International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

A Methodology for Eliciting Data Privacy Requirements and Resolving Conflicts

Asmita Manna, Anirban Sengupta, Chandan Mazumdar

Abstract*: Privacy is one of the major concerns of data protection where personal data of individuals are used by enterprises for providing services. To ensure the rights of citizens, different legal authorities, including European Union, have made it mandatory for enterprises to implement certain privacy principles. An enterprise may also have its own set of privacy principles that help provide customized privacy experience to its customers, with the motive of retaining its customer base and weaning away customers from its competitors. To ensure privacy compliance with legal policies, enterprise privacy principles and expectations of customers, the system design should consider the privacy requirements emanating from all these sources. However, the requirements are often expressed in natural languages, which are difficult to interpret for system designers. In this paper, a logic-based methodology is proposed to formally express privacy requirements emanating from all three different sources. The methodology also includes an algorithm to identify and resolve conflicts among elicited privacy requirements. The proposed approach can be considered as the first step towards ensuring privacy compliance. This would help an enterprise to identify conflicting privacy requirements, resolve conflicts as per pre-defined rules and identify implementable privacy principles to enable the management of privacy compliance.

Keywords: Privacy Policy, Privacy Requirements, Privacy Requirement Engineering, Conflict Resolution of Policies.

I. INTRODUCTION

The promulgation of data protection acts and laws in several countries, including India, has made it mandatory for almost all enterprises to implement privacy protection mechanisms throughout their information infrastructure. Besides, providing assurance to customers, regarding the protection of privacy of their sensitive data and personally identifiable information (PII), helps enterprises in expanding their business. Protection of data privacy can be achieved only if enterprises are able to identify their respective privacy requirements accurately. Privacy requirements usually emanate from three different sources: (i) privacy laws, regulations and contractual obligations; (ii) privacy principles followed by an enterprise; and (iii) customers' expectations. Hence, it is essential for an enterprise to identify privacy requirements from all three sources, represent them in a structured format and resolve conflicts, if any.

Over the years, some significant research has been carried out in the domain of privacy requirement elicitation. Researchers have proposed different methods including the following: privacy risk assessment considering actual threats and vulnerabilities of the information infrastructure; logic-based representation of privacy policies; modeling language-based representation of privacy policies; and natural language processing (NLP)-based requirement elicitation from a given set of policies. However, none of these has considered the elicitation of privacy requirements from all three sources simultaneously (as mentioned above), and the resolution of conflicts therein. This paper aims to fill this research gap by proposing a technique for consolidating privacy requirements generated from different sources and resolving conflicts among those requirements.

Before consolidating privacy requirements, we need an approach for extracting privacy requirements accurately and precisely. NLP-based approaches can lead to ambiguity, while usage of logic based description languages, like Linear Temporal Logic or Description Logic, might be too complicated to implement. Modeling language-based description may not be suitable for conflict resolution. To balance among effectiveness, expressiveness, and ease of implementation, we propose a logic-based semi-formal approach for extracting privacy requirements and performing conflict resolution thereafter.

Rest of the paper is organized as follows. Section 2 discusses some related work. Section 3 describes the proposed methodology for privacy requirements elicitation. Section 4 proposes a formalism to express privacy requirements. Section 5 describes an algorithm to resolve conflicts among elicited requirements; this has been illustrated with the help of results of a case study in Section 6. Finally, Section 7 concludes the paper.

II. RELATED WORK

In this section, we explore some existing methodologies for eliciting privacy requirements.

Deng et al. proposed a systematic approach for privacy threat modeling, named LINDDUN, to elicit the privacy requirements of software systems [1].

Manuscript published on November 30, 2019.
* Correspondence Author
Asmita Manna*, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India. Email: asmita.nag@gmail.com
Anirban Sengupta, Centre for Distributed Computing, Jadavpur University, Kolkata, India. Email: anirban.sg@gmail.com
Chandan Mazumdar, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India. Email: chandan.mazumdar@gmail.com

LINDDUN also has the provision of selecting privacy enhancing technologies accordingly. Based on the high-level system description, a DFD is created and then privacy threats are mapped to DFD elements using a pre-defined mapping table. Using the most common pre-conditions of threats, privacy threat instances are created and mapped to a threat tree pattern. Then the privacy threats, relevant to the designated system, are identified.

Risk assessment is performed to evaluate and identify the risks which need to be mitigated and then the most appropriate privacy enhancing technology is selected. Though this is a comprehensive privacy solution for enterprises, it does not consider the specific privacy requirements emanating from different sources and a DFD based method is not best suitable for automated conflict resolution.

May et al. described a methodology for extracting formal models from regulations and showed the effectiveness of the method using HIPAA [2]. Breaux et al. proposed a method using Semantic Parameterization to extract rights and obligations from legal documents [3]. They also used HIPAA for demonstrating the effectiveness of their approach. The method uses Goal-Based Requirements Analysis Method (GBRAM) and 'Semantic Parametrisation'. A formal model is constructed from natural language texts after representing the domain in first-order logic.

Hassan and Logrippo described an approach to validate compliance and consistency between legal and organizational policies (or requirements) [4]. They used Unified Modeling Language (UML)-based class model called Governance Extraction Model (GEM) to extract the requirements from natural language texts.

Young et al. presented a systematic method, named Commitment Analysis, for acquiring requirements from privacy policies of health-care organizations [5], [6]. The requirements are extracted in terms of *commitments*, *privileges* and *rights*. Barth et al. introduced a term 'contextual integrity' for reasoning about norms of transmission of personal information and presented a Linear Temporal Logic (LTL)-based framework for the same [7].

Anton et al. described a technique for aligning security and privacy policies with system requirements of e-commerce websites using inspection methodology and defining 'goal models' [8]. The basis of comparison in inspection is based on Evolutionary Prototyping with Risk Analysis and Mitigation Model (EPRAM). Though some parts are automated, it requires a lot of manual intervention and analysis.

PriS is another method for incorporating basic privacy requirements into the system design process [9]. It models privacy requirements in terms of organizational goals and uses the concept of privacy-process pattern for describing the impact.

From the above discussion, it is obvious that researchers have used different approaches like threat modeling, goal-based modeling, NLP-based modeling, linear temporal logic-based modeling etc. for eliciting privacy requirements. However, none of these approaches consider the simultaneous extraction of data privacy requirements from all relevant sources, namely legal and contractual obligations, enterprise privacy principles and customers' expectations, and resolution of ensuing conflicts (if any). This work is an attempt to fill that research gap. Here, a first order logic-based semi-formal approach is proposed to extract privacy requirements and to resolve the conflicts among those requirements.

III. PRIVACY REQUIREMENT ELICITATION

As mentioned above, privacy requirements can originate from three different sources. The legal and contractual obligations are mandated by the law of the land and/or contracts between an enterprise and relevant third parties. This should cover all privacy laws that are applicable to the business domains of a particular enterprise. Different countries usually have different privacy laws. It is important to consider the laws of all such countries where an enterprise has business interests. Contracts usually comprise of agreements with third parties including Service Level Agreements (SLAs), Non-Disclosure Agreements (NDAs) etc.

The privacy principles adopted by an enterprise are guided by its specific business interests. These are usually determined by considering different parameters like cultural practices of the land of operation, ethical issues, special privileges that may be provided to valued customers, smooth operations management etc.

The third source of privacy requirements is the expectations of customers. A customer of an enterprise may expect a certain amount of privacy of her sensitive data/PII. In case this does not match with the principles and policies of the enterprise, the customer may choose to migrate to a competitor. If the enterprise wishes to provide customized privacy experience to customers, it may classify customers into specific categories and provide separate privacy settings for each category. For example, the enterprise may choose to publish some details of customers who avail of its services free of cost; while, it may maintain complete privacy for its paid customers.

Technically, privacy refers to seven properties of data, namely confidentiality, integrity, availability, anonymity, unlinkability, undetectability and unobservability [10]. Anonymity of a data item means that the owner of the data is not identifiable from the data item within a given set. Unlinkability means that two or more data items owned by a data owner should not be linked. Undetectability is the inability to identify the existence of a data item by unauthorized subjects. Unobservability means that a data item is undetectable for unauthorized subjects and anonymous for authorized subjects.

Protection of privacy implies protection of one or more privacy properties of the data item depending on its type. So, while extracting privacy requirements, it is important to ascertain the particular privacy properties that need to be protected during various stages of the data lifecycle.

A. *Legal and Contractual Requirements*

Legal and contractual privacy requirements are embedded within legal and contractual documents written in natural language.

Analyses of laws and contracts reveal that the statements pertaining to privacy requirements state the specific actions that are permitted / denied on a particular data item by a particular actor, subject to different constraints. The privacy statements generally mention some lifecycle phase during which the requirement should be ensured; else, it should be implemented throughout the entire data lifecycle. A statement may also include some terms that mention the relative importance of privacy requirements.

If terms such as "must be", "mandatory", "compulsory", "punishable offence", "license cancellation", "legal action" etc. are used in a statement, the importance should be evaluated as *high*. If the terms "should be", "expected", "hoped", "good practice", "legal warning" etc. are mentioned, the importance may be evaluated as *medium*. In case of terms like "may be", "under the discretion of enterprise", "if enterprise has budgetary provision, may incorporate" etc., the importance is *low*. If nothing is mentioned, it should be assumed to be of highest relevance.

Since legal and contractual documents do not mention about privacy properties explicitly, experts' intervention is needed to identify them. The actionable portions of the requirement may be analyzed to extract the privacy properties.

Example 1: If the requirement contains the phrase "should not be disclosed to /published to /…", it implies that 'confidentiality' of data is being referred.

Example 2: If the requirement contains the phrase "should be always correct and updated", it refers to 'integrity' of the data item.

Example 3: If the requirement says "none should know the existence of this data", it actually refers to the 'undetectability' of data item.

As mentioned above, protection of privacy means protecting one or more privacy properties of a data item. Among the privacy properties, confidentiality, integrity, availability and undetectability are the properties of a particular data item; anonymity is related to the revelation of identity of an individual from one or more data-items; unlinkability is about linking two data items; and unobservability is a combination of undetectability and anonymity.

Therefore confidentiality, integrity, availability and undetectability requirements should be calculated for each data item whereas anonymity and unlinkability requirements are determined based on the probable privacy harm that can be caused by any combination of data items used by the enterprise. For determining anonymity, there should be a pre-defined list containing all combinations of data items that can identify an individual. If anonymity is expected for any particular report, the privacy requirement should be such that no combination from the above list is present in that report simultaneously. For determining unlinkability, there should be a pre-defined list containing all combinations of data items that may reveal any information that is harmful for an individual. In order to protect unlinkability, those data items should not be used together at any stage of the data lifecycle.

B. *Enterprise Privacy Principles*

Besides the mandatory legal and contractual obligations, enterprises often adopt additional privacy safeguards in keeping with their business strategies. This is usually done to wean away customers from competitors by offering them greater safety of their sensitive data and PII. These privacy safeguards may be influenced by cultural and ethical issues, the need to provide special privileges to valued customers, need for smooth operations management etc.

Example 4: Publication of a report with photographs of authors may be acceptable or even encouraged in some cases, while it can be deemed as a violation of privacy in others.

Example 5: People who donate a large amount to charity may not wish to publish their details citing privacy violation.

Example 6: Even if cross-border data transfer is legally permissible, it may be safer for an enterprise to transfer data only to countries that have strict data privacy rules.

Thus, consideration of such additional safeguards, along with legal and contractual requirements, is beneficial for the holistic privacy management of the enterprise. The process of elicitation of these requirements is different from that of legal obligations, as it does not involve any pre-defined documents. Relevant threats should be analyzed to check whether data privacy properties may be breached in the absence of additional safeguards. The severities of harm, which may be potentially effected by those threats, need to be computed to determine the relative importance of the privacy requirements. Therefore, a threat modeling approach along with relevant constraints would give a better insight about such privacy requirements.

C. *Customers' Expectations*

Since privacy pertains to the safety of user data, the most important source of requirements is users'/customers' expectations. Enterprises should address customers' concerns to ensure a large and stable customer base. Customers should be made aware of the specific privacy safeguards and risks involved in using the services / products of an enterprise. It is essential to obtain their consent for usage of their private data in specific manners [11]. An enterprise may also require its customers to pay certain charges in order to avail of customized privacy safeguards.

Unlike other sources of privacy requirements, customers' expectations are not generic; they are usually different for each customer. Hence, the most effective method of capturing these requirements is with the help of questionnaires. For data privacy properties like confidentiality, integrity, availability and undetectability, customers should be provided with options for choosing their relative importance (graded requirement). Anonymity and unlinkability should be handled differently. Customers may be required to list the (i) purpose(s) for which they wish to remain anonymous; and (ii) purposes that they wish not to link together.

Besides, they may be provided with options to add certain constraints for each action that can be performed on their sensitive data.

It is already stated above that the statements of legal privacy requirements are generally about specific actions that are permitted / denied on a particular data item by a particular actor, subject to different constraints. Constraints can be of different types: temporal constraint where permission or denial is valid only within certain time intervals; spatial constraint where some actions are permitted in presence or absence of certain resources; clauses regarding purpose where actions are permitted for certain purposes only; and some pre-conditions and post-conditions of events.

Example 7: The data owner should be notified within 72 hours of any disclosure of data to authorized persons (post-condition constraint); no personal data should be collected in a kiosk having CCTV (spatial constraint); data can be disclosed to unauthorized users only in case of medical emergency of data owner (pre-requisite constraint).
The privacy requirements can be specific to certain life-cycle phase of data or can be generic throughout its lifecycle within the enterprise.

Example 8: Data should be encrypted while it is being transferred outside the organization; data should be anonymized while in storage etc.
The methodology of assigning values to the *importance* parameter, in case of legal requirements, has been shown above. For enterprise privacy principles, *importance* parameter can obtain its value from the likelihood of occurrence of threat and severity of corresponding vulnerabilities, in case threat modelling approach is followed. From the above discussion, it follows that legal requirements and enterprise privacy principles can be expressed in terms of data item, lifecycle phase of data item, action on that data item, actor of the action, whether the action is permitted or prohibited and the relative importance of the requirement. From the action and the lifecycle phase, the protection need of the privacy properties related to that privacy requirement can be determined.

IV. PROPOSED METHODOLOGY

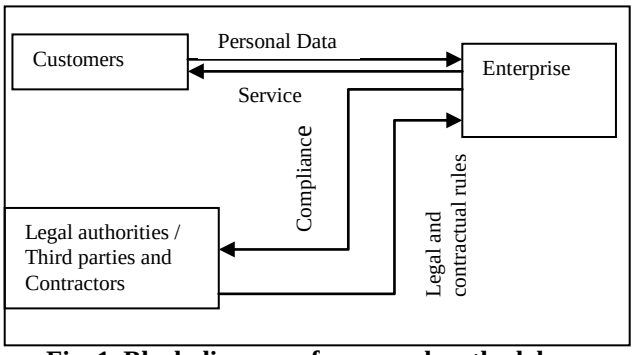

Fig. 1. Block diagram of proposed methodology

Fig. 1 represents the block diagram of interactions among enterprise, customers and legal authorities. In this section a procedure for formalization of privacy requirements is described and an algorithm for resolving conflicts among privacy requirements is also proposed.

A. *Formalization of Privacy Requirements*

In order to formally capture the privacy requirements of all three types, following entities are required:
Data item: Data item is an atomic piece of data.
Types of Data: It represents a set of types of data depending upon the relation of a data item with individuals. Data can either be submitted/collected or generated. In case of generated data, data can be generated by using data of a single individual or multiple individuals. Thus, there are four types of data: collected, generatedSingle, generatedMultiple and footprint.
Protection Levels: It denotes a set of degree of privacy needed for a data item. The related data of an individual, which can't be changed, is termed as sensitive data and it needs utmost level of protection. The data which can identify an individual uniquely, but is not sensitive, is termed as PII. The data which is personal in nature but not a PII, is termed as personal data. The data trail left by an individual is termed as footprint data. Footprint data needs least amount of protection.
For example: biometric data or date of birth is sensitive data; name, address, telephone number, email id etc. are PII data; pathological test reports or financial transactions are personal data whereas the general trail of online browsing is footprint data.
Data subject: Each data has an owner. Depending upon the type of data, the owner may vary. For the data belonging to 'collected' or 'generatedSingle' category, individual is the owner of that data and termed as data_subject whereas for the data belonging to 'generatedMultiple' category, enterprise is the data owner.
Data controller and Data Processor: If a data item is submitted to an enterprise for some purpose, the enterprise is termed as the *data controller* for that data item. If the data controller shares the data with a third party for processing, that third party is regarded as the *data processor*.
Actors: Actors represent a set of users, who perform some actions. An actor can be a data subject, a data controller or data processor or can be an employee of the data controller.
Action: Action refers to 'change of state of data'. While the state of data is changed, it may lead to loss of some privacy parameters of the data. Therefore, each action along with permission and constraints should be converted to the protection need of one or more privacy parameters for a data item. For example: disclosure of data, storing data, etc.
Constraint: Constraint refers to a condition that needs to be fulfilled while performing the action. It can be a pre-condition, a post-condition, a spatial constraint or a list of purposes for which the action is permitted or denied.
Operations: Operations are some actions along with constraints.
Permission: Permission can be +ve or –ve. It denotes whether an action is permitted or denied.
Privacy parameters: Privacy parameters are privacy properties of a data item. There are six privacy parameters: confidentiality, integrity, availability, anonymity, unlinkability and undetectability.
Lifecycle phases: Lifecycle phase refers to the phase, in which the data is currently in. Different lifecycle phases are collection, storage, processing, transfer, archiving and disposal.
Privacy Importance: It indicates the degree of importance of a privacy requirement and can have values on a 3-point scale: {high, mid and low}.

Table-I: Entities for modeling privacy requirements

Entity	Representation
Data item	d_j denotes a data item
Types of data	*type_data = {collected, generatedSingle, generatedMultiple}*
Protection Levels	*protection_level = {sensitive, PII, personal, footprint}*
Data subject	*isDataSubject*(d_j, ds_i) denotes that data d_j is owned by the data subject ds_i.

Data controller and Data Processor	*isController*(d_j, dc_x) denotes that data d_j is in possession of data controller dc_x. *isProcessor*(d_j, dp_y) denotes that data d_j is processed by data processor dp_y
Actors	$Actor_k$ represents an actor
Action	*action* represents an action
Constraint	*constraint = (prec, postc, spatialc, list_req)*
Operations	*op = (action, constriant)*
Permission	*perm = {+,-}*
Privacy	*pp = {'C', 'I', 'A', 'An', 'Ul', 'Ud')*
parameters	
Lifecycle phase	*lcp = {collection, storage, processing, transfer, archiving, disposal}*
Privacy Importance	*Importance={high,mid,low}*

Representations of the above entities are listed in Table I. Table II shows the description and examples of different constraints.

Table- II: Constraints

Name of the constraint	Description	Representation and example
Pre-condition	This action is allowed within 'n_{pre}' time interval of event. 'n' can be expressed in days and hours. E.g: 72 hours or 2 years (730 days)	prec = (event, n) prec = ('receipt of approval', '0:12')
Post-Condition	Once this action is over, an event should take place within 'n' time. E.g. the data subject should be informed within 24 hours of data disclosure.	postc = (event, n) postc =('1:00' ,'inform datasubject')
Spatial constraint	Action can take place from a place with presence or absence of certain resources. If the permission is positive, the list of resources should be present while performing the action and if the permission is negative, it indicates action should not be performed with presence of those resources in that place. The place can have values like {secured_channel, company_premises, special_zone, internet}.	spatialc={place, re} spatialc={internet, 'RFID token'} or spatialc={special_zone, 'CCTV'}
Prerequisite purposes	This is a list of purposes for which the action should be performed (in case of positive permission) or should be denied (in case of negative permission)	list_req=(list of purpsoes) list_req=(pur_1,pur_2,...,pur_m)

Formally a requirement can be written as

$requirement_l = (d_j, lcp, actor, op, perm, importance, pp)$.....(1)

If a generatedSingle type of data item is generated using multiple data items, its privacy requirements would be inherited from the privacy requirements of those data items. The default requirement would be the highest (strictest) requirement among all requirements of those data items; however, the enterprise can modify the requirement if it so desires. Similarly, the default protection level of the generated data item would be the highest protection level among all the individual data items used for the generation process. Using a representative set of rules from HIPAA [12], the expressiveness of the proposed model is shown in Table-III.

Table- III: Part of HIPAA expressed using proposed entities

Rule Name	Rule details	Formal representation
General principle for uses and disclosure	A covered entity may not use or disclose protected health information except either 1) As the privacy rule permits or 2) subject of the information authorizes in writing	i. Any d_i, if it is of protection_level PII. ii. actor – data controller iii. operation – action – use/disclosure constraint – event-triggered: permitted by any of the disclosure rules or written authorization by subject iv. permission – positive v. importance – mid vi. privacy parameter affected: confidentiality/ undetectability; lcp - disclosure
	A covered entity must disclose protected health information in only two situations: (a) to individuals (or their personal representatives) specifically when they request access to, or an accounting of disclosures of, their protected health information; and (b) to HHS when it is undertaking a compliance investigation	i. Any d_i, if it is of protection_level PII. ii. actor – data controller iii. operation – action – use/disclosure constraint – event-triggered: requested by data subject or event-triggered – requested by govt on legal ground iv. permission – positive v. importance – high vi. privacy parameter affected: confidentiality/ undetectability; lcp -disclosure
Permitted Uses and Disclosures	A covered entity is permitted, but not required, to use and disclose protected health information, without an individual's authorization in following situation: a) (1) To the Individual (unless required for access or accounting of disclosures); (2) Treatment, Payment, and Health Care Operations; (3) Opportunity to Agree or Object; (4) Incident to an otherwise permitted use and disclosure; (5) Public Interest and Benefit Activities (6) Limited Data Set for the purposes of research, public health or health care operations	i. Any d_i, if it is of protection_level PII. ii. actor- data controller iii. operation – use/disclosure constraint – list of purposes for which the action is permitted iv. permission – positive v. importance - medium (permitted but not required) vi. privacy parameter affected: confidentiality/ undetectability/ anonymity; lcp - disclosure

Authorized Uses and Disclosures	A covered entity must obtain the individual's written authorization for any use or disclosure of protected health information that is not for treatment, payment or health care operations or otherwise permitted or required by the Privacy Rule	i. Any d_i, if it is of protection_level PII. ii. actor- data controller iii. operation – use/disclosure constraint – authorization needed for purpose other than enlisted ones. iv. permission – positive v. importance - high vi. privacy parameter affected: confidentiality/ undetectability/ availability; lcp - disclosure
Authorized Uses and Disclosures	All authorizations must be in plain language, and contain specific information regarding the information to be disclosed or used, the person(s) disclosing and receiving the information, expiration, right to revoke in writing, and other data.	i. authorization request –personal data ii. actor – data controller iii. operation/ constraint – collection constraint- specification about information to be disclosed, receiver of the information, expiry time, right to revoke etc. iv. permission- positive v. importance – high vi. privacy parameter- availability, detectability; lcp - collection
Limiting Uses and Disclosures to the Minimum Necessary	A covered entity must make reasonable efforts to use, disclose, and request only the minimum amount of protected health information needed to accomplish the intended purpose of the use, disclosure, or request	i. authorization request –reports containing PII data ii. actor – data controller iii. operation/ constraint – use/disclosure constraint- if any of the PII data used for report generation is withdrawn, the report should not be generated iv. permission- positive v. importance – high vi. privacy parameter- anonymity, unlinkability; lcp -disclosure

B. *FLOWCHART OF CONFLICT RESOLUTION PROCESS*

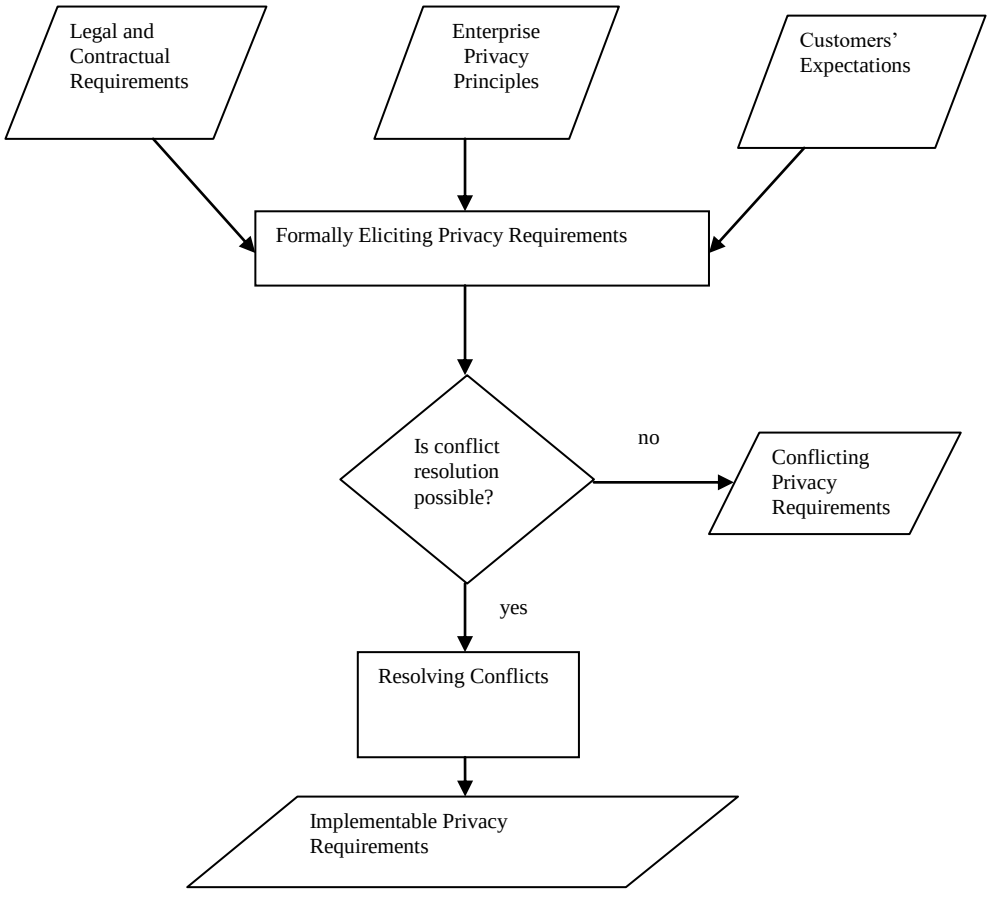

Fig. 2. Flow-chart of the conflict resolution mechanism

The privacy requirements generated from different sources may sometimes be in conflict with each other. In fact, the same source may also produce conflicting requirements.

Example 9: Law may permit cross-border data transfer to any country. However, a user may not wish to allow her private data to move to a country where no data protection law exists.

Example 10: An enterprise may have business interests in two countries whose privacy laws are in conflict with each other. Some specific categories of conflicts are as follows:

a) Conflict in the temporal specification of pre-condition or post-condition while both requirements are having same permission or both requirements have conflicting permissions having same pre-conditions or post-conditions.
b) Conflicts in the list of purposes for permitted actions and denied actions.
c) Conflicts about presence and absence of certain resources for a particular action.

Such conflicts may be resolved with the help of a set of pre-defined rules. However, it is practically impossible to predict all instances of conflicts beforehand. New categories of conflicts should be resolved manually.

Knowledge from such resolution may be used to define new rules that would enhance the existing rule-set. Examples of conflict resolution rules are: requirements having higher relative importance will get priority; or requirements having stricter temporal/spatial constraints may get priority etc. A flow chart of the proposed procedure is given in Fig. 2.

C. *ALGORITHM*

In the algorithm *validRequirements*, it is assumed that conflicts are resolved based on the relative importance of requirements; if the requirements have same relative importance, conflicts are left unresolved and error-flags are raised. Algorithm *validRequirements* considers the set of legal requirements and enterprise privacy principles, groups them as per data, lcp, actor and action of the operation, and finally identifies the conflicts among those requirements.

The checking of conflicts has been performed with the help of three different procedures as follows: procedure *condConf* checks for conflicts in pre-conditions and post-conditions; procedure *listConf* checks for conflicts in the list of permitted or denied pre-requisites of two policies; procedure *spatialConf* checks for conflicts among spatial constraints of two policies. If the checks indicate that there is no conflict, or that one policy is unambiguously stricter than the other, there is no need for conflict resolution. Otherwise, the procedure *resolve_conflict* would be applied to resolve conflicts based upon the relative importance of two policies. If the procedure fails to resolve conflicts, then the policies would be marked as "conflicting requirements"; resolution of such conflicts may require manual intervention.

Once the conflicts between legal requirements and enterprise privacy principles have been resolved, customers' expectations can be accommodated. A set of conflict resolution rules should be devised beforehand and then for each customer, requirements should be compared with the combined set of legal requirements and enterprise principles, and conflicts should be resolved using the specified rule-set.

```
Algorithm validRequirements()
1. get all 'm' requirements generated by legal requirement and enterprise privacy principles as req set.
2. group them based on (di, lcp, actor, operation.action) and get a set of 'n' requirements
3. perform step 4 to 7 for each group of 'n' requirements
4. get a flag[n] array and set every bit to 'valid'
5. for i=1 to n, take req[i] as r1, provided flag[r1]=valid
        a. for j=i+1 to n, take req[j] as r2, provided flag[r2]=valid
        b. result = checkConflict(r1,r2)
        c. if (result == 'r1') set flag[r1]=overruled
              else if (result=='r2') set flag[r2]=overruled
              else if(result==conflict) set flag[r1]=flag[r2]=invalid, raise Error
6. select all requirements having valid bit and return those as requirement set
7. select all requirements having invalid bit and return those as error_set
```

```
procedure checkConflict(r1,r2)
1. s1 = condConf(r1,r2,r1.op.prec,r2.op.prec)
2. s2 = condConf(r1,r2,r1.op.postc,r2.op.postc)
3. s3= listConf(r1, r2)
4. s4 = spatialConf(r1,r2)
5. if(s1 && s2 && s3 && s4)
        return 'valid';
   if((s1 && s2 && s3 && s4)== 'r1')
        return 'r1';
   if((s1 && s2 && s3 && s4)== 'r2')
        return 'r2';
   else return (resolve_conflict(r1,r2));
```

```
procedure resolve_conflict(r1,r2)

if (r1.importnace > r2.importance)
        return 'r2';
   if (r1.importnace > r2.importance)
        return 'r1';

   return 'conflict';
```

```
procedure condCon(r1, r2,p1,p2)
   if ((p1.event == p2.event) && (p1.n==p2.n) && (r1.perm≠r2.perm)) {
    return FALSE;}
   else if ((p1.event == p2.event) && (p1.n≠p2.n) && (r1.perm==r2.perm )){
    if(p1.n==null) return 'r1';
    else if(p2.n==null)return 'r2';
     else return FALSE;
    }
    else if(p1.event==null) return 'r1';
    else if(p2.event==null)return 'r2';
    else return 'TRUE';
```

```
procedure listConf (r1, r2 )
  if((r1.perm ≠ r2.perm) && (r1.op.list_req ∩ r2.op.list_req ≠ φ))
    return FALSE;

else return TRUE;
```

```
spatialConf (r1,r2)
    if((r1.perm ≠ r2.perm) && (r1.op.spatialc.place == r2.op.
  spatialc.place ) && (r1.op. spatialc.re == r2.op. spatialc.re ))
        return FALSE;
   return TRUE;
```

V. RESULT ANALYSIS AND DISCUSSION

To show the effectiveness and outcome of our proposed algorithm, a case study is provided in this section. An enterprise may follow privacy guidelines suggested by different standardization organizations simultaneously and can have different privacy contracts with third parties and subsidiaries. An enterprise can have its own privacy principles as well.

The number of policies of an enterprise can be numerous but for our purpose two small example sets of privacy policies of a university are considered as legal privacy requirements and enterprise privacy principles, shown in Table IV and Table V, respectively.

In order to protect the privacy of students of the university, legal authorities have mandated that any report containing PII of a student should be disclosed to outsiders only on receipt of authorization letter from the concerned data owner. However, to protect students' interests, the university decided that they would disclose the result within 24 hours of receipt of authorization and they would inform the concerned data owner after the disclosure. These two policies are represented as L1 and E1 in Tables IV and V, respectively.

As another policy, legal authorities suggested that students' personal data should not be used by the university in advertisements. This suggestion is a 'good practice' but not mandatory as per law. However, the university proposed that they would be using personal data in advertisements provided the data-owner gives consent for the same. These two policies are represented as E2 and E3 in Tables IV and V, respectively.

Table- IV: Sample set of legal requirements

	Legal Requirement	Formal Representation
L1	The report containing PII should be disclosed to specially authorized personnel only after getting authorization letter from data owner.	(PII, disclosure, univ, disclose to outsider, +, high, C, cons) *where* cons={('receipt of authorization report, null), null, null, null}
L2	It is not encouraged to transfer personal data to third parties for advertisement purpose	(Personal, transfer, univ, transfer to third party, -,mid, C, cons) *where* cons = {(null, null), (null, null), null, (advertisement)}

Table- V: Sample set of enterprise privacy policies

	Enterprise Privacy Policy	Formal Representation
E1	Any report containing PII should be disclosed to outsiders within 24 hours of receipt of approval from data owner and the data owner should be informed within 6 hours.	(PII, disclosure, univ, disclose to outsider,+, high, confidentiality, cons) *where* cons={('receipt of authorization report',00:24), ('inform data owner',00:06), null, null}
E2 and E3	Personal data can be transferred to third parties for any purpose if the consent is given by data subject in the past 6 months. It can be transferred to third parties without specific consent in case demanded by police, legal authority or ministry of foreign affairs.	(Personal, transfer, univ, transfer to third party, +,mid, C, cons) *where* cons = {('receipt of authorized consent',180:00),(null),null,(null)} (Personal, transfer, univ, transfer to third party, +,mid, C, cons) *where* cons = {(null, null), (null, null), null, (police_demand, legal_need, ministry_need)}

According to the proposed algorithm, L1, L2, E1, E2 and E3 form the total input set m. As per (d1, lcp, actor, action.operation) grouping, L1 and E1 form the group n1 and L2, E2, E3 form the group n2. ValidRequirements algorithm would be invoked with members of n1 first. The flag[L1] and flag[E1] would be set to 'valid'. CheckConflict method will be invoked for (L1, E1). The method would in turn invoke condCon method twice for identifying conflicts in pre-conditions and post-conditions and would invoke listConf and spatialConf methods for identifying conflicts in spatial conflicts and list of purposes. In this case, as E1 has a post-condition and L1 does not, E1 would be preferred; for pre-condition too, E1 would be preferred as it has a temporal constraint in the pre-condition unlike L1. No conflict arises from the other two conflict-checking procedures. Therefore, L1 would be overruled by E1.

Similarly L2 and E2 would have a conflict if the purpose is 'advertisement' and the university receives the consent from the student for that purpose.

Thus, this case study illustrates that conflicts can be identified and resolved as per pre-defined rule-sets using the proposed methodology.

VI. CONCLUSION

In this paper, it has been shown how different privacy requirements like legal and contractual requirements, enterprise privacy principles and customers' expectations can be elicited and expressed formally. It has also been shown that the proposed methodology is expressive enough by using an example subset of a privacy regulation, namely HIPAA. The importance of expressing privacy requirements in terms of privacy properties, and a method for eliciting the same, has been discussed. Finally, an algorithm for resolving conflicts within privacy requirements has been proposed, and the same has been illustrated with the help of a case study.

We intend to automate the requirements elicitation process as part of future work. A formal language may be designed for capturing the privacy requirements methodically. This may be used to design and develop a tool for the elicitation of privacy requirements and resolution of conflicts therein.

REFERENCES

1. M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen, "A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements," Requir. Eng., vol. 16, pp. 3-32, March 2011
2. M.J. May, C.A. Gunter, and I. Lee, "Privacy APIs: Access Control Techniques to Analyze and Verify Legal Privacy Policies," Proc. 19th IEEE Computer Security Foundations Workshop,pp. 85-97, 2006
3. T.D. Breaux, M.W. Vail, and A.I. Anton. "Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations." 14th IEEE International Requirements Engineering Conference (RE'06). IEEE, 2006.
4. W. Hassan, and L. Logrippo. "Governance requirements extraction model for legal compliance validation." Requirements Engineering and Law (RELAW), 2009 Second International Workshop on. IEEE, 2009.

5. J.D. Young "Commitment analysis to operationalize software requirements from privacy policies." Requirements Engineering 16.1 (2011): 33-46
6. J.D. Young and A.I. Anton. "A method for identifying software requirements based on policy commitments." Requirements Engineering Conference (RE), 2010 18th IEEE International. IEEE, 2010.
7. A. Barth, et al. "Privacy and utility in business processes." 20th IEEE Computer Security Foundations Symposium (CSF'07). IEEE, 2007
8. A.I. Antón, , J. B. Earp, and A. Reese. "Analyzing website privacy requirements using a privacy goal taxonomy." Proceedings IEEE Joint International Conference on Requirements Engineering. IEEE, 2002
9. C. Kalloniatis, E. Kavakli, and S. Gritzalis, "Addressing privacy requirements in system design: the pris method," Requir. Eng., vol. 13, pp. 241-255, August 2008.
10. A. Pfitzmann and M. Hansen, "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, unobservability, pseudonymity, and identity management - version v0.34," TU Dresden and ULD Kiel, Tech. Rep., 2011. [Online]. Available: http://dud.inf.tu-dresden.de/literatur/Anon- Terminology-v0.34.pdf
11. European Commission,"2018 reforms of EU data protection rules",May 25, 2018, Available: https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf, [Accessed November 1, 2019]
12. Act, Accountability. "Health insurance portability and accountability act of 1996." Public law 104 (1996): 191.

AUTHORS PROFILE

Asmita Manna is a research scholar in the department of Computer Science and Engineering, Jadavpur University, Kolkata, India. Prior joining PhD, she completed her Masters of Engineering in Software Engineering from Jadavpur University in 2009. Her primary research domain is Information Security and Privacy.

Dr. Anirban Sengupta is a research engineer in the Centre for Distributed Computing, Jadavpur University, Kolkata, India. His research interests include Information Security, Risk Assessment and Privacy.

Professor Chandan Mazumdar is a senior professor in the department of Computer Science and Engineering, Jadavpur University, Kolkata, India. His research interests include Information Security, Risk Assessment, Insider Threat Assessment and Privacy.