REPHRAIN

Protecting citizens online

Taking Situation-Based Privacy Decisions: Privacy Assistants Working with Humans

Nadin Kokciyan, University of Edinburgh

Pinar Yolum, Utrecht University

June 2022

UK Research and Innovation · University of Bristol · UCL · The University of Edinburgh · King's College London · University of Bath

Taking Situation-Based Privacy Decisions: Privacy Assistants Working with Humans

Nadin Kökciyan[1] and **Pınar Yolum**[2]
[1]University of Edinburgh
[2]Utrecht University
nadin.kokciyan@ed.ac.uk, p.yolum@uu.nl

Abstract

Privacy on the Web is typically managed by giving consent to individual Websites for various aspects of data usage. This paradigm requires too much human effort and thus is impractical for Internet of Things (IoT) applications where humans interact with many new devices on a daily basis. Ideally, software privacy assistants can help by making privacy decisions in different situations on behalf of the users. To realize this, we propose an agent-based model for a privacy assistant. The model identifies the contexts that a situation implies and computes the trustworthiness of these contexts. Contrary to traditional trust models that capture trust in an entity by observing large number of interactions, our proposed model can assess the trustworthiness even if the user has not interacted with the particular device before. Moreover, our model can decide which situations are inherently ambiguous and thus can request the human to make the decision. We evaluate various aspects of the model using a real-life data set and report adjustments that are needed to serve different types of users well.

1 Introduction

We are surrounded with Internet of Things (IoT) applications that enable various devices to work together. These devices operate by collecting, storing, and processing many different types of user data. The main medium for handling privacy constraints with these devices is to agree on their privacy policies, where a user is consulted to give consent, as with the policies in General Data Protection Regulations (GDPR) [Voigt and Von dem Bussche, 2017]. While giving consent enables to reflect users' preferences to some extent, it has also been shown that these approaches create an incredible decision load on the user, usually leading to users to ignore the details to be able to use the systems [Utz *et al.*, 2019]. The problem is amplified for IoT applications. The number of devices and the way they would want to use data creates an explosion in the number of situations that might arise, requiring humans to be constantly probed for consent. Previous work has shown that privacy assistants can learn the privacy preferences of the users [Squicciarini *et al.*, 2017; Kurtan and Yolum, 2021] and take actions to inform and help them appropriately [Das *et al.*, 2018; Ulusoy and Yolum, 2021]. Recent work has shown that users are welcoming to use privacy assistants beyond this; e.g., when privacy assistants take some of the decisions on behalf of the users [Colnago *et al.*, 2020]. Here, we propose a new model for privacy assistants that can make privacy decisions on behalf of the users per situation, even when the user has not explicitly been in a similar situation before. We identify three major challenges that need to be addressed to realize this:

Context. An alternative formulation of privacy is contextual integrity [Nissenbaum, 2004], where the privacy is understood to be preserved if appropriate information flows take place as designated by contexts. Inspired by this, we investigate various definitions of context [Alegre *et al.*, 2016] and follow the widely accepted definition of Dey [2001]: "any information that can be used to characterize the situation of an entity". In terms of privacy in IoT, various types of information have been identified as important. For example, it has been established that users tend to give access to sharing data if they know it will be used for a short duration but are uncomfortable to share data when they do not know how long it will be kept [Leon *et al.*, 2013]. In a similar vein, knowing the purpose of data collection or the benefit it will bring affects the privacy decisions of users. The location in which the data collection is taking place is sometimes important; e.g., restroom has been identified to be a blacklist concept [Naeini *et al.*, 2017]; while other locations, such as store or library do not have a particular connotation in user studies. Consider the following scenario [Naeini *et al.*, 2017] that expresses a situation that combines information about a user as well as an IoT device.

Example 1. *You are at a department store. This store has presence sensors to detect whether someone is present. The store management uses this data to keep track of when there are few customers in the shop to determine whether they can reduce the number of staff at these times. You are not told how long the data will be kept.*

When making a decision about Example 1, the contexts that would influence the decision can be the limited nature of personal data being collected (e.g., only presence), the purpose of collection, as well as retention of data. Thus, multiple contexts have to be formulated. Given such situations, how

can a privacy assistant formulate the context(s) of a situation in hand? Is it better to consider a dominant, single context or factor in all contexts that a situation belongs to?

Decision. When contexts are decided, the privacy assistant needs to make a decision to share information or not. Making a privacy decision in a situation requires factoring in whether the particular set of IoT devices involved are indeed trustworthy [Chung *et al.*, 2017]. Ideally, with increasing number of interactions, a privacy assistant would learn whom to trust and to what extent. Many successful trust models exist in the literature to address this [Teacy *et al.*, 2012]. However, in an IoT system, a user will interact with the same device rarely and will have to decide on whether to trust without prior interactions with that device [Kökciyan and Yolum, 2020]. How can a privacy assistant make sharing decisions based on trust in contexts, rather than trust in an entity?

Collaboration. Privacy is personalized; so are the privacy assistants. Some users are happy to allow data collection in Example 1, while some are not, requiring the privacy assistant to take personalized sharing decisions. Moreover, it is even possible for humans to make conflicting privacy decisions on occasions that are similar [Acquisti and Grossklags, 2005]. These require privacy assistant to assess whether its decision could be faulty and if so delegate the decision back to the user. However, delegating too much to the user will make the use of assistants pointless. How can a privacy assistant decide when to delegate to the user? Would different types of users benefit from different delegation frequencies from their privacy assistants?

We propose a situation-based privacy model to realize agent-based privacy assistants (PAS). PAS derives contexts automatically from the situations that the user has previously been in. Using subjective logic, PAS determines the trust for a context, based on positive and negative privacy experiences of the user in that context. This handles the problem of not having many interactions with the same device. By explicitly modeling inconsistencies between experiences of the user, PAS determines when to make a decision on behalf of the user, and when it can abstain and delegate the decision to the user. We implement the proposed agent and experimentally evaluate its workings over a case study that uses an anonymized IoT dataset [Naeini *et al.*, 2017]. We show that PAS can capture policies that pertain to multiple privacy contexts and combine them to reach a sharing decision as well as personalize it for different users.

2 Situation-Based Privacy Model

Typical approaches to trust would make use of the experiences with an individual to create a model. The more experiences there are (direct or indirect), the more accurate the model [Teacy *et al.*, 2012]. However, in an IoT setting, an agent would have few experiences with the same device, but overall many experiences with different devices. Thus, modeling the trust in a device based on the experiences of that device would create an insufficient model.

One possible way to study this question is through *categories* by identifying the device as belonging to a particular group and assigning trust based on that. Recent work has demonstrated that in certain domains assigning trust based on categories and indirect experiences have been successful [Sapienza and Falcone, 2020]. However, in IoT applications, collecting indirect experience is also difficult and would require interactions with unknown entities.

We propose to estimate trust by only considering the context that the interaction will take place in. This has the advantage that even when there is no specific prior evidence about a device, a decision can be made by considering the context. For example, in Example 1, the user might have never entered this store before. Hence, it is not possible to model trust in the sensor based on previous interactions. However, the user might have been exposed to situations where she did not know how long her data will be stored. If the user can formulate this as a context, it can use it to decide whether to share or not share information with the sensor.

2.1 Identifying Contexts

A typical way of thinking of contexts is to pre-define them at design time, so that the agent can categorize the policies based on their contextual properties [Kökciyan and Yolum, 2017; Fogues *et al.*, 2017]. However, the set of contexts would depend on the domains that the agent engages in, the granularity of experiences, and the variance between them. For some settings a context that can be associated with *medical* situations would be sufficient to capture interactions, whereas sometimes there would be a difference between *intensive care unit* and *check-up* situations, as they pertain to different information. This creates a need to derive contexts dynamically.

In our model, each privacy situation consists of a set of features that describe the IoT device (e.g., its type) as well as the dealings (e.g., purpose of interaction) (Definition 1).

Definition 1 (Privacy Situation)**.** *$P^x_y = \langle text, F \rangle$ denotes a situation of x concerning y, where x is of type PAS and y is an IoT device, $text$ is a textual representation of a situation, F denotes a set of features derived from text. P^x is the set of all situations of x, and p_i refers to the i^{th} item in P^x.*

The privacy situation in Example 1 as perceived by x concerning presence sensor ps can be represented as $P^x_{ps} = \langle t, \{location = store, data = presence, device = ps, retention = unspecified, purpose = reducestaff\}\rangle$, where t is the privacy text used in the example.

PAS interacts with IoT devices all the time, collects various privacy situations, and groups similar situations together to derive relevant contexts (Definition 2).

Definition 2 (Context)**.** *A context is a collection of similar privacy situations based on a similarity metric (Method 1). Given a context c and two situations p_i and p_j, if $p_i \in c$ and $\mathsf{sim}(p_i,p_j)$ holds then $p_j \in c$ also holds, where $c \in C^x$ and $\{p_i, p_j\} \subseteq P^x$. C^x is the set of all contexts known by x.*

Method 1 (Situation Grouping)**.** *PAS implements a clustering technique to take the set of privacy situations and a similarity metric* sim *for generating the set of contexts.*

2.2 Assigning Trust to Contexts

PAS assigns trust not to individuals but to specific contexts. To do this, PAS processes a new situation by first decid-

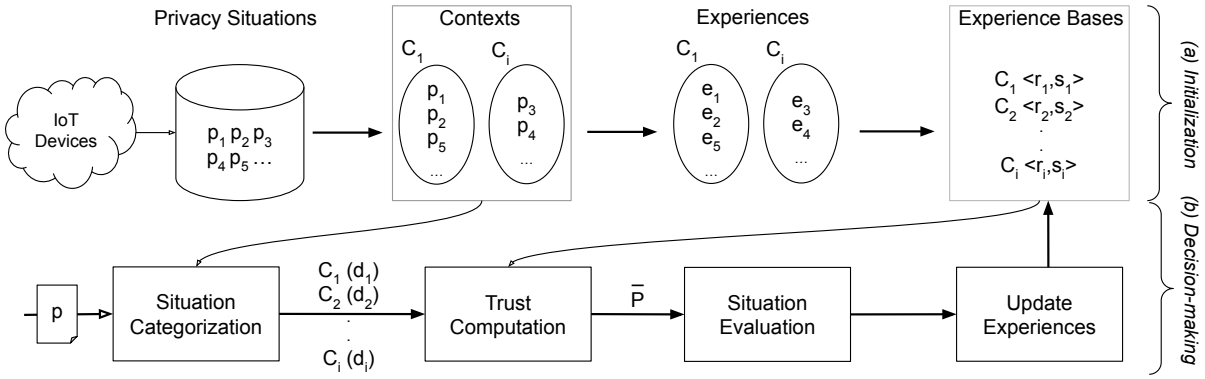

Figure 1: (a) Initialization: PAS analyzes the privacy situations (p_i in short) collected from other IoT devices during previous interactions. The similar situations are grouped into contexts according to a similarity metric (Method 1). The positive and negative experiences (e_i in short, Definition 4) associated with the privacy situations are used to compute context-based experience bases (Definition 5). (b) Decision-making: PAS computes the likelihood of an "unseen" situation belonging to various contexts (C_i) with varying degrees (d_i) (Method 2). It computes a weighted probability expectation value (Equation 2) considering opinions about contexts. PAS evaluates the privacy situation to make a decision on behalf of the user, or to delegate the decision to the user. PAS finally updates its experience.

ing on which contexts it belongs to, and then by considering how much it trusts these contexts. Our model uses subjective logic (SL) [Jøsang *et al.*, 2006] to associate trust to contexts. SL is a belief calculus to capture trust among individuals as opinions. In Definition 3, we describe an opinion (w_c^x) as the trust of PAS (i.e., x) in context c.

Definition 3 (Opinion). *$w_c^x = \langle b, d, u, \alpha \rangle$ denotes an opinion in context c, with the degrees of belief (b), disbelief (d), uncertainty (u) and a base-rate parameter (α); where $b+d+u = 1$ and $b, d, u, \alpha \in [0, 1]$.*

PAS needs to construct experience bases first before it computes context-based opinions to make a sharing decision. Each interaction with a device results in the user allowing data collection or not, which we call an experience (Definition 4). In Example 1, if x declines data collection request, $\langle P_{ps}^x, - \rangle$ would represent this negative experience.

Definition 4 (Experience). *$e_{c,i}^x = \langle P_y^x, v \rangle$ denotes ith experience in context c, where P_y^x is a privacy situation, v shows if the interaction with y was perceived as positive ($+$) or negative ($-$). E_c^x denotes all experiences of x in context c.*

In IoT applications, each PAS values privacy differently and each experience is unique to the agent. We focus on aggregating the experiences of PAS with IoT devices that are deployed in the same context (Definition 5). Figure 1(a) depicts the grouping of situations and construction of context-based experience bases.

Definition 5 (Context-based Experience Base). *$EB_c^x = \langle r, s \rangle$ denotes a context-based experience base, where r and s are the numbers of positive and negative experiences in context c; i.e., $c \in C^x$, $r = \sum_{i=1}^{|E_c^x|} e_{c,i}^x.+$ and $s = \sum_{i=1}^{|E_c^x|} e_{c,i}^x.-$. EB^x denotes the context-based experience bases of x.*

Given a context-based experience base, we derive opinions of agents about contexts using Wang and Singh's formulation, which supports that fact that uncertainty does not always decrease with more data and that uncertainty remains high when the numbers of positive and negative experiences are close to each other [Wang and Singh, 2007]. Definitions 6 and 7 show the mapping from an experience base to opinions.

Definition 6 (Certainty). *Given $EB_c^x = \langle r, s \rangle$, the certainty function $cert(r, s) = \frac{1}{2} \int_0^1 \left| \frac{x^r (1-x)^s}{\int_0^1 x^r (1-x)^s \, dx} - 1 \right| dx$.*

Definition 7 (Mapping). *Given $EB_c^x = \langle r, s \rangle$, the mapping to an opinion $w_c^x = \langle b, d, u, \alpha \rangle$ is defined as: $b = \beta \times cert(r, s), d = (1 - \beta) \times cert(r, s), u = 1 - cert(r, s)$; where α is the base-rate and $\beta = (r + 1)/(r + s + 2)$.*

2.3 Multi-context Decision-making

PAS should be able to make a decision for unseen privacy situations. First, PAS will compute which contexts an unseen situation belongs to (Method 2).

Method 2 (Situation Categorization). *PAS trains a multi-label classifier to take a privacy situation and the set of contexts for assigning a privacy situation to every context with a varying degree between 0 and 1. The sum of all degrees is 1.*

When PAS determines the opinions about multiple contexts (Definition 3), it then needs to reach a decision (i.e., allow data collection or not). PAS handles interactions in two categories. *Category 1* is the set of interactions where PAS chooses not to make a decision on behalf of the user because its knowledge about the situation is uncertain. This corresponds to a case where PAS identifies one dominant context (i.e., a context with a degree higher than 0.5) for a privacy situation, but the level of conflict in that context is high (Definition 8). Hence, PAS delegates the decision to the user.

Definition 8 (Conflict). *Given $EB_c^x = \langle r, s \rangle$, b, d, u values can be computed according to Definitions 6 and 7. The conflict is then computed as $|b - d|$, and is evaluated against a threshold ψ (a value between 0 and 1). If $|b - d| < \psi$ holds, there is a high conflict in experience in a given context c.*

Category 2 includes all other cases, where PAS makes a decision on behalf of the user. An opinion's probability expectation value can be computed according to Equation 1 (adapted

Algorithm 1: decide$(p, \psi, \theta, \gamma)$

```
   Data: EB^x, set of context-based experience bases
   Data: C^x, set of contexts
1  dec ← 0; i ← 0; wProb ← 0; cat_1 ← False
2  confSet ← categorizeSituation(p, C^x)
3  domC ← getDominantContext(confSet)
4  if domC then
5  |  b, d, u ← computeBDU(E^x_domC)
6  |  if |b − d| < ψ then
7  |  |  cat_1 ← True
8  if cat_1 then
9  |  dec ← askUser(p)
10 else
11 |  foreach c ∈ C^x do // p is of type cat_2
12 |  |  b, d, u ← computeBDU(E^x_c)
13 |  |  α = |0.6 − d|
14 |  |  prob ← getProb(b, u, α)
15 |  |  wProb ← wProb + confSet[i] * prob
16 |  |  i ← i + 1
17 |  if wProb > θ then
18 |  |  dec ← 1
19 i ← 0
20 foreach c ∈ C^x do
21 |  if confSet[i] > (max(confSet) − γ) then
22 |  |  EB^x_c ← updateExpBase(E^x_c, dec)
23 |  i ← i + 1
24 return dec
```

from [Jøsang *et al.*, 2006]). A base-rate parameter α is used to define the contribution of the uncertainty to the probability expectation value. If PAS interacts with a context few times (i.e., uncertainty is high), α ensures that an initial trust is assigned for the new context.

$$Prob(w_c^x) = b + \alpha \times u \tag{1}$$

Since PAS considers multiple contexts in making a decision, it computes a context-based weighted probability expectation value ($\overline{P}$) according to Equation 2. The weights are the degrees associated with each of the contexts. Note that the degrees are normalized values, d_c denotes the degree of belonging to context c.

$$\overline{P} = \sum_{c \in C^x} d_c \times Prob(w_c^x) \tag{2}$$

PAS then compares this value ($\overline{P}$) against a threshold θ (a value between 0 and 1) to make a final decision while dealing with a situation of type *Category 2*. The sharing decision gets a value of 1, when $\overline{P}$ is above θ (i.e., PAS allows data collection); and 0 otherwise. Higher θ values will result in less data sharing by PAS. After each decision, PAS updates its context-based experience bases according to a defined confidence threshold γ (e.g., Algorithm 1). Each interaction with an IoT device is treated as a positive experience if the data collection is allowed and negative otherwise. For each relevant context, r and s values are incremented accordingly.

3 The Formal Model in Action

The decision-making process of PAS is depicted in Figure 1(b) and realized with Algorithm 1. The algorithm takes the situation p to be categorized, the conflict ratio ψ, the decision threshold θ and the confidence threshold γ as input. The decision dec, the counter i, the combined probability expectation value $wProb$ and the category information cat_1 are initialized first (line 1). In line 2, PAS uses categorizeSituation function (Method 2) to assign the situation p to all the contexts with varying confidence scores denoted as $confSet$. The *n*th value in $confSet$ gives the confidence score for p to belong to context n. PAS first checks if there is a dominant context given $confSet$ (line 3), if so it checks the level of conflict in the dominant context regarding the specified conflict ratio ψ. In the case of a high-conflict (Definition 8), the situation p is of type *Category 1* (lines 4-7). If PAS deals with a situation of type *Category 1*, the user will be asked to decide to allow data collection or not (line 9). Otherwise, PAS will make an automated decision based on previous contextual experiences (lines 11-18). For each context, PAS computes belief, disbelief and uncertainty values for the specific context-based experience base (Definitions 6 and 7). α is set to $|0.6 - d|$ per context, which ensures that when the disbelief is low, uncertainty contributes more to the computed probability expectation value (line 15) (Equation 1). The probability value is updated for each context so that if PAS is more confident about a context assignment, the corresponding context probability will influence the resulting decision more. If $wProb$ is above the decision threshold θ, the decision dec is set to 1 to allow data sharing. Then, PAS updates the experience base for each context with acceptable confidence scores according to the decision. The confidence interval is set to $(\max(confSet) - \gamma)$, which defines the distance to the maximum confidence score for a context (lines 20-23). Finally, the algorithm returns the privacy decision (line 24).

4 A Real-World Case Study

To apply the model in real life, an IoT stack should be in place [Chow, 2017] to enable communications. We focus on the application layer and study the workings of the model using an anonymized dataset [Naeini *et al.*, 2017], which has been collected through surveys with users of IoT devices. The dataset includes 380 privacy scenarios that are split into 40 surveys, where each survey includes 14 scenarios. The participants are asked questions about particular scenarios (e.g., if the user would allow data collection) and their privacy concerns. Each survey includes responses from 20-25 participants. A scenario contains a textual description of features such as data type being collected, the purpose of data collection, the location where the data collection is happening and how long the data will be kept (e.g., Example 1). In our experimental setup, we keep one survey apart to conduct a qualitative and quantitative analysis (Section 4.2). Our supplementary material demonstrates how our approach works on two different examples for different participants as well as additional information on the participants' data. This material together with our code base is available online[1].

[1] https://git.ecdf.ed.ac.uk/nkokciya/pas-privacy

4.1 Situation Clustering, Context Classification

We use 366 scenarios from remaining surveys to: (i) generate contexts using clustering techniques, (ii) train a multi-label classifier to infer multiple contexts for unseen privacy scenarios. We use well-known Python libraries such as NLTK, Gensim and scikit-learn to implement our approach.

We train a *Doc2Vec* model to represent each scenario as a numeric vector by providing the normalized scenarios [Le and Mikolov, 2014]; such a model helps PAS to capture the semantic similarity of words in the text-based scenarios. The Elbow method suggests grouping scenarios into four clusters. In our setting, PAS uses a hierarchical clustering algorithm to discover clusters by using *ward linkage* criteria, where the similarity function sim is set to *euclidean* (Method 1). Hence, the algorithm minimizes the sum of squared differences within the clusters while merging or splitting clusters. The four clusters identified are C_0(140), C_1(27), C_2(39), C_3(160); where the number of instances belonging to that particular cluster is shown in parentheses.

Most scenarios fall into C_0 and C_3, where the former is broadly focused on non-personal data (e.g., temperature sensing in a room) and the latter unspecified data collection. C_1 spans situations where sensitive data (e.g., face) are used for identification and C_2 covers situations where the users are told how they will benefit from data collection. Each cluster corresponds to a context in our setting, where the context does not only capture location, or a topical setting but situations that differ from each other. This satisfies the points raised in Section 1. For example, library as a location is part of each context that we identify; showing that being in a library by itself is not significant for making privacy decisions.

To classify unseen situations into multiple contexts, we train a multi-label classifier. We have tried several classification models (SVM models with linear/rbf kernel, logistic regression models and so on), applied 5-fold cross-validation for model selection, and chose the model performing the best on average. In this case, this is the SVM model with a linear kernel (Method 2). The macro-averaged precision, recall and f1-score values on the test set are 0.91, 0.88, 0.9 respectively with an average accuracy of 0.86.

4.2 Experimental Results

We apply PAS for 25 participants to make a privacy decision on the scenarios specific to one survey. Each participant has only 14 scenarios labeled with privacy decisions, far less than a realistic setting. For this reason, for this case study, we populated the experience base of each participant by using similar participants' experiences, where the similar participants were chosen from the remaining 39 surveys. Two participants are considered similar if the difference between their average Internet Users' Information Privacy Concerns (IUIPC) scores is less than 1. This information was already present in the dataset. Note that in real life only the experiences of the user will be used. Since there is no dataset with this information, we are resorting to using similar users' data as well, while knowing that this will create approximation and lower accuracy in our results.

We categorize the participants into three categories based on their actual sharing behavior using the sharing decisions [Baarslag *et al.*, 2017]: *Behavioral Unconcerned* (Group 1-g_1) if they share more than 66% of the scenarios, *Behavioral Pragmatists* (Group 2-g_2) if they share between $33\% - 66\%$ of the scenarios, and *Behavioral Fundamentalists* (Group 3-g_3) if they share less than 33% of the scenarios. The dataset consists of 5 users (20%) in g_1, 16 users (64%) in g_2 and 4 users (16%) in g_3. Such a categorization is useful to understand how to set and update model parameters automatically as we discuss later.

		$\psi = 0.1$	$\psi = 0.2$	$\psi = 0.3$	$\psi = 0.4$
p_1	PAS−S	0.43	0.43	0.43	0.43
	PAS−M	0.50	0.50	0.50	0.50
	PAS	[0] 0.50	[0] 0.50	[0] 0.50	[4] 0.71
g_1	PAS−S	0.41	0.41	0.41	0.41
	PAS−M	0.41	0.41	0.41	0.41
	PAS	[2.4] 0.54	[2.4] 0.54	[2.6] 0.55	[4.8] 0.68
p_4	PAS−S	0.43	0.43	0.43	0.43
	PAS−M	0.50	0.50	0.50	0.50
	PAS	[0] 0.50	[2] 0.57	[4] 0.64	[4] 0.64
p_3	PAS−S	0.78	0.78	0.78	0.78
	PAS−M	0.71	0.71	0.71	0.71
	PAS	[4] 0.78	[5] 0.85	[5] 0.85	[5] 0.85
g_2	PAS−S	0.52	0.52	0.52	0.52
	PAS−M	0.57	0.57	0.57	0.57
	PAS	[0.7] 0.58	[1.3] 0.61	[3.4] 0.71	[4.3] 0.74
p_6	PAS−S	0.57	0.57	0.57	0.57
	PAS−M	0.78	0.78	0.78	0.78
	PAS	[0] 0.78	[0] 0.78	[4] 0.92	[4] 0.92
g_3	PAS−S	0.60	0.60	0.60	0.60
	PAS−M	0.83	0.83	0.83	0.83
	PAS	[0] 0.83	[1] 0.83	[3] 0.87	[4.5] 0.89

Table 1: The case study analysis results of participants (p_i) and groups (g_i) based on 250 experiences. We report the accuracy results based on varying conflict ratios (0.1, 0.2, 0.3, 0.4) for three different agents: PAS−S (single context + no human), PAS−M (multiple context + no human) and PAS (multiple context + human).

Table 1 shows results from our experiments with different conflict thresholds (0.1, 0.2, 0.3, 0.4) with a fixed set of 250 experiences. We include results for one participant from g_1 (p_1), two participants from g_2 (p_4 and p_3) and one participant from g_3 (p_6). We also include average values for each of the groups. For each run, we report results for three agents: PAS−S that does the computation based on single context (i.e., the context that has the highest confidence score) with no human intervention, PAS−M that does the computation based on multiple contexts with no human intervention (i.e., Algorithm 1 with $\psi = 0$), and PAS (Section 2). In Table 1, the number in the brackets is the number of times PAS delegates the decision to the user. We report accuracy values for all agents. The accuracy value for PAS is the value computed when the user cooperates with PAS to make sharing decisions. For example, in g_2, when the conflict ratio is set to 0.4; the users could see an average accuracy value of 0.74

by just making decisions for 4.3 scenarios in average. Hence, PAS would handle the remaining scenarios. Our group-based analysis could be summarized in three major themes.

Context. Deciding based on multiple contexts as opposed to a single context helps the agent to achieve better results. That is, PAS−M almost always performs either equally or better than PAS−S. For $g1$, they perform equally. For $g2$, for all values of conflict ratio ψ, PAS−M slightly outperforms PAS−S. For $g3$, the difference is more pronounced; e.g., when ψ is 0.4, PAS−M achieves 0.83, whereas PAS−S can only achieve 0.6. Note that these values do not factor in cases when the agent detects uncertainty and consults the user.

Collaboration. For any conflict ratio, when PAS consults its user, PAS outperforms the other agents (PAS−S and PAS−M) that run with no human intervention. For the previous case, when the agent consults the user for the ambiguous cases the accuracy increases to 0.89. Regardless of the group, when the conflict ratio increases, PAS accuracy values also increase. This shows that the instances that PAS consults the user for are mostly the ones it is making a mistake in. Thus, PAS identifies ambiguous cases correctly and by delegating those to the user enables PAS to reach a higher accuracy.

Personalization. We observe that different groups require different values for threshold θ and conflict ratio ψ for their PAS. g_1 has a tendency to share data more than other groups. This suggests that θ value should be set to a value less than 0.5. In other words, keeping θ high as in Table 1 requires keeping the conflict ratio high for better PAS accuracy values. For g_3, PAS performs well even for low conflict ratios; hence, human intervention could be minimized for such users by choosing low ψ values. For g_2, θ and ψ should be adjusted according to where the user stands (i.e., closer to g_1 or g_3). Thus, by observing its user's sharing pattern and adapting its parameters accordingly, PAS can help its user sufficiently while achieving a good accuracy.

5 Discussion and Conclusion

Most of the existing work on trust are based on the idea to build a model per agent over many interactions. Teacy *et al.* [2012] attack the problem by devising an algorithm that benefits from capturing hierarchical Bayesian modeling. Such an approach is successful in building trust models, but require large number of interactions and thus not immediately applicable in IoT settings. Liu and Datta [2012] develop a context aware dynamic trust, where they use a Hidden Markov Model to predict a service provider's next move. They consider interactions that are similar in certain features, rather than all interactions. Our intuition to capture context rather than previous interactions is similar but we estimate trust in devices that might never have been seen.

Burnett *et al.* [2013] study trust in short-lived interactions, similar to a setting as we have here. They propose a model of stereotypes, which enables agents to generalize their experiences with others over observable features. Fang *et al.* [2018] generalize this idea using a fuzzy semantic process and machine learning methods. In privacy, the context in which an agent resides has a tremendous effect, even when the agents exhibit similar observable properties; e.g., a user can trust a camera in a hospital but not in a bar. Thus, rather than stereotyping, our approach associates trust for each context.

Various techniques to enable agents to classify whether a content in question is private or not using various supervised machine learning algorithms or information retrieval techniques exist [Squicciarini *et al.*, 2017; Kurtan and Yolum, 2021]. These approaches either leverage big data sets or interactions with others to make privacy decisions. However, in IoT settings these are not available; hence, our focus has been on situations where there is little data and others' privacy opinions are not accessible.

Daidone *et al.* [2021] develop a blockchain-based framework to check privacy compliance of devices. They assume both user preferences and device policies are available in a structured form. Baarslag *et al.* [2017] design a negotiation strategy, where an agent makes partial offers based on utility-based heuristics to manage app permissions. Contrary to these, here we assume that the user preferences are derived based on the trust from contexts and situations are not provided in structured form but in natural language. Kökciyan and Yolum [2020] propose a multi-agent model, where an agent collects information from other agents to evaluate the trustworthiness of IoT devices before revealing its user's data. Differently here PAS does not keep individual trust values for each device but instead compute opinions about contexts.

Ajmeri *et al.* [2020] design norm-aware agents to make ethically appropriate decisions in social contexts. Mosca and Such [2021] propose an agent model to support multiuser privacy in online social networks. Our focus here has been on decision making with limited prior knowledge in new situations. The algorithm developed here is orthogonal to these works such that it would be interesting to incorporate personal values in conjunction with contexts.

We propose a novel agent-based privacy assistant PAS to handle interactions with IoT devices. PAS makes a sharing decision on behalf of the user, or it delegates the decision to the user, by modeling trust in multiple contexts. We show the applicability of our approach on an IoT case study. This work opens up interesting directions for research. Semantic relations between contexts could signal certain order to process contexts when making a decision. Obtaining more elaborate feedback from the user; e.g., "prefer not-share because the purpose is not specified." would enable PAS to assess a given situation in more depth than the textual representation alone.

Acknowledgments

This research was supported by the UKRI Strategic Priorities Fund via the REPHRAIN Research Centre, https://www.rephrain.ac.uk/; and by the Hybrid Intelligence Center, a 10-year programme funded by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research, https://hybrid-intelligence-centre.nl, grant number 024.004.022.

References

[Acquisti and Grossklags, 2005] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. *IEEE Secur. & Priv.*, 3(1):26–33, 2005.

[Ajmeri *et al.*, 2020] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Elessar: Ethics in norm-aware agents. In *AAMAS*, page 16–24, 2020.

[Alegre *et al.*, 2016] Unai Alegre, Juan Carlos Augusto, and Tony Clark. Engineering context-aware systems and applications: A survey. *J. of Sys. and Soft.*, 117:55–83, 2016.

[Baarslag *et al.*, 2017] Tim Baarslag, Alper T. Alan, Richard Gomer, Muddasser Alam, Charith Perera, Enrico Gerding, and m.c. schraefel. An automated negotiation agent for permission management. In *AAMAS*, page 380–390, 2017.

[Burnett *et al.*, 2013] Chris Burnett, Timothy J Norman, and Katia Sycara. Stereotypical trust and bias in dynamic multiagent systems. *ACM Tran. Intel. Sys. and Tech.*, 4(2):1–22, 2013.

[Chow, 2017] Richard Chow. The last mile for IoT privacy. *IEEE Security & Privacy*, 15(6):73–76, 2017.

[Chung *et al.*, 2017] Hyunji Chung, Michaela Iorga, Jeffrey Voas, and Sangjin Lee. Alexa, Can I trust you? *IEEE Computer*, 50(9):100–104, 2017.

[Colnago *et al.*, 2020] Jessica Colnago, Yuanyuan Feng, Tharangini Palanivel, Sarah Pearman, Megan Ung, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. Informing the design of a personalized privacy assistant for the internet of things. In *CHI*, page 1–13, 2020.

[Daidone *et al.*, 2021] Federico Daidone, Barbara Carminati, and Elena Ferrari. Blockchain-based privacy enforcement in the IoT domain. *IEEE Tran. on Dep. & Secure Comp.*, 2021.

[Das *et al.*, 2018] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personalized privacy assistants for the internet of things: Providing users with notice and choice. *IEEE Perv. Comp.*, 17(3):35–46, 2018.

[Dey, 2001] Anind K Dey. Understanding and using context. *Personal and Ubiquitous Computing*, 5(1):4–7, 2001.

[Fang *et al.*, 2018] Hui Fang, Jie Zhang, and Murat Şensoy. A generalized stereotype learning approach and its instantiation in trust modeling. *Electron. Commer. Res. Appl.*. 30:149–158, 2018.

[Fogues *et al.*, 2017] Ricard L Fogues, Pradeep K Murukannaiah, Jose M Such, and Munindar P Singh. SoSharP: Recommending sharing policies in multiuser privacy scenarios. *IEEE Int. Comp.*, 21(6):28–36, 2017.

[Jøsang *et al.*, 2006] Audun Jøsang, Ross F Hayward, and Simon Pope. Trust network analysis with subjective logic. In *Australasian CS Conf.*, pages 85–94, 2006.

[Kökciyan and Yolum, 2017] Nadin Kökciyan and Pınar Yolum. Context-based reasoning on privacy in Internet of Things. In *IJCAI*, pages 4738–4744, 2017.

[Kökciyan and Yolum, 2020] Nadin Kökciyan and Pınar Yolum. Turp: Managing trust for regulating privacy in Internet of Things. *IEEE Int. Comp.*, 24(6):9–16, 2020.

[Kurtan and Yolum, 2021] A Can Kurtan and Pınar Yolum. Assisting humans in privacy management: An agent-based approach. *JAAMAS*, 35(1):1–33, 2021.

[Le and Mikolov, 2014] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *ICML*, pages 1188–1196, 2014.

[Leon *et al.*, 2013] Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. What matters to users? factors that affect users' willingness to share information with online advertisers. In *SOUPS*, pages 1–12, 2013.

[Liu and Datta, 2012] Xin Liu and Anwitaman Datta. Modeling context aware dynamic trust using hidden markov model. In *AAAI*, page 1938–1944, 2012.

[Malhotra *et al.*, 2004] Naresh K Malhotra, Sung S Kim, and James Agarwal. Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information systems research*, 15(4):336–355, 2004.

[Mosca and Such, 2021] Francesca Mosca and Jose M. Such. ELVIRA: An explainable agent for value and utility-driven multiuser privacy. In *AAMAS*, page 916–924, 2021.

[Naeini *et al.*, 2017] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy expectations and preferences in an IoT world. In *SOUPS*, 399–412, 2017.

[Nissenbaum, 2004] Helen Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

[Sapienza and Falcone, 2020] Alessandro Sapienza and Rino Falcone. Evaluating agents' trustworthiness within virtual societies in case of no direct experience. *Cogn. Syst. Res.*, 64:164–173, 2020.

[Squicciarini *et al.*, 2017] Anna Squicciarini, Cornelia Caragea, and Rahul Balakavi. Toward automated online photo privacy. *ACM Tran. on the Web*, 11(1):1–29, 2017.

[Teacy *et al.*, 2012] WT Luke Teacy, Michael Luck, Alex Rogers, and Nicholas R Jennings. An efficient and versatile approach to trust and reputation using hierarchical Bayesian modelling. *Artif. Intel.*, 193:149–185, 2012.

[Ulusoy and Yolum, 2021] Onuralp Ulusoy and Pınar Yolum. PANOLA: A personal assistant for supporting users in preserving privacy. *ACM Tran. on Int. Tech.*, 22(1), 2021.

[Utz *et al.*, 2019] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (un) informed consent: Studying GDPR consent notices in the field. In *ACM SIGSAC on Comp. and Comm. Sec.*, 973–990, 2019.

[Voigt and Von dem Bussche, 2017] Paul Voigt and Axel Von dem Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*, volume 10. 2017.

[Wang and Singh, 2007] Yonghong Wang and Munindar P Singh. Formal trust model for multiagent systems. In *IJCAI*, volume 7, pages 1551–1556, 2007.

A Example Walkthroughs

Consider the following scenarios [Naeini *et al.*, 2017] that express a situation that combines information about a user (e.g., location) as well as an IoT device. When making a decision about Example 1, the contexts that would influence the decision can be the lack of personal data being collected (e.g., only presence), the purpose of collection, as well as retention of data. Thus, multiple contexts have to be formulated. Example 2 contains less information, where it is not clear whether personal information will be able to be identified. When making a decision, one needs to consider the influence of all the contexts that play a role.

Example 2. *You are at the library. This library has cameras that are recording video of the entire library. You are not told what the data is used for or how long it will be kept.*

We demonstrate how our approach works on Examples 1 and 2 when PAS considers 250 previous experiences to evaluate an unseen situation. The conflict threshold ψ is set to 0.4, and the decision threshold θ is 0.5. We will demonstrate how the examples work for two participants: p_3 and p_6. Table 2 shows the context-based experience bases and the corresponding opinions for both participants. Note that the base-rate parameter is computed dynamically according to Algorithm 1.

p_x	C_i	$EB^{p_x}_{C_i}$	$w^{p_x}_{C_i}$
p_3	C_0	$\langle 44, 4\rangle$	$\langle 0.73, 0.07, 0.20, 0.53\rangle$
	C_1	$\langle 13, 25\rangle$	$\langle 0.23, 0.43, 0.34, 0.17\rangle$
	C_2	$\langle 6, 8\rangle$	$\langle 0.23, 0.28, 0.49, 0.32\rangle$
	C_3	$\langle 70, 80\rangle$	$\langle 0.37, 0.42, 0.21, 0.18\rangle$
p_6	C_0	$\langle 42, 6\rangle$	$\langle 0.67, 0.10, 0.23, 0.5\rangle$
	C_1	$\langle 0, 38\rangle$	$\langle 0.02, 0.86, 0.11, 0.26\rangle$
	C_2	$\langle 3, 11\rangle$	$\langle 0.13, 0.45, 0.42, 0.15\rangle$
	C_3	$\langle 48, 101\rangle$	$\langle 0.26, 0.54, 0.20, 0.06\rangle$

Table 2: The initial context-based experiences and the corresponding opinions computed for the participants p_3 and p_6 when the experience base size is set to 250.

In Example 1, PAS classifies the situation as $C_0(0.40)$, $C_3(0.36)$, $C_2(0.16)$, $C_1(0.08)$, where the values in parentheses show the confidence scores. However, the expected probability values will defer for each agent since they are equipped with different context-based experience bases as shown in Table 2. p_3 computes the probability value as 0.56, which leads p_3 to allow data collection. This decision is aligned with the true label provided by the user. The same example looks quite different for p_6. p_6 does a similar probability computation for this user and decides not to allow data collection since the weighted probability value of 0.45 is below the decision threshold. This decision is also aligned with the true label provided by the user. The final decision would be a share decision for p_6, if it would only consider one single context with the highest confidence score (i.e., PAS−S). In other words, since $EB^{p_6}_{C_0}$ contains mostly positive experiences, p_6 would allow data collection. Our model considers all possible contexts that a privacy situation could belong. This leads p_6 to make a correct decision in this case by considering C_3, C_2 and C_1. This example demonstrates a typical multi-context scenario where a decision can only be reached by considering various contexts. This example is also interesting as it demonstrates aptly the personalized aspect of privacy and how our approach successfully caters to this.

In Example 2, PAS classifies the situation as $C_3(0.89)$, $C_2(0.05)$, $C_1(0.04)$, $C_0(0.02)$. This situation has one dominant context C_3. Our model ensures that, if there is a conflict that exists in a dominant context (Definition 8), PAS will consult the human to make a decision. p_3 computes the probability value as 0.41; however, p_3 detects conflict in C_3 since the difference between the belief and the disbelief values is 0.04 in $w^{p_3}_{C_3}$. It asks the user to make a decision, which is *deny data collection* in this case. PAS−M would make a not-share decision as well (which matches the true label); however, PAS chooses to consult its user in unclear cases. On the other hand, p_6 makes a probability computation that is 0.28. p_6 detects a conflict in C_3 according to $w^{p_6}_{C_3}$. p_6 asks the user to make a decision, which is *allow data collection* in this case. Note that PAS−M would make a not-share decision in this particular situation (which does not match the true label). PAS approaches the user when it is not confident about making an automated decision.

As we can see in both of these examples, the decisions made for the specific scenarios can be quite different for each user. Hence, personalized privacy assistants can be useful in managing privacy. Humans make conflicting privacy decisions even on similar occasions, which will also be the case for the agents as we observe here.

B Participants in the Survey

In our experimental setup, we report a qualitative and quantitative analysis on one survey [Naeini *et al.*, 2017] in Section 4.2. Of the 25 participants, 12 identified as female, and 12 as male. Their ages ranged from 23 to 70 with an average of 36.5 years. Thirteen participants ($\sim 50\%$) are aged between 30 and 40. Twelve participants held a high-school degree, ten held bachelor degrees, one had an associate degree and two had completed master/PhD degrees.

Participants were also asked ten questions about their privacy based on Internet Users' Information Privacy Concerns (IUIPC) scale [Malhotra *et al.*, 2004], which captures privacy concerns about data control, awareness and collection. Based on the provided responses, we computed an average privacy score for each participant, ranging from 1 to 7, which shows the level of privacy concern of a participant. The more privacy concerned participants are identified with higher scores. The average IUIPC score was computed as 6.03, with 3.9 and 7 as the lowest and the highest privacy scores respectively.