Providing Privacy for Composition Results in Web Service Using Data Anonymization

T.Manikandan¹ and K.Palanivel²

¹Department of Computer Science and Engineering, Pondicherry University, Puducherry -605014, India
²Department of Computer Centre, Pondicherry University, Puducherry -605014, India

Abstract
Web service composition is a network technology with the aim of merging information commencing from more than one resource into a solitary web application. This practice affords extraordinary type of masterpiece application that aims to incorporate data from manifold facts providers depending on the consumer’s demand. In accumulation, composition results in web service may expose isolation sensitive information or personalized data. When imposing a traditional privacy preserving model and negotiation, the composition service data would suffer from the problem of privacy attacks. In this paper we propose a new dynamic privacy model, by designing anonymization techniques for protecting the composition results from privacy attacks before the final result is returned by the mediator. Here, we use e-Epidemiological Scenario in which patient’s details are prevented from privacy attacks. We use this model for managing the trustworthiness of Web services involved in web service compositions. We introduce the generation of reputation information throughout the composition to aid all the services involved for making informed decisions about the selection of their respective component services. According to the increasing or decreasing service stature, we ensure that no service is wrongfully blamed. Our anonymization technique is effective and efficient when compared to the previous approaches.

Keywords: data privacy, DaaS, Web service composition, Data anonymization

1. INTRODUCTION
Web services have recently emerged as a popular medium for data publishing and sharing on the Web [1]. Modern enterprises across all scopes are moving towards a service-oriented architecture by putting their databases behind Web services, thereby providing a well-documented, computing platform independent and interoperable method of interacting with their data. This new type of services is known as DaaS (Data-as-a-Service) services where services correspond to calls over the data sources. DaaS eliminates redundant data and reduces associated outlay by accommodating important data in a single location, allowing data use and modification by multiple users via a single update point. Initially used in Web mashups, the DaaS strategy is often used by business organizations. Data as a Service or DaaS is a type of software as a service. The emergence of service-oriented architecture (SOA) has rendered the actual platform on which the data resides also irrelevant. This development has enabled the recent emergence of the relatively new concept of DaaS. A service provider enables data access on demand to users regardless of their geographic location. Potential disadvantages of data services include server outage duration from the data service provider, data loss in the event of a calamity, and the security of the data, both in its stored location and in the transmission of the data among users. Web service composition is a technology that combines information from more than one source into a single web application. This technique provides a special type of composition application that aims at integrating data from multiple data providers depending on the user’s request. Figure 1 shows the model diagram for web services.

![Figure 1: Model diagram for web services](image)

The basic idea of the existing model is that to locate users closely with each other who are more likely to have similar service experience than those who live far away from each other. Inspired by the accomplishment of Web 2.0 websites that highlight information sharing, collaboration, and interaction, we utilize the idea of user-collaboration in our web service recommender system. Here, we use the epidemiological scenario to illustrate the privacy challenges during service composition.
Challenge 1:
In web services users needs to invoke some information. They want to keep invocation as private since this may disclose sensitive information to others. The aforementioned first challenge puts in evidence the need for a formal model to specify private data is and how it will be defined.

Challenge 2:
Component services may require input data that cannot be disclosed by other services because of privacy concerns. They may also have conflicting privacy concerns regarding their exchanged data.

Challenge 3:
The role of the mediator is to return composite services with compatible component services with respect to privacy. The simplest way to deal with compositions with incompatible privacy policies is to reject the composition. However, a more interesting, yet challenging approach would be to try to reach a consensus among component services to solve their privacy incompatibilities hence increasing the number of composition plans returned by the mediator. In an existing system, there are two factors exacerbate the problem of privacy in DaaS. First, DaaS services collect and store a large amount of private information about users. Second, DaaS services are able to share this information with other entities. Besides, the emergence of analysis tools makes it easier to analyze and synthesize huge volumes of information, hence increasing the risk of privacy violation.

In order to overcome the tackles in the existing system, we propose a new dynamic privacy model. We aim at designing techniques for protecting the composition results from privacy attacks before the final result is returned by the mediator. We present a model for managing the trustworthiness of Web services involved in service compositions. We introduce the generation of reputation information throughout the composition to aid all the services involved, for making decisions regarding the selection of their respective component services. Our proposed dynamic model, deals with privacy in the web service composition for protecting the composition results from the privacy attacks before the final result is returned to the user. Data as a Service (DaaS) is an information provision and distribution model in which data files like text, images, sounds, and videos are made available to customers over a network, typically the Internet. In dynamic privacy model, data anonymization technique is used. The rest of the paper will be organised as follows: In section 2, we see about the related works of the paper. In section 3 we discuss about our proposed work. The conclusion of our paper is in section 4.

2. RELATED WORKS
In this section, we will see some of the related works for using different approaches:

Web service composition enables uninterrupted and dynamic integration of business applications over the web. The performance of the composed application is determined by the performance of the involved web services. Therefore, non-functional and QoS aspects are necessary for choosing the web services to take part in the composition. Finding out the best candidate web services from a set of functionally-equivalent services is a multi-criteria decision making problem. The services should optimize the overall QoS of the composed application, while satisfying all the constraints described by the client on individual QoS parameters. Data-Providing (DP) services allow query-like access to organization’s data through web services. The invocation of a Data-Providing service results in the execution of a query over data sources. In most cases, user’s queries require the composition of several services. Mashup [3] is a web technology that allows different service providers to flexibly integrate their expertise and to deliver highly personalized services to their customers. Data mashup is an important type of mashup application that aims at integrating data from multiple data providers depending on the user’s request. Service-oriented architecture (SOA) is becoming a major software framework for building complex distributed systems. Reliability of the service-oriented systems depends on the remote Web services as well as the unpredictable Internet. Designing accurate and effective reliability prediction approaches for the service-oriented systems has become an important research issue. In this paper, they propose a collaborative reliability prediction approach, which utilizes the past failure data of other similar users to predict the Web service reliability for the current user, without the need of real-world Web service invocations. They also present a user-collaborative failure data sharing mechanism and a reliability composition model for the service-oriented systems. A typical example of modeling privacy is the Platform for Privacy Preferences (P3P) [5]. However, the major focus of P3P is to enable only Web sites to convey their privacy policies. Data providers specify how to use the service (mandatory and optional data for querying the service), while individuals specify the type of access for each part of their personal data contained in the service: free, limited, or not given using the DAML-S ontology. In [6], Ran propose a discovery model that takes into account functional and QoS-related requirements, and in which QoS claims of services are checked with external components that act as certifiers. The authors refer to the privacy concern with the term confidentiality, and some questions are raised about how the service makes sure that the data are accessed and modified only by authorized personals. Some policy languages, such as XACML [7], ExPDt [8] are proposed and deployed over a variety of enforcement architectures. These languages are on the one hand syntactically expressive enough to symbolize complex policy rules, and offer on the other hand a formal semantics for operators to reason about policies, e.g., their conjunction and recently difference. Unfortunately, they do not provide solution when an incompatibility occurs. In our work, privacy resource is specified and may be related to client, Data and Service provider’s levels, and not only
to the provided data. Privacy-preserving mechanism for data mashup is represented in [9]. It aims at integrating private data from different data providers in secure manner. The authors in [10] discuss the integration and verification of privacy policies in SOA-based workflows. The previous approaches, related to data mashup and workflows, focus on using algorithms (such as k-anonymity) for preserving privacy of data in a given table, while in our work we go further and propose a model that also takes into account usage restrictions and client requirements. In comparison to the existing approaches, our privacy model defined in this paper goes beyond “traditional” data-oriented privacy approaches. Input or output data as well as operation invocation may disclose sensitive information about services and hence, should be based on privacy constraints. The proposal of [11] is based on privacy policy lattice which is created forming privacy preference-service item correlations. Utilizing this lattice, privacy policies can be visualized and privacy negotiation rules can then be generated. The Privacy Advocate approach [12] consists of three main units: the privacy policy evaluation, the signature and the entities preferences unit. The negotiation focuses on data receivers and purpose only. An extension of P3P aims at adjusting a pervasive P3P-based negotiation mechanism for a privacy control. It performs a multi-agent negotiation mechanism on top of a pervasive P3P system. The approach proposed in [13] aims at accomplishing privacy-aware access control by adding negotiation protocol and encrypting data under the classified level. In [14] they first introduce the problem of anonymization of private data: using a variety of techniques to modify the original data in such a way that the original “sensitive” data is masked. The need for anonymization is motivated by many legal and ethical requirements for protecting private, personal data. The intent is that anonymized data can be shared freely with other parties, who can perform their own analysis and investigation of the data. We will present examples which show the dangers of data release without rigorous anonymization, such as the AOL Search Data example and attacks on Netfliix data. Once the goal of anonymization is formalized, a fundamental trade-off is established between two aspects: the privacy goals of the data owners, and the utility goals of the data users. Most work in this area fixes a particular privacy requirement, and then tries to optimize the utility while guaranteeing this level of privacy. We will discuss various definitions of what is meant by both “privacy” and “utility”, using examples.

3. Proposed Work

3.1 Privacy-Enhanced Data-as-a-Service (PEDWA)

In Web service composition typically there is a mediator which combines the functionalities provided by other services usually denoted as component services to satisfy user’s requests. Several services may be able to provide the same functionality requested by the user. The service resulting from the orchestration is called composite service. We model the composition schema as a mediator model, each component service as a component service model, and all possible alternative instantiations of the schema as a service orchestration model. The service orchestration model is obtained by merging the service models associated with the mediator and all component services. In particular, we combine the service model of the outsourcer with the service model of the subcontractor by linking the goal of the former with the corresponding goal (with the same name) occurring in the service model associated with the latter. Intuitively, goals with the same name represent the same functionality and, therefore, can be considered equivalent (although they may require different data items). To complete the interaction with a Web service (composite or simple), the user has to disclose their personal information to the service. However, users may be concerned about revealing their personal data. Data protection law aims to address these user concerns. On the other side data protection law empowers users to control their data. To this end, they may define privacy preferences which specify constraints on the collection and processing of their data. On the other hand, Web service suppliers (both the orchestrator and component services) are obliged by law to publish privacy policies in which their privacy practices are declared. Here, we consider four privacy dimensions which are typically used to specify privacy policies and privacy preferences: purpose defines the reason(s) for data collection and usage; visibility defines to whom data can be disclosed; retention period defines how long data can be maintained; sensitivity represents the data subject’s perception of the harm the misuse of the data can cause. In privacy preserving data publishing, in order to prevent privacy attacks, data should be anonymized properly before it is released. Anonymization methods should take into account the privacy models of the data and the utility of the data. Generalization and perturbation are the two popular anonymization approaches for relational data. In figure 2, architecture of privacy-enhanced DaaS web service is depicted.
3.2 METHODOLOGY USED

Data anonymization [15] is the process of either encrypting or removing personally identifiable information from data sets, so that the people whom the data describe remain anonymous. The Privacy Technology Focus Group defines it as "technology that converts clear text data into a nonhuman readable and irreversible form, including but not limited to preimage resistant hashes (e.g., one-way hashes) and encryption techniques in which the decryption key has been discarded. Data anonymization enables the transfer of information across a boundary, such as between two departments within an agency or between two agencies, while reducing the risk of unintended disclosure, and in certain environments in a manner that enables evaluation and analytics post-anonymization. In the context of medical data, anonymize data refers to data from which the patient cannot be identified by the recipient of the information. The name, address, and full post code must be removed together with any other information which, in conjunction with other data held by or disclosed to the recipient, could identify the patient. De-anonymization is the reverse process in which anonymous data is cross-referenced with other data sources to re-identify the anonymous data source.

3.3 CASE STUDY

3.3.1 e-Epidemiological Scheme

The first module is e-Epidemiology scheme. We develop the scenario of epidemiology. Epidemiology is the science underlying the acquisition, maintenance and application of epidemiological knowledge and information using digital media such as the internet, mobile phones, digital paper, digital TV. E-epidemiology also refers to the large-scale epidemiological studies that are increasingly conducted through distributed global collaborations enabled by the Internet. The traditional approach in performing epidemiological trials by using paper questionnaires is both costly and time consuming. The questionnaires have to be transformed to analyzable data and a large number of personnel are needed throughout the procedure. Modern communication tools, such as the web, cell phones and other current and future communication devices, allow rapidly and cost-efficient assembly of data on determinants for lifestyle and health for broad segments of the population. The mediator selects, combines and orchestrates the DaaS services (i.e., gets input from one service and uses it to call another one) to answer received queries. It also carries out all the interactions between the composed services (i.e., relays exchanged data among interconnected services in the composition). The result of the composition process is a composition plan which consists of DaaS that must be executed in a particular order depending on their access patterns (i.e., the ordering of their input and output parameters).

3.3.2 Privacy Level

In this module we define two privacy levels: data and operation. The data level deals with data privacy. Resources refer to input and output parameters of a service (e.g., defined in WSDL). The operation level copes with the privacy about operation’s invocation. Information about operation invocation may be perceived as private independently on whether their input/output parameters are confidential or not. For instance, let us consider a scientist that has found an invention about the causes of some infectious diseases, he invokes a service operation to search if such an invention is new before he files for a patent. When conducting the query, the scientist may want to keep the invocation of this operation private, perhaps to avoid part of his idea being stolen by a competing company. We give below the definition of the privacy level.

3.3.3 Privacy Rule

The sensitivity of a resource may be defined according to several dimensions called privacy rules. We call the set of privacy rules set (RS). We define a privacy rule by a topic, domain, level and scope. The topic gives the privacy facet represented by the rule and may include for instance: the resource recipient, the purpose and the resource retention time. The “purpose” topic states the intent for which a resource collected by a service will be used; the “recipient” topic specifies to whom the collected resource can be revealed. The level represents the privacy level on which the rule is applicable. The domain of a rule depends on its level. Indeed, each rule has one single level: “data” or “operation”. The domain is a finite set that enumerates the possible values that can be taken by resources according to the rule’s topic. For instance, a subset of domain for a rule dealing with the right topic is (“no-retention”, “limited-use”). The scope of a rule defines the granularity of the resource that is subject to privacy constraints. Two rules at most are created for each topic: one for data and another for operations.
3.3.4 Privacy-aware Service Composition

We propose a compatibility matching algorithm to check privacy compatibility between component services within a composition. The compatibility matching is based on the notion of privacy subsumption and on a cost model. A matching threshold is set up by services to cater for partial and total privacy compatibility. In this module we also propose an algorithm called PCM (Privacy Compatibility Matching). The first option is to require full matching and the second is partial matching.

3.3.5 Negotiating Privacy in Service Composition

In the case when any composition plan will be incompatible in terms of privacy, we introduce a novel approach based on negotiation to reach compatibility of concerned services (i.e., services that participate in a composition which are incompatible). We aim at avoiding the empty set response for user queries by allowing a service to adapt its privacy policy without any damaging impact on privacy. Negotiation strategies are specified via state diagrams and negotiation protocol is proposed to reach compatible policy for composition.

3.3.6 Anonymization for Data

In this module, we maintain the privacy information in the web service by anonymize the user’s data. By using anonymization technique we secure the user sensitive information from the attacks.

4. CONCLUSION

In this paper, we proposed an energetic solitude model for Web services. The model deals with privacy for the sensitive data and personalized information of the user. We also planned a concession come close to tackle the incompatibilities flanked by privacy policies and necessities. Even though seclusion cannot be inaccurately bargain as emblematic data, it is impossible to consult a part of privacy policy for unambiguous purposes. In any container, privacy policies forever reflect the practice of private data as specified or arranged upon by service providers. We aimed at designing techniques for protecting the composition results from privacy attacks before the final result is returned by the mediator. Our proposed techniques are effective and efficient when compared to the previous approaches through our experimental and simulation analysis.

References


