Toward an Access Control Model for Sharing Composite Electronic Health Records

Jing Jin[1], Gail-Joon Ahn[2], Michael J. Covington[3], and Xinwen Zhang[4]

[1] University of North Carolina at Charlotte, jjin@uncc.edu
[2] Arizona State University, Gail-Joon.Ahn@asu.edu
[3] Intel Corporation, michael.j.covington@intel.com
[4] Samsung Information Systems America, xinwen.z@samsung.com

Abstract. The adoption of electronically formatted medical records, so called Electronic Health Records (EHRs), has become extremely important in healthcare systems to enable the exchange of medical information among stakeholders. An EHR generally consists of data with different types and sensitivity degrees which must be selectively shared based on the need-to-know principle. Security mechanisms are required to guarantee that only authorized users have access to specific portions of such critical record for legitimate purposes. In this paper, we propose a novel approach for modelling access control scheme for composite EHRs. Our model formulates the semantics and structural composition of an EHR document, from which we introduce a notion of *authorized zones* of the composite EHR at different granularity levels, taking into consideration of several important criteria such as data types, intended purposes and information sensitivities.

1 Introduction

Healthcare is an increasingly collaborative domain involving a wide range of individuals and organizations. Seamless electronic communication infrastructure that allows patients, physicians, hospitals, public health agencies and other authorized users to share clinical information in real-time, under stringent security and privacy protections, has become extremely important to improve the quality of healthcare while simultaneously reducing costs and administrative complexity [1]. In particular, the adoption of electronically formatted medical records, so called Electronic Health Records (EHRs), has become the primary concern for a broad range of health information technology applications and practitioners.

Critical concerns about the privacy and security of personal medical information remain high in healthcare information sharing systems. More than ever, there is a strong need to define access control models that conform to legal principles and regulations, while limiting access to information to those entities on need-to-know basis. However, an EHR includes complex health information such as the patient demographics, medical histories, examination reports, laboratory test results, radiology images (X-rays, CTs), and so on. Supporting the autho-

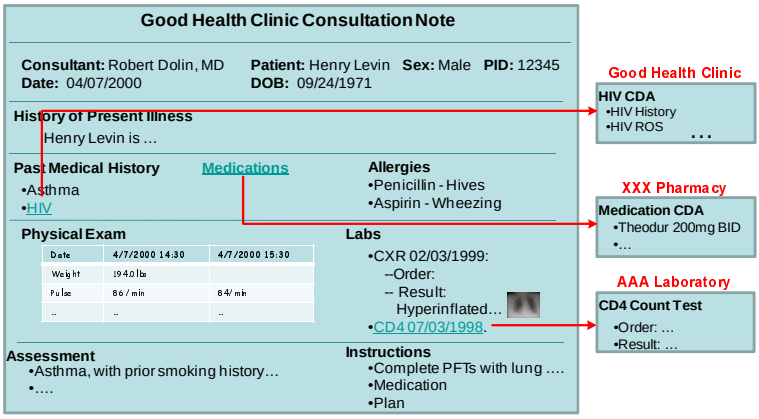


Fig. 1. Motivation EHR Document

rized and selective sharing of EHRs among several parties with different duties and objectives is indeed a great challenge.

1.1 A Motivation Scenario

In order to better illustrate access control challenges on sharing of EHRs, we consider a typical clinical EHR document, and we demonstrate our proposed approach using the same document throughout the rest of this paper.

Suppose Good Health Clinic is a member of a particular Regional Health Information Organization (RHIO) [2], where health information can be exchanged through an established infrastructure with other involved organizations. Figure 1 illustrates a sample Consultation Note in the clinic for a patient named Henry Levin [3]. The consultation note includes Henry's past medical history, medications, physical examination, labs, etc. The medical information is recorded in various data types such as texts, numbers and images. Some fields inside the document may refer to other external clinical documents. For example, Henry's HIV/AIDS disease history may be maintained in another folder of the patient, and Henry's current medications may be directly linked to the records operated by his pharmacist. Given the complexity of this EHR document, the information contained in the consultation note should be legitimately exchanged to satisfy needs of different parties in RHIO. In particular, the lab orders need to be communicated with appropriate laboratories and specific test codes are used to trigger the billing process. The doctor's prescriptions, on the other hand, are necessary to be filled by the pharmacist, and proper referrals are exchanged with specialists for complex medical problems. However, ensuring the patient's privacy and data security is still vital for the EHR exchange system. The need-to-know principle must be strictly enforced for each responsible party to obtain only the necessary information to carry out its task. For instance, only the test codes and patient's insurance information are necessary for a billing clerk to fulfill her responsibility. The document also has sensitive fields, such as Henry's HIV/AIDS

medical history, which may be hidden from general medical information sharing unless a specific treatment purpose is indicated.

The example clinical document has demonstrated several unique characteristics of an EHR, including the composition of various data types, connections among different pieces of information from multiple sources, and navigational aspects of the information linkage and exchange. We thus refer the EHRs with such features as composite EHRs. In supporting partial sharing of a composite EHR, only a portion of the document needs to be shared with authorized users. Without explicitly identifying the *protection objects* and their associations within a composite EHR, the authorization specification referring to specific protection parts is difficult. In addition, these protection objects must be classified with regard to different purposes, data types, and sensitivity levels to guide the selection of specific parts with various protection granularity levels within the document. Finally, as an EHR document may link to other EHRs, the navigation paradigm would affect the authorization model, while the navigational links serve as a visual representation of associations between the EHR documents and need to be protected in a secure manner.

In this paper we propose a novel approach for modeling access control in composite EHRs. Our model first introduces a level of abstraction to formulate the logical structure of a composite EHR in terms of its internal protection objects and relationships among them. The protection objects are categorized by three dimensional properties – sensitivity, intended purpose and object type – to facilitate the authorization model and accommodate the composition and selective sharing requirements. By manipulating the selection criteria of these properties, various *authorized zones* including different protection objects can be dynamically collected to share with recipients.

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 access control models related to conventional structured or semi-structured data. In Section 3, we present the logical composite EHR model. The proposed authorization model and specification are discussed in Section 4. We conclude the paper in Section 5 with future research directions.

2 Related Work

[**Related EHR Standards**] There are several standards currently under development to specify EHRs, such as openEHR [4] and Health Level 7 (HL7) Clinical Document Architecture (CDA) [5, 3]. These standards aim to structure and markup the clinical content of an EHR for the purpose of exchange. The most important concept introduced in openEHR is the *archetype*, which is used to model healthcare concepts such as blood pressure and lab results. These archetypes serve as fundamental building blocks to form various clinical EHR documents. Meanwhile, these archetypes and the contents contained in them are exactly what need to be protected in the process of information exchange across healthcare systems. Similarly, CDA defines the structure and semantics

of medical documents in terms of a set of coded components (called vocabulary) to model basic medical concepts. A common feature of all emerging EHR standards is that the clinical concepts are modeled and expressed independently from how the data is actually stored in underlying database. 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, instead of being carried out at the lower level in underlying database, authorization and selective sharing of medical information should be defined and enforced with common understanding of EHR standards. In our motivation example and the rest of the paper, we assume the composite EHR document conforms to CDA standard format.

[**Access Control for EHR Systems**] A number of solutions have been proposed to address the security and access control concerns associated with EHR systems. In [6], the authors propose a set of authorization policies enforcing role-based access control for the electronic transfer of prescriptions. In [7], the paper demonstrates an implementation of EHR prototype system including a basic network and role-based security infrastructure for the United Kingdom National Health Service. In [8], the authors present a trust management and role-based policy specification language, called Cassandra, for expressing access control policies in large-scale distributed systems. A case study discusses how the language can be used to specify security policies for a UK national EHR system. In [9], the paper presents a policy-based security management framework to enforce context based authorizations for federated healthcare databases. Role-based access control [10], with its superior advantages in reducing administration complexity, has become the common theme applied in these approaches. However, the EHR considered in these approaches is either a general abstract object or an isolated primitive object. None of these approaches took into account of the composition feature of EHR documents, and thus cannot support a more fine-grained access control to selectively share composite EHRs as illustrated in our motivational example.

[**Related Authorization Models for Structured Data**] Sharing of composite EHRs requires clear understanding of the internal protection objects/clinical concepts and their structural relationships. There has been a considerable amount of work in regulating access to structured or semi-structured data.

The access control models proposed in [11] and [12] are especially tailored to object-oriented databases storing conventional structured data, where information is represented in the form of objects. These models consider a rich semantic structure of objects incorporating inheritance, aggregation, and composition associations. The relationship of objects in the database is modelled as a hierarchical structure so that the validity of an authorization rule written at some level can be efficiently propagated to its descendants. Such features can be adopted in modelling the logical structure of composite EHRs. However, these models have several shortcomings in providing effective access control for information exchange of EHRs. On the one hand, EHR documents are stored and exchanged based on standards, which are defined independently from underlying database

structures. The object relationships and navigational patterns defined in standards may be totally different from the ones enforced by access control mechanisms. On the other hand, as identified in our motivation scenario, the medical information may be distributed at different sites, from which a particular composite EHR document is derived. This unique feature cannot be addressed by a localized object-oriented database.

XML has become the de facto mechanism for sharing data between disparate information systems. It is essentially adopted by HL7 to carry out its standardization efforts to describe, store and exchange health records. Regulating access to XML documents has attracted considerable attentions in recent years [13, 14, 15]. All these work represent an XML document as a hierarchical tree structure and its authorizations are propagated along with the association links to achieve different granularity levels. However, all these approaches define access control rules for particular elements and attributes of an XML document. The selection of a portion of the document requires a number of authorization rules to be defined and evaluated. This is obviously not effective and efficient in practice to authorize and share a specific part of the document to fulfill the specific functional purpose of the requesting party. In addition, an XML document itself is not semantically enough to represent a variety of data types as encountered in composite EHRs (e.g., image, audio and video). Thus the access control mechanisms proposed for XML documents cannot meet the special requirements for sharing of composite EHRs.

3 Logical Composite EHR Model

In order to enable the selective sharing of specific parts of a composite EHR, we must allow the document to be logically divided into subcomponents so that fine-grained authorization can be applied. Therefore, we consider the basic building blocks of a composite EHR as the pieces of information or clinical concepts that might be individually exchanged. A piece of information is represented as a sub-object within a composite EHR, where each sub-object should be uniquely identified. Sub-objects can be nested at any depth within the EHR and can link to other sub-objects or even other EHRs. In our example, the blood pressure can be modeled as a sub-object in the Good Health Clinic's EHR document and it is nested under the physical examination object category, while the patient's current medication is linked to another EHR document in the pharmacy. We could further differentiate these two types of links between sub-objects and/or composite EHR documents as *inclusion* link and *navigation* link, respectively. The *inclusion* link realizes the typical "is a part of" relation, and the *navigation* link represents the "reference to" relation between sub-objects within or across composite EHR documents.

To cope with the essential features of different object types and their sensitivity levels within a composite EHR, we associate such information as properties for each sub-object within the document. The properties can be categorized into three dimensions: sensitivity, intended purpose and object type. The sensitivity property is designed to label a sub-object based on the sensitivity of the

content contained in it, which eventually can prevent sensitive medical information from being disclosed unintentionally. In the practice of Iowa HISPC [16], 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 sub-objects representing Henry's *HIV* medical history and the specific *CD4* lab test should be marked with "`communicable disease`" property ("`HIV`" for simplicity). The intended purpose property is necessary to address privacy concerns to guide the exchange of data based on specific purpose(s) and it is also essential to determine necessary pieces of information to fulfill the need-to-know requirement of a specific job function. According to [17], business practices for health information exchange can be organized by 11 purposes including payment, treatment, research, etc. These purposes could be achieved by exchanging different portions of a composite EHR document. The object type property gives another dimension on sub-object selection and protection. The sub-objects can be primitive types such as plain texts, dates and time, images and reference links. They can also be a composite type in the hierarchical structure including other types of sub-objects. Considering the navigational pattern within the document, the starting point of a navigation link should always be associated with an object labelled with the type of reference link.

As a summary, a composite EHR is modeled in terms of the composition of sub-objects and their relationships as links in a hierarchical structure. Each sub-object is labelled with properties of sensitivity, intended purpose and object type. These properties are used along with authorization policies to determine whether a specific sub-object is allowed to be exchanged or not. This can be formally defined as follows.

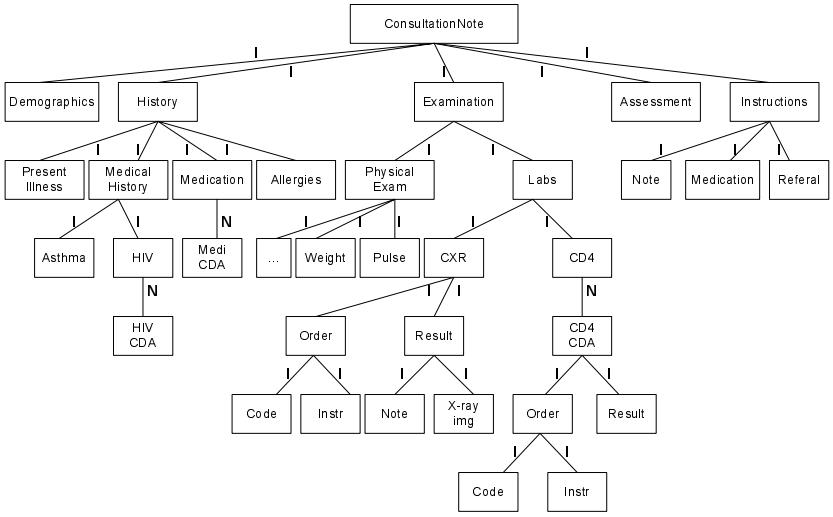
Definition 1 *(**Composite EHR**). A composite EHR is a tuple* $C = (v_c, V_o, E_o, \gamma_{E_o}, \tau_{V_o})$, *where*

- v_c *is the root representing the whole composite resource object;*
- V_o *is a set of sub-objects within the composite document under protection;*
- $E_o \subseteq V_o \times V_o$ *is a set of edges between sub-objects;*
- $\gamma_{E_o} : E_o \rightarrow \{I, N\}$ *is an edge labelling function indicating whether an edge is inclusion (*I*) or navigation (*N*) type;*
- $\tau_{V_o} : V_o \rightarrow P$ *is a sub-object labelling function to specify the property of a sub-object.* P *is a set of properties defined in Definition 2.*

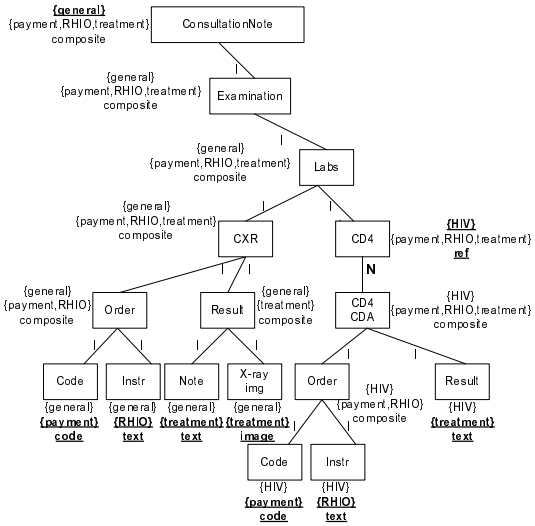
Definition 2 *(**Property**). Let* S, PU, *and* T *be the sets of sensitivity classifications, intended purposes, and object types, respectively.*

- P_s *is a collection of sensitivity classification sets,* $\{ps_1, \ldots, ps_m\}$, *where* $ps_i = \{s_1, \ldots, s_n\} \subseteq S$, $i \in [1, m]$;
- P_p *is a collection of intended purpose sets,* $\{pp_1, \ldots, pp_m\}$, *where* $pp_i = \{pu_1, \ldots, pu_n\} \subseteq PU$, $i \in [1, n]$;
- $P = P_s \times P_p \times T$ *is a set of three dimensional properties of sensitivity, intended purpose and the type.*

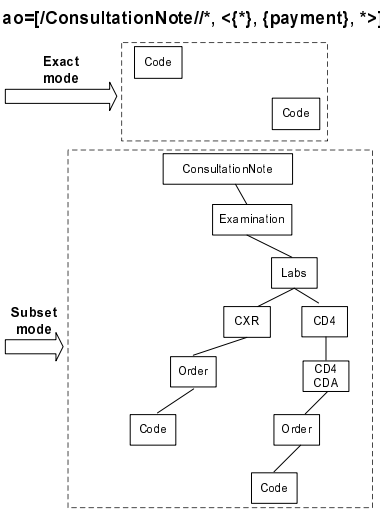
Given a property label p for a sub-object, we use the dot notation to refer to a specific property dimension. For instance, $p.ps$ refers to the sensitivity property; $p.pp$ refers to the intended purpose; and $p.t$ refers to the object type. The function $\tau(v_i)$ is used to retrieve the property label for a specific sub-object v_i inside the composition C.



(a) Composite EHR Document in a Hierarchical Structure



(b) Property Labeling and Propagation



(c) Property Match

Fig. 2. Composite EHR Document Structure

According to Definition 1, a composite EHR can be represented as a labelled hierarchical graph. The root of the tree graph indicates a particular composite EHR document. The nodes represent the sub-objects within the document and specific properties are associated with each node for authorization and selection. Edges represent the inclusion or navigational relationships between the nodes. Within the structure, nodes can be explicitly denoted by their identifiers, or can be implicitly addressed by means of *path* expressions. We apply a simplified XPath [18] expression for the path representation [1]. Simplification comes from the fact that each node is uniformly treated without a type, whereas XPath differentiates between "children" and "attributes" of an object due to the difference between elements and attributes in XML. We do not make such an explicit distinction because nodes in an EHR are the logical representation of clinical concepts under protection. By using our model, the logical structure of an EHR document in Figure 1 can be represented as a rooted tree as shown in Figure 2(a). The links inside the tree are labelled with `I` and `N` indicating types of inclusion and navigation, respectively. Figure 2(b) illustrates the example of node labelling for the section of `Labs` in the document. We use paths to select some of sub-objects in the graph as follows:

- /ConsultationNote: the whole composite EHR document;
- //`Labs/CXR`: this `CXR` lab test;
- //`Labs/CXR/*`: the child nodes of `CXR` lab test;
- //`Labs/CXR//*`: all the descendants of `CXR` lab test.

Another main design issue for the sub-object labelling scheme is the level of granularity an object should be associated with. As sub-objects are managed in a hierarchical structure of the composite EHR document, it enables us to provide a fine-grained labelling scheme yet achieves storage efficiency. In particular, we could explicitly label sub-objects with properties at a certain granularity level and allow the properties to be implicitly labelled through proper propagation and aggregation along with the hierarchical links. As properties are categorized in three dimensions, each has special characteristics for different authorization requirements. Therefore, the property propagation and aggregation for each dimension should be designed individually. We propose the following rules.

Rule 1. The property of sensitivity is automatically propagated downwards in the hierarchy until a more sensitive label is explicitly specified and overridden. We denote this as $P_s \downarrow$.

Rule 2. The property of intended purpose is aggregated upwards in the hierarchy. We denote this as $P_p \uparrow$.

Rule 3. The property of object type is aggregated upwards along with *inclusion* and *navigation* links and labelled as "`composite`" and "`ref`", respectively. We denote these types as $T \uparrow\!\!\xrightarrow{I}$"`composite`" and $T \uparrow\!\!\xrightarrow{N}$"`ref`".

In Rule 1, the structure represents an inheritance hierarchy, so that a property defined at the parent can be automatically inherited by its children, and a

[1] For brevity, we omit the formal definition of the path specification here.

child may define new properties to override the ones inherited from its parent. In our example, we assume the "`general`" label is the least-sensitive property, and other labels such as "`HIV`" and "`mental`" are more sensitive ones. As shown in Figure 2(b), the root of the clinical consultation note is labelled as "{`general`}" and this label is implicitly propagated downwards to all sub-objects within the structure. However, as *CD4* is a special lab test related to HIV/AIDS disease, the "`HIV`" sensitivity is explicitly specified to override the original "`general`" label. It is then implicitly inherited by its children nodes (e.g., *CD4 CDA*) in the hierarchy. In Rule 2, the hierarchical structure is treated as an aggregation association, where the purposes served by children nodes are aggregated by their parents. In our example, the *code* of the *CXR* lab test is used for "{`payment`}" purpose and the instruction *instr* is used for "{`RHIO`}" purpose to be communicated with the laboratory. Therefore, their parent node, *order* of *CXR* lab test, aggregates the purposes as "{`payment, RHIO`}". In Rule 3, the hierarchical structure reflects both the "is a part of" and "reference to" relations between the sub-objects. The parent node associated with *inclusion* links actually forms a type of "`composite`" to all its children nodes. And the parent node associated with *navigation* links referring to all its children nodes through a type of "`ref`". In our example, the root and all internal nodes are labelled as "`composite`", while *CD4* is labelled as "`ref`" since it is associated with a navigation link. In Figure 2(b), the properties using bold and underlined font indicate the explicitly specified properties and the ones with regular font indicate the implicitly assigned properties according to the rules.

4 Authorization Model

The fundamental question towards the selective sharing of a composite EHR is *what portion of a document can be exchanged with whom*. The role of an authorization model is then to articulate and specify policies to determine the authorized zone of a source tree that a given subject is permitted to access [2].

4.1 Authorization Subject

The role-based access control model has gained a lot of attention in healthcare security research [19, 6, 7, 8, 9] because of its ability to provide practical security administration for a large number of users. Users are authorized through their roles (e.g., patient, physician, nurse) to access EHR documents within a healthcare infrastructure. In our approach, we also adopt a notion of role, considering authorization subjects as roles directly. We assume a system-wide set of roles (R) has been established within a healthcare system and each individual user is a member of one or more roles. Access control policies are then specified as what role is authorized to access which part of an EHR document.

[2] In this paper, we mainly focus on read-only permission in our authorization model.

4.2 Authorization Objects and Property Match

The fine-grained authorization specification should support a set of protection objects with the broader coverage, ranging from a set of interrelated EHR documents to a specific portion of an EHR document. In our hierarchical composite EHR model, XPath-like path expressions can be utilized to specify the scope of the sub-objects to which an authorization policy applies. Meanwhile, properties provide the flexibility to group sub-objects and to establish authorized zones within a document scope for meeting various access control requirements. Therefore, the selection of objects can be indirectly achieved by specifying authorized properties. These authorized properties serve as the filtration criteria to be compared with labels of the sub-objects. The matched sub-objects are then selected to share with specific role(s). In specifying authorized properties, we allow patterns to be used instead of enumerating each property. Patterns are expressed by using the wildcard character. Two kinds of patterns are introduced: pattern "*" is to indicate any property type(s) within a property dimension; and pattern "{*}" is to specify any collection(s) of property sets within a property dimension. For example, $<\{*\},\{payment\},*>$ specifies the object(s) that have `any` collections of sensitivity levels, for `payment` purpose with `any` object type(s). The notion of authorized property specification is formally defined as follows.

Definition 3 *(**Authorized Property Specification**). An authorized property is specified as a tuple* $prop =< ps, pp, pt >$*, where* $ps \in P_s$ *or* $ps = \{*\}$ *is the authorized sensitivity property;* $pp \in P_p$ *or* $pp = \{*\}$ *is the authorized purpose property; and* $pt \subseteq T$ *or* $pt = *$ *is the authorized object type property.*

As each sub-object is labelled with both explicitly specified properties and implicitly inherited or aggregated properties as the result of property propagation, different semantics must be identified to accommodate such features by incorporating the cascading options to guide the matching process. We further introduce two matching modes as *exact* mode and *subset* mode. The *exact* mode can be utilized to specify access control policies for certain sub-object(s) with specific properties, while the *subset* mode can be specified to select a large collection of sub-objects related to the specified properties, considering the property propagation and aggregation along the hierarchical links.

Definition 4 *(**Property Match**). Suppose* $prop =< ps, pp, pt >$ *is an authorized property specification and* $p' = (ps', pp', t')$ *is the object property label,*

- In `exact` match mode, two properties match if the following is true: $[(prop.ps = \{*\})?true : (prop.ps = p'.ps')]\&\&[(prop.pp = \{*\})?true : (prop.pp = p'.pp')]\&\&[(prop.pt = *)?true : (p'.t' \in prop.pt)]$, that is, if patterns are not used, the sensitivity and intended purpose properties must be exactly equal in the authorized property and the object's label, and the object type must be contained in the authorized types. Otherwise, any pattern used in a property dimension returns a *true* for that property dimension.
- In `subset` match mode, the two properties match if the following is true:

$[(prop.ps = \{*\})?true : (prop.ps \supseteq p'.ps')]\&\&[(prop.pp = \{*\})?true : (prop.pp \subseteq p'.pp')]\&\&[(prop.t = *)?true : (p'.t' \in prop.pt)]$, that is, if patterns are not used, the sensitivity property of the object must be contained in the authorized sensitivity property, the authorized purpose property must be contained in the object's purpose property, and the object type must be equal.

We also define an authorization object that is used in an access control policy.

Definition 5 *(**Authorization Object Specification**). Let scp_expr be a scope expression to denote a set of authorization objects, and prop be an authorized property specification. An authorization object is specified as a tuple* $ao = (scp\_expr, prop)$.

Given Definition 4 and the example in Figure 2(b), an authorization object specified as

$ao = [/ConsultationNote//*, < \{*\}, \{payment\}, * >]$

means those sub-objects within the whole consultation note with `any` collections of sensitivities, for `payment` purpose, and for `any` object type. In the `exact` match mode, only the two *Code* objects are the matched ones, while in the `subset` match mode, all the parent nodes upwards to the root are additionally included. Figure 2(c) illustrates the property match example.

4.3 Information Sharing Privileges

Our model supports the read-only privilege which allows subjects to read the information in an EHR document and to navigate across EHR documents through navigation links. As identified in our example, navigation links serve as the visual representation of associations between EHR documents and such links should be appropriately protected. In particular, special protection mechanisms can be applied to restrict users' navigational behaviors by not revealing the existence of a navigation link, or by revealing and allowing a subject to explore the objects referenced by a navigation link. Therefore, two different sharing privileges are derived for the protection options, $\texttt{navi}^-$ and $\texttt{navi}^+$, respectively. By distinguishing these two protection options, it is possible to grant subjects the access permission to a particular EHR document without disclosing links to other EHR documents. For instance, by $\texttt{navi}^-$ privilege, a family physician may be aware of Henry's HIV/AIDS disease from his medical history documented in the consultation note. However, he cannot see the existence of the link to another EHR document for the details of Henry's *HIV/AIDS* treatment history since acquiring such information requires $\texttt{navi}^+$ being assigned. This feature provides another spectrum for the selection of information across composite EHRs.

4.4 Access Control Policy Specification

To summarize the above-mentioned approach, we introduce the definition of an access control policy as follows.

Definition 6 *(**Access Control Policy**). Let R be the system-wide set of roles in a healthcare system. An access control policy is a tuple $acp =< role, ao, match-mode, priv >$, where*
- *$role \in R$ is a specific role in the system;*
- *ao is an authorization object;*
- *match-mode $\in$ {exact, subset} is the match mode for object properties;*
- *$priv \in \{navi^{-}, navi^{+}\}$ is the sharing privilege for which the authorization is granted.*

The semantics of an access control policy is that, a certain *role* is only authorized with certain *priv* to share the sub-objects whose property labels match the *prop* using the specified *match-mode*. The followings are examples of access control policies and relative authorization zones created against Figure 2(b).

P1: *("billing clerk", [//Labs//*, <{*},{payment}, "code">], exact, $navi^{+}$)*;
P2: *("physician", [//Labs//*, <{general}, {treatment}, *>], subset, $navi^{-}$)*;
P3: *("lab technician", [//Labs//*, <{general}, {RHIO}, *>], subset, $navi^{-}$)*;

These policies select the same scope as the *Labs* category in the clinical consultation note. **P1** states that the billing clerk can only access to the two *Code* objects for both *CXR* and *CD4* lab tests. With **P2**, the physician can access to the CXR lab results, where the content of *CD4* lab test is hidden because of its sensitivity restriction. In **P3**, the lab technician can only access to the *CXR* lab test order with detailed instructions.

With given access control policies, the target scope and corresponding authorization zones are generated as illustrated in Figure 3. The authorization zones are created based on an algorithm as shown in Appendix. The algorithm takes the composite EHR source tree and an access control policy as inputs, and returns the authorized zone including only the authorized portion of the source tree for a given role. The algorithm first retrieves the target subtree from the source tree based on the scope specification and the navigation privilege. Then the properties of each object inside the subtree need to be matched against the authorized property specification in the access control policy, and unmatched ones are pruned from the subtree. Taken the property propagation and aggregation into consideration, the algorithm traverses the target tree in pre-order and post-order, respectively. Overall, the algorithm achieves a time complexity of $O(n)$ for traversing and pruning the target source tree.

5 Concluding Remarks

In this paper, we have presented an access control model for selectively sharing composite EHR documents. Essential features of the model are built with the logical abstraction of a composite EHR through a hierarchical structure, where internal sub-objects are distinguished and organized through their interrelationships. The design of three dimensional properties for each sub-object addresses the generic concerns for medical data sharing by enabling privacy protection and need-to-know principle for multiple data types, data relationships

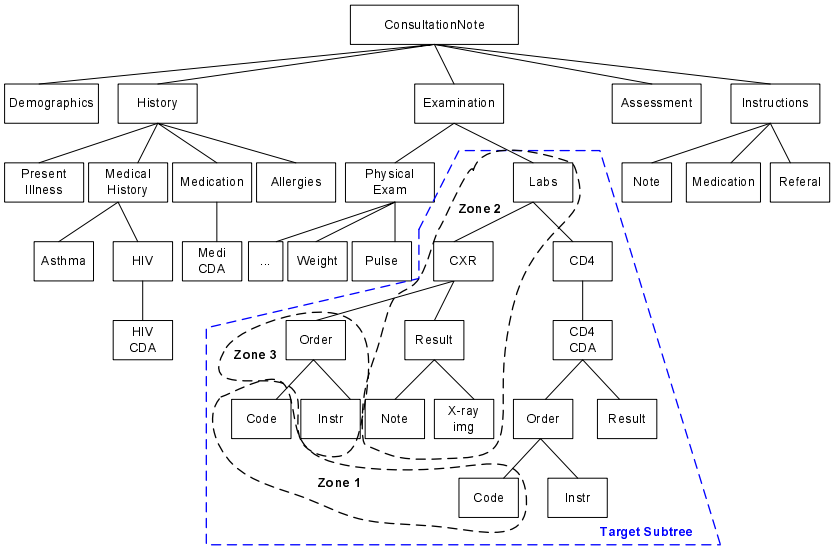


Fig. 3. Authorization Zones

and access modes. And the property-based authorization zone filtration mechanism provides a flexible yet efficient means to select and authorize a collection of sub-objects with specific property criteria.

Our future work seeks to develop a prototype policy evaluation engine based on our proposed model and standard EHR implementations. Experiments will be conducted on real healthcare systems to demonstrate the applicability and possible extension of our work. Meanwhile, performance and storage efficiency need to be measured and evaluated. Another issue concerns investigating more sophisticated authorization policies to deal with various access types in sharing composite EHRs. For example, a policy may allow a lab technician to directly submit test results to a clinic's EHR, while her access privileges on the medical record should remain intact. Finally, an effective policy propagation and enforcement scheme is necessary to maintain the control power of its original domain after an EHR is distributed and disseminated.

References

1. IEEE-USA's Medical Technology Policy Committee Interoperability Working Group, ed.: Interoperability for the National Health Information Network (NHIN). IEEE-USA EBOOKS (2006)
2. Bartschat, W., Burrington-Brown, J., Carey, S., Chen, J., Deming, S., Durkin, S.: Surveying the RHIO landscape. a description of current rhio models, with a focus on patient identification. J AHIMA **77**(1) (2007) 64A–64D
3. Dolin, R.H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F.M., Biron, P.V.: Hl7 clinical document architecture, release 2.0. ANSI Standard (2004)

4. openEHR Community: openEHR. (http://www.openehr.org)
5. HL7: Health level 7 (HL7). (http://www.hl7.org)
6. Chadwick, D.W., Mundy, D.: Policy based electronic transmission of prescriptions. In: Proceedings of the 4th International Workshop on Policyies for Distributed Systems and Networks (POLICY'03). (2003) 197–206
7. Eyers, D.M., Bacon, J., Moody, K.: OASIS role-based access control for electronic health records. In: IEE Proceedings – Software. (2006) 16–23
8. Becker, M.Y., Sewell, P.: Cassandra: flexible trust management, applied to electronic health records. In: Proceedings of IEEE 17th Computer Security Foundations Workshop. (2004) 139–154
9. Bhatti, R., Moidu, K., Ghafoor, A.: Policy-based security management for federated healthcare databases (or RHIOs). In: Proceedings of the international workshop on Healthcare information and knowledge management. (2006) 41–48
10. Ferraiolo, D., Sandhu, R., Gavrila, S., R. Kuhn, R.: Proposed NIST standard for role-based access control. ACM Transactions on Information and System Security (TISSEC) **4** (2001) 224–274
11. Fernández, E.B., Gudes, E., Song, H.: A model for evaluation and administration of security in object-oriented databases. IEEE Trans. Knowl. Data Eng. **6**(2) (1994)
12. Rabitti, F., Bertino, E., Kim, W., Woelk, D.: A model of authorization for next-generation database systems. ACM Transactions on Database Systems (TODS) **16**(1) (1991) 88–131
13. Bertino, E., Castano, S., Ferrari, E., Mesiti, M.: Specifying and enforcing access control policies for xml document sources. World Wide Web Journal **3**(3) (2000) 139–151
14. Damiani, E., di Vimercati, S.D.C., Paraboschi, S., Samarati, P.: A fine-grained access control system for XML documents. ACM Transactions on Information and System Security (TISSEC) **5**(5) (2002) 169–202
15. Gabillon, A., Bruno, E.: Regulating access to XML documents. In: Proceedings of the 15th annual working conference on Database and application security. (2001)
16. 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%20Impact%20Report_11-27-07.doc (2007)
17. Dimitropoulos, L.L.: Privacy and security solutions for interoperable health information exchange: Interim assessment of variation executive summary. http://www.rti.org/pubs/avas_execsumm.pdf (2007)
18. Clark, J., DeRose, S.: XML path language (XPath) version 1.0. World Wide Web Consortium (W3C). http://www.w3.org/TR/xpath (1999)
19. Science Applications International Corporation (SAIC): Healthcare RBAC task force charter, v1.1. http://www.va.gov/RBAC/docs/HealthcareRBACTCharterv1_1.pdf (2003)

Appendix

```
Algorithm Zone Control
  Input: C = (v_c, V_o, E_o, τ_Eo, τ_Vo)          /* C is the composite EHR source tree */
         acp = <role, ao, match-mode, priv>  /* acp is an access control policy */
  Output: Z      /* Z is the authorized zone for role including a list of nodes from the source tree*/

/* Step 1: Select the scoped subtree for evaluation */
1. LET scope = acp.ao.scp_expr     /* retrieve the scope specification from the access control policy */
2. LET Z = select (C, scope, acp.priv)          /* retrieve the subtree Z from C based on the scope and privilege spec */

/* Step 2: Traverse the subtree and match authorized properties */
3. LET prop = acp.ao.prop            /* retrieve the authorized property specification from the access control policy */
4. LET N = v_c
/* Step 2.1: Handle exact match mode */
5. IF match-mode = exact THEN  /* handle exact match mode */
6.     IF prop.ps ≠ {*} THEN     /* match sensitivity property label */
7.         WHILE preorder(N).hasnext() DO  /* traverse the subtree Z in postorder */
8.             LET ps' = τ(N).ps
9.             IF ps' ≠ prop.ps THEN
10.                remove N and all its descendant nodes from Z          /* prune unmatched nodes from the tree */
11.    ELSE IF prop.pp ≠ {*} THEN  /* match purpose of use property label */
12.        LET N = root of Z
13.        WHILE postorder(N).hasnext() DO /* traverse the remaining tree in postorder */
14.            LET pp' = τ(N).pp
15.            IF pp' ≠ prop.pp THEN
16.                remove N and all its ancestor nodes from Z       /* prune unmatched nodes from the tree */
17.    ELSE IF prop.pt ≠ * THEN    /* match object type property label */
18.        LET N = root of Z
19.        WHILE postorder(N).hasnext() DO /* traverse the remaining tree in postorder */
20.            LET pt' = τ(N).t
21.            IF pt' ∉ prop.pt THEN
22.                IF prop.pt contains "composite" THEN
23.                    remove N from Z       /* prune unmatched nodes from the tree */
24.                ELSE remove N and all its ancestor nodes from Z  /* prune unmatched nodes from the tree */
/* Step 2.2: Handle subset match mode */
25. IF match-mode=subset THEN  /* handle subset match mode */
26.    IF prop.ps ≠ {*} THEN     /* match sensitivity property label */
27.        WHILE preorder(N).hasnext() DO  /* traverse the subtree Z in postorder */
28.            LET ps' = τ(N).ps
29.            IF ps' ⊈ prop.ps THEN
30.                remove N and all its descendant nodes from Z            /* prune unmatched nodes from the tree */
31.    ELSE IF prop.pp ≠ {*} THEN  /* match purpose of use property label */
32.        LET N = root of Z
33.        WHILE postorder(N).hasnext() DO /* traverse the remaining tree in postorder */
34.            LET pp' = τ(N).pp
35.            IF pp' ⊉ prop.pp THEN
36.                remove N from Z       /* prune unmatched node from the tree */
37.    ELSE IF prop.pt ≠ * THEN    /* match object type property label */
38.        LET N = root of Z
39.        WHILE postorder(N).hasnext() DO /* traverse the remaining tree in postorder */
40.            LET pt' = τ(N).t
41.            IF pt' ∉ prop.pt THEN
42.                IF prop.pt contains "composite" THEN
43.                    remove N from Z       /* prune unmatched nodes from the tree */
44.                ELSE remove N and all its ancestor nodes from Z  /* prune unmatched nodes from the tree */
45. RETURN Z
```

Fig. 4. Zone Control Algorithm