How to Enhance Privacy within DaaS service Composition?

Salah-Eddine Tbahriti, Chirine Ghedira, Brahim Medjahed, Michael Mrissa, Djamal Benslimane

▶ To cite this version:

Salah-Eddine Tbahriti, Chirine Ghedira, Brahim Medjahed, Michael Mrissa, Djamal Benslimane. How to Enhance Privacy within DaaS service Composition?. IEEE Systems Journal, 2013, 7 (99), pp.442-454. 10.1109/JSYST.2013.2244971 . hal-00814394

HAL Id: hal-00814394

https://hal.science/hal-00814394

Submitted on 17 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

How to Enhance Privacy within DaaS service Composition?

Salah-Eddine Tbahriti, Chirine Ghedira, Brahim Medjahed, and Michael Mrissa

***Abstract* — The composition of DaaS (Data-as-a-Service) services is a powerful solution for building value-added applications on top of existing ones. However, privacy concerns are still among the key challenges that keep hampering DaaS composition. Indeed, services may follow different, conflicting privacy specification with respect to the data they use and provide. In this paper, we propose an approach for enabling privacy-aware composition of DaaS services. Our approach allows specifying the privacy requirements and policies of services and verifying their compatibility for the services involved in a composition. We propose an adaptation protocol that makes it possible to reconcile the privacy specifications of services when incompatibilities arise in a composition. We validate the applicability of our proposal through a set of experiments.**

***Index Terms*—DaaS, composition, service, adaptation, privacy.**

I. Introduction

Services of type DaaS (Data-as-a-Service) have been considered during the last few years as first-class objects that can manipulate data much like database management systems do [2][17]. They also have started to be a popular medium for data publishing and sharing on the Web. Besides, modern enterprises across all spectra are moving towards service-oriented architectures by wrapping their data sources in DaaS services for more efficient data integration [2][6][17].

DaaS Composition consists in combining several DaaS services to realize Business-to-Business (B2B) interactions described according to a business process [3][4][5][1]. While initial service composition approaches have been a powerful solution for building value-added services on top of existing ones, the issue of privacy is still considered as an important topic in the field of service computing [1][16][31][33]. Indeed, despite important efforts aimed at preserving privacy [32], privacy leakage incidents on the Web continue to make the headlines. As example, in 2011, 535 breaches, involving a combined 30.4 million sensitive records, have been identified [45]. Besides, the emergence of analysis tools makes it easier to analyze and synthesize huge volumes of information, hence increasing the risk of privacy violation. According to a recent report [44], the number of reported electronic health data breaches has increased by 32% from the year 2010 and electronic medical data breaches only cost the industry about $6.5 billion. The concept of privacy itself generates much debate. On one hand, some might argue that the term *privacy* can be applied only to humans and not to institutions. On the other hand, there is no unanimous agreement about which information should be considered private. For example, some individuals choose to publish personal information such as pictures, videos, and their phone number, while others keep this information private and under no circumstances want it to become public. In the service composition, the issue of privacy is more challenging task. Let us illustrate some privacy challenges that occur during service composition through the following scenario.

A. Scenario and Challenges

We consider the following epidemiologist's query Q (as a part of a global request *R*): "*What are the ages, genders, zips, DNA and salaries of patients infected with H1N1; and what are the global weather conditions of the areas where these patients reside*?" and a subset of services shown in Table 1.

TABLE I
A subset of DaaS services

DaaS services	Semantics services Description
$S_{1.1}$ (\$*x*, ?*s*) $S_{1.2}$ (\$*x*, ?*s*)	Return patients *s* =“*SSN*”, infected with a disease *x*
$S_{2.1}$ (\$*s*, ?*d*, ?*g*) $S_{2.2}$ (\$*s*, ?*d*, ?*g*)	Return *d* =“*DoB*”, and *g* =“*gender*” of patient identified by *s* =“*SSN”*
$S_{3.1}$ (\$*s*, ?*z*, ?*p*)	Returns *z* =“*zip*”, and *p* =“*salary*” of patient identified by *s* =“*SSN”*
$S_{4.1}$ (\$*s*, ?*n*) $S_{4.2}$ (\$*s*, ?*n*)	Return *n* =“*DNA*” of patient identified by *s* =“*SSN”*
$S_{5.1}$ (\$*z*, ?*w*)	Returns *w* = “*Weather-condition*” of address *z* =“*zip*”

We have proposed in [5] a mediator-based approach to compose services (based on a query-rewriting algorithm) and answer this kind of queries. In this approach, the mediator selects, combines and orchestrates (i.e., gets output data from a service and uses it as input data to call another service) services to answer queries. It also carries out all the interactions between composed services (i.e., relays exchanged

• Salah-Eddine Tbahriti, and Michael Mrissa are members of the LIRIS lab., UMR5205 CNRS, Université de Lyon, 69622, Villeurbanne, FRANCE
E-mail: {salah-eddine.tbahriti, michael.mrissa}@liris.cnrs.fr
• Chirine Ghedira is member of MAGELLAN lab., IAE-Université Jean-Moulin Lyon 3, 69355 Lyon cedex 08, FRANCE.
E-mail: chirine.ghedira-guegan@univ-lyon3.fr
• Brahim Medjahed is member with the Department of Computer and Information Science, University of Michigan-Dearborn, 4901 Evergreen Road, Dearborn 48128 USA
E-mail: Brahim@umd.umich.edu

.

data among interconnected services in the composition). The result of the composition process is a *composition plan*, *CP* (depicted in Figure 1), which consists of a set of services that must be executed in a particular order depending on their access patterns (i.e., the connections between their input and output parameters). Input parameters are identified with a first “\$” character and output parameters with a “?”. Hence, service $S(\$a, ?b)$ requires an input value a and provides an output value b. Then, Q can be answered as follows: First, $S_{1.1}$ is invoked with *H1N1* as input value, then for each obtained *SSN*, $S_{4.1}$, $S_{2.2}$ and $S_{3.1}$ are invoked to obtain their *DNA*, *DoB (i.e. date-of-birth)*, *zip* and *salary*. Finally, $S_{5.1}$ is invoked with the patients’ *zip* to get information about the *weather-conditions* (note that other solution *CP* can be found with the services of Table I).

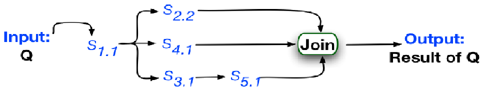

Fig. 1. Composition plan of Q.

In fact, services in *CP* may have conflicting privacy concerns regarding to their exchanged data. Some services may require some input data that other participating services cannot disclose because of their privacy specifications. For instance, let us assume that $S_{1.1}$ discloses its data (i.e., *SSN*) to a third-party service for use with a “*limited time*” restriction. $S_{3.1}$ meanwhile attests that it keeps collected data (i.e., *SSN*) for an “*unlimited time*”. $S_{1.1}$ and $S_{3.1}$ are incompatible in terms of privacy with respect to *SSN*. $S_{1.1}$ (which provides *SSN)* judges that a long retention of *SSN* by a third-party is a risk for privacy, while $S_{3.1}$ would use that data as long as possible to perform several tasks that are not considered as a privacy risk. Such a conflict invalidates the *CP* of Figure 1 in terms of privacy. Then, it becomes important on the one hand to extend service descriptions with privacy specifications, and on the other hand to insure the privacy compatibility of services selected for a composition.

B. Summary of Contributions

The previous scenario calls for a solution that must be expressive enough to capture the different needs for privacy concerns of services as well as simple and coherent with our previous service composition algorithm [5]. Since composing services is already a complex task, any target solution should involve minimal processing costs. Existing approaches based on secure multi-party computation [39] are usually characterized by their high computation time and complexity, which makes them impractical for database operations working over a large number of elements [34]. Data privacy through access control is among the classical goals of data management with countless proposals, e.g., [35]. However, our system is designed to be open, which implies that the mediator may not have preliminary knowledge on the requester. Such a circumstance makes traditional access control mechanisms less efficient, as they are mainly based on preliminary authentication of the requester, and then on validation of applicable authorizations. In this paper, we focus on the privacy issue from the point of view of data usage and expectation during the design phase of DaaS composition. We build our contribution around:

Formal Model for Privacy Specification: to capture and reason about privacy concerns from a service perspective. Our proposed model allows each service *S* to define *Privacy Policies PP* (specifying how *S* manages collected data) and *Privacy Requirements PR* (specifying how *S* expects consumers to manage the data it provides). Our privacy model is defined with both expressiveness and simplicity in mind.

Privacy Compatibility-aware Composition: detecting incompatibilities between the *PR* and *PP* of services involved in a composition is a core concept of our approach. Our matching algorithm is based on the notion of privacy subsumption and on a cost model. Then, we extend our service composition approach to take into account the privacy specifications and compatibility of services.

Privacy-aware Adaptation: our third contribution is devoted to resolve detected incompatibilities by allowing services to define adaptation sets in order to obtain valid composition plans and enhance the efficiency of our system of composition. We introduce an adaptation protocol to automatically reconcile the adaptation sets in order to make the *PR* and *PP* of conflicting services compatible. The adaptation of *PR* and *PP* of service is decided by; service reputations, individual consent and a cost function. We also devise protocols to speed up the adaptation process.

C. Paper Organization

Our paper is structured as follows. We overview the basic definitions for modeling and composing DaaS services in Section 2. Then, we describe our privacy model in Section 3. We show how our DaaS composition approach is extended within privacy compatibility in Section 4. We introduce our adaptation approach in Section 5 and detail how privacy compatibility in the composition is reached with the adaptation protocols. We present our experiments in Section 6 and discuss related work in Section 7. We discuss obtained results and future work in Section 8.

II. Background: The PAIRSE Project

The approach presented in this paper is implemented as a part of the PAIRSE project[1], which deals with the privacy issues in P2P data sharing environments in the area of epidemiological research.

[1] This research project is supported by the French National Research Agency under grant number ANR-09-SEGI-008. URL https://picoforge.int-evry.fr/cgi-bin/twiki/view/Pairse/Web/