Formal Specification and Verification of User-centric Privacy Policies for Ubiquitous Systems

BOISE STATE UNIVERSITY

Dr. Hoda Mehrpouyan, Rezvan Joshaghani, Nuhil Mehdy | DEPARTMENT OF COMPUTER SCIENCE, COLLEGE OF ENGINEERING, BOISE STATE UNIVERSITY

INTRODUCTION

A *Privacy Bill of Rights* was endorsed by the White House in 2012, a response to an increasingly loud objection of citizens on the lack of privacy and fair information practices guidelines. The predicament was not only recognized by the US government, but has also been investigated and studied at the international stage and has resulted in reports such as "*Rethinking personal data: Strengthening trust*" by the World Economic Forum (WEF) and "Recommendations for businesses and policymakers" by the Federal Trade Commission (FTC).

Despite all these efforts, ubiquitous online monitoring of users' activities and scandalous data breaches, i.e. Facebook and Cambridge Analytica, continue to haunt Online Social Network (OSN) users. These privacy breaches are often due to a lack of regulatory standardization. Hence, the onus is on the user to take control of: what types of information should be shared with whom and when. However, controlling and managing the information sharing parameters could be a cumbersome and difficult process.

RELATED WORKS

The first privacy theory emerged when newspapers started to publish personally intrusive articles and photographs[1]. This led to ***seclusion and non-intrusion theory*** of privacy that defined the user's privacy as "the right to be left alone" [2] or being free from intrusion [3]. As new technologies were introduced such as databases containing the personal information of the users [1] the information-related privacy concerns [4] emerged. To address these concerns researchers developed the ***control*** [5], ***limitation*** [6], and ***Restricted Access/Limited Control*** (RACL) [7] theories to enable users to control and limit their privacy while share information with others. In RACL theory, the user's privacy is implied as "a situation with regard to others [if] in that situation the individual. . . is protected from intrusion, interference, and information access by others." [8] The control, limitation and RACL theories assume a rigid definition of privacy, while in the current technological era the meaning of privacy changes based on the societal norms. To address this issue, Nissenbaum proposed the ***Contextual Integrity*** (CI) theory of privacy, [9] where privacy behaviors are affected by the context of the information sharing environment.

To implement the above theories, privacy languages were either created by augmentation of access control languages or have the same structure of specifying policies as a set of access roles and information categories in a structured format like Extensible Markup Language (XML).

Some well-known examples of such Languages are Platform for Privacy Preferences Project (P3P) [10], Enterprise Privacy Authorization Language (EPAL) [4], eXtensible Access Control Markup Language (XACML), and Confab. The early version of these languages lacked temporal modalities that were solved in the extended versions of them such as adding spatio-temporal attributes to XACML.

[1] Herman T Tavani. 2007. Philosophical theories of privacy: Implications for an adequate online privacy policy. Metaphilosophy 38, 1 (2007), 1–22
[2] Samuel D Warren and Louis D Brandeis. 1890. The right to privacy. Harvard law review (1890), 193–220
[3] Jamal Greene. 2009. The So-Called Right to Privacy. UC Davis L. Rev. 43 (2009), 715.
[4] Joseph Phelps, Glen Nowak, and Elizabeth Ferrell. 2000. Privacy concerns and consumer willingness to provide personal information. Journal of Public Policy & Marketing 19, 1 (2000), 27–41.
[5] Alan F Westin. 1968. Privacy and freedom. Washington and Lee Law Review 25, 1 (1968), 166.
[6] Ruth Gavison. 1980. Privacy and the Limits of Law. The Yale Law Journal 89, 3 (1980), 421–471
[7] James H Moor. 1997. Towards a theory of privacy in the information age. ACM SIGCAS Computers and Society 27, 3 (1997), 27–32.
[8] Herman T Tavani and James H Moor. 2001. Privacy protection, control of information, and privacy-enhancing technologies. ACM SIGCAS Computers and Society 31, 1 (2001), 6–11.
[9] Helen Nissenbaum. 2004. Privacy as contextual integrity. Wash. L. Rev. 79 (2004), 119.
[10] Joseph Reagle and Lorrie Faith Cranor. 1999. The platform for privacy preferences. Commun. ACM 42, 2 (1999), 48–55
[11] Paul Ashley, Satoshi Hada, Günter Karjoth, Calvin Powers, and Matthias Schunter. 2003. Enterprise privacy authorization language (EPAL). IBM Research (2003).

FORMAL MODEL FOR PRIVACY MANAGEMENT

This research extends the concept of contextual integrity to provide mathematical models and algorithms that enables the creations and management of privacy norms for individual users. The extension includes the augmentation of environmental variables, i.e. time. date, etc. as part of the privacy norms, while introducing an abstraction and a partial relation over information attributes.

The proposed framework is based on two sets of formal models: *1- User's Information Sharing Model (UISM) that represents the information sharing activities in real-time*, and *2- Privacy-Preserving Model (PPM) that formally specifies the user's privacy requirements.* Finally, the *3- privacy verification* is performed by mapping each action in UISM to its corresponding action in the PPM. In the case of not being able to map an action a privacy violation is detected and reported to user to get confirmation.

User Information Sharing Model

UISM is designed based on the formal definition of entities that construct Information Communication mechanism based on agents.

DEFINITION 1. ***(The User Information Sharing Model (UISM)*** *Let $UISMM = (K, Act, \longrightarrow, \kappa_0)$ be a 4-tuple transition system where:*

- *K is a finite set of knowledge states κ.*
- *$\kappa_0 \in K$ is the initial state $\kappa_0 = \emptyset$ (no initial disclosures).*
- *Act is a set of communication actions.*
- *$\longrightarrow \subseteq K \times Act \times K$ is a transition relation, transform the system state with actions $(a, p, \overline{t})$ as follows:*
  - $\kappa \xrightarrow{(sh,p,\overline{t})} \kappa'$, where $\kappa' = \kappa \cup \{(p, \overline{t})\}$,
  - $\kappa \xrightarrow{(st,p,\overline{t})} \kappa'$, where $\kappa' = \kappa \setminus \{(p, \overline{t}') \mid \overline{t} \cap \overline{t}' \neq \emptyset\}$.

Privacy Preserving Model

The Privacy-Preserving Model is designed to manage and govern user's information sharing activities at run-time.

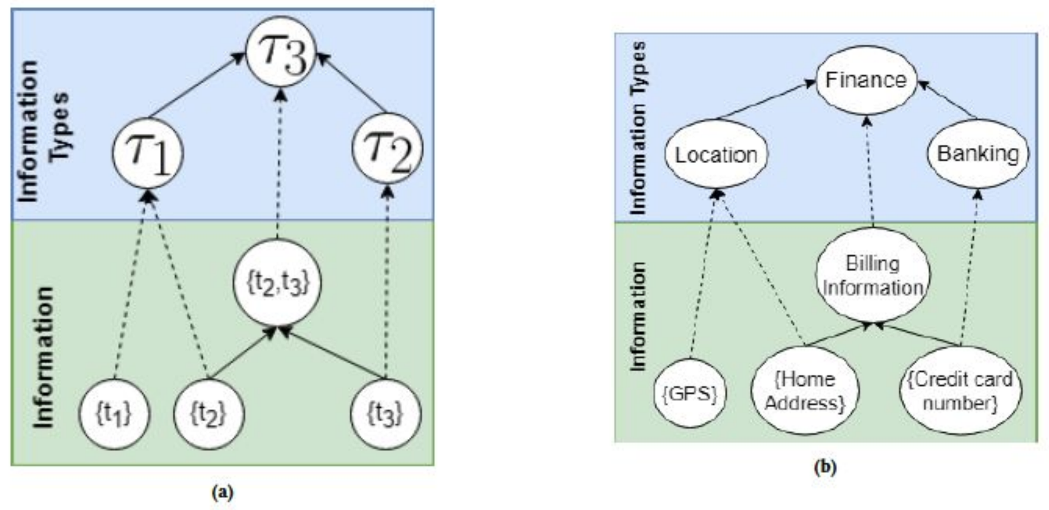


Figure 1: (a) An example of the partial order of the attributes and attribute types where the top layer show the attribute types and the bottom layer show the information themselves. (b) t1 =GPS information, t2 = home address, and t3 = credit card number. The middle layer represents the information that are used together for example the credit card number and the home address go together for billing information that is a considered as financial type.

Verification

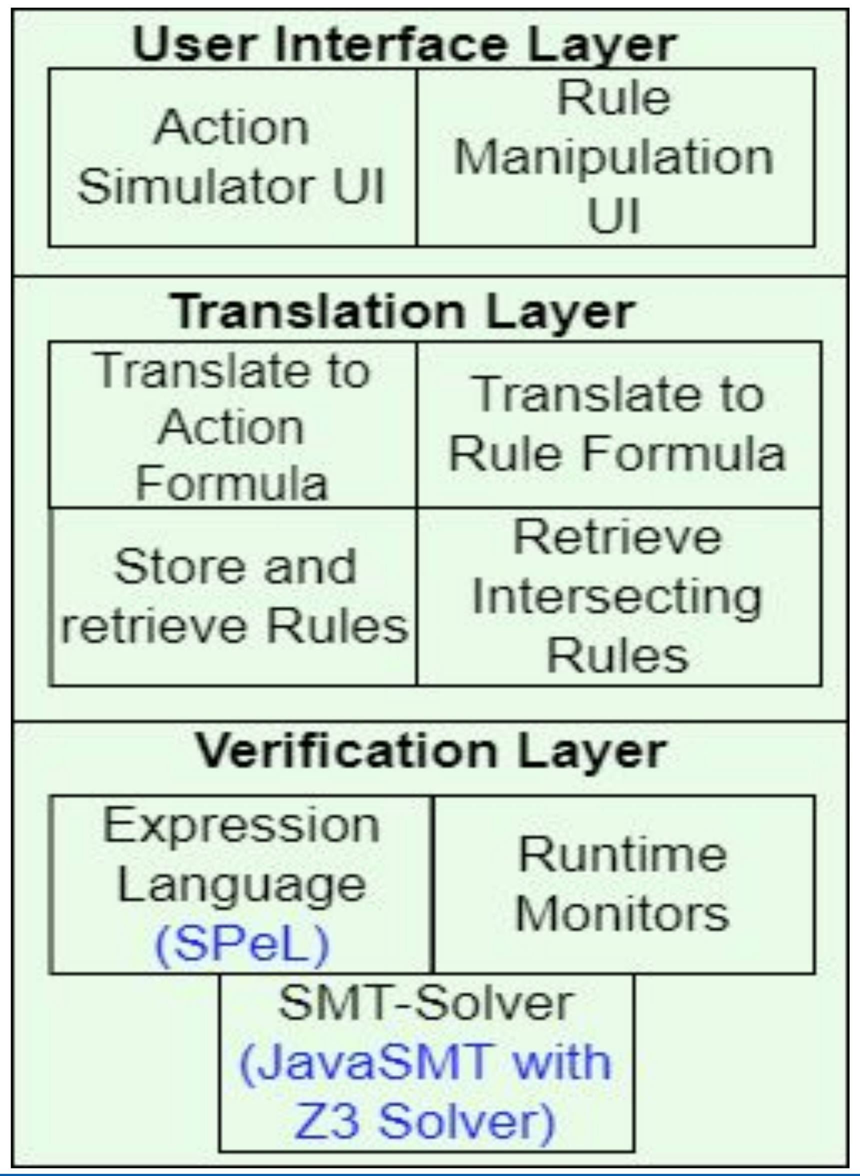
When a new norm is created, the framework checks the consistency of the new norm with the existing norms. Based on the consistency constraints in the framework first ensures that the new norm access permission does not exist in the database. Then the new norm's environmental conditions are checked for consistency. The framework parses the string of the environmental conditions and changes them to SMT solver formulas. Then the SMT solver needs to prove that the implication or equivalency relation holds and it is always valid.

Norms and Their Consistency

		1	2	3	4	5
		$\tau_1 < \tau_2$	$\tau_2 < \tau_1$	$\tau_1 = \tau_2$	$\tau_1 < e > \tau_2$	$\tau_1 < none > \tau_2$
		$Loc < Fin$	$Loc < Fin$	$Loc = Loc$	$Fin < Loc > HLth$	$Loc < none > Bank$
A	$\rho_1 < \rho_2$	$c_2 \Leftrightarrow c_1$ $\mathcal{L}(s_1) = \mathcal{L}(s_2)$	$c_2 \Longrightarrow c_1$ $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$	$c_2 \Longrightarrow c_1$ $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$	$c_2 \Longrightarrow c_1$ $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$	True
B	$Fr < BFr$	Share Loc with Fr when c1 an s1, share Fin with BFr when c2 and s2. Fin should be guarded the same or better, $c_1 \Longrightarrow c_2$, $\mathcal{L}(s_2) \subseteq \mathcal{L}(s_1)$. BFr can have less restrictive access, $c_2 \Longrightarrow c_1$, $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$	Share Fin with Fr when c1 and s1, share Loc with BFr when c2 and s2. Fin should be guarded the same or better, $c_2 \Longrightarrow c_1$, $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$. BFr can have less restrictive access $c_2 \Longrightarrow c_1$, $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$	Share Loc with Fr when c1 and s1, sare Loc with Bfr when c2 and s2. Loc should be guarded at least the same way, $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$. BFr can have less restrictive conditions, $c_2 \Longrightarrow c_1$, $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$	Share Fin with Fr and Health with BFr (or vice versa) which can share Loc. Loc should be guarded at least the same way $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$. BFr can have less restrictive condition, $c_2 \Longrightarrow c_1$, $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$	Since Loc and Bank are incomparable then those norms should always be consistent.
C	$\rho_1 = \rho_2$	$c_1 \Longrightarrow c_2$ $\mathcal{L}(s_2) \subseteq \mathcal{L}(s_1)$	$c_2 \Longrightarrow c_1$ $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$	False	$c_2 \Leftrightarrow c_1$ $\mathcal{L}(s_1) = \mathcal{L}(s_2)$	True
D	$Fr = Fr$	Share Loc with Fr when c1 and s1, share Fin with Fr when c1 and s2. Fin should be guarded the same or better way $c_1 \Longrightarrow c_2$, $\mathcal{L}(s_2) \subseteq \mathcal{L}(s_1)$. Fr should have at least the same access, $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$.	Share Fin with Fr when c1 and s1, share Loc with Frien when c2 and s1. Fin should be guarded the same or better way, $c_2 \Longrightarrow c_1$, $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$. Fr should have at least the same access $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$	There should be only one rule for the same role and attribute type - the uniqueness property	Share Fin with Fr when c1 and s1, share Health with Fr when c2 and s2, which can share the same attribute Loc. Loc should be guarded at least the same way $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$. Fr should have the same access $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$	Since Loc and Bank are incomparable then those norms should always be consistent.
E	$\rho_1 < p > \rho_2$	$c_1 \Longrightarrow c_2$ $\mathcal{L}(s_2) \subseteq \mathcal{L}(s_1)$	$c_2 \Longrightarrow c_1$ $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$	$c_2 \Leftrightarrow c_1$ $\mathcal{L}(s_1) = \mathcal{L}(s_2)$	$c_2 \Leftrightarrow c_1$ $\mathcal{L}(s_1) = \mathcal{L}(s_2)$	True
F	Fr Anna CoWr	Share Loc with Fr when c1 and s1, share Fin with CoWr when c2 and s2, which have Anna as a common agent. Fin should be guarded the same or better way $c_1 \Longrightarrow c_2$, $\mathcal{L}(s_2) \subseteq \mathcal{L}(s_1)$. Fr and CoWrk should have at least the same access to Loc $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_2) = \mathcal{L}(s_1)$, since they share an agent.	Share Fin with Fr when c1 and s1, share Loc with CoWrk when c2 and s2, which have Anna a common agent. Fin should be guarded better than Loc $c_2 \Longrightarrow c_1$, $\mathcal{L}(s_1) \subseteq \mathcal{L}(s_2)$. Fr and CoWrk should have at least the same access to Loc $c_2 \Leftrightarrow c_1$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$, since they share an agent.	Share Loc with Fr when c1 and s1, share Loc with CoWrk, when c1 and s2, which have Anna as a common agent. Loc should be guarded the same way $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$. Fr and Cowrk should have the least the same access to Loc, $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$, since they share an agent.	Share Fin with Fr when c1 and s1, share Health with CoWrk when c2 and s2, which have Anna as a common agent. Loc should be guarded at least the same way $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$. Fr and CoWrk should have the same access to Loc $c_1 \Leftrightarrow c_2$, $\mathcal{L}(s_1) = \mathcal{L}(s_2)$, since they share an agent.	Since Loc and Bank are incomparable then those norms should always be consistent.
G	$\rho_1 < none > \rho_2$	True	True	True	True	True
H	Fr, none, Fml	Since Fr and Fml are incomparable then those norms should always be consistent.	Since Fr and Fml are incomparable then those norms should always be consistent.	Since Fr and Fml are incomparable then those norms should always be consistent.	Since Fr and Fml are incomparable then those norms should always be consistent.	Since Fr and Fml are incomparable then those norms should always be consistent.

Since the norm conditions are dynamic, they cannot be hardcoded in the verification engine. Therefore to check the environmental variables a mechanism is needed to enable the verification engine to handle change in the conditions. Therefore, the conditions are formed and evaluated at run-time based on the stored environmental constraints in the database. For the implementation of such a mechanism that allows for dynamic manipulation and evaluation of conditions, the Expression Languages (EL) can be used. EL receives an object and a logical expression as a string and evaluates whether the object properties satisfy the expression or not. In our implementation,the current snapshot of the environment is given to the EL as the input object that has the environmental values and the EL expression string is the environmental constraints of the retrieved privacy norms. This framework employs Spring Expression Language (SpEL) as the EL library. EL only checks for the satisfaction of the environmental conditions and if they are not satisfied then the transition guard is not satisfied.

Architecture



CONCLUSION

The proposed framework provides a privacy formalism and verification engine to specify and model privacy from the user's perspective. Moreover, as a proof of concept, a framework was implemented and tested based on the described formalism.

The future work will eliminate the current user interface and user's privacy norms will be generated automatically utilizing text analysis, speech recognition, and AI algorithms that can infer user's privacy policies based on the user's relationships and information sharing behaviors.