Implementability of the Identity Management Part in Pfitzmann/Hansen's Terminology for a Complex Digital World

Manuela Berg, Katrin Borcea-Pfitzmann

Implementability of the Identity Management Part in Pfitzmann/Hansen's Terminology for a Complex Digital World

Manuela Berg and Katrin Borcea-Pfitzmann

Technische Universität Dresden, Faculty of Computer Science
D-01062 Dresden, Germany
{manuela.berg,katrin.borcea}@tu-dresden.de
http://dud.inf.tu-dresden.de

In memory of Andreas Pfitzmann

Abstract. Based on a widely cited terminology, this paper provides different interpretations of concepts introduced in the terminology asking for an implementable privacy model for computer-mediated interactions between individuals. A separation of the digital world and the physical world is proposed, as well as a linkage of the two worlds. The digital world contains digital representations of individuals and it consists of pure data. The physical world contains individuals and it consists of information (produced by individuals) and data. Moreover, a refined definition of privacy is being elaborated that serves as justification for identity management of individuals interested in a sophisticated perspective of privacy.

1 Introduction

Since time immemorial, relationships between individuals have been a need of mankind. Interaction, meaning mutual action and communication, between individuals is the basis for establishing a social system [1]. During the last decades, technology rapidly developed. Computer-mediated interaction became important and reached its current high point with the Web 2.0 movement. In contrast to early computer-mediated interaction, users are now more and more active in content and application production.

User profiles and content of web applications are sources of personal data. Apart from explicit publishing of personal data, users also implicitly disclose personal attitudes, opinions, or even personal statements within bulletin boards, blogs etc. Moreover, users publish data non-related to themselves. In many cases, users are not aware of the risks connected with the possibilities of linking different sources of information.

Applications are usually built to achive a certain functionality (such as communication). Privacy is usually considered as a secondary (or even tertiary) functionality, which is valuable for individuals if the primary functionality is fulfilled. Consequently, privacy aspects are not involved in many applications. An

identity management system is an application, which fulfills privacy aspects as primary functionality. This could be a reason, why identity management systems are not accepted in the public. No matter, which importance the functionality of privacy has, it is of importance. We elaborate a definition of privacy based on existing approaches.

We consider computer-mediated interaction triggered by individuals[1]. On the one hand, this interactions generate networks of individuals. We refer to the amount of individuals as *physical world*. On the other hand, computer-mediated interactions generate networks of digital representations of the individuals, which we refer to as *digital world*.[2] Of course, individuals and their digital representations are linked, so they are not independent from each other. The digital world in all its complexity and its linkage to the individuals is what we refer to when talking about a *complex digital world*.

The authors of [2] introduce the concept of privacy-enhancing identity management using partial identities, i.e., subsets of an individual's digital identity, to tackle potential privacy problems caused by linking information from different sources to the individual's identity. Related to this, Pfitzmann and Hansen continuously work on a comprehensive terminology framing the area of privacy-related terms [3]. The objective is to start an analysis of that widely accepted and cited terminology from the perspective of the complex digital world.

In this paper, we reason that two worlds have to be distinguished, the physical world and the digital world. The two worlds are linked by two functions – $Ego(\,)$ and $Oge(\,)$. While the latter maps from the physical world to the digital world and delegates the individual's communication to the digital representation, the former indicated function maps from the digital world to the physical world and carries information of the communication of digital representations back to the individuals. Moreover, we propose an extended definition of privacy, which is described as a state attained by a desire of an individual. Given computer-mediated communication, this state influences the communication of an individual and its digital representations with others in the complex digital world. Considering the building of applications and the separation of the digital and the physical world, we give a definition of privacy. We also discuss this definition regarding concepts of privacy as given in [4] and [5].

Accordingly, the following section will frame the problem space by briefly introducing the kind of complexity shaping the complex digital world from an analytical point of view and stating the research questions being the road map for the following sections. In Sect. 3, we discuss the Pfitzmann/Hansen terminology by confronting it with the setting of the complex digital world. Finally, Sect. 4 takes up the issues figured out in the previous section and discusses the separation of the physical as well as the digital worlds. Further, this section proposes an enhanced definition of privacy applied on personal interaction as well as on

[1] Organizations and computers can be imagined to trigger computer-mediated interaction, too, but we do not consider this case here.

[2] Besides digital representation, there might be other items in the digital world that we do not consider in this paper.

computer-mediated interaction, which is discussed in terms of known concepts of privacy.

2 Framing the Problem Space

Regarding the complex digital world, individuals are considered as actors in the physical world and digital representations of the individuals are considered as actors in the digital world. We call an actor *entity*. Further, we call connections between entities resulting from possible or occurred communication a *relationship*. With respect to an established relationship, we call an entity that participates in this relationship *involved entity*. Of course, relationships might be created even if there has not been any direct communication between the involved parties yet. Consequently, those relationships can be manifested by the involved entities or by non-involved entities.

Traditionally, communication is modeled using the bilateral sender-receiver model and privacy has been studied mainly in scenarios involving service providers and service consumers. We argue that communication, and so interaction, is more complex:

(a) interaction does not only take place between one particular user and one or more service provider(s). In this paper, we focus on interactions between individuals[3]. The implications on privacy protection of them have not yet been studied in detail.
(b) interaction does not only occur bilaterally. Given that at one particular point in time, there exists one sender at maximum for a message, there might, nevertheless, be several recipients (or zero or only one) for this message. Interactions constitutes a dynamic system, which develops over time. An interaction is usually describable by a bunch of message that are sent by different individuals. In the sender-receiver model, more than one message can be regarded only if the messages are considered as a totally ordered sequence while in the complex digital world, messages can be transmitted concurrently. However, there is a certain order of the messages (for example ordered by the time of sending). We call a bunch of messages *interaction*, if one recipient of a former message becomes a sender and the sender of a former message becomes a receiver. Consequently there are usually several senders in an interaction.
(c) between two entities, there might exist relationships of different quality and quantity. We defined a relationship between entities as a connection resulting from communication independent from any semantics and quantity. Further research should extend the understanding of relationships by more differentiations, i.e., by considering relationship semantics such as "is friend of" or "is colleague of" and take into account frequencies and durations of communications as well as preferred means of communication.

[3] Usually, entities comprise organizations and technical devices, apart from individuals.

(d) to be mentioned here is that entities might act differently in different situations.

The authors in [6] address the above mentioned issues. They introduce the concepts of entity, view, relationship, and context as parts of a model that touches various aspects of a complex digital world. Regarding the analysis of the Pfitzmann/Hansen terminology [3], this paper particularly focuses on entities and relationships.

In the complex digital world as outlined above, we raise the following research questions:

1. Are there concepts in [3] that make it necessary to separate individuals and their digital representation?
2. What precisely are the entities, and how are entities characterized when transcending the borders of the physical and digital world?
3. What are the interests of entities regarding the disclosure of personal information?
4. Having learned about entities and their interests, does this help us to develop an application that supports entities pursuing their interests?

3 Analyzing the Notion of Identities in the Terminology of Pfitzmann/Hansen

In 2000, Pfitzmann and Köhntopp[4] published a terminology describing the concepts of anonymity, unobservability, and pseudonymity [7]. In June 2004, they extended their terminology collection with definitions of terms related to identity management with regard to data minimization and user-control. Since then, the authors continuously extended and refined that terminology. The most recent version [3] serves as basis for our analysis regarding the questions raised in Sect. 2. In [3], the traditional sender-receiver model is assumed, where usually one message is considered and a subject is defined as a human being, a legal person or a computer. Different kinds of relationship and different situations are not considered. Table 1 compares the setting of the Pfitzmann/Hansen's terminology with the setting of the complex digital world.
We analyze the questions raised in Sect. 2:

1. Anonymity in [3] is defined as follows.

   Definition 1. *"Anonymity of a subject from an attacker's perspective means that the attacker cannot sufficiently identify the subject within a set of subjects, the anonymity set."*

   Subject is defined in [3] and covers the physical body with its feelings, needs and also its digital representation. The German law for data protection [8, §3

[4] Taking identity management serious, Hansen and Köhntopp are names for the same person.

Comparison Property	Pfitzmann/Hansen's Terminology	Complex Digital World
a)		
service providers	considered	not considered yet
b)		
number of receivers	one	arbitrary
number of senders	one or as a sequence of message sendings	arbitrary
companies and organizations	considered	not considered yet
number of messages	one or many in a sequence	arbitrary number
concurrence of messages	only sequential messages	arbitrary concurrence
c)		
different kinds of relationships	not considered	considered
d)		
different situations	not considered	considered

Table 1. The comparison of the complex digital world and the model of Pfitzmann/Hansen

(6)] defines anonymization as the act of transforming personal data in a way to make it difficult or impossible to link personal data to the individual. Since an attacker can be settled either in the physical world or in the digital world (depending on what the attacker wants to find out) and since personal data is part of the individual's digital representation, we can distinguish three types of anonymity:

(a) *An individual is anonymous in a set of individuals,* which means either that an attacker in the physical world wants to identify an entity in the physical world.

(b) *Anonymity is regarded as unlinkability between an individual and his digital representations,* which either means that an attacker in the physical world (which has control over his digital representations) wants to find out which digital representations belong to a physical entity (or which physical entity belongs to a digital representation) or it means that the attacker in the physical world wants to link some digital representations belonging to the same entity or the same group, e.g., the group of employees of one company.

(c) *A digital representation is anonymous in a set of digital representations,* which means that an attacker in the digital world wants to identify an entity in the digital world.

Depending on the respective situation, an individual has different preferences towards the kind of anonymity to be protected most. The three types of anonymity are not independent from each other. If, for example, a digital representation is not anonym and it is also linkable to its individual, then the corresponding individual can not be anonymous neither. Further, we have

to admit that representing an individual just by data might not be enough to characterize this individual (see Sect. 4).

2. Referring to individual persons, the authors of [3] define identity as follows:

   Definition 2. ***(Identity)*** *"An identity is any subset of attribute values of an individual person which sufficiently identifies this individual person within any set of persons."*

   The definition implies that every individual may have more than one identity. We will reveal different interpretations of this concept.
   (a) We can consider identity as a socio-centric concept. This means that identities are omnipresent. For example, an all-knowing entity would consider the information he knows about an individual as an individual's identity. This means that information gained from perceptions of this individual about his interaction partners might be included in his identity. Such an approach of understanding the concept of identity might be interesting for service providers, but it is not reasonable for individuals caring for their privacy because individuals should have control of their data.
   (b) The term identity can also be considered as an ego-centric concept. For an individual that wants to show himself differently to different people, this is the more appropriate approach. However, this implies even more questions: If we assume that an individual establishes an identity of himself, then this is a digital representation of himself. This approach of defining identity, however, misses the aspect of perceptions. We will refer to the individual's perception of an interaction partner as *view of an individual on an interaction partner*. The concept of a view can be realized as a set[5] in an application. On the one hand, an individual's digital representation has views on its interaction partners. So the question arises, whether these views are part of the identity established by the individual. On the other hand, the interaction partner has a particular view of the individual. This leads to the question whether the view can be compared with the identity and if they can be compared, then one may ask when do the identity and the view coincide.

3. Every individual has his own attitudes regarding his privacy. When discussing privacy aspects, Pfitzmann and Hansen refer to the definition given in [9]:

   Definition 3. ***(Privacy)*** *"Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."*

   In our view, this definition misses some aspects of variety. Nevertheless, the sentences following the cited definition in [9] contain some of our intended aspects. For example, negotiation between functionality and non-disclosure of data as well as the dependency of situations are two aspects that are mentioned and shall be included in an extended definition (see Sect. 4).

[5] A View might also be an empty set.

4. Pfitzmann and Hansen embed the definitions of their terminology in an environment aiming at achieving privacy by data minimization through user-controlled identity management. However, the terminology provides little information of how to implement the concepts of privacy protection when designing applications for the digital world.

The above-mentioned issues describe first problems that can be solved by separating the physical and the digital world. Consequently, the following section points out a first approach to model privacy and identity-related concepts in a complex digital world.

4 Approach to Model a Complex Digital World with Respect to Privacy and Identity Management

We propose a model for a complex digital world referring to the problems raised in Sect. 3. The three types of anonymity elaborated there require a separation of the digital representation from the individual of the physical world. We call the digital representation *digital entity* and the individual *physical entity*.
Similarly, Semančík distinguishes the notions of entities in the digital world and in the physical world as follows: *native digital entities* (e.g., software components), *physical entities* (persons of the physical world), and *digital proxies* (digital representations of a physical entity) [10]. Thereby, digital proxies (data structures characterizing the entity) are representations of the physical entities. Also, a physical entity may have several native digital entities and several digital proxies in "the same or in a different computer system." We neglect the notion of native digital entities here because we are interested in the (inter)action of the entities. Moreover, we think that native digital entities and digital proxies can be subsumed under the set of digital entities.
Data and information were discussed by several authors as parts of the knowledge hierarchy for information systems and knowledge management. Two possible differentiators between data and information have been identified in [11]: meaning and structure. So, information is defined either as data that have been given meaning (or a value) or as data that has been processed or organized. The choice of the differentiator implies consequences where information is embedded: either in systems or in individuals' minds or in both. Usually, information defined by meaning either is said to be produced by individuals or nothing is said about the producer of information. We refer to meaning as the differentiator and assume that information can only be produced by individuals[6].
So, the digital entity consists of pure data while the physical entity can produce information. More precisely, the physical entity has feelings and needs (e.g., regarding communication and privacy) and he is able to interpret data, i.e., produce information.

[6] If one considers organizations then they produce information only by individuals. Computers can not produce information at all. But individuals can use data out of computers to get information.

In [12], a standardization of data and information was given which has been rejected in the meanwhile. The definitions of data and information are broadly discussed as a part of the wisdom hierarchy (also known as DIKW hierarchy), e.g., in [13], [14] and [15]. We refer to the definitions of [13] as given in [11]:
Data are defined as symbols that represent properties of objects, events and their environment. They are the products of observation. But are of no use until they are in a useable (i.e. relevant) form. The difference between data and information is functional, not structural.
Information is contained in descriptions, answers to questions that begin with such words as who, what, when and how many. Information systems generate store, retrieve, and process data. Information is inferred from data.

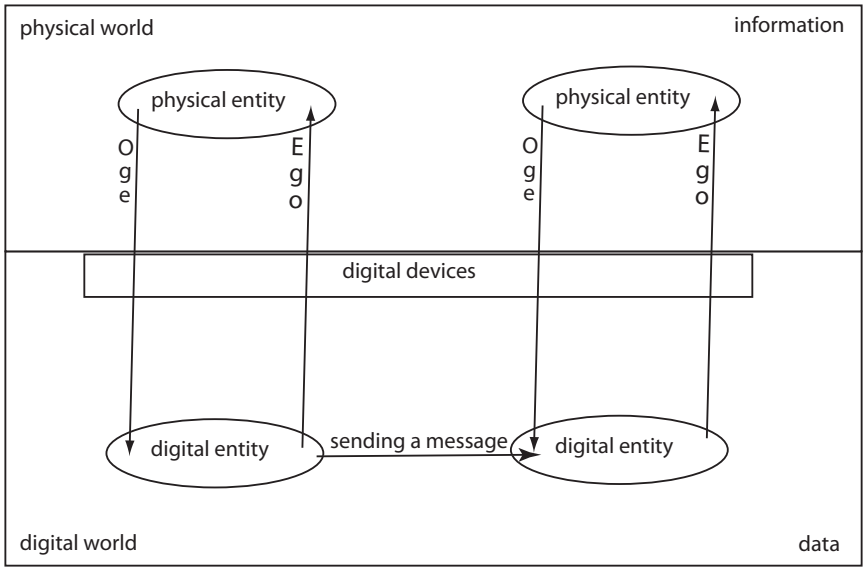

Fig. 1. The world of data and the world of information are connected by the functions Oge( ) and Ego( ).

The digital entities exist in the digital world, which is a world of pure data, and the physical entity is part of the physical world (cf. Fig. 1), where information is part of. The digital world and the physical world are not independent from each other, as it will be explained in the following.
Considering an individual A that interacts with an individual B, we assume that, first, A sends a message to B. A decides which message(s) he wants to send and how. He also decides which of his personal data he wants to reveal. But, there might be details about himself that he reveals implicitly or not intentionally. All

the matters related to the action of sending can be explicitly fixed in the digital world as data. Those matters are details A reveals about himself consciously and items that A knows (maybe or for sure) or only believes to know about B. Modeling how A delegates the communication from the physical world to the digital world, we introduce the function $Oge(\,)$. It is a function which transforms information into data by creating according data structures.
If A receives a message from B, then A probably receives new data apart from data that A already stored. Depending on historical knowledge and other parameters such as communication time, place, or the mood of A, the function $Ego(\,)$ models how A transforms data into information[7]. In general, for different physical entities, the functions $Oge(\,)$ and $Ego(\,)$ are not the same.
Depending on the derivations gotten in Sect. 3 regarding the term identity, one could consider identity as being part of the concept of entity. That means that we have *physical identities* and *digital identities*. The existence of physical identities of an individual can be justified by sociological approaches as elaborated in [16]. A digital identity can be described as subset of a digital entity which sufficiently identifies the corresponding entity.
From a social point of view, a physical entity is usually interested in whether its digital entities established in the digital world can be linked to the physical entity and whether other physical entities can conclude anything about this particular entity and what. In this respect, we propose a definition of privacy in a complex digital world:

Definition 4. ***(Privacy)*** *Privacy of a physical entity is the result of negotiating and enforcing when, how, to what extent, and in which context which of its data is disclosed to whom.*

This definition is not limited to considerations in a complex digital world and is achieved considering the following aspects:

1. The understanding of privacy in [9] includes the following sentence:
   "Thus each individual is continually engaged in a personal adjustment process in which he balances the desire for privacy with the desire of disclosure and communication of himself to others, in light of the environmental conditions and social norms set by the society in which he lives."
   It consists of the idea of negotiating the desires for privacy and communication by the physical entity with himself. We believe that negotiation also takes place between the physical entity and other physical entities. We subsume these aspects by the term "negotiating" in Definition 4.
2. Further, we add the term "enforcing" to the privacy definition because privacy is not only a desire but it is a state resulting from the negotiation process. Violation of privacy can, therefore, be explained as a lacking ability of enforcement.

[7] The functions are labelled $Oge(\,)$ and $Ego(\,)$, because "Ego" shall indicate that the way how an individual derives information from data is an essential part the (identity of the) individual.

3. We usually differentiate to whom exactly we disclose which information. We do not see all the other physical entities as a mass. Hence, we replace *others* by "whom" in Definition 4.[8]
4. The involvement of situations is already mentioned by [9] (see above). The concept of "context" is an appropriate concept for modeling situations. Context as a concept required in information technology research is typically described with respect to locations [17]. Regarding privacy, the concept needs to be elaborated with respect to user-control (cf. [18]) or integrity (cf. [19]).

The definition of privacy emphasizes the property of being a state as well as the importance of user control. Especially in respect to the latter, our definition is similar to the definition given by [9].
In [4] and [5], several concepts of privacy are described. In [4], Solove found the following conceptions:

1. The Right to Be Left Alone,
2. Limited Access to the Self,
3. Secrecy,
4. Control Over Personal Information,
5. Personhood, and
6. Intimacy,

where 3. and 4. are subsets of 2. in [4]. DeCew distinguishes informational privacy, accessibility privacy and expressive privacy [5]. These concepts of privacy are subsumed under 2., 4. and 5. by Solove. Definition 4 matches the concepts of DeCew since informational privacy is covered by the data that is processed and accessibility matches the user control in this definition. Automatic or semi-automatic decision making would not omit this point unless the user keeps control. Since this definition is valid for computer-mediated communication, the body can not be accessed, only the mental properties are concerns of accessibility in the sense of DeCew. The expressive privacy is covered due to possible modification in context, the data itself, and the negotiation. The Right to Be Left Alone is difficult to cover, since not every intrusion is an invasion of privacy. With the differentiation of an individual's privacy concerns, one can see the separation into identities as an implication of the privacy definition. Similarly, the separation of identities (resulting in partial identities) has been identified as one of the important paradigms of privacy (apart from privacy as confidentiality and privacy as self-determination) in [20].

5 Conclusion

In this paper, we proposed the separation of the physical world and the digital world as well as a linkage between the two worlds by introducing the functions $Oge(\,)$ and $Ego(\,)$. An enhanced definition of privacy is given for physical entities.

[8] If one assumes that organizations have privacy, then the definition should use the word "whom" not only for individuals but also for organizations or computer. Another, even more flexible possibility is the replacement of "whom" by "which entity".

At this point of work, we did not yet elaborate on the notion of partial identities being an essential part of the identity-related terminology by Pfitzmann/Hansen. This needs to be done in further research. Also, we did not characterize what exactly entities in the physical world could be. At the current point of research, we restrict our considerations to individuals only and did not elaborate on organizations or legal persons. This was mainly addressed in privacy considerations until now. Moreover, groups of entities represent a further challenge if they are physical entities themselves and control their digital (group) entities. Further research will address elaborations regarding the relationships between entities as well as dynamics based on time.
In our paper we apply the definition of Westin to yield a refined definition of privacy. However, scientific literature covers more definitions of privacy. Depending on the context or domain, those definitions of privacy could also serve as starting points for discussion with respect to the digital and physical worlds.
Regarding the definitions of entities (digital and physical), formal methods could be applied to allow more detailed analysis of privacy in a complex digital world. For example, social network analysis using graph theory and matrix representation are approaches to model entities and their relationships, including relationships of different kinds. Game theory would be another modeling approach.
We would like to thank Professor Andreas Pfitzmann, Stefan Köpsell and the three unknown reviewers for lots of discussions and comments.

References

1. Luhmann, N.: Soziale Systeme: Grundriss einer allgemeinen Theorie. Suhrkamp Verlag (1987)
2. Clauß, S., Köhntopp, M.: Identity management and its support of multilateral security. Computer Networks, Special Issue on Electronic Business Systems **37** (2001) 205–219
3. Pfitzmann, A., Hansen, M.: A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. `http://dud.inf.tu-dresden.de/Anon_Terminology.shtml` (2010) [Version v0.34 of August 10, 2010].
4. Solove, D.J.: Conceptualizing Privacy. California Law Review **90** (2002) 1087–1155
5. DeCew, J.W.: In Pursuit of Privacy: Law, Ethics and the Rise of Technology. Cornell University Press (1997)
6. Borcea-Pfitzmann, K., Pekarek, M., Poetzsch, S.: Model of Multilateral Interactions. Technical report, EU Project PrimeLife Heartbeat (2009)
7. Pfitzmann, A., Köhntopp, M.: Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology. In Federrath, H., ed.: Workshop on Design Issues in Anonymity and Unobservability. Volume 2009 of Lecture Notes in Computer Science., Springer (2000) 1–9
8. Bundesdatenschutzgesetz. (1990) (version 14.08.2009).
9. Westin, A.F.: Privacy and Freedom. Atheneum, New York (1967)
10. Semančík, R.: Basic Properties of the Persona Model. Computing and Informatics **26** (2007) 105–121
11. Rowley, J.: The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science **33 (2)** (2007) 163–180 DOI:10.1177/0165551506070706.

12. DIN ISO/IEC 2382-1 Information technology – Vocabulary – Part 1: Fundamental Terms (1993)
13. Ackoff, R.L.: From data to wisdom. Journal of Applied Systems Analysis **16** (1989) 3–9
14. Davenport, T.H., Prusak, L.: Working Knowledge. Harvard Business School Press, Boston (1998)
15. Bellinger, G., Castro, D., Mills, A.: Data, Information, Knowledge, and Wisdom. (2004) `www.systems-thinking.org/dikw/dikw.htm` (accessed at 1st July 2010).
16. Goffman, E.: The Presentation of Self in Everyday Life. Anchor Books (June 1959)
17. Schilit, B., Adams, R., Want, R.: Context-Aware Computing Applications. First Workshop on Mobile Computing Systems and Applications, IEEE (December 1994) 8590
18. Nissenbaum, H.: Privacy as Contextual Privacy. Washington Law Review **79** (2004) 119–158
19. Pfitzmann, A., Borcea-Pfitzmann, K., Berg, M.: Privacy 3.0 := Data Minimization + User-Control of Data Disclosure + Contextual Integrity. it – Information Technology **53 (1)** (2011) 34 – 40
20. Gürses, S.: Multilateral Privacy Requirements Analysis in Online Social Network Services. Dissertation, Katholieke Universiteit Leuven (2010)