

Cryptage et confidentialité des données médicales

Gérard Veillon

SOPRA ZIRST, 29 Chemin du Vieux Chêne, 38240 Meylan

Abstract

The generalisation of computer networks and their free access by more and more powerful personal computers have stimulated the interest for the protection of confidential data. This paper is a brief outline of classical cryptosystems, and their use for encryption, identification and authentication in medical information systems. Symetric (DES) and public key algorithm (RSA) are described. Data encryption is an element of the set of technical safeguards of data. Encryption is efficient only if all the environment has the same level of confidentiality. Physical and logical access control, password protection, flow control must be assumed by the information system. Different technical safeguards are implemented, but they must be coherent, and systematically used.

Informatique et Santé, 1991 (4) : 250-260

1. Introduction

Historiquement développé pour garantir le secret dans la messagerie, le cryptage des informations est maintenant utilisé plus largement pour interdire l'accès ou la modification des informations sensibles et garantir la confidentialité dans les applications informatiques. Cependant, le cryptage n'est qu'un élément dans l'ensemble des dispositifs d'un système complexe. La protection qu'il assure n'est valable que si elle s'insère dans un ensemble cohérent.

Après une synthèse des techniques de cryptage, cette présentation décrit sommairement les dispositifs de sécurité et de confidentialité, face aux risques spécifiques du domaine médical. Les principes généraux d'une approche de la sécurité et de l'intégration du cryptage des données sont évoqués dans la conclusion.

2. Les techniques de cryptage

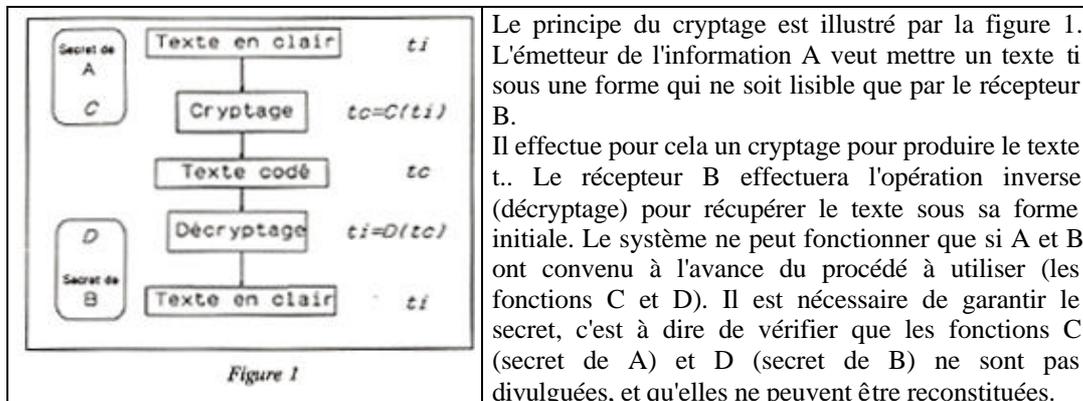
2.1 Principe général [1,2,3]

Le cryptage ou chiffrement des données est généralement décrit à partir de la communication secrète d'informations entre deux interlocuteurs. Dans un système informatique, cette confidentialité intervient en fait sous plusieurs formes, notamment dans la protection du stockage de l'information (copie de fichiers), des accès (interrogation d'une base de donnée) et de sa transmission ("écoute").

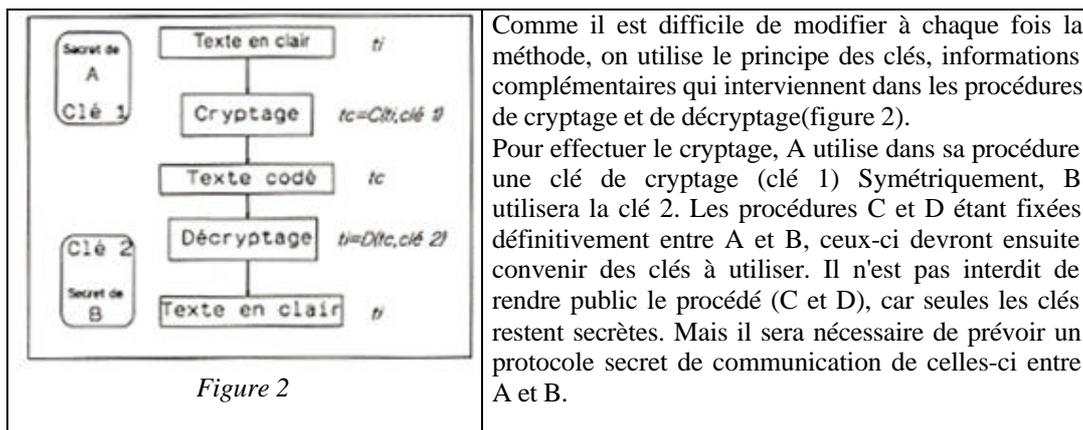
Quel que soit son support physique, l'information est toujours représentée par un codage binaire. Ce codage est conforme à un standard, qui permet de communiquer avec les dispositifs de la machine (claviers, écrans, imprimantes, traceurs ...). Lorsque les ensembles de données à stocker ou à transmettre sont très volumineux (en

particulier les images), on fait appel à des techniques de compression pour optimiser le volume et la vitesse de transmission.

Le cryptage repose sur une transformation du code vers une forme non standard, de façon en limiter l'utilisation. La technique est très ancienne, mais elle a été largement développée et transformée avec l'utilisation intensive de l'informatique et des codages numériques. Ce domaine reste, pour des raisons évidentes, très secret, et les progrès les plus récents ne sont pas divulgués. Cette présentation se limite évidemment aux techniques et outils du domaine public, qui sont ceux effectivement utilisés dans les grandes applications informatiques.



Pour limiter les risques de divulgation, on doit pouvoir changer périodiquement de technique de cryptage.



2.2 Efficacité des algorithmes de cryptage

Un procédé est décrit par des algorithmes de cryptage (calcul de C) et de décryptage (calcul de D), réalisés par des programmes. Pour être utilisable, un procédé de cryptage ne doit pas être trop élaboré : les algorithmes de codage et de décodage doivent être suffisamment rapides, donc peu complexes, pour ne pas ralentir excessivement leur exploitation.

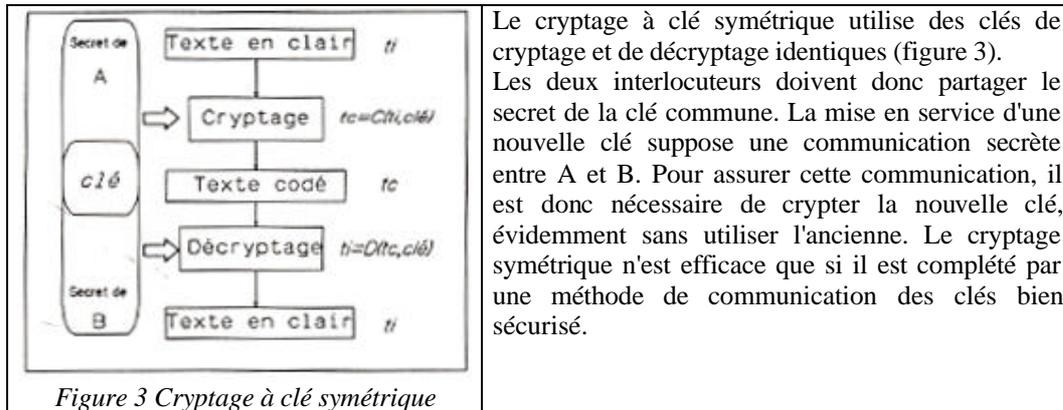
Inversement, pour garantir le secret du cryptage, il faut que celui-ci soit difficile, voir impossible à "casser". Cela suppose d'une part que le secret des clés soit bien gardé, d'autre part que les algorithmes de découverte des clés soient suffisamment complexes, donc suffisamment lents, pour décourager les recherches. La qualité d'un cryptage est donc essentiellement liée au temps d'utilisation (qui doit être court), et au temps de découverte, qui doit être long.

Le choix du cryptage doit tenir compte des intentions et des moyens de l'adversaire, et des risques encourus. Le déchiffrement d'une donnée secrète donne accès à des informations confidentielles, la connaissance de la clé de cryptage permet d'altérer des informations protégées ou d'en émettre de fausses (brouillage). La découverte d'un cryptage sera liée au temps disponible et à la puissance des ordinateurs utilisés par l'espion. La fréquence de changement des clés doit évidemment tenir compte de ces données.

Les agressions ne sont pas toujours malveillantes, les données détruites ou divulguées sont le plus souvent le résultat d'erreurs ou de négligences contre lesquelles il faut aussi se protéger.

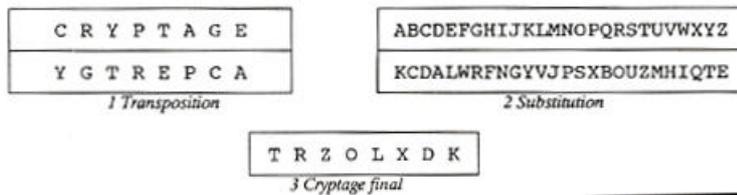
Les deux principaux procédés de cryptage connus sont les cryptages symétriques et les cryptages à clé publique (ou asymétriques) [4].

2.3 Cryptage symétrique, le DES



Les algorithmes de cryptage symétrique sont généralement fondés sur deux types d'opérations élémentaires : la transposition et la substitution. Pour illustrer d'une façon simplifiée ces deux opérations, prenons l'exemple du cryptage d'un mot français dans sa forme alphabétique. La transposition est un changement de l'ordre des caractères du mot, la substitution est un changement d'alphabet, obtenu par une transposition de l'alphabet lui même.

Exemple : on applique une transposition du mot initial "CRYPTAGE", puis une substitution d'alphabet pour obtenir le cryptage final.



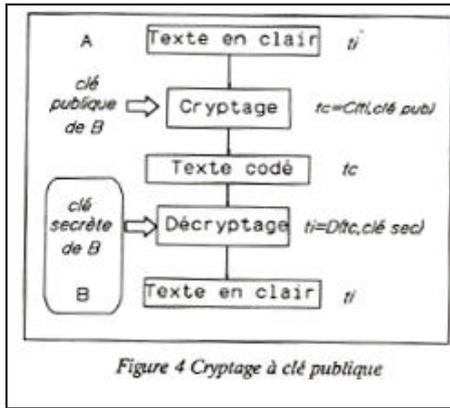
La nature et l'ordre des opérations de transposition et de substitution sont déterminés par la clé. Elles sont évidemment effectuées sur le code binaire.

Le DES ("data encryption standard"), adopté comme standard par le gouvernement américain en 1977, est le cryptage symétrique le plus connu. Il est constitué par une suite complexe d'opérations de type substitution ou transposition. Ce cryptage est très utilisé. Il est disponible sur la plupart des équipements. De nombreuses améliorations des algorithmes ont permis en outre de disposer de programmes 1, voir de composants qui le réalisent avec de bonnes performances. Cet algorithme a fait l'objet de nombreuses critiques, ne serait ce que parce que l'on en a fait un standard, mais aussi pour sa fragilité devant une attaque "professionnelle" (la clé de 56 positions binaires est peut être insuffisante). Le DES n'est plus certifié par le gouvernement américain depuis 1988, mais il reste pour longtemps le standard de fait.

2.4 Cryptages à clé publique

L'approche des algorithmes à clé publique élimine le problème de diffusion des clés.

A chaque interlocuteur B sont associées deux clés. La clé publique de B, qui est connue de tous les émetteurs potentiels, permet de crypter un message vers le destinataire B.



Mais ce message (figure 4) ne peut être décrypté que par B, au moyen d'une clé secrète qu'il est le seul à connaître [1,6].

Le texte crypté par A à l'aide de la clé publique de B ne peut être décrypté qu'au moyen de la clé secrète de B. Seule cette clé de décryptage est secrète, et ce secret n'est pas partagé entre A et B. La clé de cryptage est spécifique de chaque destinataire. Elle est disponible, par exemple dans un annuaire, fournie par le destinataire B, qui a la responsabilité unique de sa création.

2.4.1 L'algorithme RSA

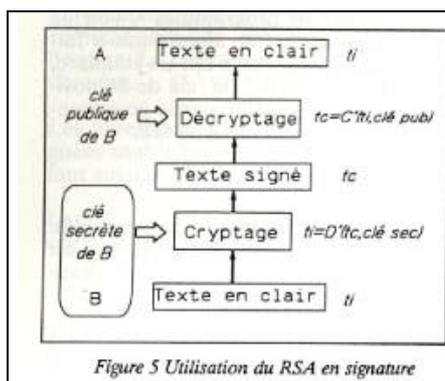
L'algorithme à clé publique le plus répandu est l'algorithme RSA, publié en 1978 [7]. Il est fondé sur les propriétés des nombres premiers. Pour casser l'algorithme, il faut factoriser un grand nombre entier (fourni par la clé publique). L'efficacité du codage sera directement liée à la taille de la clé- avec la technologie actuelle, une clé de 50 chiffres décimaux sera factorisée en quelques heures, une clé de 200 chiffres en quelques milliards d'années. Mais le développement de la puissance de calcul et l'utilisation du parallélisme améliorent chaque année les temps de factorisation. Le choix d'une longueur de clé est typiquement lié au niveau de confidentialité recherché.

L'algorithme RSA est malheureusement beaucoup plus coûteux à mettre en œuvre que le DES, et son usage est généralement utilisé pour le codage de données de taille limitée. C'est en particulier une excellente solution pour la diffusion des clés du DES.

1 Un programme de cryptage DES prend 22 micro secondes par mot (cryptage d'un mots de passe) sur un VAX 8600 [5].

2 Le principe est le suivant : on choisit deux grands nombres premiers p et q au hasard, on calcule $n = p \cdot q$ et $z = (p-1) \cdot (q-1)$, on choisit un nombre d premier avec z , et l'on cherche e tel que $e \cdot d = 1 \pmod{z}$. e et n constituent la clé publique, d la clé secrète. La fonction de cryptage est $tc = (ti)^e \pmod{n}$, le décryptage est $ti = (tc)^d \pmod{n}$. On trouvera des détails sur cet algorithme dans [1, 3, 7].

2.4.2 Autres applications du RSA: signature et authentification d'un document



Une des propriétés remarquables du RSA est de fonctionner parfaitement en sens inverse (figure 5). Si B devient émetteur et crypte un message avec sa clé secrète, A (ou tout autre correspondant) pourra décrypter ce message avec la clé publique de B, et il aura la certitude que l'émetteur est bien B. On réalise ainsi une fonction de signature inimitable, qui permet d'authentifier un message. La signature est liée au contenu, et la falsification est impossible. B ne peut mer avoir crypté, et A a la garantie que l'émetteur est bien B.

Ainsi une information engageant la responsabilité personnelle de son émetteur (ce qui peut être le cas dans un dossier médical), ne sera valable que si elle est signée et authentifiée par son émetteur. Le décryptage étant fait avec la clé publique, tous les correspondants disposant de cette clé pourront accéder au document.

Un double cryptage du RSA entre deux correspondants permettra de réaliser à la fois la confidentialité et l'authentification (figure 6). Cette précaution peut être recommandée notamment pour la transmission des clés du DES, pour garantir à la fois le secret de la clé, et se protéger contre l'émission de fausses clés.

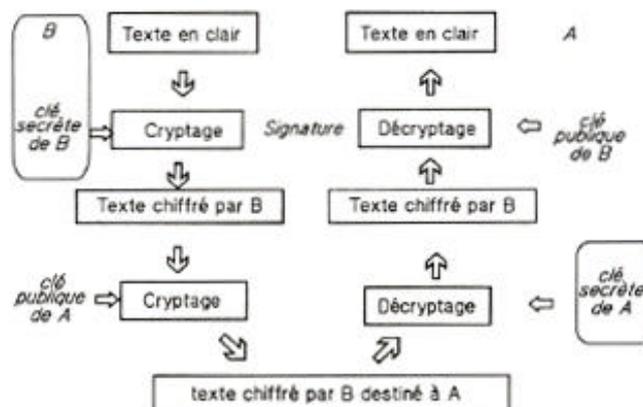


Figure 6 Utilisation du RSA signature et confidentialité

3. Cryptage et sécurité informatique [8]

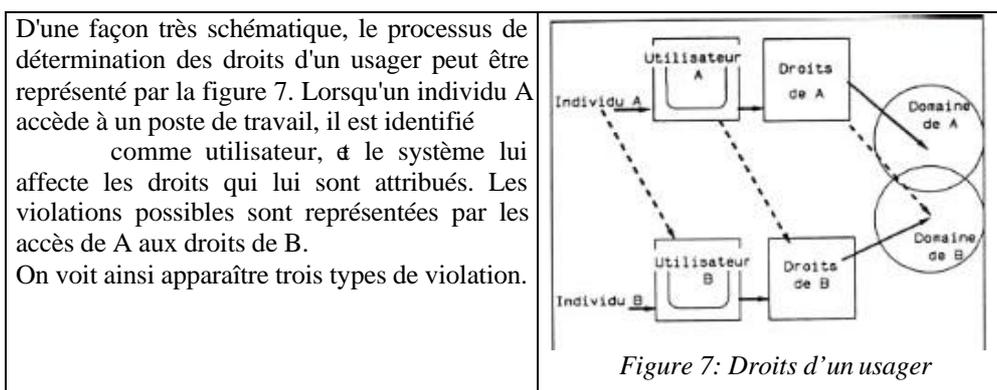
L'utilisation d'un système informatique par un usager est soumise à un ensemble de droits d'utilisation, qui limitent ses possibilités d'accès en lecture ou en écriture aux informations du système, ainsi que les utilisations de certains programmes.

3 Cela n'est pas le cas pour un document : on peut signer un chèque en blanc.

L'**écriture** illégale mettra en cause l'intégrité des données. Dans le domaine médical, cette violation peut avoir évidemment des conséquences graves. Elle peut prendre la forme d'un virus qui cause une dégradation incontrôlée. Mais elle n'est pas forcément d'origine malveillante, et peut simplement être le résultat d'une erreur ou d'une négligence. Ainsi l'utilisateur autorisé, par exemple le praticien, peut laisser un collaborateur entrer à sa place une information engageant sa responsabilité sans la certifier.

La **lecture** peut être à l'origine d'indiscrétions individuelles (violation du secret médical sur un individu) ou globales (utilisation d'un fichier).

L'utilisation illégale du système et de ses programmes peut permettre notamment de modifier les droits d'accès, donc d'autoriser des opérations interdites, ou de mettre en cause la sécurité globale du système .



3.1 Types de violation

3.1.1 L'usurpation d'identité

L'utilisateur A se fait passer pour B. La défense réside à la fois dans le contrôle d'accès physique au poste de travail, l'authentification, et la vérification anthropométrique.

3.1.2 L'utilisation des droits d'un autre usager

Ce type de violation peut intervenir sous plusieurs formes, notamment en utilisant les propriétés internes du système pour modifier les droits et se faire passer pour un autre utilisateur (technique du cheval de Troie), en particulier pour accéder directement aux informations par accès logique ou physique (copie de fichiers). La meilleure défense est la cryptographie.

3.1.3 L'accès indirect aux informations non autorisées

En utilisant les Propriétés logiques des données pour en déduire des informations confidentielles.

4 En 1988, un étudiant de l'université de Cornell, Robert Morris, a introduit un programme dont l'effet a été de paralyser 6000 ordinateurs (SUN et VAX sous UNIX) sur le réseau INTERNET. Ce "vers" a été propagé par erreur, car l'intention (non malveillante) de l'auteur était simplement d'utiliser (illégalement) les ressources des machines du réseau [9].

3.2 Protection d'accès et authentification [1]

L'utilisateur A, si il peut avoir un accès physique au poste de travail, peut tenter de se faire passer pour B. La procédure classique passe par une identification, suivi d'un contrôle d'authentification (preuve d'identité), classiquement assurée par un mot de passe. Le dialogue habituel entre l'utilisateur et le système s'effectue en deux étapes :

- 1- identification (l'utilisateur fournit son identité)
- 2- authentification (contrôle d'identité) : la forme la plus simple est l'envoi d'un mot de passe.

Cette protection n'est pas très sûre. D'une part, un mot de passe peut être divulgué voir même deviné. D'autre part, les mots de passe sont des informations du système qui sont accessibles à certains utilisateurs privilégiés [5]. La seule solution est alors d'avoir un cryptage des mots clés inaccessible à ces utilisateurs privilégiés. Les procédures d'identification et d'authentification que nous venons de décrire sont bien connues et utilisées par le grand public, par exemple dans l'utilisation des cartes bancaires. L'identification est donnée par la carte, et l'authentification est fournie par le détenteur de la carte au moyen de son code secret.

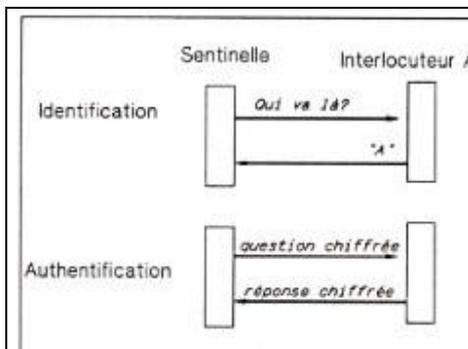


Figure 8 Authentification

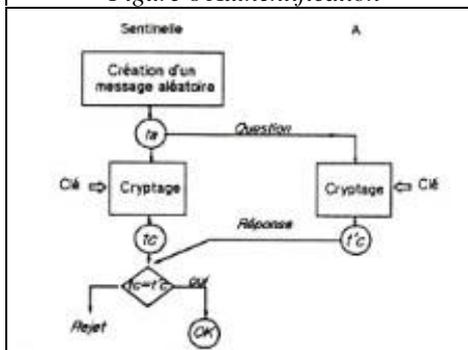
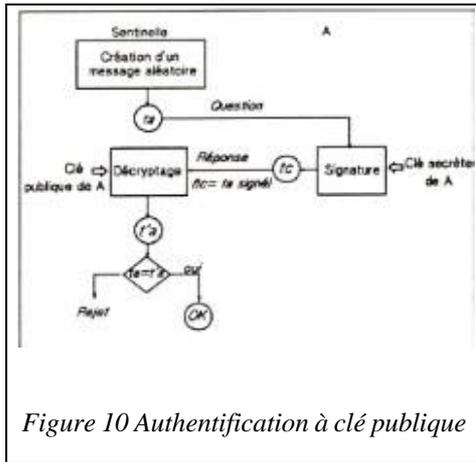


Figure 9 Authentification à clé symétrique

Un espion capable de récupérer une copie du dialogue entre le lecteur de cartes et le système peut parfaitement retrouver l'identité et le mot de passe. La procédure militaire traditionnelle d'échange de mots de passe avec la sentinelle est plus élaborée, car elle repose sur une question et une réponse conventionnelles. Le cryptage est un moyen d'effectuer une telle authentification par dialogue. Le principe est le suivant (figure 8) : les deux interlocuteurs utilisent une cryptage secret commun. Si A veut vérifier l'identité de B, il lui envoie un message quelconque. B doit lui répondre en envoyant le message crypté, et A n'a plus qu'à vérifier que la réponse de B est bien celle qu'il attendait.

L'authentification peut être réalisée avec un cryptage symétrique ou un cryptage à clé publique. Elle utilise le principe de la signature appliqué à un message aléatoire.

En cryptage symétrique (figure 9), la sentinelle envoie un message aléatoire et demande à l'interlocuteur de lui renvoyer crypté. La vérification se fait par comparaison. L'identité est prouvée si l'interlocuteur dispose de la bonne clé de cryptage.

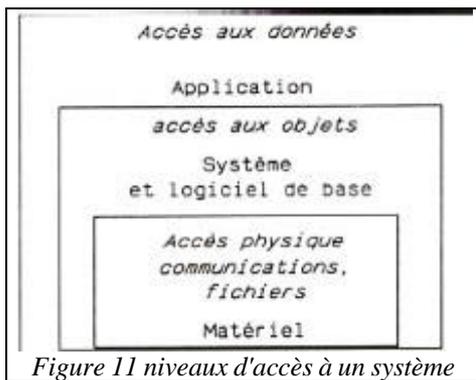


En cryptage à clé publique (figure 9), la sentinelle demande à l'interlocuteur de prouver son identité en signant un message aléatoire. Une copie du dialogue ne fournit aucune information significative : le message envoyé est aléatoire, et le secret réside dans le programme de cryptage. Les procédures d'identification et d'authentification restent liées à l'utilisation d'un mot de passe. Les seules vérifications d'identité sûres sont les contrôles biométriques (empreinte digitale, reconnaissance vocale ou analyse de la rétine) ou les analyses de signature. Ces procédés sont maintenant opérationnels, mais encore très coûteux.

3.3 Protection contre l'utilisation des droits d'un autre usager

3.3.1 Us niveaux logiques de protection

Un système informatique est utilisé par un ensemble d'intervenants de compétences très différentes. Pour simplifier à l'extrême, nous limiterons à trois niveaux ces compétences (figure 11) :



- le niveau utilisateur final, qui utilise le langage externe des applications qu'il utilise,
- le niveau système et logiciel de base, utilisé par les spécialistes (ingénieurs système : ceux-ci ont accès aux programmes et aux données qui assurent la gestion des applications, notamment la gestion des droits d'accès),
- le niveau matériel, qui comporte la partie physique du système et des supports d'information

3.3.2 Le niveau utilisateur final : sécurité logique

La sécurité et la confidentialité entre les utilisateurs finaux est spécifique de l'application. Elle est réalisée par la mise en place de droits d'accès explicites. Cependant, toutes les contraintes de confidentialité ne sont pas forcément réalisables.

3.3.3 Le niveau système et logiciel de base

C'est à ce niveau que l'on trouve les moyens de mettre en œuvre les contraintes demandées par les applications. Le système doit à la fois gérer les objets et les droits d'accès. Les limitations sont de deux natures :

- 1- Le type de contrôle que le système est capable d'effectuer
- 2- La robustesse de ces contrôles

En particulier, les accès privilégiés au système doivent être rigoureusement verrouillés, pour interdire les accès et les intrusions (c'est le procédé utilisé par les virus). Les modules gérant la sécurité doivent être protégés, notamment le fichier des mots de passe [5], les clés de cryptage et les attributions de droits.

Les études sur la sécurité des systèmes ont conduit à proposer une classification et des normes de sécurité. Ces normes reposent sur un modèle hiérarchique (modèle de Bell-La Padula [10]) dont le principe est le suivant : Reprenant les classifications militaires, on définit pour chaque programme son niveau d'habilitation (par exemple : non classifié, diffusion restreinte, confidentiel, secret, très secret) et pour chaque objet ou donnée du système une classification analogue. Un système sûr est tel qu'un utilisateur d'un certain niveau d'habilitation (par exemple "secret"), ne peut lire que les données d'une classification inférieure ou égale ("secret" ou confidentiel")

et ne peut écrire que des informations de niveau supérieur ou égal ("confidentiel" ou "très secret"). données de classification (figures 12, 13).

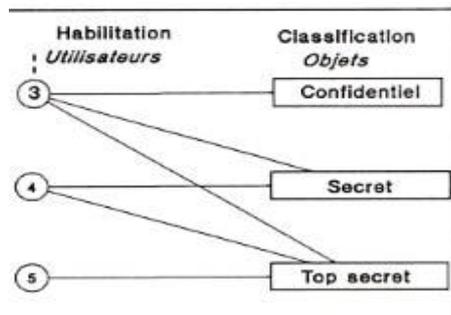


Figure 12 Habilitation en écriture

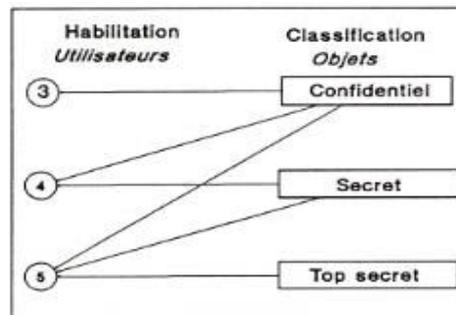


Figure 13 Habilitations en lecture

Cette hiérarchie a pour effet d'interdire tout changement de classification d'une information réduisant son niveau de confidentialité. Les limites de ce modèle sont connues, en particulier, il ne permet pas de se protéger contre certains procédés d'intrusion ("cheval de Troie").

Il existe une norme de classification ("orange book") en 5 niveaux de sécurité (de A à E) des systèmes d'exploitation, classification fondée sur le respect de ce principe.

Le niveau E correspond aux systèmes ouverts (système MS DOS des PC), qui n'offrent pratiquement aucune protection, le niveau A correspond aux systèmes de très haute sécurité.

3.3.4 L'accès indirect aux informations non autorisées

C'est le cas en particulier des applications médicales lorsqu'une utilisation globale (par exemple une étude statistique) peut permettre de retrouver des informations confidentielles par déductions. Plusieurs procédés ont été proposés pour résoudre ce problème [11] : création de base de données fictives modifiées, altération des statistiques ne modifiant pas le résultat global, ou refus de réponse lorsque le nombre de cas est trop faible.

5 Dans le cas déjà cité dans la note 3, le fichier des mots de passe, cryptés avec le DES, était accessible.

En utilisant un dictionnaire de l'anglais courant, le programme était capable de découvrir 50% des mots de passe et donc de se faire passer pour un utilisateur normal. Une erreur du système permettait de passer en mode privilégié, les fichiers et le cryptage étaient accessibles. Enfin, la plupart des mots de passe étaient des mots anglais courants faciles à essayer.[5]

3.4 Les cartes à micro processeur [12]

La carte à microprocesseur est une solution permettant de s'affranchir des limites de l'environnement d'exploitation. Cette carte se présente comme un processeur de sécurité autonome et inviolable. On dispose ainsi d'une gestion de la confidentialité indépendante et secrète. La carte est capable de conserver des informations secrètes, dont certaines sont créées à la fabrication, et d'autres sont créées ou modifiées au cours de l'utilisation.

La carte à microprocesseur Bull CP8 peut disposer de deux algorithmes de cryptage : le DES, et un algorithme de cryptage non réversible (le Télépass). Elle a la capacité de produire des nombres aléatoires. L'algorithme Télépass permet d'effectuer des opérations de signature par vérification (le décryptage n'est pas nécessaire, si l'on utilise un schéma de type signature à clé symétrique). Inviolable, elle permet notamment de stocker une clé secrète, de créer et de gérer la diffusion des clés du DES, d'effectuer les authentifications.

La production de nombres pseudo-aléatoires permet de modifier le cryptage sans communiquer la clé. Chaque cryptage utilise un nombre aléatoire, calculé à partir de la clé symétrique. Cette technique permet d'éviter une stabilité du cryptage et de réduire les risques de découverte de la clé. Le récepteur doit évidemment utiliser le même algorithme, d'une façon synchrone.

Pour utiliser pleinement les possibilités de la carte, il est nécessaire d'en généraliser l'utilisation, de façon à disposer d'une structure de famille de cartes homogènes autour d'une carte mère, véritable gestionnaire des clés. En effet, les algorithmes disponibles sont des algorithmes à clé symétrique. Il est donc nécessaire d'avoir des clés secrètes partagées avec le même niveau de sécurité, donc d'utiliser des familles de cartes pour ne plus dépendre de l'environnement système.

Par ailleurs, l'identification et l'authentification ne portent évidemment que sur la carte elle-même, et dépendent de l'utilisation d'un mot de passe, point faible qui ne peut être éliminé qu'au moyen de vérifications de type biométrique.

4. Conclusion

Les applications médicales font partie des domaines où la sécurité est critique, dans la mesure où les informations traitées concernent la vie humaine. Les risques sur la fiabilité des données interviennent aussi bien à la création (certification d'une donnée) que lors des stockages et de la transmission, où il convient de les protéger contre les modifications, qu'elles soient accidentelles ou dues à la malveillance. En permettant l'authentification et la signature, le cryptage est l'un des moyens d'assurer la fiabilité des informations.

Le deuxième aspect concerne la confidentialité des informations du dossier médical, mais aussi de certaines informations globales, car les risques d'indiscrétion peuvent aussi bien porter sur des informations individuelles que globales (vol d'un fichier pour des motifs économiques).

Pour assurer la sécurité, les techniciens offrent un arsenal dans lequel le cryptage joue un rôle important, mais comporte des inconvénients majeurs. Nous avons vu que le cryptage le plus sûr (le RSA) implique des coûts interdisant son utilisation systématique. Par contre, il peut permettre d'effectuer des cryptages ponctuels de confidentialité (par exemple sur l'identité des personnes) ou de signature (informations à certifier). Seul le DES, dont on a vu qu'il supposait une gestion très sécurisée des clés, peut être envisagé pour effectuer des cryptages systématiques. En pratique, son usage est surtout répandu dans les transmissions de données, secteur qui n'est peut-être pas le plus critique dans les applications médicales.

La sécurité d'un système ne peut être étudiée à travers une seule technique. La diffusion rapide de l'informatique, la généralisation des systèmes répartis et l'utilisation des systèmes ouverts ont fait apparaître des nombreuses possibilités d'accès à un système, et de nouveaux prédateurs. Ainsi, certains systèmes disposent encore de défenses fondées sur des hypothèses périmées, compte tenu de la puissance disponible sur les équipements individuels.

Devant cette complexité, la sécurité doit être analysée globalement, pour présenter un système de défense homogène : certains procédés coûteux (cartes à microprocesseurs, contrôles biométriques) n'ont de sens que si l'ensemble des protections est du même niveau. Cela suppose notamment une démarche systématique d'analyse des risques et de recherche des points faibles, la mise en place d'une structure sécurité, sans doute l'application d'une méthode [13].

Références

- [1] Lamere JM. In: *La sécurité des réseaux* Paris : Dunod Informatique 1987 243-311
- [2] Tanenbaum A. In: *Réseaux, architecture, protocoles, applications* Paris : Interédition 1990 636-671
- [3] Lempel A. Cryptology in transition *ACM Computing survey* Vol 11 No 4 pp. 285- 304
- [4] Simmons G. Symetric and asymmetric encryption *ACM Computing survey* Vol 11 No 4 pp. 305-330
- [5] Seeley D. Password cracking : a game of wits *Communication of the ACM* Vol 32 No 6 June 1989 pp. 689-699
- [6] Jan C. et Sabatier G. In: *La sécurité informatique* Paris : Eyrolles 1989
- [7] Rivest RL. Shamir A. Adleman L. "A method for obtaining digital signature and public key cryptosystems" *Communication of the ACM* Vol 21 No 2 February 1978 pp. 120-126
- [8] Derming DE. et Derming PJ. Data Security *ACM Computing Surveys* Vol 11 No 3 Sept 1979 pp. 227-250
- [9] Spafford E. The Internet worm : crisis and aftermath *Communication of the ACM* Vol 32 No 6 June 1989 pp. 678-688
- [10] MC Lean J. The specification and modeling of computer security *Computer* Vol 23 No 1 Jan 1990 pp.9-16
- [11] Adam NR. et Wortmann JC. Security -control methods for statistical databases A comparative *study ACM Computing Surveys* Vol 21 No 4 Dec 1989 pp. 515-556
- [12] Bull CP8 La carte MD APF 64 K EPROM Réf. TD 0143 F 01, La carte CP8 Réf. IG 0170 F 01
- [13] Lamere JM. *La sécurité informatique, approche méthodologique* Dunod 1986