CERIAS Tech Report 2007-14

ENGINEERING A POLICY-BASED SYSTEM FOR FEDERATED HEALTHCARE DATABASES

by Rafae Bhatti, Arjmand Samuel, Mohamed Y. Eltabakh, Haseeb Amjad, Arif Ghafoor

Center for Education and Research in
Information Assurance and Security,
Purdue University, West Lafayette, IN 47907-2086

Engineering a Policy-Based System for Federated Healthcare Databases

Rafae Bhatti, Arjmand Samuel, Mohamed Y. Eltabakh, Haseeb Amjad, Arif Ghafoor

Abstract: Policy-based management for federated healthcare systems have recently gained increasing attention due to strict privacy and disclosure rules. While the work on privacy languages and enforcement mechanisms, such as Hippocratic databases, has advanced our understanding of designing privacy-preserving policies for healthcare databases, the need to integrate these policies in practical healthcare framework is becoming acute. Additionally, while most work in this area has been organization-oriented, dealing with exchange of information between healthcare organizations (such as referrals), the requirements for the emerging area of personal healthcare information management have so far not been adequately addressed. These shortcomings arise from the lack of a sophisticated policy specification language and enforcement architecture that can capture the requirement for (i) integration of privacy and disclosure policies with well-known healthcare standards used in the industry in order to specify the precise requirements of a practical healthcare system, and (ii) provision of ubiquitous healthcare services to patients using the same infrastructure that enables federated healthcare management for organizations. In this paper, we have designed a policy-based system to mitigate these concerns. One, we have designed our disclosure and privacy policies using a requirements specification based on a set of use cases for the Clinical Document Architecture (CDA) standard proposed by the community. Two, we present a context-aware policy specification language which allows encoding of CDA-based requirements use-cases into privacy and disclosure policy rules. We have shown that our policy specification language is effective in terms of handling a variety of expressive constraints on CDA-encoded document contents. Our language enables specification of privacy-aware access control for federated healthcare information across organizational boundaries, while the use of contextual constraints allows the incorporation of user and environment context in the access control mechanism for personal healthcare information management. Moreover, the declarative syntax of the policy rules makes the policy adaptable to changes in privacy regulations or patient preferences. We also present an enforcement architecture for the federated healthcare framework proposed in this paper.

Index terms: Federated database security, healthcare engineering, policy-based management, role based access control.

1. Introduction

Privacy and security of sensitive healthcare data is becoming a growing concern in the healthcare industry. With the recent legislations mandating creation of electronic healthcare records (EHR) as part of a nationwide initiative for providing electronic healthcare services, it becomes imperative to ensure that the availability of medical data in electronic form adheres to the same levels of privacy and disclosure regulations as applicable to present day paper-based records accessible only from the physician's office. This introduces several challenges above and beyond data integration, and involves development of mechanisms to specify data disclosure and privacy policies that can be enforced across a network of inter-connected healthcare service providers. This network of healthcare providers gives rise to a distributed clinical database which resembles a federated system, and the problem we address in this paper is the design of a policy-based system for federated healthcare databases.

Research community has been engaged in addressing certain aspects of this problem. One notable technique called *hippocratic databases* (HDB)[3] allows applications to enforce disclosure policies on databases on arbitrary data elements in an automated fashion. HDB includes mechanisms for active enforcement (data retrieval according to disclosure policy of the institution and privacy preferences of the user), compliance auditing, (ensuring compliance with data disclosure regulations, company policies, and customer preferences), and Sovereign

Information Integration (allowing two or more autonomous entities to compute queries across their databases while only revealing the results of the queries), among others. Each of these dimensions encompasses a variety of issues that need to be considered at different levels. Our focus in this paper is the issues related to active enforcement, specifically, specification and enforcement of disclosure and privacy policies in a federated healthcare database system.

The various privacy and security approaches for database systems (including active enforcement in HDB) have relied on well-known privacy practices and languages (such as P3P [31] and EPAL [16]). While the work on privacy practices, languages and enforcement mechanisms has advanced our understanding of designing privacy-aware policies, the need to integrate these policies in a practical healthcare framework is becoming acute. No prior work has been reported on the specification of privacy and disclosure policies in relation to the clinical document storage and retrieval standards in place today. Additionally, we view the healthcare initiative involving EHR to have impact beyond the hospital boundaries within a federated network, and actually serve as an enabling technology for advancing ubiquitous and personal healthcare information management [22]. While the EHR initiative as currently envisioned is organization-oriented, i.e. it is designed to manage the patient records between hospitals (such as referrals), personal healthcare information management would add a patient-oriented dimension to it, i.e. patients equipped with mobile devices (such as PDAs, cell phones) will be able to obtain services from anywhere based on their EHRs located anywhere in the federated healthcare system. It may allow patients to obtain prescriptions without visiting a doctor's office, or get remote diagnosis from their family physician currently out of town without having to visit another physician. It may also allow requests from a physician in vicinity to access a patient's records in a remote emergency situation.

Present privacy-aware mechanisms have adequately captured the requirement for organization-oriented healthcare information management, but they do not provide adequate support for personal healthcare information management (See Related Work). Satisfying this latter requirement requires the specification and enforcement of expressive, context-aware disclosure and privacy policies. The very fact that the clinical data is now going to be available outside of its traditional boundaries raises several issues regarding the establishment of the legitimate "context" used to enforce the disclosure and privacy policies. The context has a two-fold interpretation: the context of the user, and the environmental context, and both kinds of contexts have relevance for policy design. For instance, the federated healthcare system should be able to provide an affiliated physician (*user context*) access to a patient's information from a certain location (*environment context*), and the policies related to disclosure and

privacy rules should have a context-specific restriction to incorporate this requirement in the security mechanism. In such a scenario, the user context for the physician would be used to satisfy the privacy requirements of the patient, whereas the environmental context would have implications in the disclosure rules controlling the release of sensitive information. The two contexts may even be used together in a policy: a patient may routinely permit only a particular physician to access his/her records, but if the context indicates emergency, then the patient allows any physician handling the emergency to access the records.

1.1 Contributions and Organization

Based on the preceding discussion, our contributions in this paper are two-fold. One, we highlight a missing connection in prior work on policies related to disclosure and privacy rules, vis-à-vis the integration of the policies with well-known EHR standards used in the industry. In the absence of such a connection, it is difficult to specify the precise requirements of a practical healthcare system. In this paper, we have used Clinical Document Architecture (CDA) [11] as a representative EHR standard and designed our policies based on a set of use cases for the CDA standard proposed by the community [6, 19]. We then show that our policy specification language is effective in terms of handling a variety of expressive constraints on the CDA-encoded EHRs.

Our second contribution is the use of an expressive, context-aware specification language which not only allows encoding disclosure and privacy rules using a declarative predicate-based syntax in the policy, but also allows the use of temporal and non-temporal contextual constraints to be specified in the policy. The specification of disclosure and privacy rules allows privacy-aware access control for federated EHRs across organizational boundaries, while the use of contextual constraints allows the incorporation of user and environment context in the access control mechanism for personal healthcare information management. Moreover, the declarative syntax of the policy rules makes the policy highly reconfigurable and adaptable to changes in privacy regulations or patient preferences. The design of our policy specification language builds upon the well-known Role Based Access Control (RBAC) model [25], and augments it with necessary extensions to support access management in a federated system. Our policy is supported by a well-defined administration model [9] and management framework [7] for scalable maintenance. We also present an enforcement architecture for the federated healthcare framework proposed in this paper.

The remainder of the paper is organized as follows. Section 2 provides a more detailed overview of related work and highlights the particular merits of our approach with respect to the outlined challenges. Section 3 discusses the design considerations for implementing the federated healthcare system proposed in this paper. In this regard, we

provide an overview of the federated database architecture, the CDA standard used for representing EHRs in our framework, and the access control policy rules for EHRs. Section 4 presents the details of the policy specification language, and discusses its salient features for access management in a federated system. Section 5 discusses the use of the policy specification toward providing support for federated healthcare access management. It describes the effectiveness of our language specification based on the requirement specification extracted from the CDA standard as a set of use cases proposed by the community. Some example use cases are also described in this section. Section 6 presents the enforcement architecture of our framework. Section 7 discusses the implementation of the example policy designed in the paper on our current prototype. Section 8 concludes the paper.

2. Related Work

Our framework is aimed at designing a policy-based database system that integrates the knowledge of healthcare requirement specification and context-aware policy specification language to provide comprehensive support for federated healthcare information management. There are some aspects of this approach that have been reported in prior work, but none addresses all these aspects of federated healthcare information management that we have outlined above.

Policy-based approaches have recently been proposed for access management in information systems [7, 8], and have also been applied to relational databases through notions like Hippocratic database (HDB) [2-5]. The focus of many of the policy-based approaches for database [2-5, 10, 18] has been on privacy-aware access control. HDB includes, among others, mechanism for active enforcement, which deals with data retrieval according to disclosure policy of the institution and privacy preferences of the user expressed in a well-known privacy language (such as P3P [31] or EPAL [16]). While HDB focuses on privacy preserving in relational database management systems, the work described in [18] extends the data sources to include LDAP and file system directories. The work in [10] presents a purpose-based access control model to represent the privacy policies in terms of the purpose for which the data is requested. They also give a method to determine the access purpose of a user, but they do not have a high level language that can be used to express purpose-based privacy preferences.

While HDB is the most prominent of these works, and it is designed to meet some of the requirements we have highlighted in this paper, it still lacks some features such as (i) providing a flexible and expressive specification language with well-defined administration model for scalable maintenance, and (ii) integrating these rules and policies in the current practical healthcare frameworks, e.g., the clinical document storage and retrieval standards in place today. With respect to (i), HDB does not include a policy specification language by itself but only allows

matching privacy preferences of organizations and user expressed in an existing language. The resulting policy is stored in a relational table and all queries that access user data are rewritten according to the stored policy by the active enforcement engine. Therefore, the expressiveness of the policy specification is inherently tied to the existing languages used by HDB, and there is no opportunity for scalable maintenance since no policy administration model or management framework exists to support the policy. As concerns (ii), no work has been reported that can demonstrate how the policies in HDB can be integrated with current healthcare standards. The main reason is that those policies (P3P, EPAL) are designed primarily for the online e-commerce space, and no dedicated provisions exist in them to support specific healthcare use cases, such as those designed for the CDA standard which we consider in this work.

Along another dimension, earlier efforts have been reported on policy-based access control in traditional database systems [23, 27]. Some recent work has focused on healthcare information management [26, 30], where the emphasis is to design policy-based trust frameworks for distributed healthcare applications. Standards such as Extensible Access Control Markup Language (XACML) have been used for expressing context-aware access control policies for Web-based documents [21]. The work in [15] presents a context-aware access control system where the idea is to incorporate user and environment context in the access control, much like the above cited requirement for context-sensitive protection of clinical records for personal healthcare information management. These approaches are one step forward in meeting our stated requirements, but are directed at traditional (as opposed to federated) database systems. Additionally, they do not capture the advanced context-based requirements for personal healthcare information management because they are based on very simple context abstractions (such as saying "physician d1 can access *all* records of patient p1 when *logged in from* ABC hospital network *on* Wednesday *between* 9Am to 5PM), but the federated healthcare system demands more fine-grained and expressive contextual constraints (such as saying "physician d1 can access record *of type X* of patient p1 when *accessing from or near* ABC hospital *on first* Wednesday of *first* month of *every* quarter *between* 9AM to 5PM). The distinction here is that the former authorization is not fine-grained enough to restrict the physician's access to a particular subset of records (based on specialization of the physician and/or the privacy policy of the patient), and is not expressive enough to encode the "*accessing from or near*" location condition and "*on first* day of *first* month of *every* quarter" time condition. The location conditions are more relevant in the personal healthcare scenario where either the disclosure policy of the hospital or the privacy policy of the patient may be location-sensitive. The time-based conditions become more relevant in organizational data sharing, and let hospitals allow external agencies such as insurance

staff or auditors to access healthcare data under stricter, schedule-driven temporal controls to mitigate risk of unnecessary exposure of sensitive information, and at the same time achieve regulatory compliance.

This discussion indicates the limitations of existing approaches with regards to specification and enforcement of policies related to disclosure and privacy rules for (i) organizational and (ii) personal healthcare information management. (i) is focused more on relatively stable rules, such as legislation and basic user preferences, whereas (ii) concerns itself more with provision of services against dynamic context-aware policy rules, such as fine-grained and expressive constraints described above. Additionally, the need for integrating these policies in a practical healthcare framework and analyzing the coverage of the policy in terms of real world requirements also remains unaddressed. To the best of our knowledge, ours is the first work that attempts to address these issues in federated healthcare information management.

3. Design Considerations

3.1. Federated Database Architecture

The federated healthcare database system proposed in this paper is based on the well-established Federated Database System (FDBS) architecture described in literature [12-14, 24, 27, 29]. To summarize, such an architecture consists of a multi-level schema, consisting of a local schema, a component schema, an export schema, and the overall federated schema. This schema architecture allows the resulting database system to support distribution, heterogeneity and autonomy, which are the three cardinal requirements of a federated database system. A local schema is the conceptual schema of a component database system expressed in the native data model of the component DBMS, and hence different local schemas may be expressed in different data models. A component schema is derived by translating local schemas into a data model called the canonical or common data model (CDM) of the FDBS. Two reasons for defining component schemas in a CDM are (i) they describe the divergent local schemas using a single representation and (ii) semantics that are missing in a local schema can be added to its component schema. Thus they facilitate negotiation and integration tasks performed when developing a tightly coupled FDBS. Similarly, they facilitate negotiation and specification of views and multi-database queries in a loosely coupled FDBS. The process of schema translation from a local schema to a component schema generates the mappings between component schema objects and local schema objects. An export schema represents a subset of a component schema that is available to the FDBS. The purpose of defining export schemas is to facilitate control and management of association autonomy. A federated schema is an integration of multiple export schemas. A federated schema also includes the information on data distribution that is generated when integrating export schemas.

This architecture forms the baseline for the design and operation of an FDBS. In this paper, our design of a federated healthcare database system assumes that the distributed databases form a tightly coupled federation of multiple databases. Of particular relevance to us in this architecture are the concepts of CDM, and the federated schema. We will present in Section 4 the policy definitions that act as the CDM throughout the healthcare federation, and will be used by each participating site to encode their export schemas. The integration of these schemas will then constitute the federated schema. The policy framework that we have designed can be adapted to work with any implementations of FDBS that abides by the FDBS architecture. Notable among the implementations reported in the literature are the Mermaid by Templeton et al. [27], IRO-DB [13], Disco [29], and MIRO-Web [12].

3.2. EHR Format

For representing the EHR in our framework, we choose to adopt the Clinical Document Architecture (CDA) [11] standard, which is an ANSI-certified standard from the Health Level Seven (HL7) organization. HL7 is an international community of healthcare subject matter experts and information scientists collaborating to create standards for the exchange, management and integration of electronic healthcare information. It is one of several American National Standards Institute (ANSI) -accredited Standards Developing Organizations (SDOs) operating in the healthcare arena. It explicitly deals with the clinical and administrative domain. HL7 standards are used in 90% of US healthcare facilities. The organization has affiliates in 20 countries, and in some of those, HL7 messaging has a legal status.

CDA defines an XML architecture for exchange of clinical documents. CDA specifies the syntax and supplies a framework for specifying the full semantics of a clinical document. A CDA can contain any type of clinical content. Typical types of CDA documents would be a Discharge Summary, Imaging Report, Admission & Physical, Pathology Report and so on. CDA uses XML, although it allows for a non-XML body (pdf, Word, jpg and so on) for simple implementations. CDA documents can be displayed using XML-aware Web browsers or wireless applications such as cell phones. These features of CDA and its wide-spread use in the healthcare industry[1] makes our choice of using CDA as a reference standard for EHR is a pragmatic one which facilitates our outlined goals. It may be stated that many EHR vendors already have the capability to produce CDA compliant documents [11].

CDA works by using the concept of "incremental" semantic interoperability to allow multiple organizations exchange EHRs through a more or less automated process. What this means is that there is a range of complexity

[1] As an indication of its usage, the flagship implementation of CDA in the country is the Mayo Clinic, which is producing thousands of CDAs every week and is expected to reach 50,000 notes/week in full production [11].

allowed within the specification and users must set their own level of compliance. The minimal CDA is a small number of XML-encoded metadata fields (such as provider name, document type, document identifier, and so on) and a body which can be any commonly-used MIME type such as pdf or .doc (Microsoft Word) or even a scanned image file. All documents encoded using even the minimal compliance level would be equally readable at the place of healthcare provision.