Towards Smart Privacy on the Personal Web

Reza Samavi, Mariano P. Consens

{samavi, consens}@mie.utoronto.ca, MIE, University of Toronto

Abstract

User-centric privacy management is an important component of the Personal Web, and even more so in the context of personal health applications.
In this position paper, we propose a logical framework for *smart privacy* based on a modular and extensible ontology that supports reasoning about privacy from a very broad range of perspectives. We describe the motivations behind the development of smart privacy and outline key features of the logical framework in the context of Personal Health Record applications.

1 Introduction

Personal Web is an emerging research topic driven by the transformation of the Internet and web from the way users currently interact and navigate resources on the web, to a smart paradigm mainly centered on users' experience. The main goal of smart Internet is to empower user, as an individual, to seamlessly and smartly self-manage the vast amount of web resources and services to achieve her personal goals [4]. A user-centric perspective of service utilization requires users to play an active role in the process. The promise of personal web is to make this shift socially and cognitively viable. Such a perspective brings new challenges in design and architecture of the web. For example, how to enable the underlying web architecture to support and deliver services based on the individual needs; users need to be able to customize the process according to the task they as individuals want to accomplish, and in such a way that respects their preferences. One such an imperative task is the self-management of privacy.

To meet the promises made by *Smart Internet* [4] and to make the personal web a real user-sovereign system, we propose a notion of *smart privacy* in which the essence of privacy is embodied in system design. We argue that a semantic flow model [2] can provide an unambiguous, and computer-interpretable descriptions of information flow and its relevant norms on personal web. We adopt the point of view of privacy as contextual integrity [13], and propose an end-to-end privacy model in which rather than focusing on sensitivity of data or roles, a permutation of semantically rich contextual information provides users with support to self-manage their privacy.

In this position paper, we make the case for Personal Health Records (PHRs) as an emerging personal web application that can play a key role in transforming different health care processes. From the privacy perspective of personal web, this is a perfect candidate, since the most sensitive information about one is at stake.

In the following subsections we discuss the motivations and expected contributions of this research alongside a brief overview on the state of privacy frameworks on web. In Section 2, we motivate the needs for the model with a PHR use case. Section 3 outlines the proposed privacy model based on the novel notion of semantic flow model. Finally, section 4 concludes the paper with a few remarks on future directions.

Motivation. PHRs are the type of electronic records managed by patients that may end up becoming the least fragmented picture of an individual's health [17]. PHRs, from their first appearance nearly a decade ago as being a patient-centric repository of health data, have been gradually transformed into open software platforms with published application programming interfaces (APIs) [10]. This is similar to what we are experiencing with other emerging personal information platforms like *Facebook*. With these APIs, PHRs are now viewed as *personal health information system* that enables aggregation of different types of health data across multiple healthcare providers [10]. Furthermore, it allows a growing number of innovative third-party applications to emerge around PHRs. Applications for online-health assessment, for participating in clinical trials, or for finding low cost medicine are just a few of the examples. Some futuristic application scenarios

of PHRs are also emerging from Web 2.0 communities when the power of social networking combines with patients' empowerment enabled by PHRs [16] . Studies show that patients' trust on the information received from peers with similar conditions is surpassing the trust in doctors and academic experts. This indicates the powerful role of social networking in public health with the intervention of PHRs [6].

While PHR platforms have great potential for patient empowerment, if the consequences of sharing and data usage are not clear to patients they are prone to greater risk of privacy breaches. With patients having complete control over their health information, the central issue to PHR becomes the self-administration of privacy by patients; how patients, *themselves,* can easily and effectively control their privacy, while sharing their health profile.

Existing privacy architectures for PHRs are rather primitive and the consent to share the record typically applies to the entire record. The patients' administrative role in controlling their privacy is limited to a trivial binary consent of "I agree" or "I disagree". More importantly, there is not mechanism in place for controlling the *usage* of data and understanding the consequences of sharing. Controlling usage after second and further sharing, i.e., when a receiver shares a patient's record with another party, is even more challenging.

The research outlined in this paper aims to fill these practical gaps and propose a logical framework for expressing patient's privacy expectations and mechanisms to ensure these expectations are enforceable. We believe an authenticate design of privacy system for a personal web application will be successful only if the system is semantically sound. The system should precisely capture what the user expectations are and how they affect occurring of *future* activities in a process in order to access to health information occurring *now* (obligation and promise management). Therefore, we use first order logic formulas for expressing and reasoning about these actions and norms that govern flow of information from one application to another in the PHR. We construct our logical model based on the framework of *contextual integrity* [13]. In contrast to the classical view of privacy that mainly focuses on type of data or role of objects, contextual integrity provides a systematic guidance to incorporate all elements of context (the actor, role, purpose, type of data, etc.) alongside with the norms of transmission for information.

In this research, we propose the logical framework for smart privacy as a modular and extensible ontology. This ontology is intended to support reasoning about privacy from a very broad range of perspectives, e.g. privacy preferences, third party privacy policies, delegation of privacy policies, and obligation enforcement. Instead of designing the logic from the scratch, we build the ontology using ISO 18629, Process Specification Language (PSL)[2].

Related Work. A number of frameworks for defining and enforcing privacy policies have been proposed, including P3P [5], XACML [14], EPAL [9], LPU [1], Privacy APIs [11]. These existing solutions cannot adequately address personal web privacy management requirements for the following reasons. First, these frameworks are mainly designed having institutional privacy needs in mind. In other words they are built to protect organizations from being liable in case of breaching privacy laws and regulations, not for the purpose of supporting an individual who has personal privacy preferences. Second, these frameworks either are not expressive enough to support some important features of privacy obligations such as repeating obligations, and multiple responsible agents [14], or use complex logical machinery that makes practical usage of the framework infeasible (e.g. [1]). In contrast, the approach suggested in this research by exploiting first order theories in PSL is highly expressive, while PSL constructs also can be easily and systematically extended.

Contributions. We make three contributions. First, we contribute to a specific personal web domain, i.e. PHR, by enabling PHR consumers to effectively express their privacy expectations and define their social privacy boundaries associated with their health information.

Second we design a novel logical privacy framework in which *user preferences* (in contrast to laws and regulations) form the foundation of the privacy management. These preferences dynamically regulate the flow of information in a privacy-sensitive process. The privacy constraints in the model are introduced using the same semantic constructs used to express all other process constraints (for example task ordering, concurrency, task decomposition). This allows achieving privacy goals and utility goals of a process in positive-sum paradigm as described in Privacy by Design philosophy [3].

Third, our work provides a novel ontology-based obligation model based on PSL to the privacy modeling communities in general. This approach can be applied in other domains of personal web such as social networking.

2 Use Case and Desiderata

We illustrate with an example the need for semantically rich and computer-interpretable

behaviour abstraction for process model in a personal web domain. In a simple communication between a hypothetical PHR user, Alice and a third-party service provider, MedicaSave, Alice wants to link her prescription information with the service. MedicaSave reads patient's prescription data stored in PHR and finds the generic equivalent of the drug and low-cost pharmacies for the prescribed drugs. When Alice selected which pharmacy she prefers, then MedicaSave shares Alice's personal information with the pharmacy to deliver the medicine. The Pharmacy needs to access Alice's PHR to check a potential drug interaction. An excerpt from Alice's privacy preferences is shown below:

1. I do not want my sensitive health information to be released to any other third parties except those who are in my circle of care.
2. My explicit consent is required only when the entity using my data is not a covered entity.
3. I would like to be notified monthly about the status of my data and every time my information has been viewed not later than 1 day after access.
4. I would like to receive my health relevant information by email.
5. My information on third party services should be deleted when I unlink my PHR profile to the service.

In the other side of the disclosure, MedicaSave and Pharmacy have also their own privacy policies in a form of a document that describes how they handle private information. The privacy system should be able to address the following set of high-level queries:

Decision making queries: Should Alice agree with the MedicaSave privacy agreement? In what condition is MedicaSave allowed to send Alice's information to Pharmacy? Does pharmacy need Alice's explicit consent to access her PHR?

Contextual queries: Is MedicaSave a covered entity? Is the Pharmacy a member of Alice's circle of care? Is the prescription information a sensitive health data of user? Does sending an email by an agent *q* counts as a notification?

Obligations and Provisions queries: In a given time point *t*, what type of notifications need to be sent by MedicaSave and/or Pharmacy? In a given time point *t*, should Alice's data be retained?

Exception queries: Can data item *d* be still accessed by agent *p* if the notification has not been sent?

As this simple example shows, answering these queries requires a computer-interpretable process model with rigorous and complete axiomatization that support automated reasoning about the concepts. In [15] authors showed how privacy model requirements can be extracted from scenarios, our investigation of more than twenty PHR usage scenarios (range from standard clinician- patient scenarios to futuristic usage of PHR data across social networks) led us to the following high-level desiderata for a smart privacy model.

Privacy preferences: We distinguish between two sides of disclosure, user and service. As noted above with user-centric approach, smart privacy should be able to capture first the flexibility which is inherent in *preferences* and second the privacy notions that might be of interest to a user but not specified in laws and regulations, for example concepts such as appropriateness, embarrassment, reciprocity, and deservedness.

Generic privacy concerns: Each individual service offers different privacy settings and agreements. Therefore, while privacy concerns of a user has not been changed, user is required to endlessly repeat many decisions to perform a single task of expressing privacy expectations. The privacy framework must be able to communicate the semantics of an individual's privacy preferences with multiple services.

Semantic support for privacy concepts: Users of the application are diverse. They come from different jurisdictions and with different cultural backgrounds. Therefore, the privacy framework should support different terminologies of concepts.

Granularity of privacy concerns: Privacy is a subjective property. Users of an application span from those who believe "privacy kills, openness heals" to those with very strict privacy concerns. The privacy framework should be able to capture this diversity.

Support privacy as a process: Different from access control mechanisms that an action is allowed or denied for a specific role in a specific point of time, privacy control mechanisms in general, and in Personal Web applications in particular, require reasoning about a process, a series of temporally constrained actions and occurrences.

Support temporal constraints: Reasoning about temporal constraints is essential in a privacy framework [8]. There are actions need to be performed on data objects before an access is authorized (Conditions), and/or, actions need to be performed after an access is authorized (obligations).

Support multiple responsible agents: In a distributed computing system, not always the agent responsible for an obligation activity is necessarily the same agent who performed the action caused the obligation. Therefore the logical framework must be able to reason about the agents responsible for actions.

Support context representation: In privacy as contextual integrity paradigm, transmission norms, largely depend on the context in which the communication activities occur. The privacy framework should reason about the characteristics of the context.

Support expressing of utility goals: The logical framework must be able to capture the logic of utility and privacy together [1]. In many cases reasoning about the privacy actions requires reasoning about the goal of the process itself. A privacy framework cannot reason about the privacy goals while precluding reasoning about the activities required the goal of the workflow to be achieved.

Usability for self-administration of privacy: users are neither policy makers nor system administrators; therefore, the system must be such usable that an average user can express her privacy expectations. In other words system support is required to improve comprehensibility and consciousness of an average user on controlling her privacy.

Although some of these requirements include elements of a user interface design, our focus in this research is on the infrastructure that is needed to support the required functionality. Nevertheless, the characteristics of the user interface substantially influence how the user can interact with a system.

3 Smart Privacy Model

In this research, we propose the logical framework for smart privacy as a modular and extensible ontology. This ontology is intended to support reasoning about privacy from a very broad range of perspectives, such as user privacy preferences, third party privacy policies, delegation of privacy policies, and obligation enforcement. Instead of designing the logic from the scratch, we build the ontology using ISO 18629 (Process Specification Language (PSL)[2]) upper ontology. Upper ontology is an ontology, which describes very general concepts that are the same across all domains.

In smart privacy, theories of PSL are extended to express specific privacy constructs such as pre-conditions, post-conditions, obligations, and communication behaviors as constraints over occurrences of process activities in the context of a rigorously axiomatized first-order logic framework (Figure 1). The brief review of contextual integrity reveals how the ontologies are used to gain insight with regard to privacy.

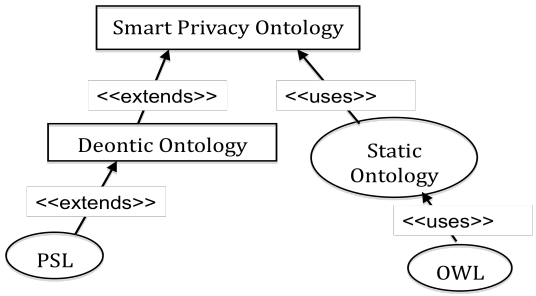

Figure 1: Ontology-based Smart privacy

Contextual integrity. Contextual integrity is a philosophical account of privacy that provides a normative model, or framework, for evaluating the flow of information between agents [13]. In contrast to the classical view of privacy as "control over information about oneself"[18], contextual integrity emphasizes on evaluating why certain patterns of flow of information provoke the sense of privacy violation, while others not. The model then defines the notion of *appropriateness* to answer this question. Disclosing information per se is not what makes us feel breach of privacy. We feel our privacy is violated when the same information communicates in a context that we feel it is embarrassing or inappropriate. For example sharing information about one's unwanted pregnancy to her family physician looks fine and appropriate while sharing the same information to her boss causes one's outcry. Therefore, features of contexts accompanied with the flow of information determine privacy. Five constructs are keys to defining contextual integrity: contexts, informational norms, actors, attributes, and transmission principles. Contexts are structured social settings characterized by the roles that actors play, by certain ends or values that a context forms around it (e.g. the value of a health care context is providing health service), and the norms that prescribe and proscribe acceptable actions and practices [13, p133]. Attributes define the nature of the information in question. If agents refrain to share any attribute, there would not be a violation of privacy. Transmission principle is a set of constraints that governs the information flow. Norms prescribe which transmission principles ought to govern the flow of information and privacy violation arises if these principles are not followed.

In our model we formalize the main concepts of contextual integrity using ontologies as shown in Figure 1. Our model consists of communicating agents (G) who take various roles

(*R*). Each role (*r*) has certain capacities (*U*). These capacities capture knowledge about the ability of an agent to pursue one or more utility goals. Agents participate in Communication Activities (*A*). For each communication activity, at least three distinct type of agents are involved; the principal agent whose information is in stake; the sender of the information; and the receiver of the information. Each receiver agent has a set of history of access to information as its property. Occurrence of each activity changes the state of an agent in terms of the history of access. Bind to each history, there exists activities that must occur and/or activities that must not occur. Occurrences of these activities are either temporally or causally constrained and are accompanied by norms of transmission specific to each context. This specification can be modeled using ontologies.

Deontic Ontology. Our semantics are based on the concept of provisions and obligation linked to the PSL *activities*. Predicates that change due to activity-occurrences are modeled using *fluents*. Pre-conditions and post-conditions modeled as activity-occurrences with temporal properties of start-point and end-point. In this way we can entail whether a privacy constraint (privacy goal) is *successful*, *failed*, or *violated;* utility goal is achieved or failed. The construct of occurrence tree in PSL captures what we call privacy contingency plans for either of these conditions. Set of all possible activity-occurrences of a process is modeled as a tree whose nodes correspond to individual activity-occurrences and where the children of one activity-occurrence correspond to the set of all possible activity-occurrences that could immediately follow it. For a fluent such as *send_notification(t, p, q)* in time *t* from agent *p* to agent *q* and *access* occurrence of *o* the value of the fluent just prior to *o* occurs is expressed by *prior(send_notification(t, p, q), o)* and the value of the fluent just after the *o* occurs is expressed by *holds(send_notification(t, p, q), o)*. By using these basic constructs, more complex behaviour execution constructs such as if-then-else can be modeled in PSL.

Deontic ontology in our logical framework captures the generic knowledge about a context and its norms of transmission. Both classes (contexts and norms) are defined in terms of activities and their occurrences using first order formulas. The novelty of our work is that, in contrast to the previous formal privacy frameworks (e.g. [1, 12]), *context* in our model is not expressed as an entity by itself, but it comes to the picture by a specific permutation of agents, roles, purposes, environmental variables and activities. Then a set of deontic constraints defines the properties and relationships of these entities along with *norms* of the context. This approach captures dynamic and extensible nature of context in our privacy model.

The second class of axioms in deontic ontology represents the *transmission norms* that govern a context's information flow. For example, the constraints such as "I would like to be notified by email every time my information has been accessed" will be presented as a deontic constraint:

for all (o1) (implies (occurance_of (o1, record access)), (exists (o2) (and (occurance_of (o2, send_email), (begine_of (o2) > (begine_of(o1))))))

We use the same semantic as the one we used for contexts to represent these principles. By this approach, while evaluating a privacy policy, the reasoning engine collects requirements from *all* applicable axioms, which restrict a *context* and its *norms,* in one step. This allows addressing an important property of contextual integrity. We can now, relate the issues of compliance and refinement of privacy policies to the logical concepts of satisfiability and entailment. Entailment is key to understand whether a context and its norms, comply with the transmission principles. Our semantic also explicitly captures the concept of past and future. Thus, reasoning is not limited to the time point of the access, but the decision of compliance can depend on what actions have occurred previously and can require occurrence of future actions. The deontic ontology in the proposed logical framework consists of two sets of first order axioms: Contexts ontology, and transmission norms ontology.

We view agent communication in PHR context as processes with specific goals. Therefore, there are many common concepts of process model in general that we are going to solicit from PSL for reasoning about a context and its privacy norms. Concepts such as *before, after, activity, activity occurrence,* are examples of some common terms that are already expressed by PSL ontology.

Static Ontology. The deontic ontology as described above works in the spirit of a static ontology. The static ontology in our model characterizes classes of entities used in deontic ontology, their properties and their relationships to each other. For example, unambiguous definition of entities such as *Data_receiver, Covered_entity, Uncovered_entity* is essential for reasoning about privacy constraints. Agents in our logical model of privacy are autonomous, so the same constraint across multiple agents may be stated differently. The static ontology con-

tributes to our logical model by providing support for interoperability and more effective use of knowledge about contexts and their information transmission norms. Furthermore, it supports neutral authoring of privacy constraints [7] in a sense that user privacy preferences can be authored in a single language with the service privacy policies. This static ontology will be formulated in description logic (DL), supported by the web ontology language OWL DL [12]. Figure 2 shows a partial representation of norms static ontology.

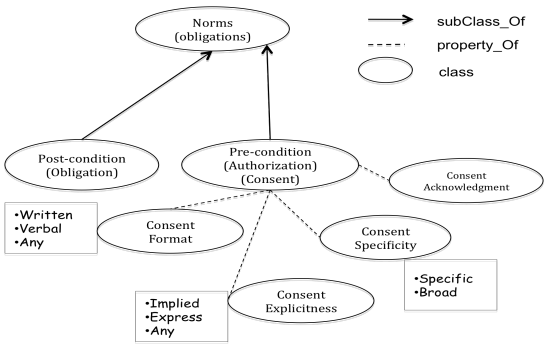

Figure 2. Static ontology (partial) for Norms

4 Closing Remarks

In this paper we introduced the concept of smart privacy for personal web. We described the motivations behind smart privacy, alongside the features that a logical model needs to adhere in order to adequately express privacy requirements of an important class of personal web applications: PHR.

We argued that existing privacy languages and frameworks are not expressive enough to capture all aspects of smart privacy. As such, we rationalized using formal ontology language to create an extensible semantic flow model for user-centric privacy management in personal web. Although this logical model per se cannot address all requirements of smart privacy mentioned in this paper, it provides a solid semantic foundation. How this logical system can architecturally be materialized, and how a user can communicate with this logical system, are unanswered questions and part of our future research.

Acknowledgements

Special thanks go to Dr. Thodoros Topaloglou for his helpful comments on the problem formulation, and Prof. Michael Gruninger for his valuable help in the designing of ontology-based privacy model. Financial support from the Natural Sciences and Engineering Research Council of Canada and IBM Privacy Award are greatly acknowledged.

References

[1] A. Barth. Design and Analysis of Privacy Policies. *Phd thesis*, Stanford U. 2008.

[2] C. Bock, M. Gruninger, PSL: A semantic domain for flow models, *Soft. and Sys.Modeling* 4:209-231. 2005.

[3] A. Cavoukian. Privacy By Design, Take the challenge. *Office of information and privacy Commissioner of Ontario*, Canada, 2009.

[4] J.R. Cordy, M. Chignell, J. Ng. SITCON:The CAS/NSERC Strategic Workshop in Smart Internet *Tech. Proc. of the CASCON 2009. Center for* Advanced Studie*s o*n Collaborative Research. Nov. 2009.

[5] L. F. Cranor, M. Langheinrich, M. Marchiori, M. Presler- Marshall, and J. R*eagle. The platform for p*rivacy preferences 1.0 (P3P1.0) specification, 2002. http://www.w3.org/TR/P3P/

[6] Edleman Trust Barometer. http://www.edelman.com/trust/2008/. 2008

[7] M. Gruninger, The Ontological Stance for a Manufacturing Scenario, to appear in Journal of Cases in Information Systems. 2009.

[8] M. Hilty, D. A. Basin, A. Pretschner, On Obligations, *in: Proc. of ESORICS'05*, Springer-.pp. 98{117. 2005.

[9] G. Karjoth, M. Schunter, M. Waidner, Platform for Enterprise Privacy Practices: Privacy-enabled Management of Customer Data, in: *Proc. Of PET'02*, Vol. 2482 of LNCS, Springer-Verlag. pp. 69{84. 2002

[10] K. D. Mandl. W. Simons. W. Crawford, and J. Abbett, Indivo: a personally controlled health record for health information exchange and communication. *BMC Med Inform Decis Mak*. vol. 25. 2007.

[11] M.J. May, C.A. Gunter, and I. Lee. Privacy apis: Access control techniques to analyze and verify legal privacy policies. *In IEEE Workshop on Computer Security Found*, pp 85{97. IEEE Comp Society, 2006.

[12] B. Motik. P. Patel-Schneider. B.C. Grau. OWL 2 Web Ont. Lang.:Direct Semantics. Tech. rep W3C 2008

[13] H. Nissenbaum. Privacy in Context: Technology, Policy, and the Integrity of Social Life. Chicago: Stanford Law Books, 2009.

[14] OASIS, eXtensible Access Control Markup Language (XACML) Version 2.0, OASIS Standard (2005). URL http://docs.oasis-open.org/xacml/2.0/access control-xacml-2.0-core-spec-os.pdf

[15] R. Samavi, and T. Topaloglou, Designing Privacy-Aware Personal Health Record Systems, *Proc. of the ER 2008 (CMLSA)*, Barcelona, Spain, 2008.

[16] Stewart, Darin. " EcontentMag.com: Socialized Medicine: How Personal Health Records and Social Networks Are Changing Healthcare." http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=56166&PageNum=2, (acc. 13, 2010).

[17] P.C. Tang, D. Lansky. The missing link: bridging the patient-provider health information gap. *Health Aff*. vol. 24. pp 1290-1295. 2005

[18] A.F. Westin. Privacy and freedom, New York, Atheneum, 1967