Proceedings of the 5th WSEAS International Conference on Applied Computer Science, Hangzhou, China, April 16-18, 2006 (pp981-986)

An Introduction to a Taxonomy of Information Privacy in Collaborative Environments

GEOFF SKINNER, SONG HAN, and ELIZABETH CHANG
Centre for Extended Enterprises and Business Intelligence
Curtin University of Technology
Perth, Western Australia (WA)
AUSTRALIA
Geoff.Skinner@newcastle.edu.au, {Song.Han and Elizabeth.Chang}@cbs.curtin.edu.au
http://www.fit.cbs.curtin.edu.au/

Abstract: - Information Privacy is becoming an increasingly important field of research with many new definitions and terminologies. Along similar rates of increase are the use, uptake and expansion of Collaborative Environments. There is a need for a better understanding and classification of information privacy concepts and terms. This is especially true in relation to their affect on the operation, creation and ongoing administration of Collaborative Environments. The knowledge provided from a information privacy taxonomy can be used to formulate better information privacy policies, practices, and privacy enhancing technologies (PET's). This paper provides a Information Privacy taxonomy for Collaborative Environments.

Key-Words: - Information Privacy, Information Security, Taxonomy, Collaborative Environments, Meta Privacy.

1 Introduction

By definition, taxonomy is '… A scheme that partitions a body of knowledge and defines the relationships among the pieces. It is used for classifying and understanding the body of knowledge.' [1]. As a field of research grows there comes a point in time that the subject matter should be at least theoretically classified into its bases, principles, procedures and rules. The area of Information Privacy, in particular, in the expanding field of Collaborative Environments, is in need of such an arrangement. This paper addresses this issue by providing an general introduction of overview of a taxonomy for Information Privacy in Collaborative Environments. The full taxonomy has not been included due to space limitations, but a much more detailed and expanded body of work has been produced by the authors of this paper.

The paper does provide a high level of detail and knowledge suitable for taxonomy of Information Privacy. Focus has been placed on Collaborative Environments (C.E.'s) due to their inherent data sharing nature and the related privacy issues they create. With the use of C.E.'s in many areas, including the health and intelligence sectors, there are numerous personal data privacy problems to address. Following on from the introduction, section 2 provides a background on the area of Information Privacy. Section 3 provides the taxonomy proposal and its three key dimensions, followed by a conclusion in Section 4, and then the References.

2 Background and Related Work

Before continuing it seems that no privacy proposal is complete without some mention of the 'type' of privacy, one is addressing. This is especially important when the subject of importance is a taxonomy. From a definition of a particular dimension of privacy one can loosely categorize the solutions aimed at each of them. Privacy in general is very subjective and means different things to different people. Common among all interpretations is the perspective that privacy is a human right but is context and environmentally dependent. A number of common privacy dimensions have been defined that have gained wide acceptance [2]. They are termed Privacy of: the person, personal behavior; personal communications, and personal data.

Personal data, also referred to as information privacy is the focus of this taxonomy. In [2] Clarke also provides a well referenced definition of information privacy after initially stating it as being a combination of personal communication privacy and personal data privacy. His formal definition of information privacy is "… the interest an individual has in controlling, or at least significantly influencing, the handling of data about themselves." [2]. The Common Criteria (CC) [3] provides a more formal requirements based definition for providing "… user protection against discovery and misuse of identity by other users.". As you can see from the CC definition, it is information systems

requirements focused, with emphasis on identity protection. Identity protection is a major component of information privacy but by no means represents the complete embodiment of its full meaning.

The Platform for Privacy Preferences (P3P) of W3C [4] is a significant effort to enable web and potentially information system users to gain control over their private information. Open to much debate as to whether it is truly a PET; it does provide automated notice and privacy policy reading from user web browsers. P3P has generated a lot of interest and naturally a lot of research and work material in the area. The scope of which is beyond this paper. What is of interest is its use of metadata to represent privacy policy settings of entities to further enhance their privacy protection. P3P can be used as an assurance mechanism for an entity to confirm the privacy policy preferences in a settings matching process. Metadata tags and document structures are used to store an entity's privacy settings and preferences. The entity requesting personal information also uses the metadata tags and document structures to represent their privacy policies and operational procedures.

3 Overview of an Information Privacy Taxonomy

We have identified three dimensions that make up the highest level of the information privacy taxonomy. These dimensions are space, time, and matter. All are inter-related and have different influences over information privacy. The dimensions translate to three corresponding views of information privacy within a collaborative environment. The views are Structural View, Computation View, and Content View respectively. Each of the dimension and corresponding views are discussed in more detail in the following sub-sections.

3.1 Dimension 1: Computation View

The time dimension, and therefore the computation view, reflects the level of privacy protection. Its time relevance relates to the amount of time and resources required to compromise the stated level of privacy protection. Three categories of privacy protection have been defined, each classified from the highest level of protection to the lowest. The three include Ideal Privacy, Computational Privacy, and Fragile Privacy, listed from highest to lowest protection respectively.

NATURE (N) = Level of Privacy Protection (pp).

3.1.1 Ideal Privacy

Privacy Protection: pp = ∞ for Ideal Privacy

With Ideal Privacy, users at all times determine when, how and what personal information is revealed. Additionally personal data owners decide to what extent others can utilize their information once access is granted. Ideal Privacy gives users complete control over their personal data and more generally all of their information privacy concerns.

Definition 1: *Ideal Privacy provides the highest level of privacy protection (theoretical and practical), providing users with complete control over all of their privacy concerns. No amount of computation can compromise ideal privacy protection.*

Equation (1): {As t->∞ AND r->∞; pp = ∞}.

This translates to: Given an infinite amount of time (t) (t -> ∞) and unlimited computational resources (r) (r -> ∞) privacy protection (pp) will always remain at the highest level and stay uncompromised (pp = ∞).

3.1.2 Computational Privacy

Privacy Protection: pp -> 0 for Computational Privacy

With Computational Privacy, users are provided with significant control over when, how and what personal information is revealed. Additionally personal data owners are the primary entities deciding to what extent others can utilize their information once access is granted. Computational Privacy gives users a high level of control over their personal data and more generally all of their information privacy concerns. However, system owners and data collectors also have a level of control over personal data collection and use, once terms have been agreed upon with personal data owners. Computational Privacy means that it is infeasible to compromise privacy protection within reasonable operational parameters. However, given a very long amount of time and a very large amount of resources, it may be possible to compromise the level privacy protection.

Definition 2: *Computational Privacy provides a medium or operational level of privacy protection, providing users with significant but not complete control over all of their privacy concerns. With an infinite or unreasonably large amount of*

computation, computational privacy protection can be compromised.

Equation (2): {As t->∞ AND r->∞; pp -> 0}.

This translates to: Given an infinite amount of time (t) (t -> ∞) and unlimited computational resources (r) (r -> ∞) privacy protection (pp) will eventually be compromised (pp -> 0).

3.1.3 Fragile Privacy

Privacy Protection: pp ->> 0 for Fragile Privacy

Given a reasonable amount of time and resources fragile privacy can be compromised. This level of privacy protection is only deemed effective against weak threats and attacks. Unfortunately, a large number of collaborative environments (C.E.'s) are of this nature, when they should be offering higher levels of privacy protection. As the adaptation and uses for C.E.'s have increased so has the need for better privacy protection. Many internet sites are still of the format that an entity either accepts the organizations stated privacy as is, or the entity is denied access to their services and resources. Additionally, it is normally the case that if the entities consent is given, control over most personal data is relinquished to the information collectors. What further exacerbates the problem is that for the majority of entities, they do not really pay attention to the finer details of the privacy policy they are agreeing to. This results in a privacy agreement that is very fragile in its nature and understanding. Either the entity had no choice but to agree to the conditions, or they did not understand what they were agreeing to.

Definition 3: *Fragile Privacy provides the lowest level of privacy protection, providing users with limited control over all of their privacy concerns. With a reasonable amount of computation, fragile privacy protection can be compromised.*

Equation (3): {As t->N_t AND r->N_r; pp = 0}.

This translates to: Given a reasonable amount (N_t a large value) of time (t) (t -> N_t) and a reasonable amount (N_r a large value) of computational resources (r) (r -> N_r) privacy protection (pp) will be compromised (pp = 0).

3.2 Dimension 2: Content View

The matter dimension, and therefore the content view, reflects the privacy of collaborative environment objects. Its matter relevance relates to the different types of data that require privacy. Three categories of objects have been defined and each classified accordingly. The three include Data Privacy, Identity Privacy, and Meta Privacy.

NATURE (N) = Objects Privacy (Obj)

3.2.1 Data Privacy

Object: Data -> knowledge and information -> {Data, Text, Emails, Documents, Files, Logs, Transcripts, etc}

Data privacy is the protection of an entities personal data that is being collected, shared, and stored. Ideally the protection is complimented with the entity having complete control over their personal data. There are a number of formal definitions for data privacy that are useful in our taxonomy representation. One such definition refers to data privacy as the '… evolving relationship between technology and the legal right to, or public expectation of privacy in the collection and sharing of data.' [7].

The problem in virtual collaborations is that they are environments made up of interconnected heterogeneous information systems that have different privacy rules and laws governing their operation. Therefore, those tasked with the management and administration of virtual collaborations must take the responsibility for securing personal data and monitoring its secure use. Further, the control and management of an entity's personal data, collected in a collaborative environment, should be tasked to that data owner.

3.2.2 Identity Privacy

Object: Entities -> individuals, groups, and organizations -> {Identity, Identification, Authentication}

One of the top five privacy issues for the year 2005 was Identity Management, specifically the need to balance privacy and security in such a context [8]. The same report states 'Identity is key to protecting personal information and privacy rights'. In an information system and in a broader virtual collaboration an identity is a set of data about an entity (individual, group, or organization) that could be used to differentiate them from other entities in the same environment.

Therefore, identity protection involves securing an entity's identity from unwanted disclosure or discovery. From a privacy preserving perspective, identity protection is concerned with entity control over how they interact with the information system

and other entities within the system. Entities should have a choice as to when, how and to who they reveal their identity to and also who has access to it. Further, transactions and interactions should be able to be carried out in any of the following ways:

- Identifiable: the ability to differentiate an entity or transaction from a group of other entities or transactions.
- Pseudo-Anonymous: the inability to differentiate an entity or transaction from a group of other entities or transactions in the normal course of events.
- Anonymous: the total inability to differentiate an entity or transaction from a group of other entities or transactions.

It is also possible for a single entity to have multiple pseudo-anonymous identities, also referred to as 'nyms'[9]. Likewise, in an ideal setting it should also be possible for many entities to use the same 'nym'. In this papers context this would allow all members of a group or organization to use the single 'nym'. Individuals using the same 'nym' would not allow true identity protection. That is, unless the entities had consented to absolute disclosure and control over each others personal data. When using multiple identities it should not be possible for other entities to deduce that any two identities represent the same entity. That is, there should be no way that a relationship can be established or shown between any two system pseudo-anonymous or anonymous identities. This includes past, present and future system and processing data, actions, and behaviors.

The use of pseudo-anonymous and anonymous identities is often in conflict with a long held misconception that a person's true identity needs to be known for authenticated access to information systems and their resources. It is now widely accepted that is most cases a suitable pseudo-anonymous identity is just effective for the majority of authorization techniques. That is, it is possible to authorize access by a form of identity rather than authenticating access through an entities true identity. Therefore, the three key aspects to identity privacy are [10]:

- Identity
- Identification
- Authentication

There is a complex relationship between all of the aspects, especially from an information systems privacy and security perspective. Further discussion of this subject is beyond the scope of this paper due to space limitations.

3.2.3 Meta Privacy

Object: Metadata and Metastructure Information -> Approaches and Purpose: -> (1) Metadata and Metastructure Content: {Personal Privacy Protection versus Privacy Policy and Preferences Representation}; (2) Unlinkable and Unobservable -> {Risks versus Benefits}.

Meta Privacy is a relatively new term, first formally defined in [11]. A common definition for the word Meta is as a prefix used in an information systems context as meaning "relating to" or "based on". More formally it is a prefix meaning "information about". So when used in conjunction with the term privacy, to formulate the term Meta Privacy, it means information about privacy. Meta privacy is concerned with the information used to support other system services and processors that impact upon an entities privacy. An entity may be an individual, group, or organization. Meta Privacy encompasses the use and management of metadata and metastructure information. It is the metadata and its implementation details, metastructure, which can be the source of either privacy enhancing benefits or privacy invasive drawbacks. This is determined by metadata use, such as P3P [Above], or abuse.

Meta Privacy is defined by the following definition: '*Meta Privacy means ensuring the security and privacy of data about privacy and personal data. Meta privacy is concerned with the security and privacy of the information used to support other system services and processors that may impact upon an entities privacy. This encompasses the protection of metadata and metastructure information that may reveal an entities identity and other personal information'* [11].

The Metastructure components are composed of the data concerned with the functioning and structural details of the information systems and their many components. This may include information on the access controls used in the systems, the system and policy frameworks which supplies rules regarding the relationships within and between the systems and their policies, and other information about the system and component structures and the interoperation. When dealing with information systems and more generally collaborative environments the management of metadata and metastructure information involves serious privacy considerations.

The controlled use, access to, and storage of metadata and metastructure information must be guided by stringent privacy protection procedures. It is the metadata and its implementation that can be

the source of either privacy enhancing benefits or privacy invasive drawbacks. This applies also to the use of metastructure information. Both types need to be protected and is the focus of this sub-section.

3.3 Dimension 3: Structural View

The space dimension, and therefore the structural view, reflects the privacy of collaborative environment entities. Its space relevance relates to the different types privacy applied to various entities and relationships within the Collaborative Environment. Three categories of entities have been defined and each classified accordingly. The three include Individual Privacy, Group Privacy, and Organizational Privacy.

NATURE (N) = Privacy of Entities and Relationships (Rln)

3.3.1 Individual Privacy

Individual Privacy is the privacy of an individual entity. In this paper the context of use refers an entity or user within the Collaborative Environment. It is an entity that has provided personal data that they wish to remain private or protected for privacy reasons. At the most fundamental level each individual entity that is a member of the collaborative environment (C.E.) should be entitled to privacy protection. This encompasses any personally identifiable information (PII) they have provided during registration and ongoing membership of the C.E.. If an entities personal data is revealed without their permission, this would constitute a privacy breach. Individual privacy includes protecting both personal data and related metadata and metastructure information. The protection supports the concept of unobservability.

Definition 4: *Individual Privacy is concerned with the protection and preservation of an individual entity's privacy.*

Individual privacy protects each user from undesirable intrusions and the maintenance of their personal space. In an information privacy context it means that a user's personal information is protected from unauthorized access and use. As individuals may be members of a number of groups, and organizations they are able to establish both Committed and Not-Committed privacy relationship.

Equation (4): {1->1 (!ε || ε); 1->n (!ε || ε); m->n (!ε || ε): where $m = n \parallel m \neq n$}

(Note: !ε means Not-Committed and ε means Committed)

This states that within a Collaborative Environment (C.E.) Individual Privacy is maintained for any entity member. The entities provided privacy protection may be a single entity, one entity to n entities, or even m to n entities with the C.E.. The entities may be either committed or not-committed to the their own privacy protection provided by the C.E..

3.3.2 Group Privacy

Groups in this context are those with no entity commitments. They are often dynamic and ad-hoc groupings in nature as a result. The members of the groups and the groups themselves need their own levels of privacy protection depending on the needs of the membership and group. These needs are also influenced by the reason for group existence and the data the group handles and produces. When members join and leave the group there needs to be at least Fragile Privacy protection maintained over the group and entity members data. Depending on the nature of the group, formal or informal, there are different levels of privacy sensitivity to the group for different members. Membership is also dependant on other factors including time, roles, requirements and personal needs.

Definition 5: *Group Privacy is concerned with the protection and preservation of a Group's and each Non-Committed individual group member's privacy.*

Group privacy protects the personal information of the group and each member in the group. The privacy protection is provided regardless of the non-committed nature of the group members. Groups may be ad-hoc, dynamic and time dependant relations. Group Privacy aims to support the varying non-committal nature of group memberships. This means that Group Privacy ensures that during the formation, duration, and after they have been dispersed the privacy of the group and each entity member of the group is protected.

Equation (5): {1->1 (!ε); 1->n (!ε); m->n (!ε): where $m = n \parallel m \neq n$}

(Note: !ε means Not-Committed).

This states that within a Collaborative Environment (C.E.) Group Privacy is maintained for any group and group member. The entities provided privacy protection may be a group with single membership, a group with one entity to n entities, or a group with m to n entities. The entities are not-committed to the groups privacy provided. This is as

a result of the Group being defined by ad-hoc, informal and non-committed entity membership.

3.3.3 Organizational Privacy

Organizations in this context are those with formal and informal commitments required by their entity membership. They are often structured, planned, and governed by a set policies and procedures. The members of an organization are committed to privacy. The privacy is maintained with when members join and leave the organization. Within a Collaborative Environment (C.E.) there may be any number of formal and informal organizations. Each of these has their own set of privacy concerns that must be addressed. An organization produces 'personal' information that should be under the management of the organization and afforded all levels of privacy protection. The organization controls who, when, where and what information is revealed to other entities.

Definition 6: *Organizational Privacy is concerned with the protection and preservation of an Organization's and each Committed individual organizational member's privacy.*

Organizational privacy protects the sensitive personal information and activities of an organization, in addition to the personal information of the organizational members. Individual entities are committed to organizational privacy protection as well as their own and each entity within the organization. Organizations generally have more stable and committed memberships that do groups. This allows formal individual commitments, privacy policy development, and privacy focused operating guidelines.

Equation (6): {1->1 (ε); 1->n (ε); m->n (ε): where m = n OR m ≠ n}
(Note: ε means Committed)

This states that within the Collaborative Environment (C.E.) Organizational Privacy is maintained for any organization and organizational member. The entities provided privacy protection may be an organization with single membership, an organization with one entity to n entities, or an organization with m to n entities. The entities are committed to the organizational privacy provided. This is as a result of the Organization being defined by committed entity membership often governed by set privacy policies.

4 Conclusion

With the increasing abundance of material in the information privacy field it has become difficult to define and classify information privacy components. This paper has addressed this issue by providing an Information Privacy taxonomy. The contribution proposed has been limited by the space restrictions of the paper, but has provided a unique arrangement of many of the important information privacy components and dimensions.Focusing the taxonomy on a Collaborative Environment Information Privacy issues has highlighted three key dimensions of the taxonomy. The dimensions are the Computation View, the Content View, and the Structural View. Within each there are additional categorizations to clearly define each of the dimensions.

References:

[1] (1986). "Standard Taxonomy for Software Engineering Standards (ANSI)". The Institute of Electrical and Electronics Engineers Inc.

[2] Clarke, R., (1999), "Introduction to Dataveillance and Information Privacy, and Definitions and Terms.". http://www.anu.edu.au/people/Roger.Clarke/DV/Intro.html (1999).

[3] Common Criteria: Common Criteria for Information Technology Evaluation. January, 2004, http:// www.commoncriteria.org (2004).

[4] W3C: The platform for privacy preferences 1.0 (P3P1.0) specification, Jan., 2002. W3C Proposed Recommendation, http://www.w3.org/TR/P3P (2002).

[7] Clarke, R. (2002), "Introducing PITs and PETs: Technologies Affecting Privacy".

[8] L. Ponemon (2004), "Top 5 Privacy Issues for 2005". Computerworld, Dec 28th, 2004.

[9] G.W van Blarkom, J.J. Borking, and J.G.E. Olk (2003), "Handbook of Privacy and Privacy-Enhancing Technologies". PISA Consortium, The Hague, 2003.

[10] Clarke, R., (19998), "The Real 'Who's Who' of the Electronic Commerce: The Identification of Organizations.". Xamax Consultancy. Prepared for Journal of Strategic Info. Systems. http://www.anu.edu.au/people/Roger.Clarke/DV/Intro.html (1999).

[11] G. Skinner, S. Han, and E. Chang (2005), "Defining and Protecting Meta Privacy: A New Conceptual Framework Within Information Privacy". Intl. Conf. on Computational Intelligence and Security, Xian, China, Dec 15-19, 2005.