A Tag-Based Data Model for Privacy-Preserving Medical Applications

Surya Nepal, John Zic, Frederic Jaccard and Gregoire Kraehenbuehl

CSIRO ICT Centre PO Box 76, Epping NSW 1710 Australia[a]

{Surya.Nepal,John.Zic,Frederic.Jaccard,Gregoire.Kraehenbuehl}
@csiro.au

Abstract. In autonomous distributed healthcare environments, patients' electronic medical records are controlled and managed by each healthcare facility. It is important to ensure that when records are accessed and transferred that it is done securely, while still respecting patients' rights on privacy and confidentiality of their personal health information. We propose a new tag-based data model for representing patients' electronic medical records as well as access and transfer policy statements. This model helps to categorize the patient information, as well as expressing patients' consent for a variety of domains (individual, health care provider and facility). Unlike most existing data models used in healthcare information systems, our model supports patients' consent expression in terms of healthcare facilities, healthcare providers, their roles, and categories of medical records or any combination of them within a single framework. Our model has been demonstrated by developing a prototype system using some trusted computing components.

1 Introduction

The coordination of individual's health care relies on the sharing of personal health information among healthcare providers such as local clinics, test laboratories and hospitals. It is well known that there are potential benefits and risks associated with sharing patients' electronic medical records [3]. One of the risks is patient's loss of privacy and confidentiality, where patients may not want to share or transfer their personal health information without their knowledge, and retain the rights to both access and transfer of this information.

The effective usage of personal health information systems is hard to achieve without addressing the patients' privacy and confidentiality concerns [2]. Different models have been proposed and demonstrated to address their concerns. An eConsent model has been developed and demonstrated in [1]. The model proposed a novel, privacy-preserving anonymous transfer protocols based on the concept of 'placeholders'.

[a] This work is completed as part of CeNTIE project that is supported by the Australian Government through the Advanced Networks Program of the Department of Communications, Information Technology and the Arts.

However, the model was based on a number of assumptions that represent a subset of real world application such as medical records are organized in nested structure to resolve the conflicts in policies [1].

This paper offers an alternate way to organize medical records and express access and transfer policies. Our approach, which we call tag-based model, extends the eConsent model so as to address the weaknesses in the current eConsent model.

In our model, an electronic medical record has a number of policy tags associated with them, which we call *eTags*. Each of these tags has two fields: *category* and *policy*. The category field categorizes records into different groups such as heart, head and AIDS. The policy field, which we call *eCo* (electronic consent), consists of rights expressions. Unlike RBAC [6] and the eConsent model [1], our approach allows definition of permission for both transfer and access in terms of (a) roles, (b) healthcare Facilities, (c) healthcare providers, and (d) categories of information.

The following summarizes the key characteristics of our tag-based model.

- **Default policy expression**: each healthcare facility, patient and their families may have different policies for different categories of medical records. Our model supports default policies for patients, their families and healthcare facilities. For example, a patient can define a default policy for his AIDS related record so that all of his AIDS related records are subjected to this default policy. A facility can define its own default policy for AIDS related records, where AIDS related records of all patients in the facility are subjected to this policy. Similarly, a patient's family can define default policies for all the members of their family.
- **Access and transfer policy**: the policy expression mechanism allows the specification of both inter- and intra-facility access and transfer rights expressions. This enables, for example, a patient to deny specific healthcare facilities for receiving their personal health information.
- **Uniform model**: our model used eTags for both categorization and policy expressions, and allows us to define a set of policies for different categories of information.
- **Flat model**: there is no nested and hierarchical structure in information representation, and provides flexible way of representing information and defining policies.
- **Categorization into multiple groups**: in our model, electronic medical records are categorized according to a common, well-defined medical ontology. For example, a prescription related to "headache" that has side effect on heart can be categorized into three different groups – heart, head and prescription- by attaching their respective eTags.
- **Prioritized conflict resolution**: the model has an underlying priority-based conflict resolution mechanism for resolving policy conflicts between varies eCos..
- **General policy expression**: the model extends the usual role-based policy expression to allow policy expressions in terms of healthcare facilities, healthcare providers and categories of information. For example, a policy expression such as "grant access to *AIDS related records* to *Dr. Smith* while working as a *heart specialist* in *North Ryde Medical Center*" is possible in our model.

The rest of the paper is organized as follows. In Section 2, we describe a motivating example and identify some of the privacy and confidentialty problems that need to be solved. The flat data model is described in Section 3. Section 4 briefly describes a prototype implemented in a trusted environment. Section 5 presents the related work and the last section presents the concluding remarks and the future work.

2 A Motivating Scenario

We consider an example distributed healthcare environment that includes two healthcare facilities: Western Sydney Hospital and North Shore Hospital. Each hospital is autonomous and has its own medical information systems. However, as is often required, these hospitals share patients' personal health information in order to provide effective services. For example, Western Sydney Hospital may move a patient to North Shore Hospital and transfer all their medical records with them.

Each hospital has its own set of privacy and confidentiality policies for patient records held within their medical information systems. For example, a doctor working in the emergency department is permitted to have access to all medical records of a patient admitted to an emergency ward. Though the basic policies are setup by government rules and legislation, each hospital may implement these policies differently. That is, each hospital may have different set of policies for different categories of information. For example, Western Sydney Hospital may have different set of policies for AIDS related records to that of North Shore hospital. Any data model must support a variety of *default access policies for hospitals* so that all medical records in the hospital are subjected to these policies.

It is a fundamental assumption that a patient owns their personal health information. A patient may have policies that differ from those of a particular hospital. For example, a patient can define a policy that all AIDS related records are accessible to their doctor, and no other doctor in the hospital can access it except in case of emergency. Similarly, a patient's family may define their own policies for family members. For example, only family doctors may be allowed to access immunization records of the family members. Any data model developed and used must support the definition of *default family and patient policies*.

The discussion so far has been on specific policies of hospitals, patients' families and patients. As we mentioned earlier, a patient is an ultimate owner of his/her health information and thus must be able to define *policies for individual electronic medical records* independently. For example, a patient's AIDS related records policy may grant access to only family doctors. Of course, the patient may require that a particular blood test to be examined by AIDS specialist. Furthermore, the patient may not want to disclose the AIDS related records to the family doctor even though the family has defined a policy that all records of family members are accessible to the family doctor.

The hospitals may share health information with each other to provide better services. Similar to access policies, *transfer policies* are also defined at hospital, family, patient and record levels. The hospital may define its own transfer policy to another

facility, but any information sharing is only possible if the patient allows it in their transfer policy

Transfer and access policies may be defined at on entities such as hospitals, patients, families and medical records. However, some policies may also need to be defined on for a group of medical records such as AIDS related records. This means medical records need to be categorized so that it is possible to define policies *for a set of records in a category*.

Hospital default policies are normally expressed on *Roles* within the facility. For example, Western Sydney Hospital may have policy that a *doctor* or *nurse* in an *emergency doctor* role can access all information. However, patient policies are less likely to depend on roles, and rather express *policies in terms of individual doctors*, such as "grant access to AIDS related records to Dr. Smith".

In order to model the above scenario, the data model must be able to express and support (a) agreed upon information categorization such as AIDS and Heart, (b) access and transfer policy rights expression, (c) default policy expressions for hospitals, patients and family, (d) default policy expression for different categories of information, and (e) policy expressions for both indentified individuals and roles.

2.1 Security features

The discussion so far has presented some of been the characteristics of the data model. This section briefly identifies the security requirements that are necessary to meet the patient's privacy and confidentiality requirements on the access and transfer of their personal health information. They include:

- no loss of privacy and confidentiality for the patient;
- medical records should be accessible to only those providers who need to know;
- access should be limited to those portions of medical records that pertain to the provider's role;
- a log or audit trail must be maintained about all access to any part of the medical records;
- anonymity must be maintained if the medical record is published into the public domain for research purpose;
- the release/transfer of data needs patients' authorization;
- any confidential data must carry the confidentiality information;
- the medical records transfer to other side must be protected;
- a secure transfer mechanism must be established;

These requirements and their impact on the required protocols to implement these requirements are further discussed in a forthcoming paper.

3 The Tag-Based Data Model

This section presents our tag-based data model and how it provides the health information requirements discussed in Section 2.

An electronic record (*eRec*) is our fundamental unit of information. An eRec could be a diagnostic report, X-ray image, or prescription as shown in Fig 1. Each eRec may have an arbitrary number of electronic tags (*eTags*) attached to it.

Each eTag has two fields: *category* and *policy expression* (as shown in Fig. 1) that we refer to as an *eCo* (electronic consent). The *category* field is a tuple, consisting of *type of category* and its associated *value*. The *type* determines whether the category is related to patient, family or facility. The *value* provides the categorization information within the category type. The *eCo* field consists of a set of *access and transfer rights policies* for the category.

Finally, each policy has a timestamp representing the time of creation of the policy. This is used for conflict resolution (as shown later) as well as for audit purposes.

Fig. 1 eTag Structure

Each eTag can be placed into one of three broad classes: (a) a special NULL category with a non-empty set of policies, used to express *record-specific policies*. (b) An eTag with a NULL policy but with category information, used for *categorization* purposes only. (c) An eTag with both category information and a set of policies that groups the set of records and specifies group-related policies.

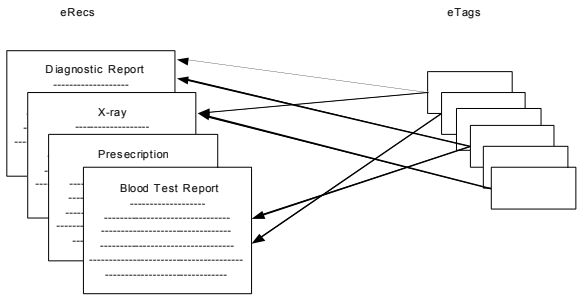

Fig. 2 eRecs and eTags in the data model

Our model allows eTags and eRecs to be related via a many-to-many relation as shown in **Fig. 2**. This allows an eRec to be categorized into multiple groups by attaching multiple eTags onto the eRec. Further, access and transfer rights are defined by each eTag, allowing complex access and transfer relationships to be defined and enforced.

The information held in the eTag catagory is also sensitive, in that poorly designed systems may result in accidental violation of privacy and confidentiality requirements. Our eTags have an eCo that applies to both an eRec and to category information. This

means a medical practitioner will not have access to the eTags if the eCo defined in the eTag denies access to the practitioner.

Our eCo expresses both the transfer and access rights of a particular eRec. One could define an eCo access rights using standard policy languages such as XACML [12] , or EPAL [13]. However, most of the current policy expression languages are primarily used for expressing access policies. Languages that may by used to express *transfer* policies are a recent development, such as those that came out of the Family Domain effort within Motorola [14] and are now in OMA 2.0 [15]. These languages are very rich and allow generalised application to any DRM application. However, we did not require their full capabilities for this application, and so we defined our own, application-specific rights policy expression language, described in the following section.

Note that we have omitted any formal description of the model due to the limited space available in this paper.

3.1 Policy Expression Language

Fig. 3 shows our rights policy expression language in BNF style.

```
policy :=
  policy_TRANSFER | policy_ACCESS
policy_TRANSFER :=
  (‘grant’ | ‘deny’) ‘transfer to’ (FACILITY)
policy_ACCESS:=
  policy_ACCESS_GRANT | policy_ACCESS_DENY
policy_ACCESS_GRANT:=
  ‘grant’ ‘access to’
  (((PRINCIPAL | ROLE)[‘with append right’]) | FACILITY)
policy_ACCESS_DENY :=
  ‘deny’ ‘access to’ (PRINCIPAL | ROLE | FACILITY)
```

Fig. 3 Policy Expression Language

The transfer policy grants or denies transfer to a certain facility (or hospital). The access policy grants or denies access to health practitioners (or principal), their roles or facilities. The access permission can be granted with append rights for healthcare practitioners or their roles defined in the facility. A doctor can access an electronic record, if and only if the access policies: (a) allow access to the subject or to the Role and (b) allow access to the Facility. That is, both the facility and practitioners need to have access permission in order to access the medical records. Similarly, a doctor can transfer an electronic record if and only if (a) the access policies allow access to the subject and his facility and (b) transfer policies grant transfer permission to the destination facility.

3.2 Policy enforcement and conflict resolution mechanism

In order to determine whether a principal can access an eRec or not, all the policies listed in the eCo of all eTags attached to the eRec must be evaluated. However, the policies defined in different eTags may conflict each other. To address this, we first define a policy priority, based on the following eTag type partial order:

$$\text{Facility} \subseteq \text{Family} \subseteq \text{Patient} \subseteq \text{NULL} \qquad (1)$$

We then use this order to define a *prioritized multi-step conflict resolution mechanism*, as follows.

1. **Higher priority wins:** A simple check is first performed according the order presented in Equation (1), above. Unfortunately, this first step cannot resolve all conflicts, since a single eCo may contain multiple policies. Similarly, an eRec can have multiple eTags from the same level of priority.
2. **More specific wins:** If eTags have identical priorities, we resolve the conflict using a principal of "more specific wins". For example, a policy concerning a doctor wins over a policy concerning a role because a doctor is more specific than the role. If this rule fails to resolve the conflict, we move onto the following step.
3. **Most recent wins**: At this point, we resolve identical priorities and specifities by use of the timestamp and a rule where "most recent wins".
4. **"Deny" wins over "grant"**: Should all the rules fail up to this point, the policy with "deny" wins over the policy with "grant".

4 Prototype Implementation

We implemented a demonstration system running on four PCs. Fig. 4 depicts the overall architecture of our "MedicClient/Server" system. The system has six major components:

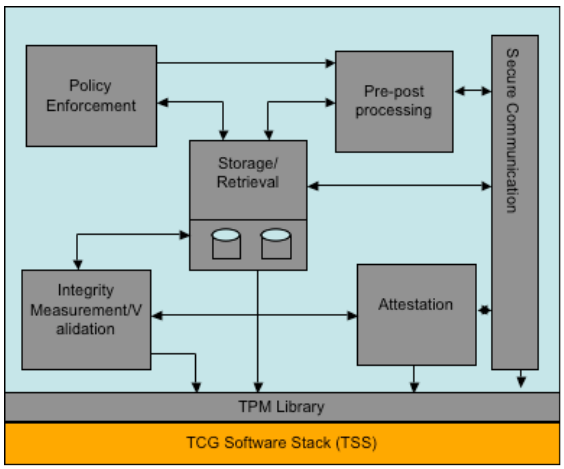

Fig. 4 MedicClient/Server Architecture

1. **Policy enforcement**: responsible for enforcing policies and resolving conflicts while accessing and transferring medical records. It is also responsible for generating and maintaining of audit logs.
2. **Integrity measurement/validation**: is used to measure the current environment of the computer where MedicServer is running and verifies that the measurements sent by other facility are as expected and so can be trusted.
3. **Secure communication**: encrypts the outgoing information and decrypts the incoming information.
4. **Attestation**: is used to determine the identity of the remote facility
5. **Pre/post processing**: processes the eTag for transfer.
6. **Storage and Retrieval**: is responsible for storing and retrieving an eRec from the SQL databases.

We consider only two components: storage and retrieval, and policy enforcement and monitoring since the scope of this paper is the data model. We present the trusted computing components (integrity measurement and validation, attestation and secure protocols) in a forthcoming paper.

Storage and retrieval: Electronic medical records and all other data needed for the functioning of components are stored in a persistent, securely encrypted store implemented using SQL server and ADO.NET classes to connect data sources and to retrieve and update stored data.

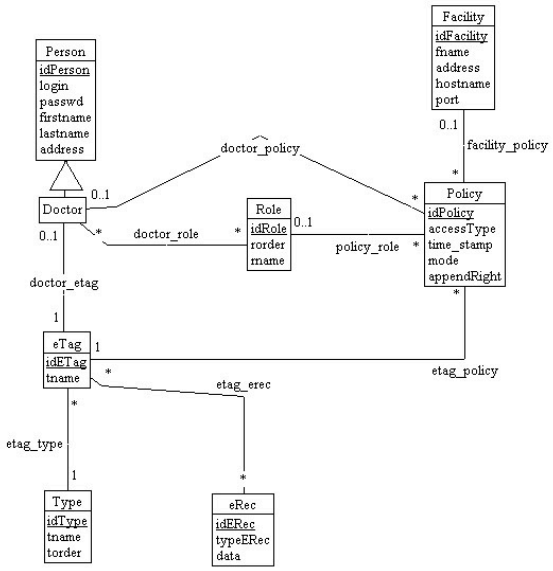

Fig. 5 E-R diagram for the tag-based model

Fig. 5 shows the relationships between the various entities in our tag-based model. It can be seen from the figure that a policy is defined on the roles and facilities as well as individual doctors. It is worthwhile noting that the doctor eTag is a NULL policy,

as it does not have any relationship with the policy entities. The entities and their relationships were used to create the SQL tables held on the server, and used trusted computing technology to ensure the privacy and security of the SQL data.

Guided by the requirement that the patient's medical record privacy must be protected, we decided to encrypt the information in an eRec table entry. This is because each table entry contains a patient's medical record. This decision then allowed a simple database implementation, as only the eRec table entry data needs to be secured. Decryption of the information in an eRec could be done without using another access to the database.

The eRec data is encrypted using Triple DES, and requires the use of a symmetric key created at the time of installation of our system. An asymmetric key is created and registered at the same time and is used for the purposes of *sealing* the symmetric key so as to prevent unauthorized access to the symmetric key. *Sealing* can be simply described as an encryption function that allows only the hardware device and its specific software environment that created the object to decrypt it. The implemented mechanisms rely on the support of the Trusted Platform Modules (TPM) and Trusted Software Stack (TSS) library, and are not discussed any further in this paper. These will be presented in a subsequent paper.

Policy enforcement and monitoring: Again, due to space constrains, we only briefly describe only one of the many implemented functions of this component, namely the access policy enforcement. The access policy enforcement mechanism consists of two steps. The first checks whether the doctor is allowed to access the information or not. The second step checks whether the facility where the doctor is trying to access the information is allowed to access it or not. For both steps, all related policies for an eRec are collected from eTags in a list and checked against the conflict resolution mechanism described in Section 3.2 to see whether the winner mode is "grant" or "deny". If both facility and doctor lists come out with the winner mode as "grant", then the access is granted to the doctor in the facility for the eRec. In all other cases, access is denied.

5 Related Work

Heath services can be improved significantly by sharing patient information, but this needs to be balanced with patient's privacy and confidentiality requirements. The electronic medical record systems enable the easy sharing and distribution of patient information. However, the disclosure of a patient's medical records without his/her permission is prohibited. In this section, we first discuss the related work in health informatics in general and then discuss the work closely related to our proposed model.

Huston [8] discusses the general security concerns on implementing e-medical records and technological and administrative tools available for safeguarding the e-medical records. Stein [7] discusses the different scenarios of electronic medical records and highlights the threats and promises. Reliability, Accountability and Privacy are considered as threats, whereas consistency, flexibility, availability, and quality are considered as promises.

One of the major privacy concerns is secure transfer of electronic medical records from one service provider to other. Task force on medical informatics [3] discusses some issues related to transfer of medical records. Chadwick and Mundy [2] look at the security requirements for electronic transfer of prescriptions from the perspectives of confidentiality, integrity and availability. It analyzed the four different transfer models published in UK: Transcript Consortium Model, Pharmacy 2U Consortium Model, SchlumbergerSema Consortium Model, and University of Salford Model.

Evered and Bogeholz [5] present a case study of the access control requirements for a health information system in a small aged care facility. The study was focused on the use of static per-method access control list. The case study found that the method is inadequate as the policy constrains become complex even for a small system taken in the case study.

Reid et al. [6] examined the RBAC as a candidate access control mechanism for health care information and found that the range of access policy expressions supported by RBAC is not adequate. The paper proposed a model where the access control is given through a consumer centric role called care team role. The advantage of this model is that a subgroup of entities within a role can be explicitly granted or denied access to health information. Motta and Furuie [10] extend the RBAC reference model by introducing contextual authorisation. The authorization module not only uses the positive and negative authorization, but also user affiliation, time and location of access, user and patient relationship, patient status, etc.

Khayat and Abdallah [4] present a formal model for flat role-based access control, which we see closer to our model. This model overcomes the some of the problems as it uses the flat model. However, this model does not consider the problem-oriented approach where the patient's medical condition is divided into a list of discrete problems such as diabetes, coronary artery disease and lower back pain as in [7]. The reason behind it is that the model considers only roles. Our approach has overcome this problem by flattening not only roles, but also problems (or categorization of medical records) as well as medical information as done in our model.

Choudhri et al. [9] presents a healthcare systems based on mobile technology. The system delivers different versions of the documents based upon their roles using dynamic trust model. The model is based on transitive trust, that is, a doctor can delegate his role to other doctors. During the delegation process, the doctor may give his full rights or limited rights. This means a doctor who was not denied by patients could have access to the patient information through delegation. We have not dealt with transitive trust in our model.

6 Conclusions and Future Work

Previous eConsent data models were developed on a set of restricted assumptions. By widening the assumptions and examining the realistic use of eConsent within the health care system, we noted that these eConsent data models needed to be changed. As a consequence, we developed a flexible data model for electronic medical records, called tag-based data model. The data model allows representing patients' medical records along their consents. The model also allows us to categorize medical records

into different group and define default policies for such categories. Unlike existing role-based access model, our data model supports both access and transfer policies on roles, and on categories of information, facilities and healthcare practitioners.

Our experience has been that the data model allows a great deal of flexibility and autonomy to the end users, and imposes a minimal set of semantic requirements on its use in specifying policies and catagories. Because eTags and eRecs are securely encrypted, indexing and searching requires the extension of the current data model to include metadata information.

Finally, we demonstrated the feasibility of the use of this data model by developing a prototype system based on .NET and trusted computing technologies. The prototype system gave us an insight of difficulties in implemented secure medical applications. One of the issues was the need for a good user interface to allow the complex relationships of the model to be accurately captured as well as easily understood. Other issues such as secure transport protocol, and the establishment of mutual trust will be presented in subsequent papers.

Acknowledgements

Discussion with Paul Greenfield of the prior e-consent model greatly assisted us in developing the tag-based data model. We would also like to thank Michael Cox from NTRU who helped on using some functionalities of NTRU TSS driver. Special thanks go to Alan Fekete from Sydney University and Andre Schiper from EPFL for encouraging Greg and Fred to take this project at CSIRO.

References

1. O'Keefe, C.M., Greenfield, P., and Goodchild, A. (2005): A Decentralised Approach to Electronic Consent and Health Information Access Control. Journal of Research and Practice in Information Technology, Vol. 37(2):161-178, May 2005.
2. Chadwick,D,, and Mundy, D (2004). The secure electronic transfer of prescriptions. Healthcare Computing, 2004.
3. Task Force on Medical Informatics (1996): Safeguard Needed in Transfer of Patient Data. PEDIATRICS Vol. 98 No. 5, pp. 984-986, Nov 1996.
4. Khayat, E.J., and Abdallah, A.E. (2003); A formal model for flat role-based access control. IEEE International Conference on Computer Systems and Applications, Tunisia, July 2003.
5. Evered, M. and Bogeholz, S. (2004); A case study in access control requirements for a health information system. Australasian Information Security Workshop 2004.
6. Reid, J., Cheong, I., Henricksen, M., and Smith, J. (2003) A Novel Use of RBAC to Protect Privacy in Distributed Health Care Information Systems. In Safavi-Naini, Rei and Seberry, Jennifer, Eds. Proceedings 8th Australasian Conference on Information Security and Privacy (ACISP 2003) 2727, pages 403-415, Wollongong
7. Stein, L.D. (1997). The Electronic Medical Record: Promises and Threats. Web Journal, 2(3), 1997.

8. Huston, T. (2001): Security Issues for Implementation of E-medical Records. Communication of the ACM, 44(9) pp. 89-94, September 2001.
9. Choudhri, A., Kagal, L., Joshi, A., Finin, T. and Yesha, Y. (2003). PatientService: Electronic Patient Record Redaction and Delivery in Pervasive Environment. Fifth International Workshop on Enterprise Networking and Computing in Healthcare Industry (Healthcom 2003).
10. Motta, G.H.M.B and Furuie, S.S. (2003). A Contextual Role-Based Access Control Authorization Model for Electronic Patient Record. IEEE Transactions on Information Technology in Biomedicine 7(3):202-207, Sept. 2003.
11. Crook, R., Ince, D. and Nuseibeh, B. (2003). Modelling access policies using roles in requirements engineering. Information and Software Technology, 45:979-991.
12. OASIS (2005). eXtensible Access Control Markup Language (XACML) Version 2.0 3, http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf, 2005.
13. Paul Ashley, Satoshi Hada, Günter Karjoth, Calvin Powers, and Matthias Schunter (2003), Enterprise Privacy Authorization Language (EPAL 1.1), IBM Technical Report, 2003.
14. Messerges, T. S. and Dabbish, E. A. 2003. Digital rights management in a 3G mobile phone and beyond. In Proceedings of the 2003 ACM Workshop on Digital Rights Management (Washington, DC, USA, October 27 - 27, 2003). DRM '03. ACM Press, New York, NY, 27-38. DOI= http://doi.acm.org/10.1145/947380.947385
15. Open Mobile Alliance, DRM Architecture, version 2.0.6, 2004 (OMA-DRM-ARCH-V2_0_6-20040820-C.zip at http://www.openmobilealliance.org