.
Bob is the superintendent on this day.
Share / Not Share

Additional Info

Problem Statement (briefly)

The Internet of Things (IoT) is an ever increasing network of heterogeneous devices interacting together towards achieving certain goals. It is commonly acknowledged that the scale of the IoT is rapidly increasing, as this concept is penetrating a number of fields, such as domotics, logistics, transportation, healthcare and more. Considering such premises, it is natural that the volume of exchanged information among devices is skyrocketing. Moreover, as devices become increasingly more autonomous, the user has less and less impact on the information exchange itself. Users' control is limited to the initial data protection policy agreement action, at the time of initializing the device. However, it is recurrent in the literature that **available data protection regulations, such as the GDPR, provide no proper guidance in context-implicit, heterogeneous and dynamic environments**, such as public surveillance, online social networks and the IoT. In these scenarios, privacy violations might occur even though not sensible information has been shared.

Proposed Approach (briefly)

According to the literature, in scenarios like public surveillance, online social net-

works and the IoT, it is convenient to reason on the privacy decision according to the context in which information is being exchanged. In particular, it is convenient to adopt the principle of Contextual Integrity (CI), according to which privacy is violated when the exchange of information is violating the norms of appropriateness of a given context. On the other hand, according to CI, privacy is maintained when information is being exchanged according to the norms of all involved contexts. In order to implement such privacy approach, it is convenient to have autonomous reasoning and decision making entities, such as agents, which can take a privacy decision according to their representation of context. It is on this representation of context that the current research focuses by trying to enhance reasoning on contexts by means of an ontology of concepts and argumentation.