Patient-centric Authorization Framework for Sharing Electronic Health Records

Jing Jin
UNC Charlotte
Charlotte, NC, USA
jjin@uncc.edu

Gail-Joon Ahn
Arizona State University
Tempe, AZ, USA
gahn@asu.edu

Hongxin Hu
Arizona State University
Tempe, AZ, USA
hxhu@asu.edu

Michael J. Covington
Intel Corporation
Hillsboro, Oregon, USA
michael.j.covington@intel.com

Xinwen Zhang
Samsung Information Systems
America, San Jose, CA, USA
xinwen.z@samsung.com

ABSTRACT

In modern healthcare environments, a fundamental requirement for achieving continuity of care is the seamless access to distributed patient health records in an integrated and unified manner, directly at the point of care. However, Electronic Health Records (EHRs) contain a significant amount of sensitive information, and allowing data to be accessible at many different sources increases concerns related to patient privacy and data theft. Access control solutions must guarantee that only authorized users have access to such critical records for legitimate purposes, and access control policies from distributed EHR sources must be accurately reflected and enforced accordingly in the integrated EHRs.

In this paper, we propose a unified access control scheme that supports patient-centric selective sharing of virtual composite EHRs using different levels of granularity, accommodating data aggregation and various privacy protection requirements. We also articulate and handle the policy anomalies that might occur in the composition of discrete access control policies from multiple data sources.

Categories and Subject Descriptors

D.4.6 [**Operating Systems**]: Security and Protection—Access controls; K.6.5 [**Management of Computing and Information Systems**]: Security and Protection—Unauthorized access

General Terms

Security

Keywords

Electronic Health Records (EHRs), Patient-centric Authorization, Selective Sharing

SACMAT'09, June 3–5, 2009, Stresa, Italy.
Copyright 2009 ACM 978-1-60558-537-6/09/06 ...$5.00.

1. INTRODUCTION

In much of the developed world, healthcare has evolved to a point where patients can have many different providers – including primary care physicians, specialists, therapists, and even alternative medicine practitioners – to service their diverse medical needs. It is not uncommon for patients to visit providers who are physically separated from one another; some are located across town, while others are across the country or on another continent. As a result, medical records can be found scattered throughout the entire healthcare sector, from primary care physician's offices and clinical laboratories to pharmacies and specialist research centers. From the clinical perspective, delivering proper patient care requires access to integrated and unified patient information that is often collected in real-time to ensure the freshness of time-sensitive data. Yet the data dispersion in current healthcare settings typically results in painstaking, time-consuming efforts to obtain a patient's complete medical history, or unnecessary duplication of tests and other investigations. There is a strong need to create an infrastructure that uniformly integrates this heterogeneous collection of medical data and delivers it to the healthcare professionals who need it at the point of care [13, 16]. The adoption of standardized Electronic Health Records (EHRs) has become an extremely important prerequisite for bringing interoperability and effective data retrieval integration to the healthcare industry [16].

Effective management of EHRs is a very complex and sensitive issue. Patient privacy concerns, along with threats that could expose medical information, highlight the need for security and privacy technologies that are well-integrated into the healthcare system and enforceable across a variety of heterogeneous systems and networks.

A shared EHR involves a complex composition of sensitive information, including patient demographic details, medical histories, examination reports, laboratory test results, radiology images (X-rays, CTs), and so on. There is a strong need for protection models that comply with legal and regulatory policies, while simultaneously ensuring that access to sensitive information is limited only to those entities who have a legitimate need-to-know and are authorized by the patient. For instance, a patient's medical information pertaining to an HIV/AIDS diagnosis may be explicitly hidden from general medical information sharing unless a specific treatment option is indicated. To support this, the patient should ultimately own his or her medical records and be responsible for maintaining access rights for the distributed EHRs [12, 6]. It is, therefore, essential that patients be provided with a secure, usable, and straightforward mechanism that allows them to quickly and easily authorize a variety of medical affiliates to access their sensitive records or a subset of the data within them. In addition, as a patient's medical records are distributed at different sites and virtually aggregated at the point of care, such an access control mechanism must be uniformly applied not only to the EHR records residing at each local site, but also to the aggregated EHR that is generated and shared on-the-fly. In this paper, we refer to the shared EHR with complex data and policy composition features as the *virtual composite EHR* and the data sources that contribute to the virtual composite EHR as *EHR instances*.

We recently proposed an access control model for selectively sharing EHRs [18]. A key characteristic of the model is that we formulate the semantics and structural composition of EHR documents in a hierarchical structure, where internal sub-objects are distinguished and associated with properties to address important criteria for medical data sharing such as data types, intended purposes and information sensitivities. Such hierarchical structure is further explored with an *authorization zone* filtration mechanism that provides a flexible and efficient means to select and authorize a portion of an EHR document to be shared with specific property criteria.

Our previous work, however, has some inherent limitations. The EHR structure in our previous work only addresses a localized EHR instance without considering the data aggregation and policy composition issues of the virtual composite EHR with multiple EHR instances from different sources. We thus need a more sophisticated information model, as well as a unified policy scheme for uniformly regulating selective sharing of both discrete EHR instances and the aggregated virtual composite EHRs at different levels of granularity. Such a refined access control model is the first contribution of this paper. In addition, the dynamic aggregation of distributed EHR instances requires the seamless integration of access control policies from multiple data sources. Our second contribution is to articulate and propose mechanisms that identify and resolve potential policy anomalies for composed access control policies at the virtual composite EHR level. Finally, our previous work lacks appropriate implementation and evaluation in practical health information sharing systems. In this paper, a virtual composite EHR sharing system is designed and implemented for integrated and federated healthcare networks, where a patient consent mechanism is incorporated to demonstrate our approach to controlling and providing only authorized "views" of patient medical information to requesters.

The rest of the paper is organized as follows. In Section 2, we provide a brief overview of the emerging EHR standards. We also review existing security solutions for EHR systems and development of e-Consent systems. In Section 3, we present our unified patient-centric authorization model and discuss the policy anomalies in policy composition. Our prototype EHR sharing system is described in Section 4. Section 5 concludes this paper with future research directions.

2. RELATED WORK

[EHR Standards]: There are several standards currently under development to structure and specify the clinical content of an EHR for the purpose of exchange, such as openEHR and HL7 Clinical Document Architecture (CDA) [20, 10]. openEHR uses a two-level methodology to model the EHR structure. In the first level, a *generic reference model* is designed to express the generic data content needed in clinical contexts and provide an explicit representation of the semantic and vocabulary that should exist in all EHR instances. In the second level, the notion of *archetype* is introduced to model specific healthcare concepts such as blood pressure and lab results. These archetypes are the fundamental building blocks to form the contents in various clinical EHR instances. Similarly, HL7 V3 standards define an underlying Reference Information Model (RIM) [14] that forms the generic information domain used across all HL7 messages, while CDA defines detailed structure and semantics of medical documents in terms of a set of coded components (called vocabulary) to model basic medical concepts.

By implementing or converting to the EHR standards, a "common language" is established between different medical information systems to communicate and share standardized medical information with each other. Therefore, authorization and selective sharing of medical information should be carried out with common understanding of EHR standards. And the two-level information modelling paradigm is also adopted in our approach to uniformly model discrete EHR instances and the aggregated virtual composite EHR.

[Access Control for EHR Systems and e-Consent]: A number of solutions have been proposed to address security and access control concerns associated with EHR systems [11, 3, 4]. All of these approaches, to some extent, utilize role-based access control (RBAC) to address organizational security management requirement and authorize access to various healthcare parties. However, selective sharing of composite EHRs requires clear understanding of the internal clinical information under protection and their structural relationships. None of these approaches took into account of structural and semantics composition of EHRs, and thus cannot support a more fine-grained access control to share composite EHRs as a whole or only partially. In this paper, we focus on the "selective" feature of an EHR system where a logical structure of a composite EHR is captured with its internal data elements being clearly distinguished and organized, so that our access control policies can be specified to select and authorize any portion of an EHR for data sharing.

Achieving privacy preservation in medical information sharing is a critical concern for an EHR system. Several purpose-based access control models have been proposed recently to protect sensitive data [5, 24]. These models associate the intended purpose information with a given data element, and access is granted when the access purpose is consistent with the data element's intended purpose. However, as healthcare is such a complex domain involving various parties with different duties and objectives, the purpose-based access control alone cannot meet all the patient's privacy protection requirements. In this paper, we incorporate more deciding factors beyond purpose to control the selective sharing of EHRs in a more flexible way.

To enable the patient control of medical information sharing, "e-Consent" mechanisms have been proposed to allow patients to issue or withhold authorization policies as electronic consents to those who wish to access their electronic health information [7, 23, 19, 22]. Several consent models with associated consent templates have been identified [7, 23], and a few e-Consent based systems have been built upon these guidelines [22, 19]. However, it is still essential to develop a systematic approach to determining how a patient's consent is expressed and at what granularity the consent is applied to the EHRs. Meanwhile, with dispersed EHR instances across many caregivers, it is also required for a patient to manage his consents in a unified and consistent manner in an online shared EHR environment.

3. PATIENT-CENTRIC AUTHORIZATION MODEL

3.1 Unified Logical EHR Model

A patient's EHR instances are typically dispersed over a wide range of distributed clinical systems and data structures. The only way to maintain a unified medical record without the need to adapt these different environments is to define a *Unified Data Schema (UDS)* for EHR instances to follow [8]. Similar to the generic reference models in openEHR and HL7, *UDS* defines generic semantics and logical relationships between patient information elements drawn from medical domains such as patient demographics, labs, medications, encounters, imaging and pathology reports, and a variety of other medical domains from primary, specialty and acute care settings. Based on these predefined categories, EHR instances are aggregated and integrated into a unified patient record as a virtual composite EHR. Since data integration is not the focus of this paper, we do not consider heterogeneity in schema integration and assume all EHR instances and the corresponding aggregated virtual composite EHR uniformly conform (or are converted to conform) to a predefined *UDS*. In Figure 1, a virtual composite EHR aggregates two EHR instances from hospitals *h1* and *h2* based on a simple *UDS* defining three categories of *Demographics*, *History* and *Labs*.

In our model, both the EHR instances and the aggregated virtual composite EHR are uniformly modelled as a labelled hierarchical structure. The nodes represent the clinical data elements that need to be protected for sharing. Their relations are captured as the association links between the nodes within the hierarchy. Each node is associated with specific properties to address essential features regarding the sources of data and their sensitivity levels. The properties can be categorized into three dimensions: *origin*, *sensitivity*, and *object type*. The *origin* property is specified to indicate the source(s) of data within the composition. As the same patient information may be duplicated in multiple EHR instances, such data elements should be merged as one element within a virtual composite EHR, and we use multiple origins to indicate such data merging, while a node with a single origin indicates that the data is unique from the respective origin. Using the category of `Illness` history in Figure 1 as an example, the `asthma` information comes from both *h1* and *h2*, while `HIV` information is uniquely from *h2* only. The *sensitivity* property is designed to label a node based on the sensitivity of the content contained in it, which eventually can be used to prevent the patient's sensitive medical information from being disclosed unintentionally. In the practice of Iowa HISPC [17], the sensitivity classifications of medical data include general medical data, drug and alcohol treatment, substance abuse treatment, mental health, communicable disease (HIV, STDs, etc.), decedent, immunizations, and so on. Based on these classifications, the data elements representing the patient's `HIV` history and `CD4` lab test should be marked with a property of "`communicable disease`" ("`HIV`" for simplicity). The *object type* property gives another dimension on data node selection and protection. The nodes can be primitive types such as plain texts, dates and images. They can also be a composite type in the hierarchical structure including other types of data nodes. Formally, an EHR can be uniformly modelled and defined as follows:

Definition 1. ***(Logical EHR Model)****. An EHR is a tuple $C = (v_c, V_o, E_o, \tau_{V_o})$, where*

- *v_c is the root representing the whole EHR object;*
- *V_o is a set of nodes within the composite structure;*
- *$E_o \subseteq V_o \times V_o$ is a set of links between nodes; and*
- *$\tau_{V_o} : V_o \rightarrow P$ is a node labelling function to specify the property of a node. P is a set of properties defined in Definition 2.*

Definition 2. ***(Property)****. Let O, S, and T be the sets of data origins, sensitivity classifications, and object types, respectively. And let $n = |V_o|$ be the number of nodes in an EHR composition C.*

- *$P_o = \{po_1, \ldots, po_n\}$ is a collection of origin sets, where $po_i \subseteq O$ is a set of origins associated with a node, $i \in [1, n]$;*
- *$P_s = \{ps_1, \ldots, ps_n\}$ is a collection of sensitivity classification sets, where $ps_i \subseteq S$ is a set of sensitivity classifications associated with a node, $i \in [1, n]$; and*
- *$P = P_o \times P_s \times T$ is a set of three dimensional properties of origin, sensitivity, and data type.*

Given a node $v_i \in V_o$ inside an EHR composition C, the function $\tau(v_i) = p$ retrieves the property label p for the node. And we use the dot notation to refer to a specific property dimension. For instance, *p.po* refers to the data origin property; *p.ps* refers to the sensitivity property; and *p.t* refers to the object type. Within a logical EHR structure, nodes can be explicitly denoted by their identifiers, or can be implicitly addressed by means of *Path Expressions*. We apply an XPath-like expression for the path representation. Table 1 describes the notions and examples we use to select nodes inside a virtual composite EHR illustrated in Figure 1(c).

3.2 Policy Specification

To enable an authorized and selective sharing of patients' EHRs, it is essential for an authorization policy to be in place to determine a subject's access privileges to specific portion(s) of an EHR instance or a virtual composite EHR. Our policy specification scheme is built upon the identified logical EHR model so that access policies can be effectively defined at different granularity levels within the structure. In this paper, we assume the data sharing happens during

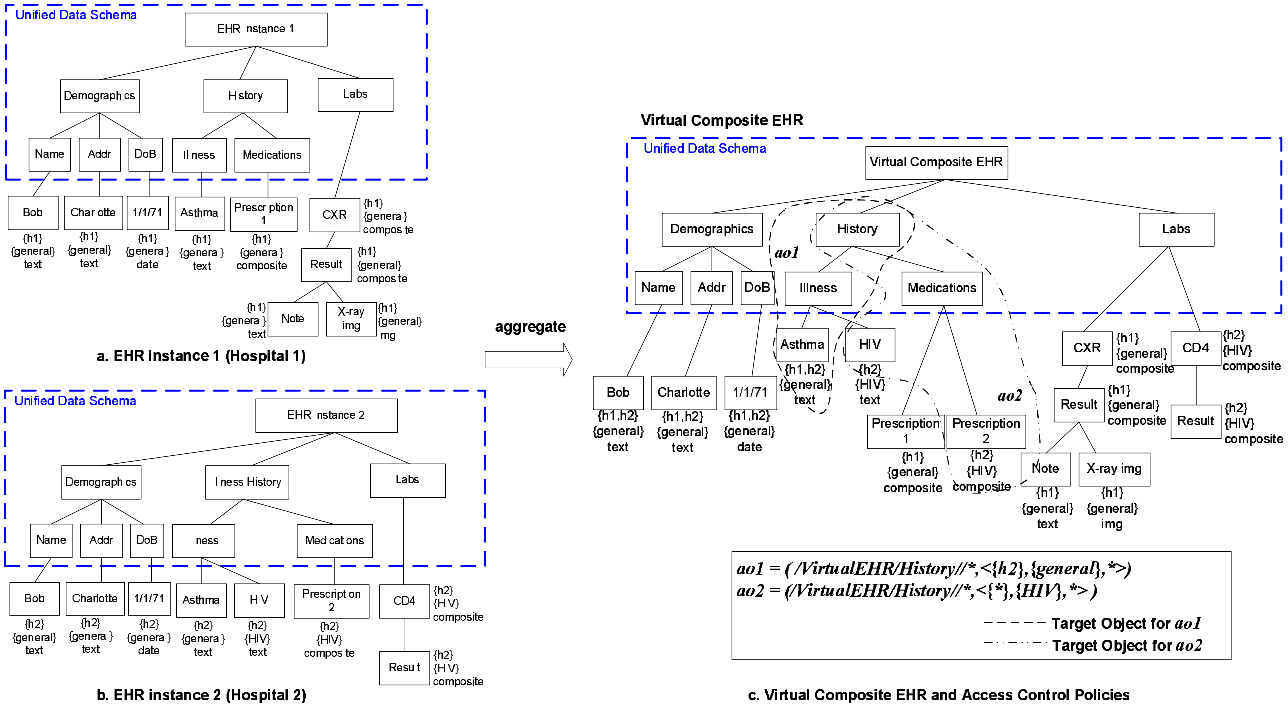


Figure 1: Virtual Composite EHR in a Hierarchical Structure

Table 1: Path Expression for Node Selection

Expression	Description	Example
nodename	Select the named nodes	*CXR*
/	Select the node through absolute path from root node	/EHR/Labs/CXR
//	Select the node through relative path	//Labs/CXR
*	Select all immediate children nodes	//Labs/CXR/*
//*	Select all descendant nodes	//Labs/CXR//*

a patient encounter when EHR instances are virtually aggregated at the point of care for a practitioner to review, therefore we mainly focus on read-only access permission.

Subjects: In the context of healthcare, an authorization may be assigned on the basis of an individual provider or providers acting in specific clinical roles within certain organizations. Some authorizations may be given by patients in relation to identified individuals. For instance, a patient may want to indicate the following intentions: "I allow Dr. Smith to access my medical data." In other circumstances, the authorization is for a role such as "general physician," "cardiologist," "nurse," and so on. As healthcare practitioners are always associated with certain organizations, such a unique property may also be articulated within the specification to further constrain the subject. Formally, the subject specification is defined as follows:

Definition 3. *(**Subject Specification**). Let E, R and O be sets of user IDs, roles, and origins, respectively. A subject sub is defined as a tuple sub=<e,so> or sub=<r,so>, where $e \in E$, $r \in R$, and optional subject origin set $so \subseteq O$. Overall, the subject set Sub is defined as $Sub = (E \times 2^O) \cup (R \times 2^O)$.*

Objects and Filtration Property: The fine-grained authorization specification should support a flexible selection of protection objects. In our hierarchical EHR model, XPath-like path expressions are utilized to specify the *scope* of data elements to which an authorization policy applies. Meanwhile, the *filtration properties* are defined to be compared with the property label of each node within the EHR, and only matched nodes are selected as the *Target Objects* of the authorization. We formally define these concepts as follows:

Definition 4. *(**Filtration Property**). Let O, S, and T be the sets of data origins, sensitivity classifications, and object types, respectively as defined in Definition 2. A filtration property is specified as a tuple prop=<po,ps,pt>, where $po \subseteq O$ is the filtration property for origins; $ps \subseteq S$ is the filtration property for sensitivity classifications; and $pt \subseteq T$ is the filtration property for object types.*

Definition 5. *(**Property Match**). Suppose prop=<po, ps, pt> is a filtration property specification, and p'=(po',ps',t') is the property label of a node, the node matches the filtration property if the following conditions are satisfied:*

1. *$p'.po' \subseteq prop.po$;*
2. *$p'.ps' \subseteq prop.ps$; and*
3. *$p'.t' \in prop.pt$.*

Definition 6. *(**Object Specification**). Let scp_expr be a scope expression to denote a set of nodes within the composition, and prop be a filtration property specification, the ob-*

ject selection specification is defined as a tuple $ao = (scp\_expr, prop)$. Given an EHR logical model $C = (v_c, V_o, E_o, \tau_{V_o})$ and an object selection specification ao, we define a function: $select(C, ao) \rightarrow V_a$, where $V_a \subseteq V_o$, to select the matched nodes within the specified scope as the Target Objects.

In specifying filtration properties, we also allow patterns to be used. Pattern "*" is to indicate *any value* within a property dimension, and pattern "{*}" is to specify *any set* within a property dimension.

Example 1. *The followings are two examples of object selection specifications against the EHR structure in Figure 1(c).*
***ao1**: ao1=(/VirtualEHR/History//*,<{h2},{general},*>); and*
***ao2**: ao2=(/VirtualEHR/History//*,<{*},{HIV},*>).*

The two object specifications select the same scope as the *History* category in the virtual composite EHR. ***ao1*** selects the nodes that come from $h2$ with *general* level of sensitivity and *any* object types. ***ao2*** selects the nodes from *any* origins with *HIV* level of sensitivity and *any* object types. Figure 1(c) illustrates the target objects being selected according to the *select*() function in Definition 6 as two dashed zones. In particular, ***ao1*** results in the `Asthma` node under `Illness` being selected, and ***ao2*** results in the nodes of `HIV` and `Prescription2` being selected under `Illness` and `Medications`, respectively.

Purposes: To further address a patient's privacy concerns in sharing his medical information, an attribute of "purpose" is necessary to be specified in the authorization policy to confine the intended purposes/reasons for data access in healthcare practice. According to [9], business practices for health information exchange can be organized by 11 purposes including payment, treatment, research, and so on. Formally, intended purpose is specified as follows:

Definition 7. ***(Intended Purpose)**. Let P be a set of purposes for business practices in healthcare domain. And let m be the total number of authorizations in the system. The intended purpose set $P_p = \{pp_1, \ldots, pp_m\}$ is a collection of possible intended purpose sets, where $pp_i \subseteq P$ specifies the intended purposes for a particular authorization, $i \in [1, m]$.*

Access Control Policy: To summarize the above-mentioned policy elements, we introduce the definition of an access control policy as follows:

Definition 8. ***(Access Control Policy)**. An access control policy is a tuple $acp =< sub, ao, pp, effect >$, where*

- *$sub \in Sub$ is a subject;*
- *ao is an object selection specification resulting in a set of nodes $V_a \subseteq V_o$ being selected as target objects;*
- *$pp \in P_p$ is the intended purposes; and*
- *$effect \in \{permit, deny\}$ is the authorization effect of the policy.*

Example 2. *Let* ao1 *and* ao2 *be specified as same as those in Example 1, the following access control policies can be articulated:*
***P1**: (<GP,{h2}>, ao1, {treatment}, permit);*
***P2**: (<SP,{h2}>, ao2, {treatment,research}, permit); and*
***P3**: (<Dr. Jones,{h2}>, ao2, {treatment,research}, deny).*

In **P1**, a patient allows all general practitioners (GP) in *h2* to view his *general* medical history for *treatment* purpose. In **P2**, the patient allows all specialists (SP) in *h2* to view his *HIV* history for *treatment* and *research* purposes. Suppose Specialist *Dr. Jones* in *h2* is a relative of the patient, the patient defines **P3** to deny his access to the *HIV* data.

In healthcare practice, a default policy may be established to satisfy most patients' most privacy requirements. Once a patient understands the default policy and agree that it meets his needs, the patient may not need to further specify any specific access control policies to control the sharing of his medical information. In particular, HIPAA regulations are widely adopted by healthcare practitioners in the United States. With the agreement of the default setting, HIPAA generally allows health care providers to share clinical information without the individual's explicit permission for treatment, payment and health care operations [22]. In addition, in order to accommodate the emergency situations, a "break-of-glass" policy ("*BG*" policy for simplicity) should be specified to allow staffs in emergency rooms to access the patient's medical information without the patient's explicit authorizations. Both the default policy and "*BG*" policy can be specified conforming to our unified policy schema.

Example 3. *The default policy and BG policy can be specified as follows:*
$\boldsymbol{P}_D$:(<HP,{}>,({*},{*},*),{treatment,payment,HCO},permit);*
$\boldsymbol{P}_{BG}$:(<ERStaff,{}>,({*},{*},*),{treatment},permit).*

3.3 Policy Composition and Anomaly Analysis

The access control policies in our model can be uniformly specified both at the EHR-instance level within different origin domains and at the aggregation level within a virtual composite EHR. However, bringing in both positive and negative policies raises problems of possible conflicts, and the flexibility in data selection may result in issues of policy overlap. We generalize such possible policy issues as *policy anomalies.* Policy anomalies may appear within one origin domain for a specific EHR instance and we call it as intra-domain policy anomaly. To make things more complicated, when data from multiple origins are aggregated and integrated into one virtual composite EHR, we have to consider the policy anomalies within the composite policies, where policies associated with different data origins should be composed together to control the data sharing of the virtual composite EHR. Especially, as the same data elements from different origins may merge, the inter-domain policy anomalies may occur over the data elements being merged. For example, the `Asthma` illness history in Figure 1 appears both in the EHR instances from *h1* and *h2*. Policy conflicts for the data element may occur between the two policy domains when a patient allows the access in *h1* but denies it in *h2*. Therefore, policy anomalies must be identified and resolved within the policy composition. In this section, we first articulate the possible policy anomalies including *policy inconsistency* and *policy inefficiency*, and then discuss how to discover and resolve the policy anomalies.

3.3.1 Anomalies in Composite Policies

We use the following examples to illustrate both *policy inconsistency* and *policy inefficiency*. All these policy anomalies are formally defined in the subsequent Section.

EXAMPLE 4. *We further define an object selection specification as*
***ao3**: ao3=(/VirtualEHR/History//*,<{*},{*}, text>)*
to select all text data elements under `History` category.

*Suppose the patient defines four policies in **h1** as follows:*
***P4**: (<GP,*>, ao2, {treatment}, deny);*
***P5**: (<Dr.Jones,{h2}>, ao2, {research}, permit);*
***P6**: (<SP,{h1}>, ao3, {research}, permit);*
***P7**: (<Dr.Jones,{h2}>, ao3, {treatment}, deny);*

*Later, the patient defines the following policies in **h2**:*
***P8**: (<Dr.Jones,{h2}>, ao3, {research}, deny);*
***P9**: (<GP,*>, ao2, {treatment}, permit);*
***P10**: (<GP,{h1}>, ao2, {treatment}, deny);*

[Policy Inconsistency]: Access control policies defined for EHRs reflect the patient's privacy protection requirements, which should be consistent within and across EHR instance origins. Inconsistent policies might result in both security and availability problems. Given the above examples, we elaborate the following policy inconsistencies:

1. *Contradictory*: Two policies are contradictory to each other if they have different *effects* (permit or deny) over the same *subjects* (*sub*), *target objects* (*ao*) and *intended purposes* (*pp*). This is often considered as a typical policy conflict. For example, **P4** is contradictory to **P9** as general practitioners (*GP*) are permitted to access the data matching *ao*2 for *treatment* purpose in **P4**, but are explicitly denied in **P9**.
2. *Exception*: A policy is an exception of another policy if they have different *effects*, but one policy is the subset of the other. Suppose *Dr. Jones* is a specialist (*SP*) in both *h1* and *h2*, then **P8** is an exception of **P6**, since all specialists in *h1* are allowed to view the data matching *ao3* for *research* purpose (**P6**), except for *Dr. Jones* (**P8**). The exception is not necessarily be a policy conflict as it is commonly used to exclude a specific access request from a general access permission.
3. *Correlation*: Two policies are correlated if they have different *effects*, but one policy intersects with the other. In this case, the intersection of the two policies is permitted by one policy, but denied by the other. This is considered as a partial policy conflict. For example, **P5** and **P8** fall in this category. The intersection of them is a data element of *HIV* history. **P5** permits *Dr. Jones* in *h2* to access this data element for *research* purpose, but **P8** denies it.

[Policy Inefficiency]: The composition of policies from multiple origins may result in a large number of policies being collected to control the access of the virtual composite EHRs. Since the response time of an access request largely depends on the number of policies to be parsed in the policy pool, inefficiencies in composite policies are not conflicts yet may still adversely affect the performance of policy evaluation. Therefore, both *redundancy* and *verbosity* in composite policies are regarded as anomalies as well.

1. *Redundancy*: A policy is redundant if there is another same or more general policy available that has the same *effect*. For example, **P10** is redundant since all general practitioners (**P4**), including those in *h1* (**P10**), are already denied to access the data matching *ao*2.
2. *Verbosity*: Similar to the data element merging in the data integration, policies from different origins may be merged in the policy composition. For example, **P7** and **P8** can be integrated into a single policy (*<Dr.Jones,{h2}>, ao3, {treatment, research}, deny*). The policy size can thus be reduced when resolving the policy verbosity.

3.3.2 Checking for Anomalies

According to Definition 8, an access control policy essentially determines a relation of $sub \times V_a \times pp$, where V_a is derived from ao. We further define such relation as an ***Authorization Zone*** (AZ). In addition, we use a utility notation $P_\alpha[\beta]$ to indicate a particular field of a policy. It implies that a field (or a set of fields) β from policy P_α. By formalizing the relationships between the policy-determined authorization zones and their effects, we could identify the policy inconsistency and inefficiency. As shown in Figure 2, there are four possible relationships between the authorization zones of two access control policies.

- *Exactly Match* ($\curlyvee_{EM}$): An authorization zone AZ_x determined by policy P_x exactly matches another authorization zone AZ_y derived from policy P_y, if and only if the fields of *sub*, *ao* and *pp* in P_x are equal to the corresponding fields in P_y. Formally,

$$\forall i : P_x[i] = P_y[i] \Rightarrow AZ_x \curlyvee_{EM} AZ_y,$$

where $i \in F = \{sub, ao, pp\}$.

- *Inclusively Match* ($\curlyvee_{IM}$): An authorization zone AZ_x determined by policy P_x inclusively matches another authorization zone AZ_y derived from policy P_y, if and only if the fields of *sub*, *ao* and *pp* in P_x do not exactly match but are a subset of the corresponding fields in P_y. Formally,

$$\forall i : P_x[i] \subseteq P_y[i] \ and \ \exists j : P_x[j] \subset P_y[j] \Rightarrow AZ_x \curlyvee_{IM} AZ_y,$$

where $i, j \in F$ and $i \neq j$.

- *Partially Match* ($\curlyvee_{PM}$): An authorization zone AZ_x determined by policy P_x partially matches another authorization zone AZ_y derived from policy P_y, if and only if the fields of *sub*, *ao* and *pp* in P_x do not exactly and inclusively match, but have intersections with the corresponding fields in P_y. Formally,

$$\forall i : P_x[i] \cap P_y[i] \neq \varnothing \ and \ \exists j : P_x[j] \nsubseteq P_y[j] \wedge P_y[j] \nsubseteq P_x[j]$$

$$\Rightarrow AZ_x \curlyvee_{PM} AZ_y,$$

where $i, j \in F$ and $i \neq j$.

- *Disjoint* ($\curlyvee_{DJ}$): An authorization zone AZ_x determined by policy P_x is disjoint with another authorization zone AZ_y derived from policy P_y, if and only if the fields *sub*, *ao* and *pp* in P_x have no intersection with the corresponding fields in P_y. Formally,

$$\forall i : P_x[i] \cap P_y[i] = \varnothing \Rightarrow AZ_x \curlyvee_{DJ} AZ_y,$$

where $i \in F$.

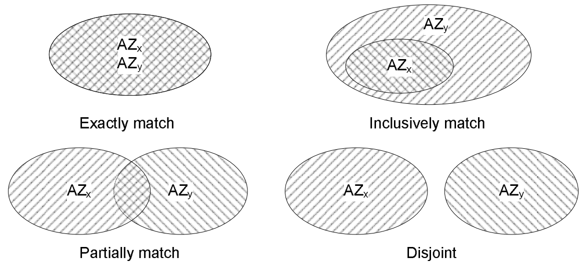


Figure 2: Relationships between Authorization Zones

Figure 3 shows a state transition diagram for policy anomaly analysis. Anomalies between two policies could be detected by first identifying the relationships of the two derived authorization zones, and then comparing the *effects* of the two policies. In particular, for two policies P_x and P_y:

1. $AZ_x \curlyvee_{EM} AZ_y$ or $AZ_x \curlyvee_{IM} AZ_y$, and $P_x[effect] = P_y[effect] \Rightarrow$ *Redundancy*: If authorization zones determined by P_x and P_y exactly match or inclusively match and P_x and P_y define the same *effect*, then P_x is a redundant policy, since all access requests that match P_x can still be matched by P_y, and get the same response from P_y.
2. $AZ_x \curlyvee_{EM} AZ_y$ and $P_x[effect] \neq P_y[effect] \Rightarrow$ *Contradictory*: When authorization zones determined by P_x and P_y exactly match and P_x and P_y define different *effects*, P_x and P_y are contradictory to each other.
3. $AZ_x \curlyvee_{IM} AZ_y$ and $P_x[effect] \neq P_y[effect] \Rightarrow$ *Exception*: If authorization zones determined by P_x and P_y inclusively match and P_x and P_y define different *effects*, P_x is then regarded as an exception of P_y.
4. $AZ_x \curlyvee_{PM} AZ_y$ and $P_x[effect] \neq P_y[effect] \Rightarrow$ *Correlation*: Authorization zones determined by P_x and P_y partially match and P_x and P_y define different *effects*. This indicates that P_x and P_y are correlated.
5. $AZ_x \curlyvee_{PM} AZ_y$ and $P_x[effect] = P_y[effect]$, or $AZ_x \curlyvee_{DJ} AZ_y \Rightarrow$ *Normal*: If authorization zones determined by P_x and P_y partially match and P_x and P_y define the same *effect*, or authorization zones determined by P_x and P_y are disjoint, there is no anomaly between P_x and P_y.

This approach can be easily adapted to detect policy anomalies among multiple policies, where the composite policies should be evaluated as a whole piece. With the *permit* or *deny* effect of a policy, we define that the determined authorization zone is tagged with "+" or "−", respectively. In order to analyze policies one by one along with the composite policies, we suppose there is a policy pool storing existing i composed policies, and the *(i+1)*th policy is being integrated with existing composite policies. We can define current entire authorization zone determined by existing composite policies as $AZ_i = AZ_i^+ \cup AZ_i^-$, where AZ_i^+ and AZ_i^- denote the entire permitted authorization zone and denied authorization zone before integrating the *(i+1)*th policy, respectively. If the *effect* of the *(i+1)*th policy is "permit", after integrating this policy with existing composite policies, the new authorization zone can be computed as the equation $AZ_{i+1}^+ = AZ_i^+ \cup AZ_{P_{i+1}}$; otherwise, $AZ_{i+1}^- = AZ_i^- \cup AZ_{P_{i+1}}$.

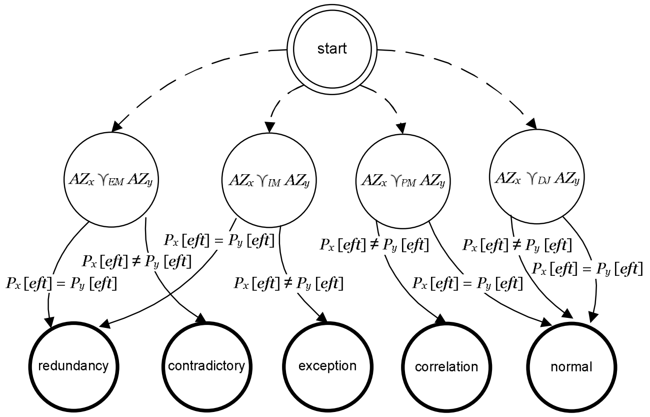


Figure 3: Simplified State Transition Diagram for Policy Anomaly Analysis

By analyzing the relationship between $AZ_{P_{i+1}}$ and AZ_i^+, or between $AZ_{P_{i+1}}$ and AZ_i^-, we can identify more accurate policy anomalies. For example, a policy may be redundant to the composition of several policies. Furthermore, this approach sequentially traverses and integrates each policy for policy analysis in the pool, thus achieves a computational complexity of *O(n)* where *n* is the number of policies for the overall anomaly analysis.

3.3.3 Resolving Anomalies

In practice, there are two stages for policy anomaly detection and resolution. The first stage is during the policy creation and update within a local EHR instance origin. If a patient inserts, modifies or removes a policy from a local policy set, policy anomaly analysis should be conducted in order to highlight any potential anomalies that may be introduced in adding and updating policies. Such anomalies can be resolved by consulting the patient to make changes to certain policy specifications against the detected anomalies.

The second stage of policy analysis is during the policy composition when the EHR instances are aggregated as a virtual composite EHR. Redundancies in composite policies could be identified and removed immediately during the stage of policy composition. However, considering a large list of composite policies, it may be very difficult, or even impossible, to resolve all identified conflicts manually. On the other hand, without a priori knowledge on the patient's authorization requirements, we cannot correct the conflict automatically either. In this case, a practical method of resolving policy conflicts is to identify which policy involved in a conflict situation will take precedence. Therefore, we introduce the following strategies which could be used to resolve policy conflicts for composite policies in a patient-centric EHR sharing system.

1. *New-authorization-overrides*: This strategy states that later authorization prevails earlier authorization. As the authorization requirements may change over a period of time, a patient may specify certain policies in one origin, and he might define new policies in other origins along with his changed authorization requirements, for example to give a special permission to a physician at the point of care. Obviously, when aggregating these policies together, newer police(s) should

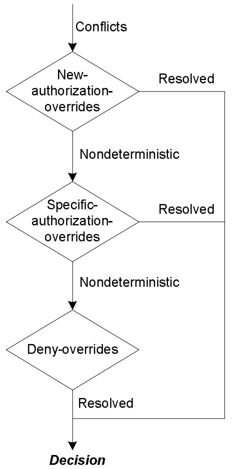


Figure 4: Composite Strategy for Policy Conflict Resolution

take precedence to respond an access request that matches policies from different policy sets. However, this strategy may not always resolve the policy conflicts since no authorization wins when two conflicting policies have the same timestamps [1]. We call a strategy that cannot always derive a solution as a *nondeterministic* strategy.

2. *Specific-authorization-overrides*: This strategy states that a more specific authorization overrides a more general authorization. As we discussed, *exception* in composite policies is not necessarily a conflict, but often used to exclude specific part of a larger set for certain effect. Thus, when *exception* occurs within several policies, the more specific policies take precedence. This strategy is also nondeterministic if there is no exception in composite policies.

3. *Deny-overrides*: This strategy indicates that the decision for an access request is "Deny" if any matched policy evaluates to "Deny". As the fundamental reason of policy conflict is that two policies specify different *effects*, it implies that at least one policy defines an *effect* as "Deny". Therefore this strategy is deterministic and always returns "Deny". In order to protect the patient's privacy and confidentiality in healthcare systems, it is plausible to apply such a strict strategy in some special cases.

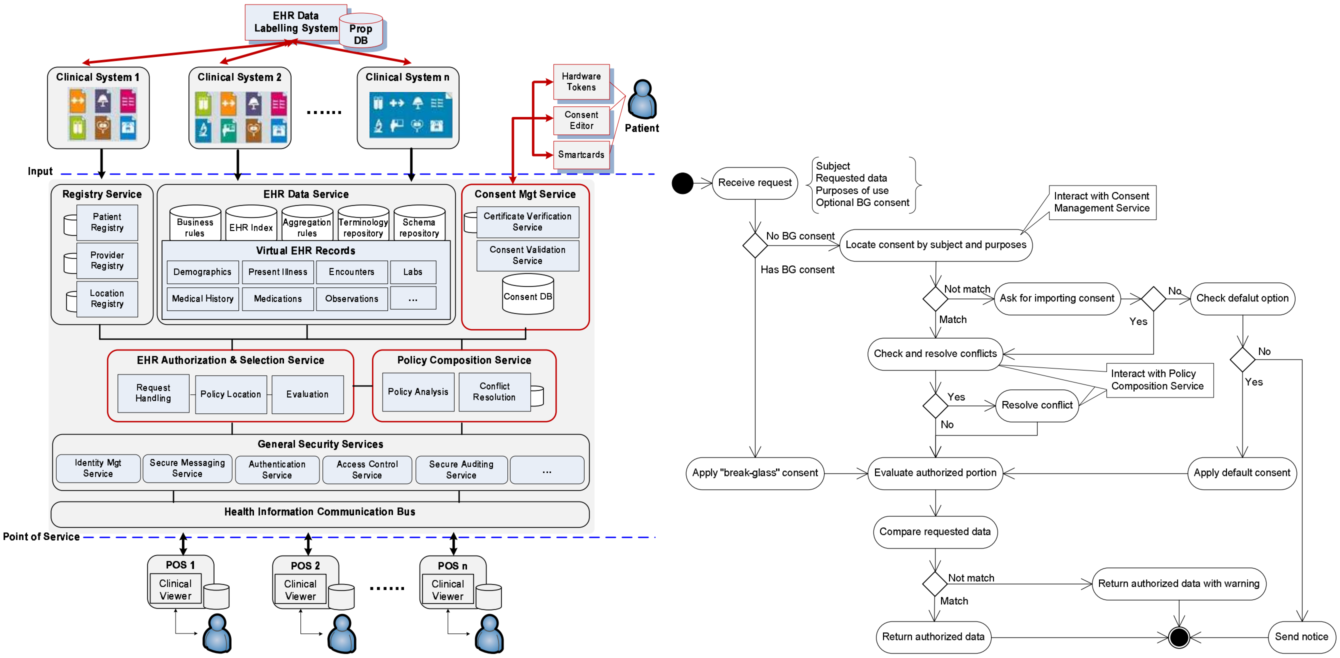
Since *new-authorization-overrides* and *specific-authorization-overrides* strategies are nondeterministic, and *deny-overrides* strategy is too restricted in general for conflict resolution, it is possible to combine these strategies together to achieve a more effective conflict resolution. Figure 4 demonstrates a composite strategy for policy conflict resolution that sequentially applies *new-authorization-overrides*, *specific-authorization-overrides*, and *deny-overrides* strategies. In Example 4, if we assume that the policy set defined in *h2* is newer than the policy set defined in *h1*, through applying *new-authorization-overrides* strategy, **P9** takes precedence over **P4** to solve the conflict by granting *GP* the access to the data matching *ao2*. In addition, **P8** takes precedence over **P6** and **P5** to solve the identified exception and correlation anomalies, respectively. Otherwise if we assume the policy sets from *h1* and *h2* have the same timestamps or corresponding timestamps could not be retrieved by the system, *specific-authorization-overrides* strategy is then applied and the exception inconsistency between **P8** and **P6** could be resolved by **P8** taking precedence, resulting in an access request from *Dr. Jones* being denied. Meanwhile, the anomalies between **P4** and **P9** and between **P5** and **P8** will be resolved by applying the *deny-overrides* strategy, where **P4** takes precedence over **P9**, and **P8** takes precedence over **P5** to deny the access requests accordingly.

4. EHR SHARING SYSTEM

As part of our ongoing research efforts, we have designed and implemented a proof-of-concept system, called *InfoShare*, which is a simplified clinical information sharing system that utilizes our proposed model for a patient to control access of his medical information. In particular, *InfoShare* collects a patient's access control policies as patient consents, and uses these consents to selectively share the patient's medical information as contained in the virtual composite EHR with different requesting parties through a point of service (POS) web application.

Architecturally, health information as EHR instances generally is maintained and managed at each geographically distributed care provider's site with the notation of federation. The data elements in the EHR instances should be associated with special properties for data filtration and authorization purposes. We therefore establish an EHR data labelling system in the federation to facilitate the unified property labelling for the care providers' EHR instances. Our *InfoShare* system serves as a middleware application to retrieve and aggregate the distributed property-labelled EHR instances on the fly at the point of care, where the resulting virtual composite EHR appears as a single integrated record logically connecting a group of care providers and organizations within the federation. Meanwhile, InfoShare also serves as a gatekeeper to protect and selectively share the integrated virtual composite EHR based on the patient's consents. Figure 5(a) illustrates the overall architecture of an integrated *InfoShare* information system. As a general clinical information sharing system, the Registry Service, EHR Data Service, General Security Service and Health Information Communication Bus are common system components to achieve the required functionalities of secure data retrieval, virtual composite EHR creation and communication with requesting POS applications. Especially, we inject the Consent Management Service, Policy Composition Service, and EHR Authorization & Selection Service as the major system modules to convey the core features of our proposed approach. In particular, the Consent Management Service enables the patient control by collecting and verifying the patient's access control policies as encapsulated in consents. A web-based consent editor tool is implemented to facilitate a patient to edit his policy consents, and interact with the Consent Management Service for the patient to store and update his consents. Besides, the patient consents can also be directly inserted into the Consent Management Service by the patient using smart cards or other hardware tokens. The Policy Composition Service handles the policy issues discussed in Section 3.3 to analyze the access control policies and resolve the identified policy

[1] Suppose a patient defines a set of policies during the same session.



(a) Overall System Architecture

(b) Consent Evaluation Procedure

Figure 5: InfoShare System

anomalies in the policy composition. The EHR Authorization & Selection Service is responsible to handle the data access requests, and evaluate relative access control policies to derive the authorized portion of data to be shared with the requester.

As discussed earlier, the expressed access control policies are encapsulated as consents. There are three types of consents in the system: the patient's specific consents, the default consent, and the "*BG*" consent, where the default consent and "*BG*" consent specify the default policy P_D and "*BG*" policy P_{BG}, respectively as shown in Example 3. The precedence order of evaluating these consents is defined as $BG\_consent \succeq patient\_consent \succeq default\_consent$. Figure 5(b) illustrates the procedures for the EHR Authorization & Selection Service to handle access requests and derive the authorized data to be shared. An access request includes information of the requester subject, the requested data, the intended purposes of use, and an optional "*BG*" consent in emergency situations. The "*BG*" consent in emergency has the highest priority in execution, therefore such consents are directly evaluated to get the authorized data. In other situations, the authorization service interacts with the Consent Management Service to locate the related patient consents based on the specified subject and intended purposes. If certain matched consents are located, the Policy Composition Service is invoked to check and resolve any possible policy conflicts, and then the policies are evaluated to derive the authorized portion of data within a virtual composite EHR. If there are no patient consents being located, our system asks for the patient to insert a patient consent at run-time and the consent is evaluated accordingly. Otherwise, the default consent is evaluated to derive the authorized data. After the authorized data portion is determined, it is compared with the requester's requested data and only matched data portion is returned to the requester. If the requester is not authorized to access all the data he requested, the requester is notified with a warning, so that the requester may further ask for new patient consents to access the data for the need of his practice. Such an effective mechanism is utilized to balance the data integrity concern of the practitioners and the privacy concern of the patients for shared EHRs.

In terms of implementation, the XML-based HL7 Clinical Document Architecture (CDA) [10] is utilized in our *InfoShare* system for the formal representation of EHR instances as well as the virtual composite EHR. We use Jaxe XML editor [1] as the EHR data labelling service to associate properties with data elements in CDA EHRs. We implement the patient consents as X.509 attribute certificates [15], where the access control policies are encapsulated as attributes within the certificate. The *InfoShare* system implements a Java Servlet based web portal as the POS application for a healthcare practitioner to query and view the authorized medical information of a patient.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed an innovative approach to supporting authorized and selective sharing of virtual composite EHRs. The access control policies are specified around the unified logical EHR model taking into consideration critical issues such as distributed data integration and privacy protection concerns. We also proposed a mechanism to identify and resolve the policy anomalies in the process of policy composition. Our approach was demonstrated in a prototype *InfoShare* system that uses e-Consent mechanism to enable the patient-centric medical information sharing with different parties in the healthcare environment.

For the future work, rigorous experiments need to be conducted to evaluate the performance and storage efficiency of our *InfoShare* system. One of critical prerequisites for patient-centric healthcare systems is how easily a common patient can maintain his access control and privacy preferences for such a huge amount of sensitive and complex information across sites while making the information highly usable for healthcare professionals. Therefore, a variety of analytical and empirical methods from the area of usability engineering could be adapted to investigate usability of our system. In addition, a patient may wish to delegate the capability to nominated representatives or medical practitioners, who may further wish to delegate the consent privilege to other health professionals. Practical consent delegation and control mechanisms are crucial while ensuring the patient's control power on his medical data with proper privacy protections. Finally, our approach is complementary to and can be adapted to other existing security solutions (i.e., RBAC [2, 11] and situation-based access control [21]) for providing a fine-grained access control in healthcare systems.

Acknowledgments

This work was partially supported by the grants from National Science Foundation (NSF-IIS-0242393) and Department of Energy Early Career Principal Investigator Award (DE-FG02-03ER25565).

6. REFERENCES

[1] Jaxe XML editor. http://jaxe.sourceforge.net/.

[2] J. Barkley and K. Beznosov. Supporting relationships in access control using role based access control. In *Proc. of 4th ACM Workshop on Role-Based Access Control*, pages 55–65, 1999.

[3] M. Y. Becker and P. Sewell. Cassandra: flexible trust management, applied to electronic health records. In *Proc. of IEEE 17th Computer Security Foundations Workshop*, pages 139–154, 2004.

[4] R. Bhatti, K. Moidu, and A. Ghafoor. Policy-based security management for federated healthcare databases (or RHIOs). In *Proc. of the international workshop on Healthcare information and knowledge management*, pages 41–48, 2006.

[5] J.-W. Byun, E. Bertino, and N. Li. Purpose based access control of complex data for privacy protection. In *Proc. of 10th ACM symposium on Access control models and technologies (SACMAT)*, pages 102–110, 2005.

[6] Ciena. The national health information network creating a new vision. White Paper, Healthcare Information and Management Systems Society (HIMSS) Conference 2008, 2008.

[7] E. Coiera and R. Clarke. e-consent: the design and implementation of consumer consent mechanisms in an electronic environment. *Journal of the American Medical Informatics Association*, 11(2):129–140, 2004.

[8] dbMotion. White paper: The critical role of integrated patient information in the delivery of high quality healthcare, January 2008.

[9] L. L. Dimitropoulos. Privacy and security solutions for interoperable health information exchange: Interim assessment of variation executive summary. http://www.rti.org/pubs/avas_execsumm.pdf, July 2007. RTI Project Number 0209825.000.009.

[10] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, and P. V. Biron. Hl7 clinical document architecture, release 2.0. ANSI Standard, 2004.

[11] D. M. Eyers, J. Bacon, and K. Moody. OASIS role-based access control for electronic health records. In *IEEE Proceedings – Software*, pages 16–23, 2006.

[12] C. Gates and J. Slonim. Owner-controlled information. In *Proc. of the 2003 workshop on New security paradigms*, pages 103–111, 2003.

[13] J. Grimson, G. Stephens, B. Jung, W. Grimson, D. Berry, and S. Pardon. Sharing health-care records over the internet. *IEEE Internet Computing*, 5(3):49–58, 2002.

[14] HL7. Hl7 reference information model. http://www.hl7.org/Library/data-model/RIM/modelpage_mem.htm.

[15] R. Housley, W. Polk, W. Ford, and D. Solo. Internet x.509 public key infrastructure certificate and certificate revocation list (crl) profile. RFC3280, http://rfc.net/rfc3280.html, 2002.

[16] IEEE-USA's Medical Technology Policy Committee Interoperability Working Group, editor. *Interoperability for the National Health Information Network (NHIN)*. IEEE-USA EBOOKS, 2006.

[17] Iowa Foundation for Medical Care. HISPC state implementation project summary and impact analysis report for the state of Iowa. http://www.ifmc.org/news/State Impact Report_11-27-07.doc, 2007.

[18] J. Jin, G.-J. Ahn, M. J. Covington, and X. Zhang. Toward an access control model for sharing composite electronic health record. In *Proc. of 4th International Conference on Collaborative Computing*, 2008.

[19] C. M. O'Keefe, P. Greenfield, and A. Goodchild. A decentralised approach to electronic consent and health information access control. *Journal of Research and Practice in Information Technology*, 37(2):161–178, 2005.

[20] openEHR Community. openEHR. http://www.openehr.org.

[21] M. Peleg, D. Beimel, D. Dori, and Y. Denekamp. Situation-based access control: Privacy management via modeling of patient data access scenarios. *Journal of Biomedical Informatics*, 41(6):1028–1040, 2008.

[22] J. Pritts and K. Connor. The implementation of e-consent mechanisms in three countries: Canada, england, and the netherlands. SAMHSA report, http://ihcrp.georgetown.edu/pdfs/prittse-consent.pdf, 2007.

[23] C. Ruan and V. Varadharajan. An authorization model for e-consent requirement in a health care application. *Applied Cryptography and Network Security, LNCS*, 2846:191–205, 2003.

[24] N. Yang, H. Barringer, and N. Zhang. A purpose-based access control model. In *Proc. of 3rd International Symposium on Information Assurance and Security (IAS)*, pages 143–148, 2007.