

SECURITY MODELS FOR MEDICAL AND GENETIC INFORMATION

Eduardo B. Fernandez, María M. Larrondo Petrie, and Tami Sorgente
Florida Atlantic University
777 Glades Road, Boca Raton, FL 33191-0991, USA

ABSTRACT

In the past, medical information was physically stored in hospitals, laboratories, and doctors' offices. Access to this sensitive data was limited, and it was protected by its physical isolation and ignorance of its existence. With the digitization of medical data, this information is becoming accessible through distributed systems, including the Internet. This has increased the numbers of people that can potentially access medical information by orders of magnitude, often providing more efficient transfer of medical records and related information. Misuse of a person's medical and genetic data could potentially negatively impact his ability to be hired, and limit his career path and his insurability. Clearly medical information is one of the most sensitive types of information and requires strong security measures. We discuss the requirements and policies required for an access control model suitable for medical and genetic information. We indicate the general structure of such a model and conclude that it requires a layered structure. We then show its highest level. We use the Unified Modeling Language (UML) to model a patient record and we make it more precise by defining constraints using the Object Constraint Language (OCL).

KEYWORDS

Access Control, Genetic Data, Medical Data, Object-Oriented Modeling, Security Engineering.

1. INTRODUCTION

Medical information is one of the most sensitive types of information. Its misuse could have a very serious effect on an individual's life; leakage of information about a psychiatric treatment could ruin a career, an incorrect change in a medical record may result in a wrong prescription with damage to the patient. In past times this information was collected and stored at physicians' offices and hospitals and relatively few people even knew it existed. In most instances it was not computerized and was protected by its isolation and the ignorance of its existence. All this is fast changing, most or all of the doctor offices use computers, hospitals have large information systems, and a good part of this information is becoming accessible through distributed systems, including the Internet. This means that the number of people that can potentially access information about patients has increased by orders of magnitude.

Due to technological advances there is also more information about an individual; for example there is a whole set of genetic information, which was not available a few years ago. This information could affect a person's ability to be hired, his career path, possible promotions, salary, and continued employment.

The ease of use of the Internet for accessing patient treatment records, procedures, and lab results provides physicians with appropriate information for treatments as well as making convenient related functions such as order processing and billing for labs and pharmacies. It also may allow patients to interact with their physicians, insurance, and labs. Finally, remote access to patient records is valuable when a person gets sick while traveling. However, all this also opens a window of opportunity for misuse of this information.

When building a system that maintains private and sensitive information, security should be forefront in the analysis and design phases of development. We need new access control models that can describe the

specific requirements of medical and genetic information systems. Most of the systems built until now use ad hoc solutions that cannot assure security. Such a model should include several architectural levels. We have developed the first stage of a model that can satisfy those requirements. We discuss the requirements and their effect on the model in the next section, followed by a description of the model. We end with some conclusions.

2. REQUIREMENT AND POLICIES FOR THE MODEL

An access control model suitable for medical records must implement general security policies as well as more specific policies oriented to this type of application. The general security policies that apply to these models, interpreted in this context are:

- It is necessary to apply a “need to know” policy, providing only the information the authorized medical users need for their work and no more.
- Access for the users of this system should be defined by their roles but individual access must also be defined.
- There is a strong emphasis on privacy, which implies a large amount of control by the people about whom we keep information.
- The system should be a closed system, where the lack of an authorization rule implies no access.

To these we can add more specific policies for medical information including:

- Different types of roles have specific access constraints, e.g., patients can see their records and doctors can modify their patients’ information.
- Patients give their consent to the use of their records and have the right to be notified of their actual use.
- A doctor or other clinician serves as the record custodian and is responsible for the use of the patient record.
- Rights may need to be overridden in exceptional situations.
- Rights may need to be delegated for expediency, e.g., a doctor on vacation..
- Records of patients with genetic or infectious diseases need to be linked to records of their relatives or persons with whom they had contact..
- Each patient has one or more medical records, normally distributed, which together can be seen as one logical record.
- For research or statistical purposes there must be aggregate types of access that do not reveal an individuals’ personal data.

Even more specific policies can be defined; for example, for mental health or infectious diseases, etc. To these policies we must add that the context is a loosely-coupled distributed system, including local area networks as well as the Internet, with applications where records must be frequently exchanged.

General security models fall short of what is needed. From the policies and from the environment where this information is kept we can deduce some requirements for the security model:

Attribute and credential-based authorization—In an environment where not all the users that may need access to a document are known in advance, we need to have authorization models that can consider user attributes and credentials to determine access rights.

Content-dependent authorization—The granularity of access should be to the record level to separate individual information.

Context-dependent access modes—There are occasions where the standard predefined authorization must be overridden. For example, if a patient is unconscious and needs immediate attention, it is possible that the authorized users of her record may not be present and someone must access the record to decide about

treatment. Patients may move around different units of the hospital for tests or treatments and authorization should depend on their location.

Delegation of rights—Any authorization model must contain policies on how the rights of a subject are delegated to other subjects. This is especially important in models where privacy is a major objective.

Administration of security — We need to have traditional security administrators that define roles, assign users to roles, create groups, and perform similar global functions. We also need a special type of administrator for each individual record (custodian), a clinician assigned perhaps by the patient, who is in charge of the specific record. Due to the characteristics of this data, the users or their rights change frequently.

Temporal restrictions—A doctor may have access to a record when the patient is under treatment by her but not after the treatment is finished. He may still retain access to her treatment data but if the patient changes doctors she does not have access to the new additions to the record. This means that the model must be able to apply temporal restrictions.

Multimedia objects—Medical records are a combination of text (medicines, treatments, annotations), audio (dictation), and images (X-rays, CAT scans, ultrasound images), as well as other documents related by hypertext links. The model must then include as protection objects all the aspects associated with a record as well as its links to related documents.

Representation of documents using Extensible Markup Language (XML)s— For convenience and flexibility we think that patient records should be embodied in XML documents; in fact, we can see a clear tendency towards this direction. We can access these documents in two ways: using Remote Procedure Calls (RPC) style to apply specific operations, and document style, where the unit of access is a whole document, and portions of the document may be controlled independently. Any security restrictions defined by medical policies should be reflected in the documents.

Inference control—Access to some information could allow one to infer other aspects and we need to control at least basic inferential associations.

Need for coordinated authentication and encryption — The access control model should be tightly coupled to other aspects such as the authentication model and rely on it. The model should also have a way to indicate when a record or a part of a record should be encrypted for transmission or possible storage.

Consideration of web standards—There is a variety of standards that apply to web services and Internet access. These include standards such as WS-Security, SAML, and others. The model should be consistent with these standards in order to be of practical guidance for real systems.

Consideration of different architectural levels—Enforcement of the security model requires the consideration of all architectural levels. The model must indicate how the enforcement mechanisms at these levels relate to each other.

Compliance with laws protecting security and privacy of health care information—The Health Insurance Portability and Accountability Act (HIPAA) Public Law 104-191, enacted in 1996 by the U.S. Department of Health and Human Services, established standards and requirements for the maintenance and transmission of health care information to protect the security and confidentiality of electronic patient health information and cut the cost of print-based record transactions. It requires ensuring integrity and confidentiality at all stages of transmission and storage of health care information.

Explicit audit—The model should make explicit those aspects that need to be logged for future audit. Audit is particularly important when we have context-dependent authorization because of the possible legal implications of overriding or adding authorizations.

It is clear that no single model can satisfy all these requirements. We need several related models at different abstraction levels. We show here a model that satisfies part of these requirements and that is part of a set of models at different levels of abstraction that can cover all the requirements.

3. MEDICAL INFORMATION

Medical information is collected from the moment a person is born until her death (maybe before she is born in sonogram images and after her death in autopsy records). Some of the actors involved in handling medical information include health institutions (hospital, clinics), physicians' offices, insurers, research facilities and regulating bodies (Department of Health and Human Services (DHHS), in the US). At the present time there is not one medical record for each individual that is kept in some sort of central registry. Each consortium or clinician keeps their own record for each patient and information is passed between groups or individuals through referrals and discharge letters [Anderson, 1996].

Medical information typically includes aspects of a person's physical health such as treatments, medicines and diagnoses. In addition to this health information, a medical record may also include information about substance abuse, sexual behavior, family relationships, and private thoughts expressed through psychotherapy. This is very private information that is often keyed to a social security number. With an increasing aged population, information about long-term care has joined the typical clinical information about patients. There is a lack of consistent privacy protection in using Social Security Numbers [EPC, 2002]. Information about infections and genetic diseases may include additional aspects such as race, ethnicity, habits, etc. These two types of information also require linking between records, e.g., family trees.

While several authorization models have been proposed for general use, few models are specifically intended to represent access constraints in medical environments. One of the earliest discussions of unique security needs for medical systems is a paper by T.C.Ting discussing the requirements of mental health security [Ting, 1990]. J. Biskup did some significant work on privacy aspects of medical systems [Biskup, 1990]. G. Pangalos developed several design models for medical databases [Pangalos, 1994]. R. Anderson did a systematic work of identifying policies for general clinical records [Anderson, 1996]. None of these models tries to find a model to describe a variety of policies. Anderson's work comes closest to our ideas but he did not try to unify his study of policies into a general model or to formalize the policies.

There are also several studies by medical informatics researchers on the issues and requirements of patient records, including some actual implementations [Chalmers, 2003; Denley, 1999; Dugas., 2001; Rossig, 1995]. Their studies are valuable to understand experience in implementing and using medical information but they do not attempt to develop new security approaches but mostly apply known approaches to their specific needs. For example, the SEISMED project of the European Union relied completely on cryptography [Rossig, 1995].

4. AN ACCESS CONTROL MODEL FOR MEDICAL INFORMATION

We use a hybrid model that combines the access matrix and RBAC models to include aspects such as those discussed above. This model is represented using object-oriented diagrams, where authorizations are superimposed on the medical information. We integrate semi-formal and formal specification techniques, combining the Unified Modeling Language (UML) [Rumbaugh, 1999] with OCL (Object Constraint Language) [Warmer, 1999], to produce an unambiguous model that is easy to understand. We are also developing a secure methodology to build and configure this type of system. For this, we are adapting and extending our secure systems methodology [Femandez, 2004] to suit these types of applications. This methodology makes heavy use of patterns and uses all the architectural levels of the system. The results are being tested on real medical environments, including a hospital and a medical laboratory, having the models and scenarios checked by doctors and nurses.

When building operating systems and other system software, it has long been agreed that security must be an integral part of the design, never an add-on feature or patched as an afterthought. What is not so obvious is that the same principle is valid for general applications, although some authors [Ting, 1990] have indicated this need. Another important aspect is that many models express well security constraints but do not indicate how to enforce these constraints. We believe it is important to provide an abstract architecture to enforce the constraints. Our approach is to develop the appropriate requirements and policies, define an object-oriented model for the security policies, identify patterns in the model, develop an abstract implementation, test the model and implementation, use the model to create a protection profile, and build a prototype to validate some aspects of the models.

Our work has identified some basic aspects for a model that satisfies the requirements just discussed:

Use of object-oriented models—Health care records and genetic data contain information that has complex relationships with other information. The object-oriented approach enables us to capture these complex associations in a visual and intuitive manner. Use cases allow us to define role-based access requirements.

Use of formal constraints — Using the OCL language we add formality, thus reducing ambiguity while retaining the understandability of the model. The OCL also allows the expression of complex time-sensitive or context-sensitive access rights.

Consideration of static and dynamic aspects—An object oriented approach includes two types of models: A static model, normally a class diagram, that describes the information and a dynamic model composed of state, collaboration, and activity diagrams, This means that we can describe in our model static and dynamic aspects, including information, states, collaborations, and workflows. In particular, collaboration diagrams are useful to understand a system and to explain the functions to a lay audience.

Use of different levels of abstraction—The requirements of the model imply mostly application-oriented aspects that reflect patient policies. However, some requirements are about architectural aspects, e.g., access control of XML documents. We can build several models at different levels of abstraction, including middleware, DBMS, and operating system aspects[Fernandez, 2002].

Use of patterns—Specific combinations of policies can be defined as patterns. A pattern is a recurrent submodel that describes a solution to a specific problem in a given context [Gamma, 1994]. Complex policies can be expressed by combining patterns and adding ad hoc parts.

Use of implied authorization—Patient records are logically aggregates of a variety of information, including aspects such as treatments, medications, visits, schedules, and annotations. We have developed policies for the propagation of authorization rules along aggregation hierarchies [Larrondo-Petrie, 1990]; we are extending those policies to consider the specific aspects and constraints of this model.

Content-dependent authorization — We are investigating two possibilities:

- Extended roles, where the role rights are filtered by content-dependent predicates.
- Attribute-based models, where access depends not only on the subject but also on the satisfaction of assertions including attribute values.

We have started work on the second approach, defining a pattern for access control based on metadata [Pribe 2004]. This will be used in conjunction with the model proposed here.

Context-dependent authorization—We are combining our authorization models for databases from [Fernandez, 1994] with state and workflow conditions, such as those described by BPEL4Ws and ebXML. Today's environment of terrorist threats, in particular germ warfare, requires tracking of certain medical profiles and at times quickly locating individuals for containment of diseases. This threat may require certain authorities to have unlimited access to medical data and patient identity. The model would limit this type of authorization to confirmed threats or confirmed life-threatening situations. The need to override individual consent needs to be linked to an authorized, ascertained emergency situation.

Emphasis on privacy—This requires support for specific policies, such as patient control over their records. Another important aspect is to take into account the provisions of the Policies for Privacy Preferences (P3P) [W3C].

Evaluation of security requests—When a user sends a request to access information there must be a way to validate the request and decide what information (if any) will be returned to the user. The request is handled by an abstract reference monitor. The evaluation algorithm must consider both explicit and implicit rules and be reasonably efficient. Starting from a reference monitor pattern [Fernandez, 2002], we have extended the evaluation algorithms of [Fernandez, 1994] to this kind of models.

Figure 1 shows a pattern for the RBAC model of [Fernandez, 2001], tailored for use with medical information. This RBAC can be used to define precise access rights to these roles according to a need-to-know policy [Fernandez, 1997]. A role corresponds to a job or function within a job, or an individual. Rights are assigned to roles, not to each individual. Figure 1 shows how the model can represent some of the policies, including:

- A **Patient** role that has the rights to read his own record and authorize the use of this record. Rights are represented in the model as association classes.
- A **Doctor** role showing that a given doctor may act as custodian for a patient record.
- The **Medical Record**, that includes the constraint that any reading of a record must be notified to the corresponding patient.
- Specific medical records may be associated (linked) with other records to describe, for example, family relationships, physical contact, etc.

Because there are a large variety of policy combinations, specific class/association combinations can be described and catalogued in the form of patterns. For example, we could have several models like the one of Figure 1 to represent different combinations of policies. These patterns can be combined to describe more complex sets of policies. They can also be combined with other related aspects such as billing and others. In fact, some of this pattern comes from another pattern, the Patient Treatment pattern of [Sorgente and Fernandez, 2004]. Figure 1 shows also the addition of OCL constraints for greater precision.

We are building a catalog of atomic medical security patterns as well as combinations of patterns. We have already developed some patterns for secure patient treatment [Sorgente and Fernandez, 2004]. For the lower levels, security patterns can be used to define the architecture of the enforcement structure. We have produced several patterns for secure system design, including patterns for authorization models [Fernandez, 2001], for operating systems security [Fernandez, 2002], for remote authentication [Fernandez, 2003], and for firewalls [Fernandez, 2003]. We need to add patterns for secure distribution, e.g., a secure Broker.

5. CONCLUSIONS AND FUTURE WORK

We have found that using UML/OCL we can represent complex combinations of medical policies in a precise way that is also convenient for implementation. Patterns for specific combinations of policies, for example for HIPAA use, can be built and catalogued to be used when building secure systems that are HIPAA-compliant. Future work includes expanding this model to include more policy combinations. We also need to develop the lower level models corresponding to distributed architectures that implement and enforce the model. Security is a multilayer problem, one cannot secure just one architectural layer. The lower layers are needed to enforce the application model constraints. Without the enforcement aspect there is not much chance that the model will be used in practice.

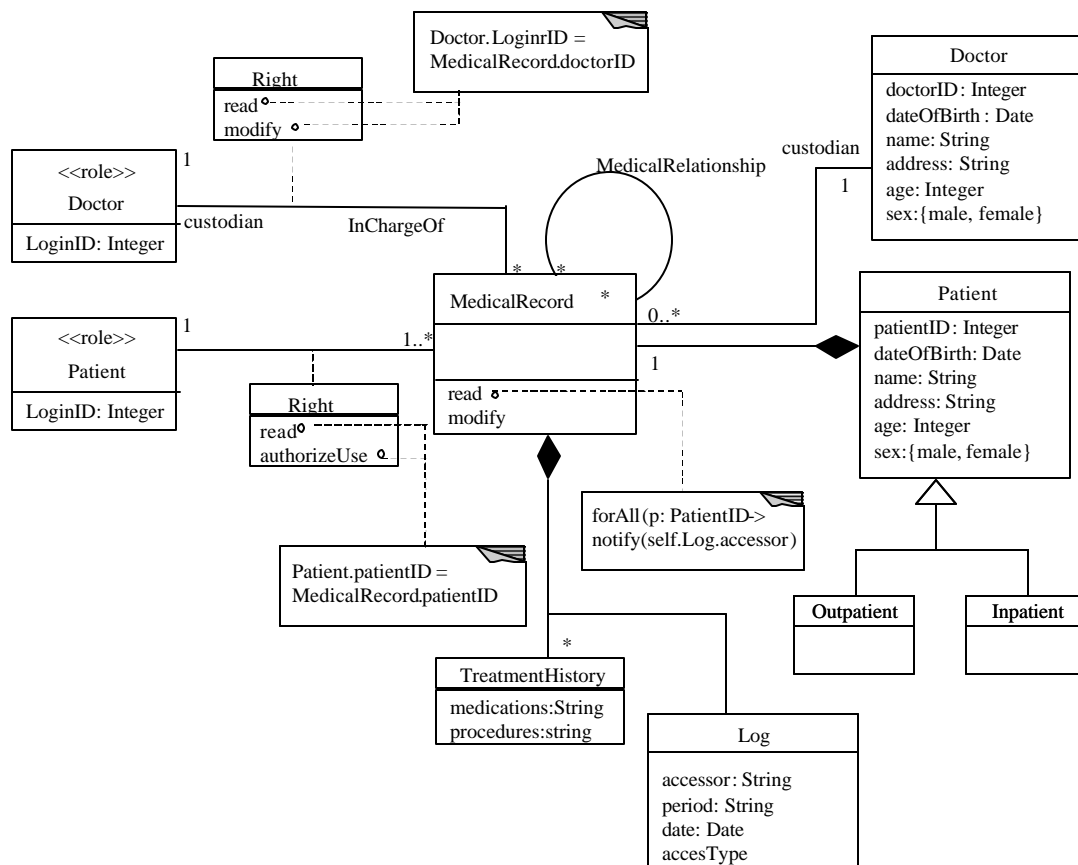


Figure 1. A model for a patient record

REFERENCES

- Anderson, R., 1996. Security in Clinical Information Systems. Computer Laboratory, U niv. of Cambridge, Version 1.1.
- Biskup, J., 1990. Protection of privacy and confidentiality in medical information systems: Problems and guidelines. In *Database Security III, Status and Prospects*, D.L. Spooner and C. Landwehr (Eds.), Elsevier Science Publishers B.V., IFIP, pp. 13-23.
- Chalmers, J. and R. Muir, 2003. Patient privacy and confidentiality. *British Medical Journal*, Vol. 326, pp. 725-726. <http://bmj.bmjournals.com/cgi/content/full/326/7392/725>
- Denley, I. and S. W. Smith, 1999. Privacy in clinical information systems in secondary care. *British Medical Journal*, Vol. 317, No. 7194, pp. 1328-1331. <http://bmj.bmjournals.com/cgi/content/short/318/7194/1328>
- Dugas, M. et al., 2001. Impact of integrating clinical and genomic information. University of Munich, <http://www.bioinfo.de/isb/gcb01/talks/dugas/main.html>
- EPC, 2002. 2002. Medical Privacy. Electronic Privacy Information Center. <http://www.epic.org/privacy/medical>

- Fernandez, E. B., E. Gudes, and H. Song, 1994. A model for evaluation and administration of security in object-oriented databases *IEEE Transactions on Knowledge and Database Engineering*, Vol. 6, No. 2, pp. 275-292.
- Fernandez, E. B. and J.C. Hawkins, 1997. Determining Role Rights from Use Cases. In *Procs. 2nd ACM Workshop on Role-Based Access Control*, ACM, 121-125. <http://www.cse.fau.edu/~ed/RBAC.pdf>
- Fernandez, E. B. and R. Pan, 2001. A Pattern Language for security models. In *Procs. of PLoP 2001*. http://jerry.cs.uiuc.edu/~plop/plop2001/accepted_submissions
- Fernandez, E. B., 2002. Patterns for operating systems access control, *Procs. of PLoP 2002*. <http://jerry.cs.uiuc.edu/~plop/plop2002/proceedings.html>
- Fernandez, E. B. and R. Warriar, 2003. Remote Authenticator/Authorizer, *Procs. of PLoP 2003*.
- Fernandez, E. B., M. M. Larrondo-Petrie, N. Seliya, N. Delessy, and A. Herzberg, ., 2003. A pattern language for firewalls, *Procs. of PLoP 2003*.
- Fernandez, E. B., 2004. A methodology for secure software design, submitted for publication.
- Gamma, E. R. Helm, R. Johnson, J. Vlissides, 1994. *Design Patterns: Elements of Reusable Object-Oriented Software*, Addison-Wesley, Boston, Mass.
- HIPPA. <http://www.hipaa.org/>
- Larrondo-Petrie, M. M. E. Gudes, H. Song, E. B. Fernandez, 1990. Security Policies in Object-Oriented Databases. In *Database Security III: Status and Prospectus*, D.L. Spooner and C. Landwehr (Eds.), Elsevier Science Publishers (North-Holland), pp. 257-268.
- Pangalos, G. A. Pomportsis, L. Bozios, and M. Khair., 1994. Development of secure medical database systems. In *Procs. of DEXA'94*, pp. 680-689.
- Priebe, T. E.B.Fernandez, J.I.Mehlau, and G. Pernul, 2004. A pattern system for access control. To appear in *Procs. of the 18th IFIP WG 11.3 Conference on Data and Applications Security*, Sitges, Spain, July 2004.
- Rossig, N., 1995. SEISMED, A Secure Environment for Information Systems in Medicine. Presentation note of the *Programme AIM (Advanced Informatics in Medicine) of DGXIII*, Brussels.
- Rumbaugh, J. et al., 1999. *The Unified Modeling Language Reference Manual*. Addison-Wesley, Boston, Mass.
- Sorgente, T. and E.B.Fernandez, 2004. Analysis patterns for patient treatment, January 2004.
- Ting, T.C., 1990. Application information security semantics: A case of mental health delivery. In *Database Security III, Status and Prospects*, D.L. Spooner and C. Landwehr (Eds.), Elsevier Science Publishers B.V., IFIP 1990, 1-12.
- W3C, <http://www.w3.org>
- Warmer, J. and A. Kleppe, 1999. OCL: The Constraint Language of the UML. *JOOP* (May 1999).