A View on Privacy & Trust in IoT

Jörg Daubert\textsuperscript{a,b} and Alexander Wiesmaier\textsuperscript{a} and Panayotis Kikiras\textsuperscript{a}
\textsuperscript{a}AGT International, Germany
\textsuperscript{b}Telecooperation Group, Technische Universität Darmstadt / CASED, Germany

Abstract—Internet of Things (IoT) technology is rapidly gaining popularity, not only in industrial and commercial environments, but also in our personal life by means of smart devices at home. Such devices often interconnect with cloud services that promise easy usage and global access. However, managing the balance between trust in the service provider and need for privacy of individuals becomes a major challenge considering automatic exchange of manifold personal information. In this paper, we propose a formal model that establishes a relation between information, privacy, as well as trust, and that automatically maps between these terms while maintaining user control.

Keywords—Internet of Things, Trust, Privacy

I. INTRODUCTION

The Internet of Things (IoT) started to gain new momentum in recent years as the consequence of the rapid growth of internet connected devices. However, security, in particular privacy and trust, remain major challenges as Sicari et al. \cite{1} recently pointed out. IoT applications, such as smart home, which for instance use sensors to recognize room presence to control the light, collect information in the most sensitive and personal domains of our life. In such automated communication the owner takes no active role in communication, but relies on devices and services to act on her behalf. Nevertheless, such personally identifiable information (PII) is not always processed by the owner of the information, but rather by an external service. For instance, smart home equipment suppliers also offer cloud-based services in conjunction with their devices. Moreover, IoT marketplaces have emerged and will continue to emerge \cite{2}. Such marketplace facilitate the trade not only of services but also PII as well. While services in the IoT domain may benefit from analyzing large amounts of PII, the privacy need of PII owners has to be considered as well. Compared with other technological revolutions, IoT has a more profound impact on privacy: first, the exchange of PII in IoT is not as precisely regulated as for instance medical records in healthcare scenarios. Second, IoT penetrates our personal life by means of sensing and acting devices at home and wearables. Prior research tackles the challenges of trust assessment and Privacy Enhancing Technologys (PETs) independently. However, both challenges seem to be interconnected as trusted parties may process private data, whereas our need for privacy seems to increase when an un-trusted party is involved.

This paper analyzes definitions and relations among trust and privacy in Section II, followed by a proposal for a new model to link privacy and trust in a formal way. For that, a user-centric perspective is used, where the owner of PII shares information within an IoT environment while protecting her privacy. Section III discusses multi-dimensional privacy and how it can be mapped to our model. Section IV extends the model to multi-dimensional trust. Section V concludes this paper and provides an outlook onto future work.

II. PRIVACY, TRUST, AND INFORMATION

This section presents a novel view on the relation between privacy and trust. Manifold definitions of the terms privacy and trust exist. Hence, we first elaboration on the terminology at hand in Section II-A. Second, we propose a model for the relation among these terms in Section II-B

A. Definition of Terminology

Commonly agreed definitions of terms such as trust and privacy do not exist. For instance, Uslaner introduces trust as “chicken soup of social life” \cite{3}, and thus highlights the mere flavors and mysterious definition of this term. Likewise, multiple definitions for privacy exist. Therefore we summarize definitions and extract a common interpretation from an information centric perspective, i.e., the perspective of a subject that owns the information and wants to share it with a service provider. The focus of information is personally identifiable information (PII) \cite{4} as term summarizing all information that could potentially relate to a person, and is therefore relevant for privacy.

a) Privacy: Maslow identifies a “need for privacy” as a core property of self-actualizing, the tip of his hierarchy of needs \cite{5}. Other definitions categorize privacy as requirement in the context of PII, e.g., “personally identifiable information (PII) leaving the control of the person whose PII it is” \cite{6}. More detailed, Lessig defines privacy as the combination of “empowerment to control”, “utility to protect”, “dignity to establish an equilibrium”, and “regulating agent to balance power” \cite{7}.

Alongside privacy as property that can be either fulfilled or not, privacy can be measured by a metric. Aimeur et al. \cite{8} measure privacy on a discrete scale from no privacy over soft privacy and hard privacy to full privacy depending on how much PII is cryptographically protected from an attacker. Bohli and Pashalidis \cite{9} model relations between privacy notions, e.g., “strong unlinkability with participant hiding”, and map other definitions of anonymity, pseudonymity, and unlinkability to their model. The remainder of this paper uses privacy as a measurement for the need of privacy (protection). This definition harmonizes with existing ones as it quantifies the need for privacy (Maslow \cite{5}) while establishing compatibility with the technical definition as a requirement (Poore \cite{6}). Moreover, this definition puts privacy into perspective with the notion PII.

Formally, we assume a metric $P$ for privacy that can be normalized to the interval $P \in [0, 1]$, indicating $P = 0$ for no
privacy and \( P = 1 \) for absolute privacy in the corresponding metric. For example, k-anonymity [10] can be normalized to \( P_{k-\text{anon}} = k/n \) where \( n \) is the full population of records (information) and \( k \) the size of indistinguishable records, and thus the anonymous subset.

b) Trust: Uslaner uses the chicken soup of social life metaphor to describe trust [3]. For Bao and Chen trust is a compound of the properties “honesty, cooperativeness, and community interests” [11]. According to Poore, trust follows the definition of privacy and is “the assurance that PII will be only used as agreed and will be protected against unauthorized access” [6]. Leister and Schulz follow the notion of trust as a scalar metric [12].

Following the information centric perspective, we define trust as a *measurement for the need of trust*, similar to privacy. This definition mirrors trust assessment approaches, e.g., recommendation and reputation systems, which calculate the trustworthiness of one subject to match it against the need for trust of another subject.

Formally, we define (analogously to privacy) trust \( T \) as a metric normalized to \([0, 1]\) where \( T = 0 \) means no trust at all and \( T = 1 \) stands for universal trust. For example, in [13] the authors show that a probabilistic model can extend a no/partial/yes model in GnuPG to a continuous \([0, 1]\) trust scale for the authenticity of public keys.

c) Sensitivity: Existing definitions of privacy and trust relate to the term PII, however definitions of PII itself differ as well. European law categorizes information into two classes, information covered by Directive 95/46/EG [14] (PII), and all other information (non-PII). Narayanan et al. further subdivide PII into directly identifying, e.g., a unique citizen ID number, compared to conjunctive identifying, e.g., name, place, and date of birth [15]. Jabeur et al. distinguish minimum, medium, and maximum sensitivity for PII in such a hierarchy [16]. Role-based access control mechanisms establish relations among roles [17] to reflect a multitude subdivisions regarding the sensitivity of the controlled information.

To generalize these concepts, we define sensitivity of information \( S \) as a metric normalized to \([0, 1]\). For example following Directive 95/46/EG, we measure a directly identifying information with \( S = 1 \) whereas a conjunctive identifying information is measures as \( S = 1/n \) with \( n \) denoting the number of required items of information for unique identification.

**B. Relations among privacy and trust notions**

With the definitions of privacy, trust, and sensitivity, this section elaborates on the relations among these notions.

We propose the relation given by Equation 1 to connect trust to privacy, privacy to sensitivity, and sensitivity to PII. This relation follows the insight that a higher need for trust is caused by a higher need of privacy. The need for privacy is consequently higher the more sensitive information is. Finally, the sensitivity of information depends the actual information (PII) itself.

\[
\begin{align*}
T & \leftrightarrow P \leftrightarrow S \leftrightarrow PII \\
\text{privacy: } & S \rightarrow P \quad (3)
\end{align*}
\]

Via this relation, it becomes algorithmically decidable how much trust is required to process given PII. Likewise, it becomes decidable what PII can be provided given a certain level of trust. The next paragraphs discuss how such a mapping can be implemented such that it becomes algorithmically decidable.

a) Sensitivity: The sensitivity \( S \) is a metric for PII and therefore is defined by a function \( \text{sensitivity} \) that takes PII as input domain and provides \( S \) as output domain (Function 2).

\[
\text{sensitivity: } PII \rightarrow S 
\]

To illustrate this mapping, we assume 20 elements of PII in partial order according to their sensitivity as an example. Figure 1 depicts sensitivity \( S \) over PII as given by some regulation (circles, red) versus PII owner specified preferences (triangles, blue). Elements 1 . . . 4 are hardly suitable to identify a person, elements 5 . . . 16 require many to few other items to identify a person, whereas elements 17 . . . 20 uniquely identify a person. As elements 5 . . . 8 illustrate, the owners’ judgement regarding sensitive may deviate from regulation.

The mapping to sensitivity rather than dealing with raw PII has the following advantages: first, a unified representation incorporating all laws, regulations, and policies. Second, the possibility for the owner of the PII to express personal preferences. Third, the function \( \text{sensitivity} \) can be based solely on how sensitive the PII is considered without the need to consider sharing of information and trust in service providers yet. In Figure 1, rules, regulation, and policies are considered as a default, a minimal consensus. The user may override this default with his own utility function for sensitivity.

![Figure 1. Information sensitivity over PII. The function models the personal view on how sensitive information is.](image-url)
Figure 2. Privacy over sensitivity. The function models the owners’ need for privacy. The dots show the application to the 20 example items.

Figure 2 illustrates the mapping from sensitivity to privacy using the sample data introduced in the previous paragraph. We use the owner defined Function 4 as an example to model a particular instance of this mapping. Here, the owner always desires a minimum of privacy (constant component +0.1) and considers sensitive information exponentially more private (component \(S^2\)).

\[
privacy(S)_{\text{example}} := S^2 + 0.1
\] (4)

Our model enables the formalization of privacy without having to consider PII directly, i.e., it provides a layer of abstraction. Furthermore, privacy can be easily customized to fit the needs, e.g., a regulation that applies to multiple PII owners as well as personal preferences of PII owners as in the example.

c) Trust: The notion of privacy so far represents the view of one owner of PII. Trust finally incorporates a second party, the service provider that desires PII. The need of trust \(T\) expresses how much trust the PII owner has to invest in the service provider—or how trustworthy the service provider has to be.

For example, the owner wants to process PII using smart home cloud service providers. Then the owner first uses a computational trust framework such as from Habib et al. [18]. Second, the owner obtains the trust scores of cloud services providers according to own criteria and weight. Third, the owner specifies a \(trust_{\text{service}}X\) function for every cloud service provider \(X\) according to scores’ uncertainty and personal beliefs. Finally, the model can calculate, compare, and thus automatically decide if the cloud service provider \(X\) fulfills the desired level of trust.

Function 5 formalizes the mapping between privacy and trust.

\[
trust: P \rightarrow T
\] (5)

Figure 3 illustrates such a mapping for our example with the mapping Function 6 as an instance of an owner defined function. This function requires the service provider to show significant trustworthiness, even for hardly private data. The function also acknowledges that a trust desire of 100% can be rarely fulfilled and thus a reasonably trustworthy service provider may handle all PII.

\[
trust(P)_{\text{example}} := \log_3(P + 1.2)
\] (6)

Our model only requires a loose coupling between privacy and trust via an owner customizable function. As a result, the trustworthiness assessment can be obtained from an arbitrary system, while the the calculation of sensitivity and privacy can be performed independently.

d) PII to Privacy: The formal representation of our model can be used to directly calculate privacy desire given PII. By combining Functions 2 and 3, we obtain Function 7, as well as Function 8 for the example.

\[
privacy'(PII) : PII \rightarrow P
\] (7)

\[
privacy'(PII)_{\text{example}} := sensitivity(PII)^2 + 0.1
\] (8)

Figure 4 illustrates this mapping for our example. Such a representation empowers the data owner to explore and understand the influence of PII classification on privacy.

e) Sensitivity to Trust: Similar to the previous paragraph, we also use this model to explore the relation between sensitivity and trust. Functions 9, as well as Function 10 for the example, formalize this relation.

\[
trust' : S \rightarrow T
\] (9)

\[
trust'(S)_{\text{example}} := \log_3(S^2 + 1.3)
\] (10)

Figure 5 illustrates this mapping for our example. Such a representation empowers the data owner to explore and understand how sensitive information might be to be handled by a service provider given the trustworthyness.
f) Summary: PII to trust: The presented three mappings, from PII to sensitivity, from sensitivity to privacy, and from privacy to trust allow the owner to model personal preferences individually by defining the mapping functions. By combining all three functions, this formalization translates from PII to trust need while remaining customizable and transparent. Function 11 formalized this relation, and Function 12 completes the example.

\[ \text{trust}'' : \text{PII} \rightarrow \mathcal{T} \]  
\[ \text{trust}'(\text{PII})_{\text{example}} := \log_3(\text{sensitivity}(\text{PII})^2 + 1.3) \]  

Figure 6 illustrates this mapping for our example. Such a representation empowers the data owner to explore and understand exactly what PII can be handled by a service provider given a certain trustworthiness.

III. DIMENSIONS OF PRIVACY

In Section II-A, we elaborate on definitions for the term privacy and establish privacy as a scalar metric \( \mathcal{P} \) for the mapping in between sensitivity and trust. However, literature suggests that privacy is perceived as a multi-dimensional metric rather than a scalar. Therefore, we discuss our model in the context of multi-dimensional privacy. Based on that model, we propose a multi-dimensional version of the mapping between trust and and PII, i.e., we introduce the metric \( \mathcal{P}' \) as an extension of \( \mathcal{P} \).

Barker et al. categorize privacy of PII alongside the dimensions purpose, visibility, as well as granularity, and propose a 4-dimensional representation of privacy that includes the duration of the permitted usage as well [19]. Domingo-Ferrer establishes a 3-dimensional representation of privacy with respect to the involved entities: the owner who controls the PII, the service provider (called respondent) who leaks information by answering a questions, and the user who leaks information by asking the service provider a question [20].

These definitions express privacy in a context, e.g., where the data is shared, who accesses it, and for which purpose. To complement these definitions, this papers discusses the question “How private is this piece of PII?” from the owners’ perspective. Hence we measure sensitivity and privacy along dimensions of the origination of PII rather than their destination.

Privacy can be categorized along several dimensions [21]. The next paragraphs briefly discuss four of these dimensions, namely identity privacy, location privacy, footprint privacy, and query privacy. Other sets of privacy dimensions may be argued and are also applicable to our model.

A. Identity Privacy

Identity privacy refers to the need of privacy for information that can identify a person. Many approaches [22], [23], [10] have been proposed to comply to this need, e.g., by performing pseudonymization or anonymization of such PII. Likewise, approaches to de-anonymize persons from such protected information have been proposed [24], [15].

B. Location Privacy

Location privacy refers to a specific and widely researched form of footprint privacy. Location can reveal manyfold PII, e.g., points of interest [25] and thus even the religion given by the location of a church. Therefore approaches to comply to the need of location privacy such as MIX-zones [26] and sharing of location slices [27] have been proposed.
C. Footprint Privacy

Footprint privacy refers to the privacy need for all PII that gets leaks unintentionally, e.g., preferred language and operating system when browsing the web. Such data is also referred as metadata and micro data. Like identity privacy, approaches to fulfill this need exist [28]. However, the sole fact of acting may leave a footprint without the possibility of avoidance.

D. Query Privacy

Query privacy refers to need for privacy regarding the PII contained in a query, e.g., weather forecasts require location and date. Methods for answering queries while respecting the need for privacy exist [29], however the fact that a query has been made is hardly hideable, e.g., the fact someone is interested in weather.

E. Multidimensional Privacy

We extend the privacy metric $\mathcal{P}$ to $\vec{\mathcal{P}}$ as defined by Equation 13 to reflect the multiple dimensions of privacy, and to allow the owner to specify personal preferences in each dimension independently.

$$\text{privacy} : S \longrightarrow \vec{\mathcal{P}} \quad (13)$$

IV. DIMENSIONS OF TRUST

In Section II-A, we deal with some definitions of trust and establish trust as a scalar metric $\mathcal{T}$ that is later on mapped to privacy, sensitivity and PII. As with privacy, literature suggests that trust is rather a multi-dimensional metric and we introduce the notion of multi-dimensional trust and expand our model by $\vec{T}$.

Roman et al. split trust into the two dimensions of “trust in the interaction between entities” and “trust in the system from the point of view of the user” [30]. Likewise, Hochleitner et al. distinguish between system trust and interaction trust [31]. Bao and Chen identify three trust properties honesty, cooperativeness, and community interest [11].

As with privacy, we follow a user centric approach and aim for answering the question “How trustworthy is the service provider for this PII?”. Hence we measure trust along dimensions of the handling of the PII and compare the trust assessment with the trust need.

The next paragraphs briefly discuss four of these dimensions, namely device trust, processing trust (service provider), connection trust and system trust (overall perspective). Other sets of trust dimensions may be argued and are also applicable to our model.

A. Device Trust

Device trust refers to the need to interact with reliable devices such as sensors and actuators. Common approaches to meet this goal are trusted computing [32] and trusted software [33].

B. Processing Trust

Processing trust expresses the need to deal with correct and meaningful data. This is usually achieved by accurate data gathering combined with suitable data analytics. The results may be further improved by data fusion.

C. Connection Trust

Connection trust stands for the requirement to exchange the right data with right service providers and only with them. This is achieved by ensuring canonical security goals such as confidentiality, authenticity, and non-repudiation.

D. System Trust

System trust refers to the desire to leverage a dependable overall system. This kind of trust is commonly achieved by providing transparency to all involved subjects regarding workflows, processes, underlying technology an so on, e.g., by passing respective certifications.

E. Multidimensional Trust

We extend the trust metric $\mathcal{T}$ to $\vec{T}$ as defined by Equation 14 to reflect the multiple dimensions of trust. Note that we make use of the previously introduced multidimensional privacy metric $\vec{\mathcal{P}}$. Altogether we then come to the multi-dimensional model as expressed by Equation 15.

$$\text{trust} : \vec{\mathcal{P}} \longrightarrow \vec{T} \quad (14)$$

$$\vec{T} \longleftrightarrow \vec{\mathcal{P}} \longleftrightarrow S \longleftrightarrow \text{PII} \quad (15)$$

Multidimensional trust can be visualized to user for instance via radar charts. Even variations that incorporate uncertainty in case trust cannot be assessed absolutely exist [34] and seem to be easy to assess by users.

V. CONCLUSION

This paper presents a novel view on the relations between PII, sensitivity of information, privacy, and trust. Furthermore, these relations can be expressed in a formal way, but yet abstracts complexity by decoupling the need for trust from PII. We also showed that this formal mapping can be expressed in a multi-dimensional space to express the need for privacy fine-grained, e.g., according to identity, location, query, and footprint.

We expect that this model will easy new application scenarios in the IoT, Machine 2 Machine (M2M), as well as the Big Data domain where sharing of PII is rapidly evolving towards an automated process. Future work will address the challenge of leveling gaps between the need for trust provided by the data owner and the trustworthiness of a service provider. In particular, we will model the type of service provider as another dimension to express multiple service providers and personal preferences towards them in one model. This work can be included into automatic policy negotiation frameworks in conjunction with PETs.
REFERENCES


