Towards Compliance and Accountability: a Framework for Privacy Online

Huanchun Peng and Jun Gu
School of Software, Tsinghua University, Beijing, China
Email: {phc07, j-gu07}@ mails.tsinghua.edu.cn

Xiaojun Ye
School of Software, Tsinghua University, Beijing, China
Email: yexj@mail.tsinghua.edu.cn

Abstract — **Over the last twenty years, there has been a tremendous growth in the amount of data collected about individuals. Most existing privacy enhancing technologies could not prevent privacy breach effectively, since the real threat is not the control of private data access but the control of usage. While "access control" is well understood, how to achieve "usage control" is still unclear. In the online environment, information is easily copied or delivered. $UCON_{ABC}$, as the next generation of access control, is inadequate to cover the entire privacy information life cycle. As an alternative, accountability may become a candidate means to judge the correctness of individual data's usage. In this paper, we give a framework with the goal of privacy promise compliance and accountability, which may help to such kind of situation before sound privacy answers may be realized. Besides, we discuss some relevant technical and non-technical components which are needed in the privacy scenario. In the end, we state several research challenges towards the implementation of our framework.**

Index Terms — **privacy, privacy policy, usage control, compliance, accountability**

I. INTRODUCTION

A short while ago, the common worry about using the Internet was security, especially for business applications. The news media was filled with dire warnings about the risks of using a credit card online, identity theft, etc. The fact that consumers worry about online security was ranked as the biggest obstacle to e-commerce on the Internet. Today, however, consumers are more concerned about another issue: privacy.

The Internet makes it easy to collect information about what people are doing, writing or viewing. Furthermore, the fact that personal information can be collected, stored, used and disclosed without owners' consents or awareness creates the fear of privacy violation and distrust between customers and enterprises. Federal Trade Commission (FTC) reported the results of privacy "sweep", in which the agency visited more than 1,400 commercial Web sites to search clearly displayed privacy policies. The FTC reported that 85% of the sites collect personal information from consumers, only 14% had posted any privacy-related notices, and only 2% had posted comprehensive privacy policies [7]. To protect their privacy, users abort transactions, falsify personal details, or maintain several email accounts. Such practices may deprive Web enterprises of information to meet customers' need and sustain competitive advantages. To encourage users to participate in online transactions, Web enterprises must ease people's concerns about data misuse and data loss. To earn users' trust, a Web site should make it explicit that customer's data are treated in a fair and responsible manner [4]. It has therefore become a common practice for Web enterprises to post privacy policies on their Web sites to inform users about data handling practices. Another factor conducive to user trust is to allow users to control over their data by means of *opt-in* or *opt-out* facilities [9].

For aforementioned introduction, it is crucial that users could perceive the organizations' commitments to their privacy as credible [13]. To achieve this, organizations supplement their privacy policies with privacy seals or make their Web sites P3P-compliant [22]. However, P3P could not solve the privacy problem online, it lack formal and unambiguous semantics, limited in expressive power, and lack enforcement and auditing support [11]. Moreover, end user privacy management tools are limited in capability or difficult to use. To provide effective online privacy protection, a framework that covers the entire privacy identifying information (PII) life cycle with the goal of both privacy policy compliance and information accountability is needed. This life cycle should include the phase of PII collection, storage, usage and disclosure. We consider the key to achieve policy promise compliance in our framework is to have a semantic privacy police specification language and an elaborating privacy policy enforcement mechanism. Enterprises need enforcement mechanisms to ensure that their IT systems are compliant with both the policies they articulate and regulations or laws in practice. Moreover, they need to understand how to specify, deploy, communicate, and enforce privacy policies. Legislators and regulatory bodies require enforcement mechanisms to verify how privacy-related laws are actually enforced by

enterprises in their software systems. Besides, privacy policies must be easily understood by end users. Finally, data owners should know whether their personal information has been handled properly or not by some effective, transparent, usable and comprehensible online privacy-protection mechanisms. Information accountability [15, 20] seems to attract more and more attention on the solution of privacy issues. Our framework tries to attribute the solution of accountability to a solid and secure audit logging mechanism and privacy policy reasoning methods.

In this paper, we present a comprehensive architectural framework that supports the whole PII life cycle privacy management. We identify the relevant technical and non-technical components required to support this framework. The relationship and interaction between these components are also discussed. We enunciate key problems to solve the privacy online and identify some technical challenges and problems to implement the framework. We also outline some approaches that may lead to the final solution. We recognize that technology alone cannot address all of the concerns surrounding a multidimensional issue like privacy. The total solution has to be a goulash of laws, societal norms, markets, economics and technologies. However, by determining which parts are technically realizable in advance, we can influence the overall quality of the solution.

The remainder of this paper is organized as follows. Section 2 introduces some recent privacy-enhancing technologies about concealing data. Section 3 proposes our framework for the privacy online. Section 4 gives several research challenges towards the implementation of our framework and proposes potential solution approaches. We conclude with some closing remarks and in the final section.

II. PRIVACY-ENHANCING TECHNOLOGY

Alan Westin published his landmark study *Privacy and Freedom* in 1967 [21]. Still in the age of mainframe computers, it set the stage for thinking about privacy over the next four decades. Westin presented what has become a classic definition of privacy, emphasizing the individual's right to control how personal information "is communicated to others." We could follow Westin's idea and give our definition that privacy is the claim of individuals, groups, and institutions to determine for themselves when, how, and to what extent information about them is used lawfully and appropriately by others.

In a privacy scenario, each person is presented as an *actor*, and each actor can assume different roles. One actor can send (a copy of) data to another actor. In this case, the former is called *data provider* and the latter is called *data consumer*. The roles changes dynamically. Each data item has a *data owner* who the data involved. As aforementioned, consumers have little knowledge about or pose no control over the usage of their personal information on today's Internet websites. Therefore, the essential goal of Internet privacy is to make data owners have the abilities to control which websites can collect their personal data and how to collect, store, use and disclose their personal identifiable information [3].

Privacy policy languages such as P3P and EPAL [5, 22] are proposed to specify the privacy requirements in a machine-readable format such as XML. The question, however, is that although privacy policy language written in XML could be both human-readable and machine-readable, the interpretation could be ambiguous and the language itself lacks of clear semantics. This ambiguity derives from the direct translation from a nature language privacy policy to XML-format privacy policy. To eliminate this ambiguity, what we need is to design a precise semantic formal model for privacy policy to describe the nature language privacy policy, and then automatically translate the formal model to an IT-enable model to implement the enforcement mechanism of privacy policies.

On the enforcement mechanism of the privacy policy, two widely recognized fair information practices in database systems are: (1) extend access control models to support privacy policies, and translate the privacy-related policies to access control policies that could be enforced. The typical examples are task-based access control (TBAC) proposed by Fischer Hubner [10] and purpose-based access control (PBAC) proposed by Elisa Bertino [2]. (2) Translate privacy policies to the constraints of finer grained access control (FGAC) and rewrite the SQL statements to wrapper the privacy constrains[1]. These methods, however, are all based on the access control concept. While "access control" section of security and privacy is well understood, how to achieve "usage control" is still unclear. We conclude that most existing privacy enhancing technologies based on access control will fail, since the real threat is not the control of access, but the control of usage of collected data. For example, since a data owner has already disclosed his PII to the data provider, then he cannot control how data provider handler his PII by using access control mechanism.

Recently, usage control (UCON) proposed by Sandhu et al. do significant contributions to the concept, logic model, usage control policy, enforcement mechanism and safe analysis of usage control [14, 16, 23, 24]. The usage control concerns what happens to the data once it has been released to the data consumer: how the data consumer may, must, and must not use it, which is very similar to the data privacy requirements. The fundamental problem of both is that data providers want to impose control on how data consumers' processing devices or information systems handler data. To implement usage control enforcement, we consider a consumer-side enforcement mechanism would be appropriate and could lead to the final solution. A UCON policy enforcement based on XACML [6] for collaborative computing environment was proposed by Zhang et al. [24]. The difficulty to implement the usage control enforcement lies on the implementation of customer-side enforcement mechanism, which may based on trust computing technology and a form model of obligation, which still remains some thorns on the way to a reasonable solution .

TABLE I. PRIVACY-ENHANCING TECHNOLOGY

	Privacy Guidelines	Privacy Mechanisms	Example	Privacy Approach
Access Control	Agreement on data collection	Policies, Seals, Certificates	P3P, EPAL, Privacy Aware System	Limited collection
	Purpose-binding	Privacy-centric Access Control	TBAC,PBAC,PRBAC,FGAC	Transparency based upon past
	Controlled disclosure of data	Statistical Disclosure Control	k-Anonimity, differential privacy	generalization, suppression and randomization
Usage Control	Transparent processing and usage	Monitoring processing of personal data	Authorizations ,Obligations and Conditions	Transparency based upon past and the present
	Enforcing policy compliance	Evidence Creation	Secure Logging and Auditing	Transparency based upon past and present with ex post enforcement

While most of privacy enhancing technologies are based on concealing data, yet it is increasingly inadequate for a connected world where information is easily copied or aggregated. Even if it is not revealed explicitly, information may be uncovered by automated correlations and inferences across multiple public databases. Recently, several researchers proposed that as an alternative, *accountability* may become a primary means through which society addresses appropriate use. Information accountability means the use of information should be transparent so it is possible to determine whether a particular use is appropriate under a given set of predefined rules and that the system enables individuals and institutions to be held accountable for misuse. Transparency and accountability make bad acts and misusage of data visible to all concerned. However, visibility alone does not guarantee compliance. Then again, the vast majority of legal and social rules that form the fabric of our societies are not enforced perfectly or automatically, somehow most of us still manage to follow most of them most of the time. This could be done because social systems built up over thousands of years encourage us, often making compliance easier than violation. For those rare cases where rules are broken, we are all aware that we may be held accountable through a process that looks back through the records of our actions and assesses them against the rules. The implement of data accountability may need three basic features [20]. 1) *Policy-aware transaction logs* with the responsibility of recording information-use events that may be relevant to the assessment of accountability to some set of policies. 2) *Policy-language framework*, a common framework for describing policy rules which are used to assess policy compliance over a lager set of transactions logged at a heterogeneous set of endpoints. 3) *Policy-reasoning tools* which assist users in answering questions such as: Is this data allowed to be used for a given purpose? And can a given string of inferences be used in a given context, in light of the provenance of the data and the applicable rules? Data accountability is a promising and challenging direction for solving the problem of privacy.

Finally, we give a conclusion to the classification of current privacy preserving mechanisms, which is given in Table I. In the horizontal columns, the mechanisms are classified according to what they control: access or usage. While access control is well understood as defined authentication and authorization, usage control extends access control and encompasses all those mechanisms that actually deal with the runtime detection of privacy violations. Besides, in the vertical columns, guidelines, mechanisms, approaches and typical examples for privacy preserving are distinguished. We believe that a multi-dimensional issue like privacy could not be solved by means of only one of them, but a comprehensive and synthetic solution, which take most of them into consideration.

III. A FRAMEWORK FOR PRIVACY ONLINE

In this section, we propose a general framework towards privacy compromise compliance and information accountability for the privacy preserving on the public Internet. Fig 1 shows the architectural of the framework, including the function models not only on the data provider side, but also on data owner and data consumer side. We attribute the solution of privacy promise compliance to a formal semantic privacy policy language and a privacy policy enforcement mechanism which contains not only access control mechanism, but also usage control, privacy policy negotiation mechanisms. Furthermore, we use an accountability audit component based on secure logging audit mechanism and PII misuse evidence creation components according to privacy policy as the solution of data accountability.

A. Data Provider Side

Following the PEI method proposed by Sandhu [17], the privacy policy management at data provider's side should be organized as a three-tier model.

Top tier (Privacy policy specification): To earn users' trust, data provider's Web sites should make it explicit that customer data are treated in a fair and responsible manner. Then Web sites begin to post privacy policies on their Web sites to inform users about data handling practices. Traditional privacy policies are specified in natural language at first; research, however, has shown that privacy policies tend to intensify privacy concerns rather than engender trust. The reason may lies on that online privacy policies have been drafted with the threat of privacy litigations in mind rather than commitment to fair data handling practices. Then Web sites make their privacy policy P3P-compliant or EPAL-compliant. However, existing languages for specifying privacy policies lack formal and unambiguous semantics, limited in expressive power, and lack enforcement and auditing support. What we really need is a formal semantics privacy policy language with precise mathematical

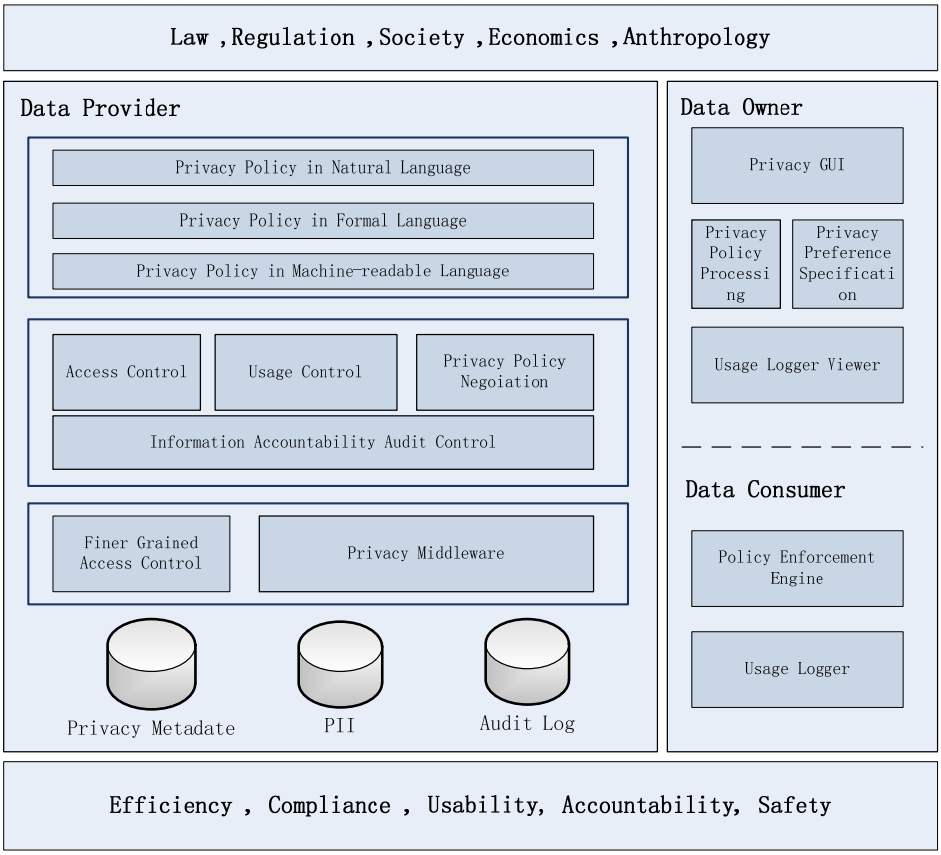

Figure 1. Framework for privacy online

definition, and a mechanism could automatically translate formal language policies to both natural language privacy policies, which are needed for the presentation on web sites for legal requirements; and machine readable privacy meta-data, which are needed for the mechanism for privacy police enforcement implementation. Policy languages for this tier should focus on which privacy goals are to be achieved, rather than how to achieve them.

Middle tier (Privacy policy enforcement): In this layer, machine readable privacy policies could be enforced directly or could be automatically translated to existing traditional security control models. For example, those governing authentication, access control, usage control, information flow and audit policies are needed to enforce high-level privacy policies. The machine readable privacy policy must be semantics unique and expressive for the reason that it should be automatically transformed from the formal privacy policies. The fundamental of privacy policies appears to be an obligation to privacy litigations in mind rather than commitment to fair data handling practices. Then Web sites need not only to make their privacy policy to be well understood, but also an efficacious mechanism to enforce their privacy promise. Privacy centric access control has been well studied in the recent years. Access control alone, however, could not solve the privacy online; to implement the management throughout the whole PII life cycle, promising research are proposed to be carried out in the areas of the followings:

- *Privacy-centric usage control*, which control the usage of PII. $UCON_{ABC}$ model is a suitable reference model for the privacy-centric usage control. However it poses no runtime analysis for the evaluating of policy compliance and is free of policy syntax and semantics.
- *Privacy policy negotiation*, which controls the disclosure of PII from data provider to shared third parties. Since most of these third parties are often interconnected, the information flows among these parties must be properly controlled to prevent privacy breaches.
- *Privacy-centric audit control*, which creates the evidence of PII abuse and misuse, not only the misuse of data consumers, but also the execution of privacy policies at data provider side.

Enforcement mechanism in this tier is application specific, and is usually independent of application implementation details. Different levels of abstractions in an application are commonly exploited to simplify the management overhead. The separation between logical information flow and its physical storage, exchange promotes the dependence between enforcement mechanism and its implementation.

Bottom tier (Privacy policy implementation): The implementation tier focuses on specific technical solution for privacy enforcement models. These issues need to be resolved in a detail level such as the pseudocode. We propose this layer could be implemented as an internal mechanism embedded into an existing database or as a privacy middleware. For the first place, if privacy policy

is implemented as an internal security mechanism embedded into database, the nature of privacy policies tends to be a fine-grained access control or usage control constraints. It may require row-level, column-lever or even cell-level access control or usage control to support privacy promises; for the second place, if privacy policy is implemented as a privacy middleware, the nature of privacy policies tends to be a filter of requested PII and a monitor of PII usage. As existing DBMSs are hard to control the usage of released data, current trends are to implement these components as a privacy middleware.. Furthermore, the efficiency of policy evaluation and enforcement in the bottom tier is an important issue that must be addressed.

B. Data Owner Side

The data owner side components include user agents for *privacy preference specification* which personalize data owners' privacy preferences, *policy processing and presentation*, which control the collection of PII, and a *usage log viewer*, which represent the usage log of data owner's PII in a friendly manner.

The *preference specification* part interacts with the user through a paradigm that is close to the user's privacy protection objectives and generates privacy preferences made by the data owner in a formal language, so that the matching between enterprises' privacy policies and users' privacy preferences can be processed automatically, which is the main function of the *policy processing and presentation* part. Additionally, the user agent provides an expressive user interaction model. When necessary, it presents the policies in an accurate and accessible manner and interacts with the user to help achieve privacy protection objectives.

The *usage log viewer* is a new feature in this framework, whose purpose is to solve the problem that make data owners to know the precise information about usage of their PII after the disclosure through the collection process. How to implement usage log viewer is intimately relative to the privacy-centric auditing technology, which is also a challenging research area.

C. Data Consumer Side

The data consumer side components include a consumer side usage control policy enforcement engine (called *Policy Enforcement Engine*), which enforce the decision made by privacy usage policy decision point towards privacy policy compliance; and a *secure logging point* which responses for logging the usage of PII at the data consumer side within the requirements of data accountability.

Policy Enforcement Engine is a critical component to enforce privacy policy decision at data consumer side. And how consumer side policy enforcement mechanism could be implemented? How can we control the use of PII after it has already disseminated to data consumers? Up to now, trusted computing technologies enabled by Trusted Platform Module (TPM) [19] specified by Trusted Computing Group (TCG) provides mechanisms for this purpose and, generally, can support our requirements.

In a decentralized data consumer side, each endpoint must assume the responsibility of recording information-usage events, which is relevant to the assessment of data accountability to some set of privacy policies. *Secure logging mechanism* is the primary and basic requisite for creating privacy evidence. However, standard logging mechanisms cannot be used for evidence creation, as they fail to ensure the necessary authenticity guarantees of log data. Authenticity of log data means: 1) confidentiality, log entries cannot be visualized or accessed by unauthorized individuals; 2) integrity, the log entries are accurate (log entries have not been modified), 3) completeness, log entries have not been deleted, and 4) compactness, log entries have not been illegally added to the log file; and 5) uniqueness, log data shall not allow for parallel realities [18].

Ⅳ. RESEARCH CHALLENGES

In this section, we invest some interesting issues we identified in implementing the framework. We have also hinted at potential approaches to each issue. This list is by no means exhaustive; its purpose is to initiate discussions.

A. Formal Language Privacy Policies and Preferences

Existing privacy policy specification languages have three shortages: Firstly, none of existing privacy policy specification languages has formal semantics. As a result, policies written in these languages are often ambiguous. Secondly, although Web sites start to post their privacy policies, the majority of online privacy policies are published in natural language. Currently, most enterprises use only textual policies to state privacy policies. Natural language privacy policies cover a much broader scope of enterprises' practices than existing XML-language based policies. Moreover, natural language policies tend to be more ambiguous and incomplete, which makes it difficult to maintain consistency between policies and their policies in machine readable language. Thirdly, existing privacy policy specification languages does not address the enforcement or auditing mechanism of the policy. Enterprises are uncertain of whether published privacy policies are actually enforced in their information systems. Also they can hardly prove to other parties that adequate procedures have been followed to ensure compliance with their privacy policies. Furthermore, the problem is exacerbated by the fact that enterprises are sharing customers' data with other business partners, who may have different privacy practices. What we need is a formal semantics and expressive privacy policy specification language along with an automatic transform mechanism among formal privacy policy, natural language privacy policy and machine executable privacy policy and constraints.

The formal methods community has proposed a number of languages for security with formal semantics. Perhaps the most popular one is based on some extension of Datalog [12]. These extensions are tractable fragments of first-order logic. They allow a limited use of function symbols and negation. Unfortunately, the extensions do not seem to have expressive power to capture a number

of policies that are currently written in English. Recently researchers from Cornell University propose a formal language called *Lithium* to cover the former difficulties [18]. As a fragment of first-order logic, this language has a clear syntax and semantics. It is expressive enough to capture the policies that people want to write in an easy and natural way. It is also tractable enough to allow interesting queries about policies. We think that it is possible to describe our privacy policy in *Lithium* to overcome the shortage of semantic ambiguity and the lack of expression. With regard to tractability, we focus on three questions as follows:

(1) Given a privacy policy constituted of a group of privacy rules and the environment (the context in which policies are applied, including subject and system attributes assignment that provides all relevant facts, etc.), does the access or the usage request is permitted or not according to the privacy preference of the data owner?

(2) Is one set of privacy rules in a privacy policy consistent? In other words, are there any actions that are both permitted and forbidden by the policy in the set?

(3) When the PII is disclosed from data provider to a shared third party, are their privacy policies consistent? In other words, are there any actions are permitted by one and forbidden by the other simultaneously?

The answers to these questions could reflect the essential problems of privacy-centric access control, usage control and information flow. More importantly, we believe that the answers to these questions provide a good understanding of privacy policies. They increase our confidence that the formal statements could capture informal rules and policy creators' intents.

To the second question, we can take a simplistic approach to combining rules, for example combination algorithms in XACML [6]. For the rest two questions, we first define the notation of policy, privacy preference and environment, and then answer the questions in first order logic formulas. An *environment* **E** includes specific statements about the current state of the data consumer attributes and system attributes. *Privacy preference* **PP** includes all the data owners' privacy preferences which are represented as PIIs' attributes. A *Policy* **P** is a closed first-order formula of the form:

$$\forall x_1 \cdots \forall x_m (f \Rightarrow (\neg) Permitted(s,o,r)) \quad (1)$$

Where f is any first-order formula. *Permitted(s,o,r)* shows the result of decision, which is a value of Boolean type, and *s*, *o*, *r* represents subject, object and action respectively. Therefore, the first question could be answered by the following logic:

$$E \wedge PP \wedge P \Rightarrow (\neg) Permitted(s,o,r) \quad (2)$$

For the last question, we assume that a data provider's privacy policy is P_1, and the privacy policy of a third party is P_2. A privacy policy negotiation is to check whether $E \wedge PP \wedge P_1 \wedge P_2$ is valid.

As we all know that the validity problem for first-order formulas is to be undecidable. If the privacy rules are written in first order logic, the answers to the former questions are undecidable. However, policies in a restrict form of first order logic (*Lithium*) have been proved to be computed in the polynomial time. If we write the policy in *Lithium*, the solution can be feasible and efficient.

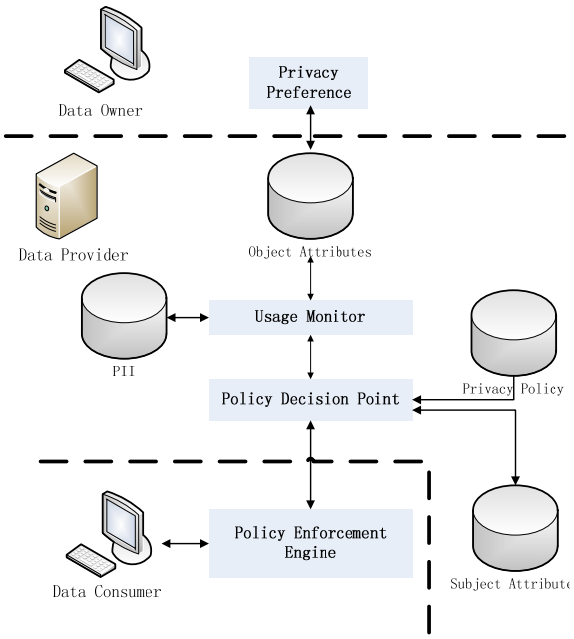

Figure 2. Privacy-centric UCON enforcement

B. Privacy Preserving Usage Control

Park and Sandhu systematically treat the usage control concept and develop a comprehensive and expressive model [17]. They introduced $UCON_{ABC}$ model integrates Authorization (A), oBligation (B) and Condition (C) components in usage control decisions. It has two outstanding features that distinguish it from traditional access control models: decision continuity and attribute mutability. Continuity means that control decisions are determined and enforced not only before the access, but during the period of the access. Mutability means that subject and/or object attributes can be updated as the results of an access decision. The model is divided into three core models: (1) Authorizations, in which the access decision is made based on attributes of the requesting subjects and target objects. (2) Obligations, in which the fulfillment of some actions by specific subjects on specific objects (the obligations subjects and objects may differ from the authorizations pair) should be checked. (3) Conditions where environmental information (also called environment attributes or system attributes) are checked. With respect to the decision continuity factor, it can be distinguished between decisions made before the access session (pre-) and decisions made during the access session (on-). And according to the mutability factor, update actions can be carried out before, during, or after an access session.

Several years later, Zhang et al. [23] have defined a formal model and policy specification for $UCON_{ABC}$ based on an extension of the temporal logic of actions (TLA). The usage is defined as a state-transition machine. The following states have been defined in original UCON during a usage process: initial, requesting, accessing, denied, revoked and end. Initial state means that access request is not generated; requesting state indicates that

the access has been generated and is waiting for the system's usage decision; denied state refers to the state when the system has denied access; the accessing state means that the system has permitted the access and the subject is accessing the object. The termination of an access happens either when the system revokes the access or it is ended normally by the user. The UCON policy is a TLA formula, which includes control rules (CRs) and update rules (URs). And they give an approximately implementation of UCON in collaborative computing systems several years later.

Privacy preserving usage control aims to control the usage of PII. It must follow privacy policies made by data providers and privacy preferences made by data owners. We propose to translate privacy preferences to object attributes to implement the compatibility between UCON policies and the privacy policies. As UCON policy and the privacy policy are all described in *Lithium*, the translation process would be feasible. We give a flexible and effective enforcement mechanism to implement privacy preserving usage control through our framework as in Fig 2.

To control the usage of PII, we build *Policy Enforcement Engineer* (PEE) into data consumer side's agent. The responsibility is to execute the decision made by *Policy Decision Point* (PDP) in data provider's side and delete disseminated PII to prevent from further disclosure.

In data provider side, PDP is a core component. It gives the decisions of the usage according to privacy policies depend on the request, subject attributes stored on a LDAP sever and object and system attributes such as time and occurrence of certain event.

We use a *Usage Monitor* to monitor the access and usage of PII as well as the modification to the object attributes. A PII's attributes can only be modified by its data owner. If the data provider want to modify the object attributes, he need to attain data owner's consent firstly. In order to translate privacy preferences to object attributes, a friend GUI for data owner to explain his privacy preference is needed.

C. Accountability

Information accountability requires that the use of information should be transparent. It is possible to determine whether a particular use is appropriate under a given set of rules. As a result, the system enables individuals and institutions to hold accountable against misuse [20]. Process transparency and data accountability make bad acts visible to all concerned. Protecting privacy is more challenging than ever due to the proliferation of personal information on the Web and the increasing analytical power available to large institutions through Web search engines and other facilities. Access control and collection limits over a single instance of personal data are insufficient to guarantee the protection of privacy. For instance, the same information is publicly available elsewhere on the Web. It is possible to infer private details from other information which is public. What make thing worse, many privacy protections (such as lengthy online privacy policy statements in health care

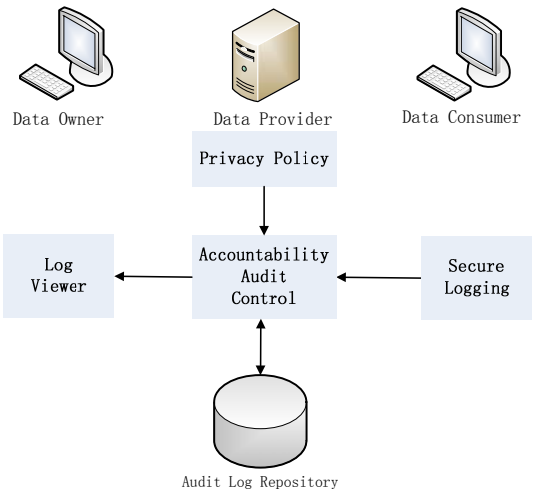

Figure 3. Accountability Audit

and financial services) do not consider the increasing exposure of information due to the social or commercial behaviors. Even people agree to provider their personal information on the public Internet; they do not wish their personal data to be used for any unintended purpose. Mostly, privacy breach is caused not by the disclosure of information, but by the inappropriate, discriminatory, and possibly illegal use of the information. Therefore, accountability should be integrated into the privacy preserving solution as a complementary technology.

The objective of information accountability is to make information usage more transparent. In order to discover individuals and institutions who misuse PII, three basic features are required. First, a *secure logging mechanism* responsible for recording information use events; second, a *privacy log reasoning mechanism* which is responsible for analyzing audit log to discover the pattern of the usage log and determine the access and usage purpose; third, a *log viewer* responsible for display the usage log of PII. A feasible solution to achieving accountability can be shown in Fig 3.

D. Efficiency & Usability

Since the addition of the newly privacy components in the framework, efficiency is a challenging objective to implement the framework. The performance bottleneck lies on the PII usage permission and PII disclosure to third parties. Thus we focus on the efficiency analysis on the theoretic analysis of formal privacy policy validation and index structure to optimize the search performance in the process of finding correspond privacy policies and attributes.

Firstly, the theoretic analysis of efficiency recurs to the two questions we answer in the former section. As we know, the validity problem of first-order formulas is undecidable. In [8] researchers have proved that policies in *Lithium* can be computed in polynomial time. Therefore, the validation of the privacy policies written in *Lithium* could be solved in the ideal time complexity. Besides, in order to improve the search performance of privacy policies and attributes, index structure should not only based on policy id or attributes id, but a more

flexible structure. For example, an index structure based on *target* concept in XACML is suitable for the distributed computing environment.

This framework tries to enable end users to take an active role in protecting their privacy online. Thus from some perspective, *usability* is a key component. Usually, keeping security and privacy is heavily reliant on users' cooperation (users need to specify their preferences). The benefit of preference specification methods cannot be got unless interactions between users and systems are simple and friendly. A friendly GUI used to present privacy policies, privacy preferences and usage logs is needed. The basis is comprehensive study of human-computer interaction.

V. CONCLUSION AND FUTURE WORK

Privacy is increasingly become a major public concern that prevents users from fully enjoying the convenience, effective, and flexibility offered by online services. Luckily, a variety of privacy-enhancing technologies have been proposed. We focus on achieving privacy preserving by means of privacy policy specification and enforcement mechanism. Privacy policy technologies ask for a cooperative relationship among data owners, data providers and data consumers in the privacy scenario. However， existing privacy policies tend to intensify privacy concerns rather than engender trust. We attribute the solution to (1) a formal semantic privacy policy language; (2) a privacy policy enforcement mechanism containing not only access control, but also usage control and privacy policy negotiation; (3) a secure usage logging mechanism achieving data accountability. In this paper we propose a framework to manage the whole life cycle of PII on consider the above objectives. The framework includes the function models not only on the data provider side, but also on data owner and data consumer side. We also point out several research challenges in implementing this framework. To some interesting issues, we suggest some potential approaches to the final solution. Due to the limited length of the article, we only give the coarse grained blueprint to achieving privacy online. Still there remain a lot of thorns on the way to the final goal.

ACKNOWLEDGMENT

This work was supported by NSFC 60673140 and NHTP (2007AA01Z156, 2008ZX01045, 2009CB320706).

REFERENCES

[1] R. Agrawal, P. Birdz, et al. Extending Relational Database Systems to Automatically Enforce Privacy Policies, In Proc. of ICDE 2005.
[2] J.-W. Byun and N. Li. Purpose based access control for privacy protection in relational database systems. VLDB J.,17(4):603–619, 2008.
[3] L.F.CARNOR, Internet Privacy: A Puclic Concern, Communications of the ACM 2,3(Feb. 1999) .
[4] Culnan, M.J. and Armstrong, P.K. Information privacy concerns, procedural fairness, and impersonal trust. An empirical investigation Organization Science 10, 1 (Jan. 1999), 104–115.
[5] IBM, Enterprise Privacy Authorization Language, available at http://www.zurich.ibm.com/security/enterpriseprivacy/epal.
[6] eXtensible Access Control Markup Language (XACML) Version 2. Standard, OASIS, February 2005.
[7] The FTC's privacy Web site, available at http://www.ftc.gov.
[8] J. Halpern and V. Weissman. Using First-Order Logic to Reason about Policies. ACM Transactions on Information and System Security 11, 4, (July 2008).
[9] Hoffman, D.L., Novak, T.P., and Peralta, M. Building consumer trust online. Commun. ACM 42, 4 (Apr. 1999), 80–85.
[10] S. Fischer-Hubner. IT-security and privacy: design and use of privacy-enhancing security mechanisms. Springer-Verlag New York, Inc, New York, NY, USA, 2001.
[11] Li, N., Yu, T., and Antón, A.I. A semantics-based approach to privacy languages. CERIAS Technical Report TR 2003–28. Purdue University.
[12] G, Molina, H, Ullman, J. D., and Widom, J. 2002. Database Systems: The Complete Book. Prentice Hall, New Jersey.
[13] Olivero, N., and Lunt, P. Privacy versus willingness to disclose in e-commerce exchanges: The effect of risk awareness on the relative role of trust and control. J. Economic Psychology 25, 2 (Apr. 2004)
[14] J. Park and R. Sandhu. The UCON_abc usage control model. ACM Transactions of Information and System Security, 7(1):128–174, 2004.
[15] S. Sackmann, J.Struker, R.Accorsi, Personalization in privacy-aware highly dynamic systems, Communications of ACM 49,9 2,3(Sep. 2006).
[16] R. Sandhu and J. Park. Usage control: A vision for the next generation access control. Inter. Workshop on Mathematical Methods,Models and Architectures for Computer Networks Security, 2003.
[17] R. Sandhu, K. Ranganathan, and X. Zhang. Secure information sharing enabled by trusted computing and PEI models. Proc. of ASIACCS 2006, pages 2–12,2006.
[18] Schneier, B. and Kelsey, J. Security audit logs to support computer forensics. ACM Transactions on Information and System Security 2, 2 (May 1999), 159–176.
[19] TCG TPM. 2003. Main part 1 design principles specification version 1.2, available at https://www.trustedcomputinggroup.org
[20] Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, Gerald Jay Sussman. Information accountability. Communications of the ACM 51,6(Jun. 2008)
[21] Westin, A. Privacy and Freedom. Atheneum Press, New York, 1967.
[22] W3C. Platform for Privacy Preferences (P3P) project; available at www.w3.org/P3P/ .
[23] X. Zhang, F. Parisi-Presicce, R. Sandhu, and J. Park. Formal model and policy specification of usage control. ACM Transactions on Information and System Security, 8(4):351–387, 2005.
[24] X. Zhang, M. Nakae, M. J. Convington, and R. Sandhu. A usage-based authorization framework for collaborative computing systems. In Proc. of ACM Symposium on Access Control Models and Technologies, 2006.