Concilier Ubiquité et Sécurité des Données Médicales

Tristan Allard, Nicolas Anciaux, Luc Bouganim, Philippe Pucheral, Romuald Thion

CHAPITRE IX

CONCILIER UBIQUITÉ ET SÉCURITÉ DES DONNÉES MÉDICALES

Tristan Allard
Université de Versailles, France
Nicolas Anciaux
INRIA Rocquencourt, France
Luc Bouganim
INRIA Rocquencourt, France
Philippe Pucheral
Université de Versailles, & INRIA Rocquencourt, France
Romuald Thion
INRIA Grenoble, France

Résumé

Au cours de la dernière décennie, de nombreux pays ont lancé d'ambitieux programmes de dossiers médicaux électroniques (Electronic Health Record – EHR) avec pour objectif d'accroître la qualité des soins tout en réduisant les coûts. L'accès ubiquitaire aux données médicales vise à rendre les informations de santé disponibles en toute sécurité n'importe où et n'importe quand, même dans des environnements déconnectés (par exemple, chez le patient). Les solutions existantes d'EHR basées sur des serveurs s'adaptent mal aux situations déconnectées et peinent à apporter les garanties de sécurité attendues par les patients. La solution proposée dans le présent document s'appuie sur un nouveau dispositif matériel combinant un microcontrôleur sécurisé (similaire à une carte à puce) avec une grande mémoire Flash externe sur un facteur de forme de type clé USB. En embarquant le dossier du patient, un système de gestion de bases de données et un serveur Web dans un tel dispositif, il devient possible de gérer en toute sécurité un dossier médical en autonomie totale. Le présent document propose également une nouvelle manière de personnaliser les politiques de contrôle d'accès pour répondre aux attentes des patients en termes de respect de la vie privée, avec une assistance minimale de la part des praticiens. Bien que les deux propositions soient orthogonales, leur combinaison au sein de la même infrastructure apporte une solution nouvelle et convaincante à la construction de dossiers médicaux ubiquitaire sécurisés.

Introduction

Le besoin d'améliorer la qualité des soins tout en réduisant leurs coûts a poussé de nombreux pays à mettre en place des systèmes de gestion de dossiers médicaux électroniques (Electronic Health Record systems – EHR) qui recueillent l'historique médical des individus. L'interopérabilité entre les systèmes d'information de santé, souvent hétérogènes, et la

protection de la vie privée constituent deux défis majeurs à relever. Par ailleurs, l'accès ubiquitaire aux données médicales cherche à éliminer les contraintes de lieu et de temps pour accéder aux dossiers des patients. Les soins dispensés au domicile des personnes âgées ou des personnes handicapées illustrent bien ce besoin. Dans ce contexte, les données de santé sont principalement recueillies et consultées au domicile des patients par des praticiens ayant des droits différents et intervenant à différents instants. Les informations de santé doivent alors être échangées en toute sécurité entre les praticiens pour améliorer la coordination des soins, sans toujours disposer d'une connexion à Internet. Des données peuvent également être fournies par des établissements externes à la coordination des soins (par exemple, un laboratoire médical) et être ajoutées au dossier du patient. Enfin, l'accès aux données peut être effectué par des praticiens qui ne sont pas chez le patient (par exemple, dans un cabinet médical ou à l'hôpital). Dans cet article, nous proposons d'utiliser des dispositifs portables intelligents pour permettre un accès efficace et ubiquitaire aux dossiers médicaux sans violation de la vie privée.

Les systèmes EHR visent à répondre à la plupart des exigences mentionnées ci-dessus. L'objectif de centralisation des informations médicales dans des Systèmes de Gestion de Bases de Données (SGBD) est multiple (1) : complétude (les données sont complètes et à jour), disponibilité (elles sont accessibles via Internet 24 heures sur 24 et sept jours sur sept), facilité d'utilisation (elles sont organisées, leur interrogation est aisée), cohérence (elles satisfont les contraintes d'intégrité ; les mises à jour sont atomiques et sans conflit) et

(1) Le terme centralisation signifie ici que les données sont stockées, organisées, mises à disposition et contrôlées par des serveurs de bases de données, quelle que soit l'infrastructure du système informatique (centralisée ou distribuée).

durabilité (elles sont protégées contre les défaillances). Un rapport récent a identifié plus de 100 projets EHR en cours à travers le monde à l'échelle d'un pays ou d'une région en 2007 (Door, 2008). D'autres rapports indiquent que près de 25 % des professionnels de santé aux États-Unis utilisent des systèmes EHR. En Europe, ces chiffres varient considérablement d'un pays à l'autre, allant de 15 % en Grèce jusqu'à 90 % aux Pays-Bas actuellement.

En ce qui concerne l'accès ubiquitaire à ces données, les dossiers médicaux sont parfois rendus accessibles via Internet par le biais de dispositifs mobiles (par exemple, ordinateurs portables, PDA, PC tablette). Néanmoins, cela nécessite que chaque point du territoire soit connecté par un réseau sécurisé, rapide, fiable et bon marché, ce qui n'est pas le cas dans de nombreux pays et de nombreuses régions de nos jours.

De plus, malgré l'avantage indéniable des systèmes EHR en termes de qualité de soins, des études conduites dans différents pays ont révélé que les patients sont souvent réticents à utiliser ces systèmes EHR face à l'augmentation des menaces sur leur vie privée (The Times, 2008 ; The International Council on Medical & Care Compunetics, 2009). Ces soupçons sont avivés par des études sur la sécurité informatique mettant en exergue la vulnérabilité des serveurs de bases de données face aux attaques externes et internes (Gordon et al, 2006). En effet, la centralisation et l'organisation des informations augmentent leur valeur, ce qui motive les attaques et les utilisations abusives. Indépendamment de la législation protégeant l'utilisation des données médicales et des procédures de sécurité mises en place au niveau des serveurs, le patient a le sentiment de perdre le contrôle sur ses données.

Ainsi, la mise en œuvre de dossiers médicaux ubiquitaires nécessite de répondre à quatre questions majeures :

1. Comment accéder au dossier d'un patient en mode déconnecté (par exemple, à son domicile) ?
2. Comment accéder au dossier d'un patient dans une zone connectée ?
3. Comment amener le patient à faire confiance à la sécurité du système EHR ?
4. Comment obtenir le consentement du patient pour l'usage de son dossier ?

Comme cela a été abordé ci-dessus, les systèmes EHR existants répondent bien à la question 2, mais n'apportent aucune réponse à la question 1 et peinent à répondre à la question 3. Les systèmes EHR ne répondent pas non plus avec précision à la question 4, de par le manque de confiance des patients dans les serveurs.

Le présent document suggère une nouvelle manière d'organiser un système EHR pour répondre simultanément à l'ensemble de ces questions. La solution proposée s'appuie sur un nouveau dispositif matériel appelé ci-après « dispositif portable sécurisé » (Secure Portable Token – SPT). Globalement, un SPT combine un microcontrôleur sécurisé (similaire à une puce de carte à puce) avec une grande mémoire Flash externe (de l'ordre du giga-octets) sur un facteur de forme de type clé USB (Eurosmart, 2008). Un SPT peut héberger des données embarquées et exécuter un code embarqué avec de fortes garanties de sécurité. L'ajout d'un SGBD et d'un serveur Web dans un SPT permet de gérer en toute sécurité un dossier médical en autonomie totale. L'accès au dossier embarqué chez le patient nécessite un simple terminal (par exemple, un ordinateur portable ou un PDA) muni d'un port USB et d'un navigateur Web. La réponse à la question 1 est donc apportée par construction. Les propriétés de sécurité du SPT (matériel protégé contre les manipulations illicites, logiciel incorporé certifié) répondent à la question 3 avec un niveau

de confiance largement supérieur à celui de n'importe quel serveur traditionnel.

Il devient néanmoins plus difficile de répondre à la question 2. En effet, le dossier du patient n'est pas accessible sans être physiquement en possession du SPT du patient. Dans le cas où l'accès à distance au dossier est obligatoire, la solution proposée consiste à réintroduire un serveur dans l'architecture de manière à ce qu'un échange sécurisé d'informations puisse être organisé entre le patient et un cercle de personnes dignes de confiance. La solution est telle que les données du patient ne sont jamais stockées en clair sur le serveur et les clés de chiffrement ne sont connues que des SPT participant au cercle de confiance défini par le patient. Par conséquent, la réponse à la question 2 est apportée d'une manière pratique sans remettre en cause la réponse à la question 3.

On pourrait également considérer que répondre à la question 3 conduit à répondre à la question 4. Malheureusement, tel n'est pas le cas. La confiance dans la sécurité du système EHR est en effet une condition nécessaire au consentement du patient, mais en aucun cas une condition suffisante. L'expression d'un consentement éclairé passe par la compréhension et l'acceptation d'une politique de contrôle d'accès spécifiant à qui (individus ou rôles) et à quelle partie de son dossier l'accès est accordé. Le grand nombre de personnes interagissant avec le dossier, la diversité de leurs rôles, la complexité des informations médicales et les difficultés intrinsèques à déterminer quelles données (ou associations de données) révèlent une pathologie rendent cet objectif extrêmement difficile à atteindre. Le présent document propose une nouvelle alternative pragmatique pour définir des politiques de contrôle d'accès gérables par un patient avec une assistance minimale de la part des praticiens. Cette solution complète bien l'architecture EHR basée sur SPT en répondant à la

question 4. Cette solution est néanmoins orthogonale à l'architecture basée sur SPT et nous sommes convaincus qu'elle pourrait s'appliquer à de nombreux systèmes d'information de santé.

En conclusion, l'objectif du présent document est double. Premièrement, il discute dans quelle mesure les architectures EHR existantes peuvent répondre aux quatre exigences susmentionnées (**Section 2**) et propose une alternative basée sur le dispositif matériel SPT (**Sections 3 et 4**). Deuxièmement, il discute si les politiques de contrôle d'accès et les mécanismes de confidentialité existants permettent au patient d'exprimer un consentement éclairé **(Section 5**) et propose une solution reposant sur un nouveau modèle de contrôle d'accès (**Section 6**). La **Section 7** présente une expérimentation sur le terrain effectuée avec une architecture EHR ubiquitaire, combinant les deux propositions, dans le contexte de soins à domicile apportés aux personnes âgées.

Section 1. – Contexte

Un système de gestion de dossiers médicaux électroniques constitue une collection de dossiers de patients, contenant chacun l'historique médical complet d'un individu, géré et consulté par des professionnels de santé autorisés (Alliance, 2008). La construction d'un système EHR nécessite l'interconnexion de multiples systèmes d'information de santé, généralement hétérogènes, pour agréger les données médicales maintenues localement (par exemple, des données d'hôpitaux, de praticiens et de pharmaciens concernant le même individu).

Par conséquent, le premier défi à relever par les programmes EHR est d'assurer l'interopérabilité entre systèmes hétérogènes. Comme cela a été mis en évidence dans l'introduction, garantir la disponibilité des données même dans des

environnements déconnectés et faire respecter la sécurité des données sont deux défis complémentaires et obligatoires. Les sections suivantes présentent l'état de l'art relatif à ces trois défis.

§ 1. Interopérabilité des EHR

Trois approches principales peuvent être distinguées en fonction du niveau d'intégration visé entre systèmes d'information de santé existants.

La première approche consiste à interconnecter des systèmes autonomes existants dans une plus grande infrastructure sans aucune centralisation de données et avec un contrôle central minimal. Le réseau danois de données de santé (Danish Healthcare Data Network - Pedersen, 2006) est représentatif de cette catégorie. Il connecte les intranets déjà sécurisés d'organisations de santé par l'intermédiaire de VPN (Réseau Privé Virtuel) sur Internet, en allant progressivement des organisations aux départements, des départements aux régions et des régions à la nation. L'effort danois en matière d'EHR consiste principalement à définir un modèle commun de données représentant des données cliniques. Les États-Unis ont adopté une approche fédérale pour la construction d'un EHR. Au niveau régional, les organisations d'informations de santé (Regional Health Information Organizations - RHIO) permettent l'échange d'informations entre organisations de santé locales (par exemple, CalRHIO pour RHIO en Californie). Au niveau national, le projet de réseau d'informations de santé (Nationwide Health Information Network – NHIN), supervisé par le bureau du coordinateur national de l'informatique de santé (Office of the National Coordinator for Health IT – ONC), permet l'échange sécurisé d'informations de santé à travers les États-Unis en utilisant les RHIO en tant que blocs régionaux de construction. Le NHIN est donc un « réseau de réseaux » construit sur Internet.

La deuxième approche renforce l'intégration grâce à des index centralisés et/ou des résumés de données. L'organisation nationale de la santé (National Health Society – NHS) au Royaume-Uni a lancé un projet EHR intitulé CRS (Care Record Service). Premièrement, le CRS vise à lier les enregistrements médicaux électroniques (Electronic Medical Record - EMR) d'un individu en constituant un dossier de santé unique virtuel. La navigation entre les EMR d'un individu et le recueil de données détaillées s'en trouve facilitée. En outre, en partageant des données à travers les EMR, la duplication de données, par exemple administratives, est éliminée. Deuxièmement, CRS vise à stocker des résumés de données détaillées sur « l'épine dorsale » (Spine) du système, à savoir un système central déjà existant qui fournit des services associés à la santé (par exemple, ePrescriptions pour les ordonnances électroniques, eReservations pour les réservations électroniques). Les données résumées serviront à élaborer des rapports et à analyser les informations de santé collectées. Avec son architecture fonctionnelle à deux niveaux, le projet Diraya en Andalousie est similaire au projet CRS du Royaume-Uni. Les données détaillées dans les EMR sont maintenues là où elles sont produites (par exemple, à l'hôpital) et le système central Diraya les indexe. Diraya centralise ce que l'on appelle les « données principales », c'est-à-dire les données auxquelles on accède le plus fréquemment. Aux Pays-Bas, le projet du centre national d'information de santé (National Healthcare Information Hub - LSP en hollandais), dirigé par Nictiz, est principalement un index central stockant l'emplacement des EMR de chaque individu. L'initiative autrichienne ELGA (Husek, 2008) est similaire au projet LSP. Le programme national canadien Infoway-Inforoute finance les projets EHR provinciaux dont la plupart se concentrent sur l'interopérabilité entre les organisations de soins.

Par exemple, l'EHR Alberta Netcare centralise régionalement les données récapitulatives du patient. Le projet Yukon Telehealth rend les EMR locaux accessibles aux praticiens spécialistes à distance.

L'approche la plus intégrée cherche à réunir tous les EMR portant sur la même personne au sein d'un dossier de santé centralisé. Aux États-Unis, certaines organisations privées avaient déjà ressenti le besoin d'agréger les données de tous leurs patients en un seul dossier avant l'avènement des RHIO. Par exemple, le Veteran Administration Medical Center a développé le système VistA (Brown et al, 2003), un système d'information de santé (Health Information System - HIS) permettant aux centres de santé munis de VistA de partager les données de leurs patients. Le programme national français de Dossier Médical Personnel (DMP) vise également à centraliser les dossiers médicaux via des hébergeurs de données accrédités. Dans un autre esprit, des systèmes comme Google Health™ et Microsoft HealthVault™ proposent aux individus de centraliser leurs enregistrements de santé personnels (PHR) de leur propre initiative. Ces deux systèmes téléchargent les données médicales directement des centres de santé avec lesquels ils ont des agréments, mettent des outils pratiques à disposition des patients (par exemple, interactions entre médicaments, recherches d'hôpitaux), et peuvent fournir un accès contrôlé au PHR à un éventail sélectionné de personnes. Les deux systèmes sont gratuits, il est simplement imposé aux utilisateurs de faire confiance à leur politique de respect de la vie privée.

§ 2. Disponibilité d'EHR

Tous les systèmes EHR précédemment mentionnés sont disponibles 24 heures sur 24 et sept jours sur sept, avec la supposition qu'une connexion Internet puisse toujours être établie. Cela n'est malheureusement pas le cas de partout et dans toutes les situations, ce qui introduit la nécessité d'un accès déconnecté aux dossiers médicaux.

Au Royaume-Uni, la carte électronique de santé (Health eCard) est une initiative privée qui propose de stocker des copies chiffrées d'EMR complets dans des cartes à puce spécifiquement conçues, en mettant les données de santé du patient à disposition dans des situations déconnectées (par exemple, des situations d'urgence, des consultations à la maison).

L'organisation allemande Gematik dirige l'eGK, un projet ambitieux combinant une infrastructure traditionnelle à des cartes à puce pour traiter les situations connectées et déconnectées (Smart Card Alliance-b, 2006). Les patients et les professionnels sont munis d'une carte à puce, les cartes à puce des patients stockant les EHR alors que les cartes à puce des professionnels sont utilisées pour l'authentification forte, la signature numérique et le chiffrement/déchiffrement des documents. L'infrastructure maintient une copie centralisée des EHR, accessible par Internet. Ce projet en est toujours au stade préliminaire.

Aux États-Unis, de nombreuses initiatives privées lancées par des centres de soins traitent l'exigence de « l'accès déconnecté » (Smart Card Alliance-a, 2006), par exemple le projet de passeport de santé (Health Passport Project – HPP) du centre médical de l'Université de Pittsburgh (University of Pittsburgh Medical Center), l'eLife-Card en Floride, le Queens Health Network, et la carte de santé personnelle (Personal Health Card) du Mount Sinai Medical

Center. Toutes ces initiatives stockent une copie des informations de santé importantes chiffrées sur une carte à puce pour les mettre à disposition en cas d'urgence.

En 2001, Taïwan a lancé un projet pour remplacer les dossiers de santé traditionnels sur papier par des cartes à puce (Smart Card Alliance, 2005). Les cartes à puce sont utilisées exactement comme les dossiers papier l'étaient. Les cartes stockent de façon permanente les données de santé personnelles, les données récapitulatives administratives et elles stockent temporairement les données médicales relatives aux six dernières visites. Toutes les six visites, les données médicales temporaires sont chargées dans l'infrastructure de santé taïwanaise. Le projet de cartes à puce de santé est intégré de manière transparente à l'infrastructure de santé précédente, en fournissant une authentification forte du patient et une gestion de données sans papier.

Bien que de nombreuses initiatives portent sur le défi de l'accès déconnecté, les faibles capacités de stockage des cartes à puces utilisées dans les projets susmentionnés (c'est-à-dire, au mieux de quelques centaines de kilo-octets) limitent considérablement la quantité des données embarquées et donc l'avantage de l'approche.

§ 3. Sécurité des EHR

Une authentification forte est généralement requise pour se connecter aux serveurs EHR. Les professionnels de santé s'authentifient avec une carte à puce (par exemple, CRS au Royaume-Uni, LSP aux Pays-Bas), tout comme les patients qui accèdent à leur dossier médical (par exemple, Diraya en Andalousie). De plus, les canaux de communication sont habituellement protégés par des techniques cryptographiques basées sur des protocoles comme TLS (Internet Engineering Task Force, 2008), qui

permettent aux entités d'échanger des messages en toute sécurité (c'est-à-dire, chiffrement, protection d'intégrité et non-répudiation de messages) et des mesures complémentaires de sécurité sont mises en œuvre sur les serveurs centraux. Tous ces principes sont nécessaires mais pas suffisants pour assurer la confiance dans le système.

La suspicion est avivée par des études sur la sécurité informatique qui démontrent la vulnérabilité de serveurs de bases de données aux attaques externes et internes. Les systèmes de bases de données sont identifiés comme cible principale de la criminalité informatique, et même les serveurs les mieux défendus, y compris ceux du Pentagone (The Financial Times, 2007 ; Liebert, 2008), du FBI (The Washington Post, 2007) et de la NASA (Computer World, 2003), ont été attaqués avec succès. De plus, près de la moitié des attaques (Gordon et al, 2006) sont internes, c'est-à-dire conduites par des employés des sociétés ou des organisations concernées. Il existe également de nombreux exemples dans lesquels la négligence conduit à des fuites de données personnelles. Pour n'en citer que quelques-uns, des milliers de dossiers de patients de Medicare et Medicaid dans huit États aux États-Unis ont été perdus (FierceHealthIT news, 2006) et Hospitals County a publié accidentellement des enregistrements médicaux sur le Web (FierceHealthIT news, 2008 ; WFTV, 2008), y compris des notes de docteurs, des diagnostics, des procédures médicales voire même le nom et l'âge des patients. Une étude récente a révélé que 81 % des sociétés américaines déclarent que certains employés ont perdu leurs ordinateurs portables avec des données sensibles (Computer World, 2006). Les pertes de données sont si fréquentes qu'un projet de recherche appelé DataLossDB a été créé pour signaler de tels incidents.

En pratique, les EHR sont donc très difficiles à protéger. Cela confirme les réserves exprimées par

les praticiens et les patients en ce qui concerne les programmes EHR (The Times, 2008 ; eHealth Insider, 2008). Aux Pays-Bas, les inquiétudes liées au respect de la vie privée constituent des arguments majeurs pour le retrait du projet national d'EHR (The International Council on Medical & Care Compunetics, 2009). En particulier, le manque de mesures de sécurité limitant l'accès aux données pour les prestataires de services et la perte de contrôle sur leurs propres données ont été identifiés comme les raisons principales pour lesquelles les citoyens rejettent ce projet national d'EHR.

Seuls les EMR stockés dans du matériel personnel sécurisé comme des cartes à puce (cf. **Section 2.2**) peuvent bénéficier d'un respect véritable de la vie privée. Néanmoins, (1) la capacité de stockage des cartes à puce utilisées par les projets actuels (de quelques Ko à Mo) est trop faible pour stocker un EHR complet, limitant la disponibilité des données en situation déconnectée, (2) leur faible connectivité a pour conséquence que les données hébergées sont rarement disponibles, et (3) leur nature portable les rend sujettes aux pertes et aux destructions. En outre, pour assurer la disponibilité, ces projets reposent sur des serveurs centraux dont le niveau de sécurité est très éloigné de celui des cartes à puce (Eurosmart, 2008).

Pour autant, nous pensons qu'il est possible d'assurer à la fois la protection et la disponibilité des données. La solution que nous proposons est centrée sur un dispositif personnel basé sur carte à puce qui stocke le (ou la partie la plus significative du) dossier du patient et qui étend la sphère de sécurité du matériel sécurisé aux serveurs centraux traditionnels.

Section 2. – Un dossier médical sécurisé et portable

Les recherches effectuées dans le projet PlugDB (2) nous ont amené à concevoir un SGBD embarqué dans une nouvelle forme de SPT : un dispositif protégé contre les manipulations illicites. Globalement, un SPT combine un microcontrôleur sécurisé (similaire à la puce d'une carte à puce) à une mémoire Flash externe volumineuse (potentiellement quelques giga-octets) sur un facteur de forme de type clé USB (Eurosmart, 2008). Son architecture matérielle protégée contre les manipulations illicites et son système d'exploitation certifié (Eurosmart, 2008) permettent aux SPT d'héberger des données et d'exécuter du code avec des propriétés de sécurité prouvées. L'objectif principal de la technologie PlugDB est la gestion de dossiers personnels sécurisés et portables. Les dossiers médicaux sont représentatifs de cette classe de dossiers personnels volumineux ayant de fortes exigences de sécurité et de portabilité.

La capacité de stockage d'un SPT est supérieure de quatre ordres de grandeur aux cartes à puce utilisées dans les autres projets EHR (cf. **Section 2**), le rendant capable d'embarquer le dossier complet d'un patient afin d'en assurer l'accès en mode déconnecté. En plus des données, une chaîne logicielle complète est embarquée au sein du SPT : (1) un serveur Web, (2) des Servlets mettant en œuvre l'application, (3) un pont JDBC, et (4) un SGBD gérant la base de données embarquée et assurant le contrôle d'accès. Par conséquent, en embarquant données et code, le SPT peut être considéré comme un serveur complet, accessible à travers un navigateur Web fonctionnant sur tout dispositif muni d'un port USB (par exemple, un ordinateur portable, un tablet-PC, un PDA, voire un téléphone portable). Comparé à

(2) PlugDB est un projet financé par l'ANR, l'Agence Nationale pour la Recherche : http://www-smis.inria.fr/~DMSP

un serveur traditionnel, le SPT est personnel, ne nécessite pas de connexion réseau, et fournit des garanties de sécurité sans précédent.

L'architecture matérielle spécifique du SPT introduit néanmoins de nombreux défis techniques. Les plus importants d'entre eux sont discutés ci-après.

§ 1. Architecture matérielle et système d'exploitation du SPT

Un SPT combine sur la même plate-forme matérielle une puce sécurisée et une mémoire Flash NAND de stockage de masse. La puce sécurisée est du type carte à puce; elle comprend une unité centrale RISC de 32 bits cadencée à 50 MHz, des modules de mémoire composés d'une ROM, de dizaines de kilo-octets de RAM statique, d'une petite quantité de stockage stable interne (FLASH NOR), et de modules de sécurité. La mémoire Flash NAND de stockage de masse est à l'extérieur de la puce sécurisée à laquelle elle est connectée par un bus; elle ne bénéficie donc pas de la protection matérielle de la puce.

Gemalto, le leader mondial de la carte à puce, a développé une plate-forme SPT expérimentale. Cette plate-forme comprend un nouveau système d'exploitation multi-tâches permettant le développement d'applications Web basées sur les technologies Java et Servlet, et proposant ainsi un moyen standardisé d'intégrer des services ou applications web incorporées au SPT. Le système d'exploitation comprend nativement : les protocoles de communication USB 2.0 et IP (Vandewalle, 2004) ; des applications Java multi-threaded ; des primitives cryptographiques ; la gestion de mémoire ; la gestion de Servlets et un serveur Web. Le lecteur intéressé par plus de détails techniques sur la plate-forme matérielle et le système d'exploitation pourra consulter http://www-smis.inria.fr/~DMSP.

§ 2. Système de gestion de bases de données embarqué

Afin de pouvoir gérer des dossiers nomades et sécurisés, le SPT intègre un véritable moteur de SGBD. Ce moteur embarqué a pour objet le stockage de données personnelles dans la Flash NAND externe, leur indexation pour une recherche efficace, la journalisation des mises à jour pour assurer la propriété transactionnelle d'atomicité, l'évaluation de requêtes assertionnelles (i.e. à base de prédicats) et le contrôle de droits d'accès sophistiqués (également assertionnels comme dans toute base de données, par exemple, seules les données satisfaisant la qualification Q1 sont accessibles aux usagers satisfaisant la qualification Q2). Il est impératif que le contrôle d'accès soit embarqué dans le microcontrôleur sécurisé pour assurer sa résistance aux attaques lorsque le dossier est accédé via un terminal vulnérable ou hostile. Le contrôle d'accès s'appuyant sur le gestionnaire de requêtes pour évaluer les prédicats présents dans les règles de droits d'accès, et ce dernier s'appuyant à son tour sur le moteur de stockage et d'indexation, c'est l'intégralité du SGBD qui doit être embarqué dans la puce sécurisée.

§ 3. Disponibilité et sécurité des données

Un individu peut interagir avec le SPT et obtenir les données auxquelles il a légitimement accès à partir de tout terminal muni d'un port USB et d'un navigateur Web. Par conséquent, en cas d'indisponibilité de connexion Internet (par exemple, situations d'urgence, intervention à domicile), les SPT garantissent la disponibilité des données des patients. Ceci répond à la question 1 identifiée dans l'introduction. En outre, les connexions locales aux SPT ne pâtissent pas de performances imprévisibles dues à des serveurs

distants surchargés ou à des connexions de mauvaise qualité : le serveur embarqué est mono-utilisateur et le débit de la communication USB-2 est garanti.

Les données du patient résident dans la mémoire Flash NAND externe. Comme indiqué en **Section 3.1**, la mémoire FLASH NAND du SPT n'est pas protégée matériellement contre les attaques et doit donc l'être par des méthodes cryptographiques. Ces méthodes (chiffrement, hachage, contrôle de version) doivent être compatibles en granularité et en coût avec la grande quantité d'accès aléatoires aux données générée par l'évaluation de requêtes aux bases de données. Les clés cryptographiques résident dans la Flash NOR, protégée par la puce sécurisée contre les manipulations illicites. Les processus de chiffrement, de déchiffrement et de hachage interviennent physiquement dans la puce sécurisée et disposent également de cette protection matérielle. Plus généralement, la chaîne logicielle complète (serveur Web, servlets, SGBD) s'exécute dans la puce sécurisée et bénéficie de sa protection matérielle: les étapes d'authentification, de contrôle d'accès et d'interrogation sont protégées par la puce. La sécurité d'un système reposant sur le fait que le coût de la conduite d'une attaque dépasse son bénéfice, la sécurité de notre architecture est renforcée par la différence entre le coût extrême des attaques et leur bénéfice réduit (divulgation du dossier d'un seul patient). En conséquence, cette architecture répond bien à la question 4 identifiée dans l'introduction.

§ 4. Architecture du SPT

L'architecture du SPT et l'organisation des composants logiciels embarqués sont illustrées sur la Figure 1. Le code embarqué et les données de protection (par exemple, les clés cryptographiques) résident dans la puce sécurisée ; les données du patient résident dans la mémoire externe non

sécurisée, précédemment chiffrées par l'environnement d'exécution sécurisé.

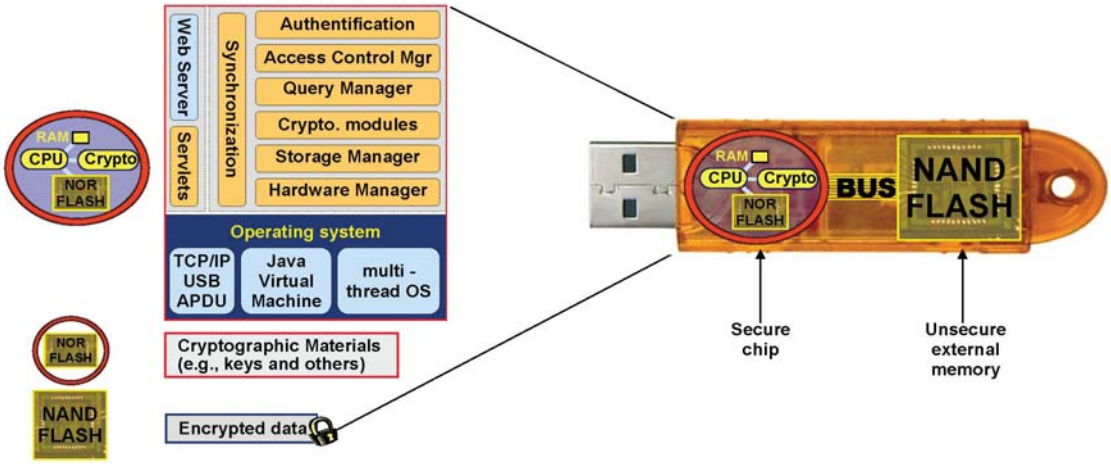

Figure 1 : SPT (Secure Portable Token)

Section 3. – Un dossier médical ubiquitaire et sécurisé

La **Section 2** a traité des questions 1 et 3 ; la présente section se concentre sur la question 2 : l'accès à distance aux dossiers médicaux. Comment le médecin traitant peut-il donner son avis dans une situation d'urgence sans avoir le SPT du patient ? Nous introduisons un serveur central dans l'architecture comme moyen de rétablir la disponibilité des données, c'est-à-dire la capacité d'accéder immédiatement aux données du patient malgré l'absence de son SPT.

Réintroduire un serveur ne doit pas faire perdre les bénéfices du SPT en termes de sécurité et de contrôle par son propriétaire: le serveur central ne doit en aucun cas avoir accès à une information considérée sensible. Deux règles découlent de ce postulat : (1) les données sensibles doivent être stockées sous forme chiffrée sur le serveur, et (2) les données sensibles doivent être chiffrées et

déchiffrées dans un environnement d'exécution sécurisé, c'est-à-dire un SPT.

Pour être à la disposition des professionnels de santé, les dossiers médicaux résident (chiffrés) sur le serveur central et ils peuvent être téléchargés sur leurs SPT à la demande. Pour obtenir l'accès au dossier, le SPT du professionnel doit contenir les clés cryptographiques du patient et détenir les privilèges d'accès à son dossier. Au patient de définir le cercle de professionnels de confiance détenant une copie de sa clé.

Pour sécuriser les communications, nous utilisons des protocoles comme Transport Layer Security (TLS) (Internet Engineering Task Force, 2008). TLS repose sur une autorité de certificats pour émettre des certificats de confiance liant une identité à une clé publique. Les professionnels, les patients et le serveur central communiquent en toute sécurité après avoir échangé et contrôlé leurs certificats respectifs. Les certificats sont insérés dans la mémoire interne sécurisée du SPT au début de leur cycle de vie. Le serveur est chargé de sécuriser son propre certificat. À noter que les SPT ne sont pas durables : ils peuvent être perdus ou cassés. Afin d'assurer la durabilité des certificats et des paires de clés publiques et de clés privées, ceux-ci doivent être dupliqués via un tiers de confiance (Trusted Third party – TTP). Pour des raisons de concision, nous ne détaillons pas ces protocoles dans le présent document.

Dans la suite de cette section, nous classifions d'abord les données en fonction de leurs besoins en termes de confidentialité et nous illustrons une architecture centrée SPT qui répond à ces besoins. Ensuite, nous nous concentrons sur les questions de synchronisation entre le SPT du patient, le serveur central et les entités externes (par exemple, un laboratoire). Puis, nous définissons comment partager des clés cryptographiques à l'intérieur

d'un cercle de personnes de confiance donné. Enfin, nous décrivons un cas complet d'utilisation du système.

§ 1. Confidentialité et classification des données

Nous appelons *données normales* (Regular Data – RD), les informations dont le patient accepte qu'elles soient dupliquées sur un serveur distant sans protection cryptographique. De telles données sont protégées par la politique de sécurité du serveur et sont accessibles en ligne par tout praticien ayant les privilèges requis. Si la politique de contrôle d'accès que le serveur et le SPT font respecter sont censées être identiques, le serveur est incapable d'assurer le même niveau de sécurité que le SPT : d'une part le serveur ne bénéficie pas de la sécurité matérielle du SPT et d'autre part les données normales sont accessibles sur le serveur sans que le patient n'ait connaissance de tous les accès. Seules les données normales peuvent être exportées en clair sur le serveur. Ce qui est considéré comme donnée normale est à l'appréciation du patient; cela peut être par exemple des données administratives, des médicaments non sensibles (par exemple, l'aspirine) et des diagnostics non sensibles (par exemple, la grippe). Les données normales dupliquées sur le serveur profitent de propriétés de disponibilité et de durabilité intrinsèques au serveur.

Nous appelons *données secrètes* (Secret Data – SD), les informations que le patient considère comme très sensibles (par exemple, analyse psychologique) et dont il n'accepte pas qu'elles soient stockées sur un serveur distant. Les données secrètes sont stockées exclusivement dans le SPT du patient. De ce fait, leur durabilité relève de la responsabilité du patient ; d'autre part leur disponibilité nécessite la présence du SPT du patient.

Enfin, nous appelons *données confinées* (Confined Data – CD), les informations qui sont trop sensibles pour être gérées en tant que données normales mais dont la disponibilité en ligne et/ou la durabilité sont obligatoires pour la pratique des soins (par exemple médicaments de chimiothérapie, images IRM). Les données confinées sont dupliquées sous forme chiffrée sur le serveur mais les clés de chiffrement ne sont jamais présentes sur le serveur. Ainsi, les données confinées sont protégées contre les attaques du serveur. Les clés de chiffrement sont stockées et gérées uniquement par les SPT. Pour assurer la disponibilité en ligne des données confinées, le patient sélectionne un ensemble de personnes de confiance (par exemple, le médecin traitant et certains spécialistes) dont les SPT contiendront une copie des clés de chiffrement correspondantes. Nous appelons cet ensemble de personnes le cercle de confiance. Ses membres peuvent accéder aux données confinées, stockées sous forme chiffrée sur le serveur, et utiliser leur SPT pour les déchiffrer. C'est au patient de définir le cercle de confiance. La durabilité est garantie par le serveur comme pour les données en clair. Néanmoins, la récupération des données confinées après une panne ou une perte d'un SPT implique la récupération préalable des clés de chiffrement associées. Cela peut être réalisé soit par une *passphrase* soit par l'enregistrement des clés de chiffrement auprès d'un tiers de confiance. À noter que les données secrètes peuvent être rendues durables en les déclarant en tant que données confinées et en ne partageant pas les clés de chiffrement.

Le classement des données relève de la responsabilité du patient, pouvant éventuellement bénéficier d'une aide externe (par exemple, de son médecin traitant). Le patient peut changer d'avis par la suite (par exemple, en suivant l'avis de son médecin) d'après la hiérarchie suivante : données secrètes → données confinées → données normales.

Tout autre changement est incertain ; par exemple en passant de données normales en données secrètes, les données normales, en clair, peuvent avoir été interrogées ou copiées au préalable.

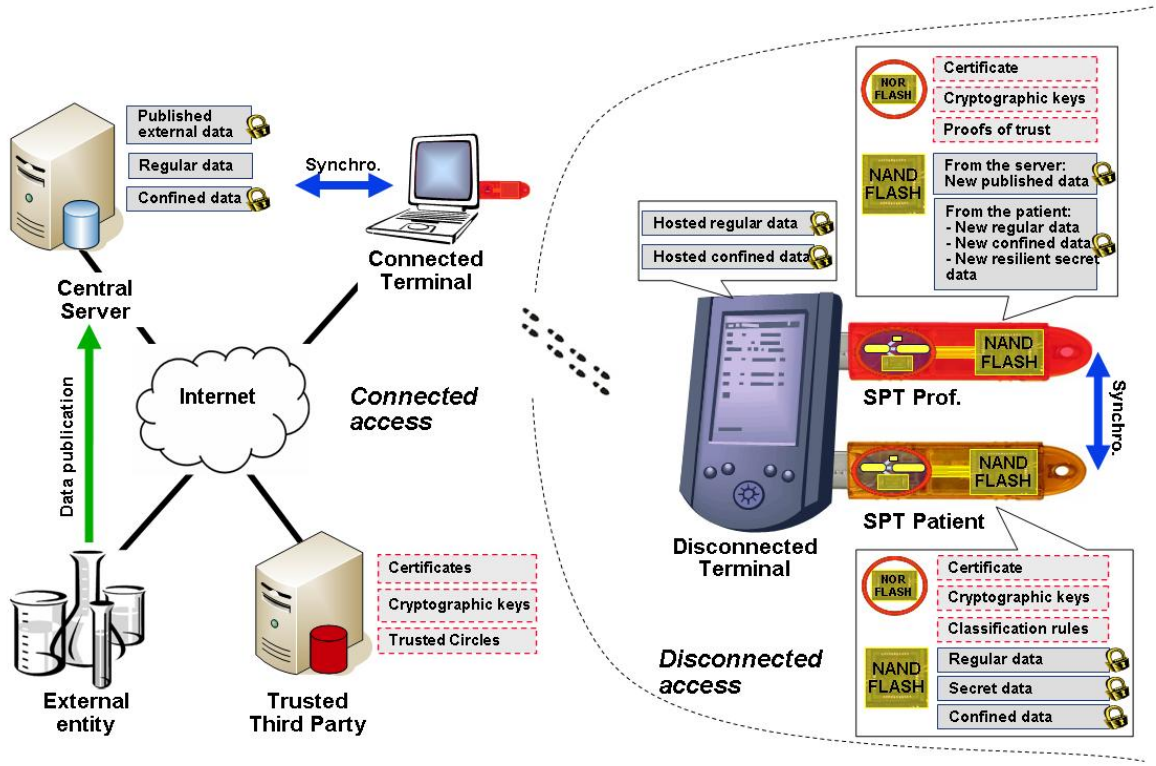

Figure 2 : architecture fonctionnelle

La figure 2 illustre l'architecture fonctionnelle globale, montrant le lieu de stockage des différents types d'information, et précisant leur forme de stockage, chiffrée ou en clair. Les données situées dans les rectangles en ligne pointillée résident dans un stockage de confiance (soit la mémoire interne du SPT soit la TTP) contrairement aux données situées dans les rectangles en ligne continue. Les données situées à côté d'un cadenas sont chiffrées. Cette architecture fournit plus de garanties de protection de la vie privée que n'importe quel EHR traditionnel. Les attaques conduites contre le serveur (en contournant les

mesures de sécurité traditionnelles) ne peuvent que révéler des données normales, car les données secrètes sont absentes du serveur et des données confinées sont chiffrées avec des clés laissées sous le contrôle des SPT et TTP. Les attaques conduites contre le SPT d'un patient sont rendues très difficiles de par son matériel sécurisé.

§ 2. Synchronisation

La duplication des données sur le serveur central assure la disponibilité et la durabilité, mais soulève un problème de synchronisation. Lorsqu'un serveur et un SPT sont directement connectés l'un à l'autre, les méthodes de synchronisation traditionnelles s'appliquent. Néanmoins, un SPT peut ne jamais se connecter directement au serveur central (par exemple, un patient qui ne sort pas de chez lui). Dans ce cas, les SPT des professionnels de santé doivent se comporter comme des *proxys* en portant des messages de synchronisation chiffrés des SPT des patients jusqu'au serveur central, et vice versa.

Les professionnels de santé portent les messages de synchronisation chiffrés des SPT des patients jusqu'au serveur central lorsqu'ils rendent visite aux patients. Au cours de la visite, le professionnel peut insérer de nouvelles données dans le SPT du patient. À la fin de la visite, les données normales et confinées qui viennent d'être créées (c'est-à-dire qui ne se trouvent pas encore dans le serveur) sont copiées dans le SPT du professionnel. Le serveur central est actualisé chaque fois qu'un professionnel s'y connecte. Réciproquement, le SPT du professionnel porte des données chiffrées qui viennent d'être créées sur le serveur pour actualiser la version du SPT du patient. Tel est le cas lorsque des entités externes produisent des données médicales directement sur le serveur central, par exemple un laboratoire qui

remet des résultats d'examens. Ces données ne peuvent toutefois pas être produites en clair et elles ne sont pas encore classées par le patient. Pour surmonter ce problème, les entités externes doivent chiffrer les données avec la clé publique du patient avant de les publier sur le serveur central. Au moment de la synchronisation, le patient est alors en mesure de déchiffrer ces données, de les classer et de les stocker en fonction de leur classe de confidentialité.

§ 3. Scénario d'usage

Illustrons le comportement du système par un scénario impliquant quatre participants : un patient âgé appelé Patrick, son médecin traitant David, une infirmière Nora, et une espionne Sandra. Patrick, David, et Nora ont chacun leur propre SPT. Plusieurs examens médicaux sont prescrits à Patrick qui les classe en tant que données normales, confinées et secrètes. Patrick a récemment subi des examens sanguins dans un laboratoire d'analyses médicales. Le laboratoire d'analyses médicales ayant effectué les examens a publié les résultats chiffrés sur le serveur central. Les résultats ont été chiffrés avec la clé publique de Patrick qui a été obtenue auprès d'une instance de certification.

Nora rend fréquemment visite à Patrick à son domicile. Patrick n'a pas de connexion Internet et sort rarement de chez lui. Nora sert donc de moyen de synchronisation pour le dossier de Patrick (tout comme toute autre personne rendant visite à Patrick et détenant un SPT). Avant la visite, Nora télécharge du serveur central les dernières mises à jour effectuées dans le dossier de Patrick, qui ont été chiffrées avec la clé publique de Patrick, notamment les résultats des examens récents. Au cours de la visite, le SPT de Nora et celui de Patrick se synchronisent : le SPT de Nora envoie à celui de Patrick les résultats d'examens chiffrés,

le SPT de Patrick les déchiffre avec sa clé privée, les classe et les chiffre selon leur catégorie, les résultats de laboratoire étant par exemple des données confinées, puis les renvoie au SPT de Nora qui actualisera le serveur central la prochaine fois qu'elle se connectera à celui-ci. Le SPT de Nora copie également les dernières mises à jour effectuées dans le dossier local de Patrick, le cas échéant. Nora n'a pas accès à ces données bien qu'elles soient transportées par son SPT.

Au cours d'une visite précédente, Patrick a demandé à David de se joindre à son cercle de confiance. Le SPT de Patrick a haché et signé le certificat de David, qui a chargé cette preuve de confiance sur le SPT. Après la visite de Nora, à son bureau, il peut se connecter au serveur central et voir le dossier à jour de Patrick, y compris les résultats des examens récents (classés en tant que données confinées) et les éventuelles mises à jour ramenées par Nora (dans la limite des droits d'accès de David). En rendant visite à Patrick chez lui, David obtient les mêmes informations en accédant aux données directement par le biais du SPT de Patrick.

Un jour, Patrick perd son SPT qui est découvert par Sandra. En l'absence du code secret, Sandra ne peut pas s'authentifier au SPT. Elle peut essayer d'ouvrir le SPT et espionner le contenu de la mémoire Flash NAND mais les données sont chiffrées. Si elle tente toute manipulation illicite sur la puce sécurisée pour obtenir les clés de déchiffrement, les contre-mesures matérielles de sécurité détruiront les composants du SPT. La seule attaque qui pourrait être conduite avec succès consisterait à récupérer les données normales sur le serveur central. Toutes les données sensibles sont elles stockées sous forme chiffrée et aucune clé n'est présente sur le serveur (la clé étant dans les SPT ou chez le tiers de confiance).

Section 4. – Expression du consentement du patient

Le patient doit avoir confiance dans la sécurité de l'EHR pour donner son accord à un usage ubiquitaire de son dossier médical mais ceci n'est pas une condition suffisante. L'expression d'un consentement éclairé implique la compréhension et l'acceptation d'une politique de contrôle d'accès spécifiant qui (individus ou rôles) a accès à quelle partie de son dossier. Cette section approfondit la notion de confidentialité de l'usager et étudie les modèles et mécanismes actuels pour y parvenir.

§ 1. Protection de la confidentialié

A. – Approche juridique

Le respect de la vie privée correspond à la protection des informations personnelles identifiables (PII), au moyen de la restriction de l'accès, du transfert, du stockage, etc. des PII. Le concept de « consentement informé » constitue la clé de voûte de la plupart des règlementations de respect de la vie privée. Le consentement à l'utilisation de données personnelles doit être un acte éclairé, libre, sans équivoque et unilatéral. La protection des PII revêt la plus haute importance pour le déploiement des systèmes informatiques ouverts comme les systèmes EHR (Langheinrich, 2005).

La Directive de l'Union Européenne 95/46/CE fixe la protection des individus par rapport au traitement des données personnelles. L'article 29 de cette directive établit un ensemble de principes de base de confidentialité, qui sont assez proches des dix principes fondamentaux des systèmes de bases de données hippocratiques (Agrawal et al., 2002) :

1. le principe de limitation d'objet : les données doivent être traitées pour un objet spécifique et déclaré.

2. le principe de qualité et de proportionnalité des données : les données doivent être précises, adéquates et pertinentes par rapport à l'objet déclaré.
3. le principe de transparence : des informations doivent être fournies en ce qui concerne l'objet du traitement, l'identité du contrôleur de données doit être assurée.
4. le principe de sécurité : des mesures appropriées de sécurité doivent être prises.
5. les droits d'accès, de rectification et d'opposition.
6. les restrictions de transfert ultérieur.

Comme cela est stipulé par l'article 29 (Data Protection Working Party, 2007), l'un des principes essentiels concernant les systèmes EHR est de limiter l'accès à un dossier aux professionnels de santé qui sont impliqués dans le traitement du patient. La protection de données peut être améliorée par des droits d'accès modulaires : il doit être donné au patient la possibilité d'empêcher l'accès à ses données, en s'assurant que celui-ci comprend les conséquences de ses choix.

B. – Préférence de confidentialité

Beaucoup d'attention a été portée à l'expression des préférences de confidentialité, à savoir le consentement à l'utilisation des PII d'après les principes susmentionnés. La plate-forme P3P (Platform for Privacy Preferences – P3P) définie par le W3C est un format électronique d'expression des préférences de l'utilisateur en terme de confidentialité. La plate-forme E-P3P (Platform for Enterprise Privacy Practices - E-P3P) (Karjoth et al., 2002) ou EPAL (Ashley et al., 2003) définissent quant à elles des règles d'application pour les politiques internes aux entreprises. En d'autres termes, P3P est utilisée pour exprimer la politique de confidentialité des clients vis-à-vis des

entreprises qui collectent ces PII alors qu'E-P3P est utilisée pour appliquer en interne le contrôle de l'accès aux PII collectées.

Comme le suggère le titre de l'article (Massacci & Zannone, 2004) « Privacy Is Linking Permission to Purpose », l'objectif de l'usage des données collectées (par exemple, marketing, études, paiements, etc.) est au cœur de l'expression d'un consentement éclairé. L'intégration de l'objectif (et des obligations relatives à celui-ci) dans le contrôle d'accès est l'un des principaux défis de recherche abordés dans la protection de la confidentialité des données.

§ 2. Mécanisme de contrôle de la confidentialité

A. – Contrôle d'accès traditionnel

Une politique de contrôle d'accès est une forme de politique de sécurité spécialisée pour la gestion des permissions (Samarati & Di Vimercati, 2000). Une politique de contrôle d'accès est structurée selon un modèle. Le modèle décrit formellement le langage dans lequel les politiques sont exprimées ainsi que la prise de décision (accord ou refus) d'une demande d'accès. Les modèles traditionnels de contrôle d'accès sont le contrôle d'accès par mandat (Mandatory Access Control - MAC) et le contrôle d'accès discrétionnaire (Discretionary Access Control - DAC). MAC est un mécanisme de contrôle d'accès basé sur étiquette (par exemple, Non-classifié, Confidentiel, Secret, Top secret). À chaque utilisateur et chaque donnée est associée une étiquette unique définissant le niveau d'autorisation de l'utilisateur et le niveau de sécurité de la donnée. L'accès à une donnée est autorisé si l'utilisateur dispose d'un niveau d'autorisation suffisant. DAC est un mécanisme décentralisé basé sur l'utilisateur dans lequel le créateur d'une donnée définit les autorisations.

Des concepts intermédiaires entre les données et les utilisateurs ont été introduits pour simplifier l'administration des politiques de contrôle d'accès. Dans la famille des modèles de contrôle d'accès basés sur les rôles (Role-Based Access Control - RBAC), des rôles sont affectés aux utilisateurs et des permissions sont accordées à ces rôles (Ferraiolo et al., 2003). Ainsi, une politique RBAC est un ensemble d'attributions entre des utilisateurs et des rôles et entre des rôles et des permissions. La règle de base de RBAC stipule qu'une demande d'accès est accordée si l'émetteur joue un rôle auquel est associé ce privilège.

À partir de l'initiative RBAC, plusieurs modèles ont été étudiés dans la littérature. Ces modèles peuvent soit étendre RBAC (par exemple, avec des contraintes temporelles ou géographiques) soit organiser des politiques au moyen de concepts supplémentaires (par exemple, équipe, tâche, organisation) pour améliorer leur pouvoir expressif et leur flexibilité. La logique du premier ordre en tant que cadre général a été montrée apte à formaliser les modèles et les politiques de contrôle d'accès (Halpern & Weissman, 2008).

B. – Contrôle d'accès pour la protection des données personnelles

Les modèles traditionnels de contrôle d'accès comme RBAC sont couramment utilisés pour organiser des droits d'accès. Néanmoins, ils ne sont pas adéquats pour exprimer un contrôle plus fin intégrant l'usage des données. Le modèle de contrôle d'usage (Usage CONtrol model - UCON) est un fondement des modèles de contrôle d'accès de prochaine génération. Dans ce modèle, une décision de contrôle d'usage est déterminée en combinant des autorisations, des obligations et des conditions (Zhang et al., 2005). Le contrôle d'usage est une manière de mettre en œuvre une gestion des droits

numériques, par exemple en fournissant des garanties de restriction de transfert ultérieur.

Pour garantir le principe de limitation d'objectif, une politique doit garantir que les données ne peuvent être utilisées que pour l'objectif auquel elles sont destinées, et l'objectif de l'accès doit être conforme à l'objectif prévu des données. Les auteurs (Yang et al., 2007) ont proposé un modèle de contrôle d'accès basé sur l'objectif d'après ce fondement. (Ni et al., 2007) ont lié les objets à RBAC dans un modèle intégré « Privacy-Aware RBAC » (P-RBAC). Ce modèle a été affiné pour inclure la définition d'obligations conditionnelles. L'intégration du contrôle d'objectif et de RBAC dans le cadre des bases de données relationnelles a également été étudiée (Byun & Li, 2008).

C. – Modèles spécialisés de contrôle pour les données médicales

Plusieurs modèles ont été définis pour organiser les droits relatifs aux données médicales. Alhaqbani et Fidge proposent une architecture de contrôle d'accès en cascade qui se compose de trois couches de contrôle d'accès (Alhaqbani & Fidge, 2007). La première couche est basée sur DAC, la deuxième couche est basée sur RBAC et la troisième couche est basée sur MAC. Lorsque les droits des politiques s'accordent, l'accès est autorisé. Les auteurs de (Røstad & Nytrø, 2008) ont combiné plus étroitement DAC et RBAC dans un système de contrôle d'accès des enregistrements de santé contrôlé personnellement (Personally Controlled Health Record - PCHR). Dans leur approche, deux politiques sont définies : une politique commune et une politique personnalisée. Uniquement la politique personnalisée est définie par le propriétaire de l'EHR. Une règle de résolution de conflit (par exemple, refuser ou permettre une prise de contrôle manuelle) est

définie et est utilisée en cas de désaccord entre les deux politiques au sujet de la décision d'accès.

Les auteurs de (Alhaqbani & Fidge, 2007) et (Røstad & Nytrø, 2008) se concentrent sur des politiques définies par le propriétaire de l'EHR, les données étant centralisées et maintenues dans un seul dispositif. Becker et Sewell se concentrent sur des réglementations de haut niveau exprimées au niveau national (Becker & Sewell, 2004). Ils proposent un langage logique appelé Cassandra. Ce langage est basé sur Datalog avec contraintes, qui est un sous-ensemble de la logique du premier ordre bénéficiant de bonnes propriétés de décidabilité.

§ 3. Vers un modèle de masquage de données de santé

Les recherches sur le contrôle d'usage, le contrôle d'accès basé sur objectif, les préférences de confidentialité et les pratiques de confidentialité fournissent de nombreux résultats précieux pour traiter de la confidentialité et de l'expression du consentement. Des recherches actuelles abordent un large éventail de questions de confidentialité (par exemple, pratiques de confidentialité au sein de l'entreprise, langage d'expression des préférences de confidentialité) pour traiter des règles complexes (par exemple, des obligations conditionnelles, un usage restreint dans le temps, des règles logiques et des contraintes). Néanmoins, les politiques de contrôle d'accès généralement définies pour réguler l'accès aux systèmes EHR sont beaucoup trop complexes pour espérer obtenir un consentement éclairé de la part des patients à leur égard, comme la loi l'exige. Cela s'explique par deux caractéristiques principales de ces politiques :

C1. Le très grand nombre de règles de contrôle d'accès. Ceci est dû au grand nombre de personnes

interagissant avec un dossier, combiné à une grande diversité de rôles et de privilèges.

C2. La complexité des données à protéger. Ceci est dû à la terminologie spécialisée généralement utilisée pour décrire les données, combinée à la difficulté intrinsèque à déterminer quelles données (ou quelle association de données) révèlent une pathologie donnée.

La politique de contrôle d'accès par défaut définie pour le futur DMP français illustre assez bien cette complexité. Comme cela est représenté sur la figure 3, cette politique basée sur RBAC est exprimée sous la forme d'une matrice Document × Rôle, où les éléments de Document sont les catégories de documents constituant un dossier de santé, les éléments de Rôle sont les rôles qui peuvent être joués par des praticiens et chaque entrée donne les privilèges correspondants de lecture et d'écriture. Dans sa forme actuelle, cette matrice contient déjà plus de 400 entrées malgré la largeur des catégories de documents. L'aspect « gros grain » de cette catégorisation a pour but de faciliter le contrôle d'accès au détriment de la précision des catégories (par exemple, des radiographies peuvent révéler des pathologies très différentes en fonction de l'organe radiographié). Ce qui est finalement révélé reste obscur pour le patient.

À la lumière de cet exemple, la politique de contrôle d'accès par défaut doit être considérée comme l'expression du principe du *besoin-d'en-connaître* (c'est-à-dire qu'un utilisateur doit obtenir l'accès aux seules informations strictement nécessaires pour accomplir les tâches liées à son rôle) au lieu d'un outil qui peut être configuré par le patient pour mieux protéger sa propre vie privée. Le droit de cacher une partie de son historique médical a néanmoins été reconnu aux patients par la législation, pourvu qu'ils soient préalablement

informés des conséquences possibles de cette action. Nous sommes convaincus que l'obtention du consentement éclairé du patient exige de lui fournir des outils efficaces et complets pour masquer les informations sensibles dans son dossier. À cet effet, nous avons conçu un modèle de masquage qui consiste à définir des règles supplémentaires ayant une sémantique pour chaque patient (sur la base de termes définis par le patient), et qui est prioritaire sur la politique de contrôle d'accès par défaut qui n'est pas contrôlable par le patient.

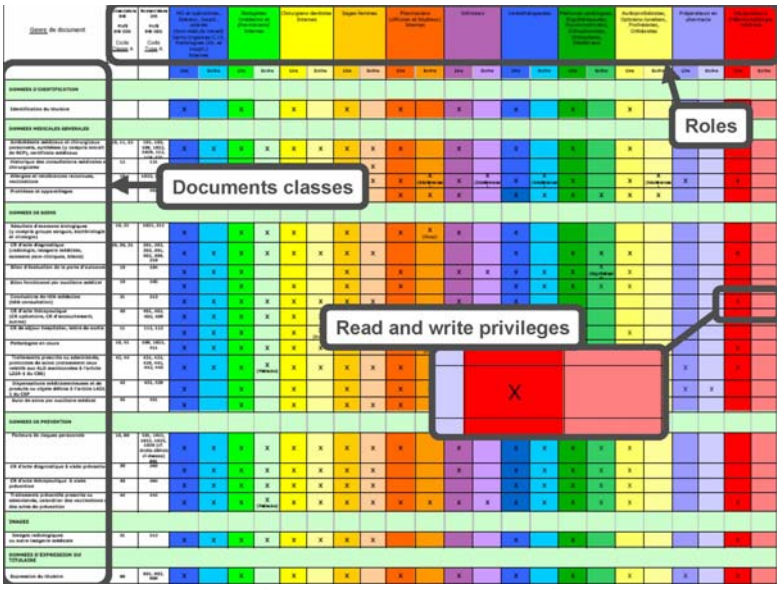

Figure 3 : matrice par défaut du DMP (dossier médical personnel) français
http://www.d-m-p.org/docs/TabCxPS.pdf

Section 5. – EBAC : un modèle de contrôle d'accès basé sur les événements

Le modèle de contrôle d'accès basé sur les événements (Event-Based Access Control model - EBAC) a été conçu pour aider le patient à masquer des enregistrements de santé sensibles dans son dossier. Les fondements de la conception de EBAC sont la simplicité et la précision. Le modèle est organisé selon les concepts principaux d'événements, d'épisodes et de relation de confiance.

- Événement : tout document ajouté à un dossier médical est associé à un événement. Les événements sont dotés de propriétés, parmi lesquelles l'auteur du document (c'est-à-dire un praticien) et l'*épisode* médical auquel il appartient.
- Épisode : un épisode est un ensemble d'événements liés sémantiquement et pour lequel le patient souhaite définir une politique de masquage commune. Par exemple, le patient peut définir des épisodes « MonAvortement2008 », « MaSecondeDépression » et les associer à des événements entrants, potentiellement avec l'aide de son médecin traitant. Le patient définit sa politique de masquage, épisode par épisode, en définissant qui (rôle ou utilisateurs identifiés) est autorisé à participer à quel épisode.
- Relation de confiance : la participation d'un praticien P à un épisode est régulée par une relation de confiance avec le patient stipulant (1) quel événement de cet épisode P peut voir et (2) qui peut voir les événements produits par P dans cet épisode (par exemple, Dr. Guru ne peut voir que les événements qu'il produit et personne d'autre que Guru et le patient ne peuvent voir ces mêmes événements). En d'autres termes, les participants d'un épisode constituent un cercle de confiance selon la définition de la **Section 3.1** et la relation de confiance définit le périmètre de leurs actions respectives dans cet épisode. Pour rendre le modèle simple et intuitif, nous introduisons deux périmètres pour les actions de lecture et d'écriture selon les termes partagé (dénoté par S) et exclusif (dénoté par X). La combinaison de ces périmètres engendre quatre relations de confiance possibles :

SS : le praticien P peut lire les événements partagés dans l'épisode et produit lui-même des événements partagés pour l'épisode ;

SX : le praticien P peut lire les événements partagés dans l'épisode et produit des événements exclusifs pour l'épisode ;

XS : le praticien P peut lire les événements exclusifs produits par lui-même dans l'épisode et produit des événements partagés pour l'épisode ;

XX : le praticien P peut lire les événements exclusifs produits par lui-même dans l'épisode et produit des événements exclusifs pour l'épisode.

Les idées principales du modèle sont donc :

- Il existe une matrice de contrôle d'accès par défaut (basée sur les rôles), définie au niveau de la réglementation et non modifiable par le propriétaire de l'EHR;
- Chaque enregistrement de santé est associé à un événement, lui-même associé à (au plus) un épisode qui est le niveau auquel le propriétaire définit sa politique de masquage;
- La décision d'accès est prise en fonction de l'identité du demandeur, de l'auteur de l'événement et de l'épisode auquel l'événement appartient, la priorité étant donnée à la règle de masquage en cas de conflit avec la matrice de contrôle d'accès;
- Seule la permission de lecture est considérée : le but du modèle EBAC est uniquement d'empêcher toute infraction à la confidentialité, les autres actions sont contrôlées au niveau de la régulation.

Pour illustrer l'approche, nous avons défini un échantillon de politique EBAC. Quatre professionnels nommés Guru (un adepte des médecines alternatives), MyPhysician (mon médecin), MyNurse (mon infirmière),

et AnotherPhysician (un autre médecin) interviennent dans le système. Le patient considéré a défini deux épisodes, l'un pour un cancer et un autre pour un avortement, avec les règles suivantes :

- *E1*, "Cancer": *XX(E1)={Guru}, SS(E1)={ MyPhysician, MyNurse }*.
  Cette règle stipule que Guru, MyPhysician et MyNurse constituent le cercle de confiance du patient pour l'épisode E1. Aucun autre utilisateur ne peut lire des événements dans cet épisode, quelle que soit la politique de contrôle d'accès par défaut. MyPhysician et MyNurse partagent les documents produits dans cet épisode, à l'exception de ceux produits par Guru, parce que le patient souhaite cacher le fait qu'il consulte Guru pour son cancer.
- *E2*, "*Abortion*": *SX(E2)={ MyPhysician, AnotherPhysician }, SS(E2)={ MyNurse }*
  MyPhysician, AnotherPhysician et MyNurse constituent le cercle de confiance de l'épisode E2. MyNurse, qui a effectué l'avortement, produit des événements partagés. MyPhysician et AnotherPhysician partagent ces événements mais ce qu'ils produisent eux-mêmes restent invisibles l'un pour l'autre. Une telle règle peut être définie par le patient après avoir consulté MyPhysican au sujet d'un problème à la suite de l'avortement, et avant de consulter AnotherPhysician pour une deuxième opinion, si le patient n'a pas complètement confiance dans le diagnostic de MyPhysician.

Section 6. – Le projet DMSP

La plate-forme matérielle du SPT est aujourd'hui opérationnelle et les principaux composants logiciels décrits dans les sections précédentes ont été développés et intégrés : serveur Web et SGBD

embarqués. Les principaux composants de l'architecture sont aussi opérationnels: serveur central et protocole de synchronisation. L'application proprement dite est en cours de développement et sera expérimentée sur le terrain d'ici la fin 2009 sur une population d'environ 150 patients âgés et praticiens. Le vieillissement de la population rend cruciale la surveillance de la santé des personnes âgées à leur domicile. Dans ce contexte, des données sensibles doivent être partagées entre tous les participants des réseaux médico-sociaux (docteurs, infirmières, assistantes sociales, aides à domicile et cercle familial) avec différents droits d'accès. Les données doivent être disponibles au chevet du patient pour une meilleure surveillance de sa santé. À cet effet, le département des Yvelines en France a décidé d'effectuer un projet expérimental de Dossier Médico-Social Partagé (DMSP). La première étape de ce projet cible les personnes âgées de deux réseaux de gérontologie. Une étape ultérieure l'étendra à d'autres personnes dans des situations d'instabilité ou de handicap.

L'association de soins à domicile ALDS a déjà réalisé un « dossier médical commun » au format papier, qui permet aux professionnels et aux participants du secteur médico-social de consigner des faits importants relatifs à la surveillance des personnes âgées. Bien que l'utilisation au quotidien de ce dossier papier ait fait preuve de son efficacité, deux questions brûlantes n'ont toujours pas été résolues :

- Absence de confidentialité : tous les participants (les praticiens mais également les assistantes sociales, les aides à domicile et le cercle familial) peuvent lire tous les enregistrements dans le dossier papier du patient alors que certains patients font face à des situations humaines complexes

(diagnostic de maladie en phase terminale, dépendance, difficultés financières, etc.).

- Pas d'accès à distance au dossier : en conséquence, le dossier n'est pas mis à jour de manière homogène et opportune, ce qui conduit à une surveillance moins précise.

L'objectif de cette expérience sur le terrain est de démontrer la pertinence de la technologie proposée pour répondre à ces deux questions. Ce projet implique l'INRIA (Institut National de Recherche en Informatique et en Automatique), l'Université de Versailles, SANTEOS (un fournisseur français d'EHR), Gemalto (le leader mondial de la carte à puce), l'ALDS (une association de soins à domicile) et COGITEY (une clinique pour personnes âgées).

Conclusion

Des projets d'EHR sont en train d'être lancés dans la plupart des pays développés. Certes, les avantages découlant de la centralisation des informations de santé dans des serveurs de bases de données en termes de qualité de l'information, de disponibilité et de protection contre les défaillances sont indiscutables. Reste que les patients sont souvent réfractaires à l'idée d'abandonner le contrôle sur leurs données les plus sensibles (par exemple, données révélant une maladie grave) à un serveur distant. De plus, l'accès au dossier dépend de l'existence d'une connexion Internet haut débit sécurisée en tout lieu et à tout moment.

Le présent document s'appuie sur un nouveau dispositif portable associant la sécurité d'une carte à puce à la capacité de stockage d'une clé USB pour rendre au patient la maîtrise de son historique médical. Nous avons montré comment ce dispositif peut compléter un serveur EHR traditionnel (1) pour

protéger et partager des données hautement sensibles entre des parties de confiance et (2) pour fournir un accès transparent aux données même en mode déconnecté. Du point de vue architectural, le point essentiel est l'incorporation dans une puce sécurisée de la chaîne logicielle complète s'exécutant généralement sur des serveurs traditionnels : serveur Web, servlets, SGBD et enfin la base de données proprement dite. Du point de vue de l'utilisation, le point essentiel est une nouvelle manière de personnaliser les politiques de contrôle d'accès avec une assistance minimale de la part des praticiens. Bien que les deux contributions soient orthogonales, leur intégration à la même infrastructure permet de construire des dossiers médicaux ubiquitaires sécurisés de façon convaincante.

La solution proposée va être expérimentée dans le contexte d'un réseau médico-social apportant des soins médicaux et des services sociaux au domicile de personnes âgées. Le résultat attendu de cette expérience est la démonstration de l'efficacité de la technologie proposée avec un impact positif sur la coordination des intervenants médicaux et sociaux et sur l'acceptation par les patients d'une utilisation électronique de leur historique médical.

Remerciements

Ce travail est partiellement financé par le Conseil Général des Yvelines dans le cadre du projet DMSP et par l'Agence Nationale pour la Recherche (ANR) dans le cadre du projet PlugDB. Les auteurs tiennent à remercier Laurent Braconnier et Jean-François Navarre (Conseil Général des Yvelines), Philippe Kesmarszky (ALDS), Sophie Lartigue (COGITEY), Morgane Berthelot (SANTEOS), Jean-Jacques Vandewalle (Gemalto) et Karine Zeitouni (Université de Versailles) pour leur participation active à ces deux projets.

Références

Agrawal, R.; Kiernan, J.; Srikant, R. & Xu, Y. (2002). Hippocratic Databases. *International Conference on Very Large Data Bases (VLDB)*, 143-154.

Alhaqbani, B. & Fidge, C. J. (2007). Access Control Requirements for Processing Electronic Health Records Business. *Process Management Workshops*, 371-382.

Alliance. (2008). The National Alliance for Health Information Technology, on Defining Key Health Information Technology Term. Report to the Office of the National Coordinator for Health Information Technology. Retrieved February 17, 2009, from http://www.hhs.gov/healthit/documents/m20080603/10_2_hit_terms.pdf.

Anciaux, N., Benzine, M., Bouganim, L., Pucheral, P, & Shasha, D. (2007). GhostDB: querying visible and hidden data without leaks, *ACM SIGMOD Conference*, 677-688.

Anciaux, N., Bobineau, C., Bouganim, L., Pucheral, P., & Valduriez, P. (2001). PicoDBMS: Validation and Experience. *International Conference on Very Large Data Bases (VLDB)*, 709-710.

Anciaux, N., Bouganim, L., & Pucheral, P. (2006). Data Confidentiality: to which extent cryptography and secured hardware can help. *Annals of Telecommunications*, 61 (3-4), 267-283.

Anciaux, N., Bouganim, L., & Pucheral, P. (2007). Future Trends in Secure Chip Data Management. *IEEE Data Eng. Bull.*, 30 (3), 49-57.

Data Protection Working Party (2007). *Working Document on the processing of personal data relating to health in electronic health records (EHR).* Technical report 00323/07/EN WP 131. European Commission.

Ashley, P.; Hada, S.; Karjoth, G.; Powers, C. & Schunter, M. (2003). *Enterprise Privacy*

Authorization Language (EPAL 1.2). Technical report. IBM Tivoli Software, IBM Research.

Becker, M. Y. & Sewell, P. (2004). Cassandra: Flexible Trust Management, Applied to Electronic Health Records. *Computer Security Foundations Workshop,* 139-154.

Brown, S.H., Lincoln, M.J., Groen, P. J., & Kolodner, R. M. (2003). VistA, U.S. Department of Veterans Affairs national scale HIS. *International Journal of Medical Informatics*, 69, 135-156.

Byun, J. & Li, N. (2008). Purpose based access control for privacy protection in relational database systems. *The VLDB Journal*, 17, 603-619.

Computer World. (2003). NASA Sites Hacked. Retrieved February 17, 2009, from http://www.computerworld.com/securitytopics/security/cybercrime/story/0,10801,88348,00.html

Computer World. (2006, August). Survey: 81% of U.S. firms lost laptops with sensitive data in the past year. Retrieved February 17, 2009, from http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9002493&source=NLT_PM&nlid=8.

Dahl, M.R. (2006). *Status and perspective of personal health informatics in Denmark*. Denmark: University of Aarhus, Section for Health Informatics, Institute of Public Health. Retrieved February 17, 2009, from http://www.ieee2407.org/files/ws01_mads01.pps.

Door, J.-P. (2008). *Le dossier médical personnel* (Information Rep. No. 659). France: Assemblée Nationale.

eHealth Insider. (2008, January 16). German doctors say no to centrally stored patient records. Retrieved February 17, 2009, from http://www.e-health-insider.com/news/3384/.

Eurosmart. (2008, April). Smart USB Token. White Paper. Retrieved February 17, 2009, from

http://www.eurosmart.com/images/doc/WorkingGroups/NewFF/Papers/eurosmart_smart_usb_token_wp_april08.pdf

Ferraiolo, D. F.; Kuhn, R. D. & Chandramouli, R. (2003). *Role-Based Access Control*. Artech House Publishers.

FierceHealthIT news. (2006, August 20). Massive data loss at HCA. Retrieved February 17, 2009, from http://www.fiercehealthit.com/story/massive-data-loss-at-hca/2006-08-21.

FierceHealthIT news. (2008, September). GA hospital health data breach due to outsourcing error. Retrieved February 17, 2009, from http://www.fiercehealthit.com/story/ga-hospital-health-data-breach-due-outsourcing-error/2008-09-28.

Gordon, L. A., Loeb, M. P., Lucyshin, W., & Richardson, R. (2006). 2006 CSI/FBI Computer Crime and Security Survey. *Computer Security Institute*. Retrieved February 17, 2009, from http://i.cmpnet.com/gocsi/db_area/pdfs/fbi/FBI2006.pdf.

Halpern, J. Y. & Weissman, V. (2008). Using First-Order Logic to Reason about Policies. *ACM Trans. Inf. Syst. Secur.*, 11, 1-41.

Husek, C. (2008, August). ELGA The Electronic Health Record in Austria. *International Conference of Society for Medical Innovation and Technology*. Vienna, Austria.

Internet Engineering Task Force. (2008). The Transport Layer Security (TLS) Protocol Version 1.2. Retrieved February 17, 2009, from http://tools.ietf.org/html/rfc5246.

Karjoth, G.; Schunter, M. & Waidner, M. (2002). Platform for Enterprise Privacy Practices: Privacy-Enabled Management of Customer Data. *Privacy Enhancing Technologies*, 69-84.

Langheinrich, M. (2005). *Personal Privacy in Ubiquitous Computing*. Doctoral dissertation. ETH Zurich.

Liebert, T. (2008, March). Ongoing concern over Pentagon network attack. *IT News Digest*. Retrieved February 17, 2009, from http://blogs.techrepublic.com.com/tech-news/?p=2098.

Massacci, F. & Zannone, N. (2004). Privacy Is Linking Permission to Purpose. *Security Protocols Workshop,* 179-191.

MasterCard International. (2002). •asterCard Open Data Storage Version 2.0. Technical Specifications.

Ni, Q.; Trombetta, A.; Bertino, E. & Lobo, J. (2007).Privacy-aware role based access. *Proceedings of the 12th ACM symposium on Access control models and technologie,* 41-50.

Pedersen, C.D. (2006, September 8). MedCom - the Danish Healthcare Data Network. *Meeting with AGFA*.

Pucheral, P., & Yin, S. (2007). System and Method of Managing Indexation of Flash Memory. *European Patent by Gemalto and INRIA*, N° 07290567.2.

Pucheral, P., Bouganim, L., Valduriez, P., & Bobineau, C. (2001). PicoDBMS: Scaling down database techniques for the smartcard. *Very Large Data Bases Journal (VLDBJ)*, 10 (2-3), 120-132.

Røstad, L. & Nytrø, O. (2008). Personalized access control for a personally controlled health record. *Proceedings of the 2nd ACM workshop on Computer security architectures,* 9-16.

Samarati, P. & di Vimercati, S. D. C. (2000). Access Control: Policies, Models, and Mechanisms. *Foundations of Security Analysis and Design on Foundations of Security Analysis and Design 2000,* 137-196.

Smart Card Alliance. (2005). The Taiwan Health Care Smart Card Project. Retrieved February 17, 2009, from http://www.smartcardalliance.org/resources/pdf/Taiwan_Health_Card_Profile.pdf.

Smart Card Alliance. (2006-a). Smart Card Applications in the U.S. Healthcare Industry. White Paper N°HC-06001. Retrieved February 17, 2009, from

http://www.smartcardalliance.org/resources/hc/Smart_Card_Healthcare_Applications_FINAL.pdf

Smart Card Alliance. (2006-b). German Health Card. Retrieved February 17, 2009, from http://www.smartcardalliance.org/resources/pdf/German_Health_Card.pdf.

The Financial Times. (2007). Chinese military hacked into Pentagon. Retrieved February 17, 2009, from http://www.ft.com/cms/s/0/9dba9ba2-5a3b-11dc-9bcd-0000779fd2ac.html.

The International Council on Medical & Care Compunetics. (2009). Dutch nationwide EHR postponed. Are they in good company? Retrieved February 17, 2009, from http://articles.icmcc.org/2009/01/23/dutch-ehr-postponed-are-they-in-good-company/.

The Times. (2008, December 26). Patients avoid NHS database blunders by keeping cards close to their chest.. Retrieved February 17, 2009, from http://www.timesonline.co.uk/tol/life_and_style/health/article5397883.ece.

The Washington Post. (2007). Consultant Breached FBI's Computers. Retrieved February 17, 2009, from http://www.washingtonpost.com/wp-dyn/content/article/2006/07/05/AR2006070501489_pf.html.

Vandewalle, J.-J (2004). Smart Card Research Perspectives. *LNCS Construction and Analysis of Safe, Secure and Interoperable Smart devices.*

WFTV. (2008, August 14). Medical Center Patient Records Posted On Internet. Retrieved February 17, 2009, from http://www.wftv.com/news/17188045/detail.html?taf=orlc.

Yang, N.; Barringer, H. & Zhang, N. (2007). A Purpose-Based Access Control Model. *Symposium in Information Assurance and Security,* 143-148.

Yin, S.; Pucheral, P. & Meng, X. (2009). A Sequential Indexing Scheme for Flash-Based Embedded

Systems. *International Conference on Extending Database Technology (EDBT)*, 588-599

Zhang, X.; Parisi-Presicce, F.; Sandhu, R. & Park, J. (2005). Formal model and policy specification of usage control. *ACM Trans. Inf. Syst. Secur*, 8, 351-387.