CCETT
CENTRE COMMUN D'ÉTUDES DE TÉLÉDIFFUSION
ET TÉLÉCOMMUNICATIONS
Groupement d'Intérêt Économique
régi par l'ordonnance du 23 septembre 1967

EPT/02/057/88/LG

Enjeux de la normalisation internationale et perspectives en matière de sécurité

par

Louis C. GUILLOU
CCETT (Rennes)

****

Ce document fait partie de notre série :

le masque et la puce

1. Introduction

Le monde extérieur négocie des transactions avec la puce dans la carte par le biais d'une **interface** à six contacts : MAS, VCC, VPP, HOR, RAZ et E/S. Par rapport à un voltage de référence fixé par le contact MAS, des signaux électriques convenables doivent être fournis sur quatre contacts : RAZ pour l'initialisation, HOR pour l'horloge, VCC pour la tension de fonctionnement et VPP pour la tension d'écriture, afin que la puce puisse échanger des données dans les deux sens, à l'alternat sur le contact E/S.

Les performances des cartes évoluent à l'abri de cette interface : de nouveaux masques sont développés pour une même puce, et de nouvelles puces sont mises au point, tout cela sans que soient remis en cause les terminaux et les infrastructures des systèmes utilisant les cartes. Cette évolution des puces va de pair avec l'évolution générale de l'industrie des circuits intégrés : microprocesseurs et mémoires.
Soit un facteur dix tous les cinq ans !

Note : Les cartes passives ne jouissent pas de cette évolutivité ; ainsi, les normes de pistes magnétiques définissent les moindres détails : les niveaux de magnétisation, le nombre de bits et leur signification ... Toute évolution technologique y est alors impossible sans remettre en cause les normes.

**Pour la réussite commerciale de nos développements,
la normalisation de l'interface est bien un enjeu stratégique ;
cela, nos adversaires l'ont bien compris.**

Les gains en puissance des puces viennent renforcer des sécurités qui restent l'apanage des ordinateurs, telles que choix et calculs complexes. Au contraire, avec seulement des caractères embossés, des pistes magnétiques ou des enregistrements optiques, les cartes passives ne disposent pas de ces possibilités.

* * La **sécurité physique** des cartes à puce repose sur la difficulté qu'il y a à en lire le contenu et à l'altérer. Ceci est intimement lié aux technologies et aux interconnections des mémoires dans la puce : la ROM (Read Only Memory) où sont inscrits les programmes à la fabrication de la puce, la RAM (Random Access Memory) utilisée pour exécuter ces programmes et la PROM (Programmable ROM, Electrically PROM, Electrically Erasable PROM) recelant les évènements, les paramètres et les secrets de la puce.

Seule une conception soignée de la puce permet la sécurité physique des cartes à puce.

* * La **sécurité logique** des cartes à puce repose sur la puissance de traitement du microprocesseur de la puce et sur les algorithmes mis en œuvre dans le masque de la puce. La cryptologie est un élément essentiel dans l'évolution de la sécurité des cartes à puce.

Seule une conception soignée du masque permet la sécurité logique des cartes à puce.

Cependant, **la sécurité totale n'existe pas**, dans les cartes à puce, pas plus que dans n'importe quel autre ordinateur. Dans la mise au point d'une application, le concepteur doit considérer attentivement les conséquences du viol d'une carte. Les secrets dans chaque carte doivent être, autant que faire se peut, individualisés et diversifiés, liés à l'identité de la carte et de son porteur ; ainsi, le viol d'une carte entraine une attaque contre un usager sans remettre en cause le système tout entier. En restreignant les espérances de gain du fraudeur, une telle stratégie réduit d'autant les menaces contre le système.

Aujourd'hui, la normalisation des cartes ne prend pas encore en compte la sécurité des puces. Aussi, dans les batailles internationales, une grande part de nos principaux avantages ne figure pas encore à notre crédit. Si nous ne les exploitons pas correctement, ces avantages se retourneront contre nous. Nos concurrents ne se gênent pas pour ironiser sur nos solutions trop compliquées, voire incompréhensibles. Nous avons un grave problème de communication pour expliquer en langage simple ces avantages décisifs.

**Il faut que nous sachions tirer avantage de notre considérable avance
en matière de sécurité ; nos adversaires ne les ont pas encore compris.**

2. Les puces et les enjeux de la normalisation

21. Les puces et leur évolution

Deux puces en nMOS (3.5 µm) sont aujourd'hui largement répandues dans les cartes bancaires produites par BULL CP8 et PHILIPS (plus exactement TRT-TI) suite à des accords croisés entre les deux sociétés :

-Depuis 1981, MOTOROLA Inc. produit une puce de 18 mm^2 dans son usine d'EAST KILBRIDE près de GLASGOW : UC de 6805, 36 octets de RAM, 1.6 koctet de ROM, 1 koctet d'EPROM.

-Depuis 1985, THOMSON EUROTECHNIQUE produit une autre puce de 18.5 mm^2 dans son usine du ROUSSET près d'AIX-en-Provence : UC de 8048, 44 octets de RAM, 2 koctet de ROM, 1 koctet d'EPROM.

Deux versions de ces puces ont été produites en améliorant la densité d'intégration des mémoires :

-pour MOTOROLA, UC de 6805, 52 octets de RAM, 2 koctet de ROM et 2 koctet d'EPROM ;

-pour THOMSON, UC de 8048, 64 octets de RAM, 3 koctet de ROM et 4 koctet d'EPROM.

Cependant, ces deux versions arrivent commercialement un peu tard. Il vaut mieux miser sur la nouvelle technologie CMOS en finalisation dans ces mêmes sociétés.

Note : Dans les circuits Européens, la RAM est dynamique, il y a des caractères témoins dans la mémoire programmable électriquement, et la ROM est écrite par implantation ionique.

A ces circuits Européens, il faut ajouter deux circuits Japonais en CMOS avec EEPROM :

-HITACHI a développé en 1986 le 65901. Cette puce de 29 mm^2 est trop grande pour une production de masse : c'est plutôt une maquette de faisabilité. L'UC est un RISC (Reduced Instruction Set Computer) avec : 128 octets de RAM, 3 koctet de ROM et 2 koctet d'EEPROM.

-OKI dispose du 62580 conçu aux Etats Unis par CATALYST. OKI peut vendre directement ce circuit de 22,5 mm^2 qui a les mêmes capacités : 128 octets de RAM, 3 koctet de ROM et 2 koctet d'EEPROM.

Note : La sécurité de ces puces Japonaises laisse à désirer : la RAM est statique (l'analyse pas à pas est possible) ; il n'y a pas de caractère témoin dans l'EEPROM (on peut effacer et "recycler" les puces) ; la ROM est dessinée par masque (elle se lit donc aisément).

La technologie Européenne évolue aussi vers le CMOS à 1.5 µm, voire à 1.2 µm. La réduction en consommation d'énergie et l'augmentation en densité d'intégration sont considérables.

Trois autres puces Européennes en CMOS arrivent cette année sur le marché :

-MOTOROLA a mis au point dans son usine d'ECOSSE une puce de 18.5 mm^2 en fin d'homologation par BULL CP8. Elle comporte une UC de 6805, 128 octets de RAM, 4 koctet de ROM et 8 koctet d'EPROM.

-SGS a produit l'automne dernier un premier silicium en collaboration avec SCHLUMBERGER Industries qui, avec la carte M64, entre ainsi dans le club des fabricants de cartes à microprocesseur. Ce composant comporte une UC de Z9, 256 octets de RAM, 4 koctet de ROM et 8 koctet d'EPROM.

-THOMSON SEMI-CONDUCTEURS (anciennement EUROTECHNIQUE) finalise un premier silicium pour Avril prochain : UC de 6805 cette fois, 224 octets de RAM, 4 koctet de ROM et 8 koctet d'EPROM, du moins pour ce premier silicium. La surface de silicium atteindra 20 mm^2 pour une gravure à 1.5 µm ; après une réduction (shrinking) à 1.2 µm, cette surface tomberait à 16 mm^2.

Note : Cet ambitieux projet de THOMSON consiste à proposer une gamme de circuits avec des compromis sur la répartition des tailles de mémoires pour rester au dessous des 20 mm^2 avec les valeurs maximum suivantes : jusqu'à 1 koctet de RAM, jusqu'à 16 koctet de ROM, jusqu'à 24 koctet d'EPROM,et jusqu'à 2 koctet d'EEPROM. Le mariage d'EPROM et d'EEPROM sur la même puce est prévu en 1990.

22. La bataille technologique

Les enjeux des batailles technologiques sur les puces sont considérables. Et les deux points suivants, soulignés par la Normalisation à l'ISO, méritent une attention toute particulière dans le déroulement de cette bataille : technologies de mémoire programmable et fréquences de référence.

Dans cette bataille, les Japonais risquent peu : ils n'ont pas encore développé grand chose à échelle industrielle. **Quels sont nos atouts, à nous qui avons tout à y perdre ?**

Les technologies de mémoire programmable
Les Japonais nous accusent de pratiquer des technologies "obsolètes", car, disent-ils, l'avenir est exclusivement à l'EEPROM qui seule permet de produire en interne dans la puce la tension de programmation. Dans les spécifications d'interface, ceci se traduit par la disparition du problème de la gestion du contact VPP par les terminaux. Avec une volonté farouche de rejeter les technologies EPROM au nom d'une simplification des terminaux, les Japonais espèrent bien à moyen terme nous expulser de la compétition.

Les puces Japonaises ont une mémoire EEPROM avec production interne de la tension de programmation : cela séduit aujourd'hui les Banques et PTT Allemandes qui lancent leurs premiers développements avec le circuit HITACHI. Ils vont rencontrer quelques problèmes notoires du fait d'une pollution aléatoire de la mémoire EEPROM due à un comportement incontrôlé tant à la mise sous tension qu'à la mise hors tension.

Le choix exclusif pour l'EEPROM est discutable. Et même dans ce cas, il faut se méfier de la suppression de la gestion de VPP : la nouvelle technologie EEPROM-FLASH exige une production externe de la tension de programmation. Quoi qu'il en soit, la technologie EPROM est beaucoup plus mûre, car mieux maitrisée industriellement. Dans ce contexte, les recherches pour mieux utiliser l'EPROM sont très intéressantes (procédés de porte-monnaie à Caen et de tire lire à Rennes). Ne vaut-il pas mieux une grande mémoire EPROM plutôt qu'une petite mémoire EEPROM ? En fait le bon choix, ne réside-t-il pas dans un mélange des technologies sur la même puce ? Les développements annoncés par THOMSON et les mêmes développements non annoncés par la discrète MOTOROLA sont donc également très importants.

La fréquence de référence
C'est la fréquence à fournir à la puce sur le contact HOR pour échanger à 9600 bits par seconde sur le contact E/S. L'AFNOR a toujours proposé 3.57 MHZ, fréquence produite par un quartz utilisé dans les téléviseurs NTSC. Les Japonais et les Allemands proposent 4.91 MHZ, c'est à dire exactement 512 fois 9600 Hertz en nous accusant d'utiliser une fréquence de télévision et non pas une fréquence informatique. Par une remise en cause de nos choix, là encore, les Japonais espèrent à moyen terme nous expulser de la compétition internationale. Jusqu'à Octobre 1986, les deux fréquences 3.57 et 4.91 MHZ figuraient dans les documents normatifs en cours d'élaboration. A la demande pressante de l'ANSI, 4.91 MHZ a été retirée au cours d'une réunion du Sous Comité à Bruges à cette date.

Le choix de 4.91 MHZ est très discutable. Le seul argument ayant quelque base technique est de simplifier l'utilisation d'un circuit sérialisateur-désérialisateur (UART) pour gérer les échanges entre la carte et le monde extérieur. Pourtant, des UARTs très largement utilisées exigent d'autres fréquences : 4.02 et 3.68 MHZ. Les contributions Japonaises demandant à l'ISO des contraintes nouvelles dans la gestion de l'interface démontrent que les UARTs sont très mal adaptées à la gestion d'un contact à l'alternat. Dans cette affaire, ce sont les UARTs qui relèvent d'une technologie "obsolète" : ils ont été conçus il y a plus de dix ans pour un tout autre usage. Aujourd'hui, il vaut mieux utiliser un microprocesseur masqué.

Le choix de 4.91 MHZ pénalise la vitesse des échanges quand on compare deux cartes recevant la même horloge. En effet, une puce à basse fréquence de référence échange à 9600 bits/s dès 3.57 MHZ, alors qu'une puce à haute fréquence de référence n'atteint ce débit qu'à 4.91 MHZ. Autrement dit, même pour une puce apte à fonctionner jusqu'à 5 MHZ, il vaut beaucoup mieux choisir 3.57 MHZ comme fréquence de référence : on fait 40 % de mieux sur le contact E/S, toute chose étant égale par ailleurs.

MOTOROLA considère que 4 MHZ est une étape importante dans les rendements de fabrication, donc dans les coûts des puces ; or 3.57 MHZ est la fréquence la plus répandue au monde à 10 % sous cette barre. C'est sans doute pour cette raison que beaucoup d'UARTs ne sont pas réglées pour 4.91 MHZ.

23. La bataille des protocoles

L'organisation des protocoles d'échange mérite également quelques éclaircissements. Le principe retenu aujourd'hui par tous les pays consiste à organiser les échanges en caractères asynchrones sur le contact E/S. Chaque caractère est constitué d'un moment de départ à l'état A, suivi par huit moments codant un octet, puis un moment codant la parité de l'octet, et enfin deux moments de repos à l'état Z.

La transmission à l'alternat suppose une synchronisation précise des échanges sur le contact E/S. La carte indique dans la réponse à la remise à zéro quelle est sa fréquence maximum de fonctionnement, comment elle va modifier son débit ultérieur, et combien de moments supplémentaires elle souhaite (pour calculer) entre les caractères émis successivement par le coupleur. Ces dispositions sont valables quel que soit le protocole utilisé ultérieurement par la carte. Ce sont des paramètres généraux de l'interface.

L'Addendum 1

Dans notre proposition (Addendum 1), le récepteur peut, durant ces deux moments de repos, émettre un signal d'erreur à l'état A pour demander la répétition immédiate du caractère qui aurait été reçu avec une mauvaise parité. Cette procédure est simple, rapide et efficace, compte tenu du faible taux d'erreur rencontré dans la vaste gamme des systèmes que nous avons déjà largement développés.

Il est évident que, en raison du signal d'erreur, notre proposition empêche l'utilisation des UARTs traditionnelles pour assurer les échanges. Aussi, est-elle vigoureusement contrée par les Japonais qui veulent absolument pouvoir les utiliser. Les Allemands (SIEMENS et GAO) et, à un degré moindre, les Anglais leur emboîtent le pas. L'opinion des Etats Unis est plus contrastée, mais depuis que l'ATT commence à s'intéresser à la normalisation, quelques voix s'y élèvent aussi pour réclamer la possibilité d'utiliser ces UARTs. Pourtant, au vu des problèmes soulevés par les Japonais, on peut se douter que notre choix d'exclure les UARTs est bien le bon.

Quand la carte est en attente de commande, elle accepte cinq caractères successifs pour initialiser la commande. Ensuite, elle prend la main pour répondre par un caractère de procédure afin d'asservir :
-d'une part, l'état du contact VPP en demandant d'établir ou de couper la tension de programmation,
-d'autre part, la transmission d'éventuels caractères ultérieurs pour transférer des données.

Durant une commande, la chaîne de caractères de données est soit entrante (dans la carte), soit sortante (de la carte). Ceci est ressenti par certains de nos collègues comme une intolérable limitation justifiant la mise au point d'un autre protocole où les caractères seraient organisés en blocs (projet d'Addendum 2). Ces mêmes collègues insistent aussi sur la nécessité, à leurs yeux, d'introduire des redondances supplémentaires pour mieux assurer l'intégrité des données.

La seule contrainte que supporte la carte durant une commande est d'émettre un nouveau caractère avant qu'une temporisation WT ne soit dépassée dans le coupleur qui déduirait alors que la carte est muette. Tout caractère reçu par le coupleur réarme cette temporisation. La carte ne tombe jamais en time-out.

L'Addendum 2

Les Japonais ont donc proposé dès 1983 un protocole par bloc, avec des blocs marchant au pas à l'alternat. Ils ont introduit trois temporisations :
•une temporisation CW entre les caractères dans un bloc, pour détecter des anomalies de fin de blocs,
•une temporisation BW entre blocs, pour détecter des mutismes et des blocages,
•un très mystérieux temps de commutation MS, au renversement de la direction de transmission.
Note : Il n'y a jamais eu moyen de savoir officiellement quelle temporisation était gérée par quel temporisateur dans quel sens de transmission. Il est rare d'obtenir des éclaircissements de la part des Japonais, soit parce que la question n'est pas comprise, soit parce que la réponse est incompréhensible.

La temporisation CW ne peut servir au coupleur qu'à détecter un incident qui doit mener à une remise à zéro. Si un autre traitement est envisagé, il y a risque de collision. Une telle temporisation, courte pour être exploitée rapidement, est donc inutile, voire dangereuse. Si on allonge CW, il se confond avec BW. L'AFNOR propose donc de se limiter au cas CW = BW = WT, c'est à dire "temps d'attente du prochain élément de procédure". Les Japonais tiennent pourtant beaucoup à distinguer BW et CW.

Note : Les experts Allemands tiennent aussi beaucoup à CW. Nous venons de comprendre qu'ils utilisent CW pour détecter la fin d'un bloc, car ils n'ont pas confiance dans l'indication de longueur donnée dans le bloc lui-même en raison d' éventuelles erreurs doubles non détectées par la parité du caractère. Cette procédure est pour le moins surprenante de la part d'experts si pointilleux sur la pureté des protocoles.

La temporisation MS n'a toujours pas reçu de justification. Elle provient certainement de défauts de conception du terminal dans la gestion des UARTs. Cela est totalement inacceptable ; car par ce biais, les terminaux Japonais pourraient rejeter les cartes qui ne connaîtraient pas a priori la valeur à utiliser. A part les Japonais, aucun expert à l'ISO ne voit l'utilité de cette mystérieuse temporisation.

Dans le protocole Japonais, nous avons mis en évidence des boucles fatales (dead-locks) : si un bloc de demande de répétition était victime d'une erreur, la carte et le lecteur échangeraient éternellement des demandes de répétition. Les Japonais ont alors proposé de gérer une pile des demandes de répétition. La mise en évidence de ce problème éclaira de nombreux experts sur la qualité des propositions Japonaises.

Nous avons également démontré la lourdeur de la gestion de VPP par un bloc dans un sens suivi par un acquittement dans l'autre sens : le temps nécessaire pour demander d'établir et de couper l'état actif de VPP (quelques dizaines de ms) serait alors très supérieur au temps nécessaire pour écrire (quelques ms).

Cependant le refus français de discuter les propositions de l'AFNOR entrainait un sérieux risque de coalition IBM-RFA-Japon. L'émergence d'un protocole par bloc devenant inévitable, nous avons donc décidé d'y contribuer en faisant présenter une contribution nouvelle à Tokyo en Juillet 1987 par MCTI, la filiale de BULL CP8 aux Etats Unis. Cette contribution qui plait bien à IBM n'est absolument pas un aveu de quelque faiblesse dans nos propositions, mais bien la volonté de mettre au point un protocole "haut de gamme", inspiré du modèle OSI.

Le résultat pratique obtenu aujourd'hui est que les propositions d'IBM, des Japonais et des Allemands sont remises à plat. Tout le monde veut construire un beau protocole en guise d'Addendum 2 en se disant que le temps ne presse pas trop : l'Addendum 1 existe et est suffisamment efficace. En outre, personne ne discute plus les détails de l'Addendum 1, puisqu'il y aura un jour une autre proposition venant compléter les protocoles au cas où apparaitraient des applications non satisfaites par ce premier Addendum.

A San Francisco, à la dernière réunion ISO, nous avons créé une autre menace en faisant une provision pour envisager une évolution de l'Addendum 1 qui concurrencerait directement le projet d'Addendum 2.

Note : Les Allemands ont mis au point entre temps un protocole "national" en attendant que quelque chose soit normalisé par l'ISO. Aujourd'hui, ils veulent introduire une négociation de protocole afin qu'une migration ultérieure soit possible. C'est bien la première fois que les Allemands sont demandeurs ; cela les amènera peut être à un peu plus de compréhension pour nos propres demandes.
Note : Il y a un risque de voir maintenant les Japonais soutenir le protocole national Allemand.

3. La sécurité

Dès la conception des premières puces, il est apparu que deux solutions étaient envisageables pour tester les puces à l'issue de l'élaboration des galettes (wafers) de silicium : ou bien chaque puce porterait une vingtaine de contacts pour que les tests soient menés par un programme extérieur à la puce, ou bien un programme d'autotest serait implanté dans le silicium avec seulement un ou deux contacts supplémentaires pour les tests. Un bilan des surfaces de silicium a fait opter en faveur de l'autotest. Il est heureux que ces considérations économiques aillent dans le sens d'une plus grande sécurité des puces. C'est à mon avis un des rares cas où l'option la moins chère se trouve être justement la plus sécuritaire.

Nous allons examiner quelques uns de ces dispositifs sécuritaires dans les puces qui pour leur plus grande part relèvent du savoir faire des constructeurs, ou même de leurs secrets de fabrication.

Les algorithmes cryptographiques jouent aussi un rôle essentiel dans les cartes à puce. Les puces en nMOS disponibles aujourd'hui ne peuvent supporter que des algorithmes à clé secrète. Dans ce domaine, nous avons une avance considérable sur nos concurrents. Cependant, les puces en cours de mise au point sont tout à fait aptes à supporter aussi des algorithmes à clé publique. Donc, dans notre analyse, nous allons également considérer l'impact des clés publiques, et leur évolution.

En dehors des applications bancaires, il faut que nous sachions mettre en évidence d'autres applications dans lesquelles la cryptologie joue un rôle irremplaçable pour **gérer plus efficacement le système.**
Note : Les Japonais nient tout intérêt à la sécurité et à la cryptologie dans les cartes à puce !!!

31. Les sécurités prises durant la vie de la carte

Les puces étant encore réunies sur la galette de silicium, chaque puce est donc testée par une machine qui active un programme d'autotest dans la puce grâce aux six contacts et à un ou deux autres contacts réservés au test. Quand la puce est déclarée mauvaise, le programme n'écrit rien du tout. Quand la puce est déclarée bonne, le programme coopère avec la machine pour écrire un code de fabrication, un numéro de série, ainsi que des éléments mémoires utilisés ensuite en témoins de non effacement de l'EPROM. Et de toute façon, les contacts de test sont ensuite détruits définitivement (claquage d'un fusible enterré en polysilicium) par la machine de test avant la découpe de la galette. Ainsi, les puces non satisfaisantes se retrouvent définitivement inutilisables : il n'y a pas lieu de faire les poubelles des fabricants de puces.

Les caractères témoins sont des éléments mémoires répartis dans l'EPROM qui surveillent une éventuelle modification (accidentelle ou malveillante) de l'état des mémoires EPROM. Ils sont réglés pour être les plus sensibles de la mémoire EPROM, de manière à être effacés les premiers en cas d'exposition de la puce à des rayonnements X ou ultra-violets. On ne peut plus les rétablir ensuite, car le programme d'autotest n'est plus accessible depuis la destruction du ou des deux contact(s) de test.

Enfin, les bonnes puces sont insérées dans des cartes. Elaboré par une carte lot, le code de fabrication permet de se prémunir contre le détournement d'un lot de puces. Durant toute la vie de la carte, certains dispositifs dans le silicium surveillent systématiquement les conditions de fonctionnement de la carte pour signaler aux programmes de la puce des alertes majeures : maximum ou minimum du voltage admissible sur VCC, maximum ou minimum du voltage actif admissible sur VPP, minimum de la fréquence admissible sur HOR. Certaines autres alertes sont de moindre intérêt : maxima de l'intensité lumineuse et de la température admissibles à la surface de la puce.

Le plan d'adressage des mémoires est contrôlé par un codage précisant les conditions d'accès aux différentes zones en fonction des différentes parties du programme.

32. Les utilisations d'algorithmes à clé secrète

Plusieurs algorithmes à clé secrète sont utilisés aujourd'hui dans les différentes cartes disponibles sur le marché : examinons successivement Télépass 1, TDF (Double Corps), Télépass 2, Vidéopass et DES.
Note : Notre objectif n'est pas de décrire ces algorithmes, mais bien de préciser leurs fonctionnalités et la manière dont ils sont utilisés dans les masques qui les contiennent.

TÉLÉPASS 1 a été mis au point par BULL en 1980. Cet algorithme est à sens unique, en ce sens qu'il n'a pas d'inverse. Il tient sur environ 200 octets de code ROM.

Un résultat R (64 bits) est calculé à partir d'une clé secrète S (96 bits) et du contenu d'un mot ~~d'un mot~~ non secret M (32 bits) écrits dans la mémoire EPROM de la carte, et d'un argument E (64 bits) fourni à la carte. L'adresse du mot M est également fournie en argument à la carte.

Cet algorithme est utilisé dans le **masque M4** qui réalise les cartes bancaires quand il est personnalisé en format B0 et les cartes "publiphone" à microprocesseur quand il est personnalisé en format B03. Chaque carte contient seulement deux clés cryptographiques : une pour la banque, une pour le fournisseur de services. Chaque clé est généralement diversifiée en fonction du numéro de série de la puce. Des cartes mères et des modules de sécurité (MCS) savent reconstituer le secret diversifié des cartes filles à partir d'un secret de base et du diversifiant propre à la carte fille.

Cet algorithme permet de vérifier à distance, de manière interactive, qu'une carte contient bien un mot à une adresse donnée, par exemple pour contrôler qu'une carte porte bien un droit tel qu'elle le proclame. Cet algorithme permet aussi de vérifier qu'une écriture s'est bien effectuée et d'en garder une trace sous la forme d'un certificat qui peut être vérifié en temps différé en cas de contestation.

Cependant, pour contrôler le résultat d'un calcul fait avec Télépass 1, il faut refaire exactement le même calcul en utilisant donc une carte mère ou un module de sécurité MCS. Puisqu'elles sont capables de simuler n'importe quelle carte fille du système considéré, ces cartes mères et ces modules de sécurité méritent des protections très fortes : leur détournement met en péril la sécurité de tout le système.

TDF (Twisted Double Field ; en Français, Algorithme à deux corps entrelaçés) a été mis au point au CCETT en 1981. Cet algorithme inverse un autre algorithme exécutable par un autre microprocesseur à l'extérieur de la carte. L'algorithme dans la carte est "l'inverse à gauche" de l'algorithme externe. Il tient sur environ 300 octets de code ROM.

Un résultat R (61 bits) est calculé à partir d'une clé secrète S (127 bits) inscrite dans la mémoire EPROM de la carte, ainsi que d'un cryptogramme C (127 bits) et d'un paramètre P (23 bits) fournis à la carte comme arguments.

Cet algorithme est utilisé dans le **masque porte-clés PC0** qui sert à controler l'accès à des informations diffusées par ANTIOPE, tel que les magazines boursiers CHRONOVAL et CHRONOPTION commercialisés par les agents de change regroupés dans la société SDIB.

Dans les cartes porte-clés, il faut distinguer clairement les **clés de service** et la **clé émetteur**.

Chaque carte contient plusieurs **clés de service**. Par l'algorithme externe, le diffuseur de services calcule des cryptogrammes d'un mot de contrôle et les diffuse avec les composantes embrouillées du service. Repérée par un identificateur, chaque clé de service voit son usage limité dans chaque carte par des conditions définissant l'état de l'autorisation. Quand la carte contient une clé convenant au service, et quand les conditions indiquées par le paramètre sont compatibles avec l'état de l'autorisation (par exemple, la date indiquée figure bien dans une période d'abonnement), alors la carte reconstruit le mot de contrôle. Ce mot de contrôle sert ensuite à désembrouiller les composantes du service diffusé.
Note : Grâce à la semi réversibilité de l'algorithme, le même mot de contrôle peut être décrit par autant de cryptogrammes qu'il y a de catégories d'audiences autorisées à accéder au service diffusé.

En outre, chaque carte porte une **clé émetteur** qui est diversifiée. Par l'algorithme externe, l'émetteur de cartes calcule des cryptogrammes personnalisés destinés à produire une action particulière dans une carte, telle que le renouvellement d'un abonnement pour le mois suivant. La carte traite ce cryptogramme en utilisant sa clé émetteur, puis considère le résultat : si le résultat se compose de la répétition d'un motif, cette redondance persuade la carte d'avoir affaire à son émetteur. Elle effectue alors le traitement indiqué par le motif. Chaque carte sait reconnaître son maître, lequel est en réalité l'émetteur de la carte, et non pas son porteur.
Note : Grâce à la semi réversibilité de l'algorithme, on peut ainsi gérer efficacement et de manière totalement sécurisée certains droits à distance dans les cartes, sans qu'il y ait lieu de faire d'hypothèse sur la sécurité du réseau et des terminaux utilisés.

Des dispositifs mères exécutent l'algorithme externe pour calculer les cryptogrammes tant de mots de contrôle que de gestion de droits. Suivant la terminologie développée au sein de l'UER (Union Européenne de Radiodiffusion), les cryptogrammes de mots de contrôle constituent la messagerie de contrôle de titres d'accès, alors que les cryptogrammes de gestion de droits constituent la messagerie de gestion des titres d'accès. La messagerie de contrôle accompagne bien sûr le signal diffusé, mais la messagerie de gestion peut aussi bien emprunter d'autres voies, telles que le Minitel ou la Poste.
Note : Cependant, en matière de certificats, l'algorithme TDF n'est pas plus avantageux que Télépass 1.

TÉLÉPASS 2 a été mis au point par BULL CP8 en 1984 en tenant compte des enseignements procurés par les deux développements précédents : Télépass 1 et TDF. Cet algorithme réversible développé pour le masque bancaire **B1** est la propriété du GIE des Cartes Bancaires ; et de plus, la fonction inverse est programmée directement dans le masque bancaire **B2** (une version étendue de B1).

VIDÉOPASS a été également mis au point par BULL CP8 en 1984. Egalement réversible, la fonction inverse est programmée directement dans le masque des cartes porte-clés **PC1**.

Ces deux algorithmes, Télépass 2 et Vidéopass, offrent suffisamment de similarités pour être considérés ensemble dans notre analyse. Ils permettent à la fois :

-les services de Télépass 1 pour la gestion de certificats,

-les services de TDF pour la gestion de mots de contrôle et de droits dans les cartes.

Décrivons les interfaces de l'algorithme Vidéopass : un résultat R (64 bits) est calculé à partir d'une clé secrète S (128 bits) inscrite en mémoire EPROM, d'un cryptogramme C (64 bits) fourni en argument, et d'un paramètre P (32 bits) qui est : -soit un argument fourni à la carte, -soit un mot non secret inscrit dans la carte figurant à une adresse fournie en argument.

Ces deux algorithmes sont **dissymétrisés**. Pour que la même carte puisse exécuter les deux sens de l'algorithme avec la même clé, il faut que la carte porte deux enregistrements de la clé : dans un enregistrement, les octets sont écrits dans l'ordre inverse de l'autre enregistrement. Cette dissymétrisation donne des propriétés très intéressantes : une carte mère peut vérifier les certificats fournis par des cartes filles sans être capable pour autant de simuler ces cartes filles en forgeant des faux certificats.
Note : Il est surprenant à première vue d'obtenir une telle propriété pour un algorithme à clé secrète, car cela rappelle fort les clés publiques. En fait, dans les deux cas, il y a un problème complexe et difficile à résoudre. Dans les chiffres à clés publiques, la sécurité est basée sur la difficulté qu'il y a à résoudre certains problèmes (la factorisation des grands entiers, par exemple) ; dans ce système à carte à puce avec algorithme dissymétrisé, la sécurité est basée sur la difficulté qu'il y a à investiguer le silicium pour y lire les secrets écrits dans l'EPROM.

Un **verrou d'orientation** écrit en EPROM restreint les cartes filles B2 et PC2 à un seul sens d'algorithme, alors que les cartes mères gardent les deux possibilités. Les cartes B1 sont exclusivement filles.

DES a été implanté en 1986 dans un masque **D1** par la Société TRT-TI du groupe PHILIPS sur environ 800 octets de code ROM. L'algorithme est également dissymétrisé pour obtenir les propriétés mentionnées ci-dessus. Ce masque D1 est plutôt de la famille porte-clés, bien que les modalités de gestion des droits dans les cartes ne soient pas aussi élaborées que dans PC0 ou PC1. Par contre, la hiérarchie des clés est bien plus élaborée dans D1 que dans B1 ou PC1.

DES a été implanté l'an dernier sur la puce SGS dans un masque **M64** par SCHLUMBERGER Industries.

Note : DES est bien moins adapté à la gestion des cartes, car il n'admet pas de paramètre supplémentaire tel que celui qui fait l'avantage de Télépass 2 et de Vidéopass. DES est ainsi moins efficace pour gérer des droits dans une carte : il faut alors l'utiliser deux fois pour gérer un paramètre supplémentaire. Cette remarque est très importante, et dans les réalisations utilisant DES, il faut considérer très attentivement la manière dont ce problème a été résolu.

TÉLÉPASS 3 versus DES ??

Un nouveau masque **MP** est en cours de développement chez BULL CP8. Et à la demande du consortium Européen EUROMAC préparant la diffusion directe par satellite, dans le cadre de la mise au point de l'accès conditionnel en D2-MAC Paquets, un nouveau masque **PC2** pourrait être développé par BULL CP8 dans la famille des cartes porte-clés. Ce masque PC2 pourrait aussi servir aux besoins de l'accès conditionnel sur les réseaux de vidéocommunication. Ces deux masques MP et PC2 sont développés pour la génération de puces en CMOS ; le DES pourrait très bien y être proposé, suivant les spécifications des clients, bien que je préfère nettement y voir utiliser un algorithme secret dans la suite de Télépass 2.

Note : Dans tout nouveau masque à produire, comme par exemple PC2 et MP, on pourrait introduire un ***verrou de désarmement*** *qui rendrait systématique le test de présence de la redondance à l'issue de tout calcul cryptographique dans la carte. C'est déjà le cas pour la gestion de titres d'accès. Ce serait aussi le cas pour la distribution de mots de contrôle. Si la redondance n'est pas là, la carte devient muette, ce qui peut s'interpréter : soit comme un résultat nul (concept de fonction nulle presque partout), soit comme la reproduction de la valeur d'entrée (permutation constante presque partout). Ainsi, une carte fille ne pourrait être utilisée qu'en communication avec une carte mère. Ne pouvant plus communiquer entre elles, les cartes filles ne peuvent plus servir à des mises à la clé entre elles.*

Le positionnement de ce verrou peut se comparer au bris du percuteur d'un fusil, la carte fille perdant ainsi sa qualité d'arme de guerre. Une fois le verrou activé, les cartes filles pourraient donc être déclassifiées, la classification étant réservée aux cartes mères et aux cartes filles non désarmées.

33. La première utilisation d'algorithme à clé publique

Des cartes bancaires à puce sont aujourd'hui émises massivement en France et en Norvège. Chacune de ces cartes porte une **valeur d'authentification** : c'est un entier n de 320 bits calculé par un dispositif CAMELIAS et inscrit dans la carte durant la personnalisation par l'émetteur de cartes.

Note : La valeur d'authentification est également utilisée sur les cartes d'abonnés "publiphone".

Durant toute transaction bancaire, tant sur terminal point de vente que sur certificateur, la machine du commerçant lit la valeur d'authentification de la carte et l'élève au cube dans Zn, l'anneau des entiers modulo un nombre n de 320 bits publié par l'autorité bancaire responsable. Le test est réussi quand le résultat obtenu répète un motif de 160 bits constitué par les **paramètres d'identification** bancaire figurant en clair dans la carte : ce sont le numéro de série de la puce sur 44 bits, le numéro de compte bancaire sur 76 bits, un code usage sur 8 bits et une période de validité sur 32 bits.

Il fallait disposer des facteurs du nombre n pour **habiliter** ainsi la carte, et cette factorisation est précieusement gardée secrète par l'autorité bancaire responsable de l'émission des cartes. Les dispositifs CAMELIAS permettent aux organismes émettant des cartes bancaires de protéger et d'utiliser les facteurs de ce nombre n pour calculer les valeurs d'authentification. Par contre, tous les terminaux bancaires connaissent le nombre composé n. On considère que la factorisation d'un nombre de 320 bits est encore aujourd'hui pratiquement impossible.

La carte contient donc un nombre A (320 bits), la valeur d'authentification. Dans l'anneau Zn, le cube de A est un nombre J (320 bits) constitué par la répétition d'un motif I (160 bits) décrivant les paramètres d'identification de la carte bancaire. Le motif I est appelé **Identité** de la carte, et l'entier J (le motif I doublé par une redondance appelée **ombre**) est appelé **Identité ombragée** de la carte.

$$\text{Soit } A^3 \bmod n = J = I \cdot (2^{160}+1)$$

Cette méthode appelle trois premières remarques plutôt triviales :

-Compte tenu des progrès de la factorisation, les nombres utilisés deviennent un peu courts : un nombre de 297 bits a été factorisé l'été dernier en 24 mois d'UC sur des stations de travail SUN 3.

-L'exposant 2 pourrait avantageusement remplacer l'exposant 3, moyennant quelques précautions techniques pour choisir n et définir une permutation à trappe sur un sous ensemble de l'anneau Zn.

-La structure de la redondance pourrait être améliorée, afin d'éviter une structure multiplicative.

Dans le cadre d'une normalisation internationale d'un schéma de **signature avec ombre**, tous ces aspects sont actuellement examinés au sein du comité technique ISO-IEC/JTC1 par le groupe de travail SC20/WG2 traitant des techniques à clé publique. On peut donc considérer que les spécifications des schémas de signature avec ombre sont en bonne voie de trouver une solution normalisée au niveau international au sein de l'ISO, ce qui résout les trois remarques triviales précédentes. Ces travaux sont décrits dans un avant projet de norme ISO identifié par l'appellation DP9796.

Note : Les schémas de signature avec ombre sont définis par la conjonction d'une structure de redondance et d'une permutation à trappe. Grâce à des règles de redondance, les paramètres d'identification sont transformés en un élément d'un ensemble sur lequel est défini une permutation à trappe. La valeur d'authentification qui en résulte "habilite" la carte, car en vérifiant ce résultat "signifiant", on y révèle les paramètres d'identification munis d'une redondance appelée "ombre".

Mais il y a une autre remarque bien plus subtile : si le procédé décrit ci-dessus gêne bien la production d'identités frauduleuses, il ne suffit pas à empêcher la production de clones de cartes existantes. Comment convaincre la machine du commerçant que la carte dispose de la bonne valeur d'authentification sans que la machine n'apprenne rien sur cette valeur qui resterait alors secrète dans la carte ?

Les ingrédients nécessaires à la mise en œuvre d'un procédé à "apport nul de connaissance" sont réunis :

-D'une part, il y a l'énoncé d'un problème complexe : il est constitué par les paramètres d'identification de la carte, par le nombre composé n publié par l'autorité responsable, ainsi que par le schéma normalisé de signature avec ombre (règles de redondance et exposant public) ;

-D'autre part, il y a la solution du problème complexe : elle est inscrite dans la carte sous la forme d'une valeur d'authentification.

Un procédé à **apport nul de connaissance** ferait disparaître la menace de production de clones.

Note : L'autorité est la seule à disposer de la factorisation du nombre composé. Les dispositifs CAMELIAS permettent de protéger et d'utiliser ces facteurs. Bien entendu, s'il y a plusieurs autorités en concurrence, chacune dispose de son propre nombre composé ; et une même carte peut porter plusieurs valeurs d'authentification correspondant à des objectifs divers, voire même à des autorités concurrentes.
Note : L'opération d'habilitation d'une carte n'est pas systématiquement liée à l'émission de la carte.

4. L'orientation prévisible de l'authentification des cartes

41. Une technique interactive à apport nul de connaissance

Le procédé suivant qui "approfondit" la valeur d'authentification a été découvert suite à divers travaux menés récemment au CCETT à Rennes et au Laboratoire de Recherche de PHILIPS à Bruxelles. Une demande de brevet français a été déposée conjointement par le CCETT et le Laboratoire de Recherches de PHILIPS. Ce brevet doit être ensuite étendu à l'étranger.

Note : Par rapport au célèbre procédé de FIAT et SHAMIR où le compromis implique l'enregistrement de plusieurs valeurs d'authentification dans chaque carte et plusieurs répétitions de la procédure, notre procédé se déroule en une seule séquence et exige l'enregistrement d'une seule valeur d'authentification.

Un nombre composé n est choisi par la banque qui en garde précieusement secrète la factorisation. Un exposant public p est choisi assez grand, de l'ordre du niveau de sécurité requis par l'application. Le nombre J est l'identité ombragée de la carte. Chaque carte porte sa valeur d'authentification B, inverse et profonde, qui vérifie la relation simple :

$$B^p \cdot J \bmod n = 1$$

La procédure de vérification se déroule en trois phases, sans répétition :

1 - ***La carte proclame son identité et donne une valeur test.***

A chaque vérification, la carte tire au hasard dans l'anneau Zn un nouvel élément r qu'elle garde secrète et en calcule la puissance $p^{ième}$ qu'elle transmet au vérificateur en guise de valeur test T (= r^p mod n).

2 - ***Le vérificateur pose une question.***

Le vérificateur tire au hasard un entier d entre 0 et p-1, et le transmet à la carte. La carte calcule le produit dans Zn du nouvel élément r par la $d^{ième}$ puissance de la valeur d'authentification B, et le transmet au vérificateur en guise de valeur témoin t (= $r \cdot B^d$ mod n).

3 - ***La carte dévoile une valeur témoin.***

Pour vérifier le témoin t, le vérificateur doit retrouver la valeur test T en calculant dans Zn le produit de la puissance $p^{ième}$ du témoin t par la $d^{ième}$ puissance de l'identité ombragée J (soit : $J^d \cdot t^p$ mod n).

La procédure est bien à **apport nul de connaissance sur la valeur d'authentification**, car :

-Le fraudeur chanceux qui devine le tirage de l'exposant d a une stratégie gagnante évidente : il tire t au hasard et fait à l'avance les calculs de vérification pour en déduire T. Il peut ainsi tromper le vérificateur s'il a bien deviné le tirage de la question d.

-La connaissance de deux valeurs possibles t' et t" du témoin pour la même itération implique la connaissance d'une puissance de B de rang [d'-d"], inférieur à p, ce qui équivaut à la connaissance de B à la puissance PGCD(p, d'-d"), c'est à dire à la connaissance de B si p et d'-d" sont premiers entre eux (ce qui est toujours le cas quand p est premier). Le fraudeur ne peut préparer anticiper qu'une seule question.

-Un juge ne peut pas distinguer entre des enregistrements de transactions de vérification et des mascarades produites en choisissant d'abord les valeurs témoin t, puis les questions d, avant d'en déduire les valeurs test T. Le vérificateur récupère des données indiscernables de données qu'il aurait très bien pu fabriquer tout seul, sans aucune intéraction avec la carte !

Pour passer avec succès la procédure, un imposteur doit deviner un exposant parmi p : la valeur de p est aussi appelée la "profondeur" de la valeur d'authentification. Et un imposteur n'ayant pas de connaissance (même partielle) du secret ne pourra pas préparer à l'avance plus d'une valeur possible pour le témoin t. Puisqu'un imposteur n'a qu'une chance de succès sur p à chaque transaction, on peut se contenter d'un seul traitement quand l'exposant p est de l'ordre de grandeur du **niveau de sécurité** souhaité.

Pour une transaction de vérification, le niveau de sécurité dépend de l'application : personne ne présentera de faux papiers au gendarme avec une chance sur 1000 de ne pas être pris ; par contre, certains n'hésiteront pas à tester à distance et incognito l'accès à un ordinateur même en entreprenant 1000 essais, pourtant ils ne tenteront pas un million de connexions ; enfin, avec un PC, on peut tout à fait tenter un million d'essais pour chercher au hasard à forger une signature, mais même avec les ordinateurs les plus puissants, il est illusoire d'envisager de tenter 10^{18} essais.

Dans cette optimisation, la mémoire programmable utilisée dans la carte et les données échangées à l'interface de la carte sont réduites au "minimum minimorum" au détriment de quelques calculs supplémentaires dans la carte (un peu moins de trois fois plus !) par rapport au célèbre procédé proposé par Amos FIAT et Adi SHAMIR consistant à multiplier les valeurs d'authentification et les itérations.

L'approfondissement d'une valeur d'authentification est donc particulièrement bien adapté aux cartes à puce, en minimisant les ressources requises en mémoire et en transmission.

42. Une méthode de signature basée sur l'identité

Dans un processus d'authentification, le vérificateur a un comportement probabiliste ; il est donc assez surprenant d'envisager de lui substituer une fonction de compression à sens unique "anti-collision", car une telle fonction est par essence déterministe. Bien que divers travaux aient porté sur les familles de fonctions à comportement statistiquement indiscernable de celui de variables aléatoires, ainsi que sur les fonctions résistant à la mise en collision, il faut bien reconnaître que l'on n'a pas aujourd'hui sur la table une proposition de fonction de compression qui soit prête à être normalisée.

Supposons pourtant que l'on dispose d'une bonne fonction de compression qui résiste à la mise en collision. On peut alors laisser l'usager s'interroger lui-même, en calculant directement la question grâce à la fonction de compression appliquée à la valeur de test et au message. Ainsi, on obtient les spécifications d'une nouvelle méthode de signature à partir des techniques décrites précédemment avec apport nul de connaissance sur une valeur d'authentification dans la carte. Peut-on parler de méthode non interactive à apport nul de connaissance ? Ce qui suit éclaircit ce point.

Pour signer le message M, on le complète par un appendice constitué de la manière suivante :

-1-Dans l'anneau Zn, la carte calcule en guise de **test T** la puissance $p^{ième}$ d'un nouvel élément r qu'elle tire au hasard à chaque traitement ;

-2-Le terminal de l'usager (ou la carte, suivant des contraintes propres à l'application) applique la fonction de compression h au message M et au test T pour obtenir une **question d** entre 0 et p-1 ;

-3-Enfin, conformément au mot d, la carte calcule en guise de **témoin t** le produit dans Zn du nouvel élément r par la $d^{ième}$ puissance de la valeur d'authentification B.

Soit : $T = r^P \bmod n$; puis : $d = h(M, T)$; et enfin : $t = r \cdot B^d \bmod n$.

Le message signé est constitué par le message M suivi par un appendice consistant en la concaténation de l'identité I, de la question d et du témoin t. Soit : M, I, d, t.

*Note : Les schémas de **signature avec appendice** diffèrent sensiblement des schémas de signature avec ombre. Le message M doit y être transmis en clair, et il y a usage d'une fonction de compression.*

Pour vérifier une telle signature, il faut d'abord reconstituer la valeur de test T à partir de :
•quelques données fournies par l'usager : le message M, l'identité I, la question d et le témoin t,
•et quelques données publiées : p et n, avec les règles pour passer de l'identité I à l'identité ombragée J.
Il faut ensuite appliquer la fonction de compression h au message M et à T qui vient d'être reconstitué. L'authentification est réussie quand ce dernier résultat reproduit la question d figurant dans l'appendice.

La question d est elle reproduite par $h(M, t^P \cdot J^d \bmod n)$??

Pour démontrer que ce procédé est encore "à apport nul de connaissance" sur la valeur d'authentification, considérons d'abord un usager qui utilise : -d'abord sa carte à puce pour élaborer le test T, -puis son PC pour appliquer la fonction de compression au message M et au test T afin d'obtenir ainsi la question d, et -enfin à nouveau sa carte pour calculer le témoin t pour la question d. La procédure à l'interface de la carte ne transmet pas de connaissance sur la valeur d'authentification, car si on remplace l'usager par un ennemi qui a le libre choix de la question à poser à la carte, on ne modifie pas la sécurité de la carte. Tout comme l'ennemi, l'usager n'apprend rien sur la valeur d'authentification qui reste secrète dans la carte. Et donc, même quand la carte pratique elle-même la compression, la propriété se maintient, car **cette compression aurait pu être pratiquée à l'extérieur.**

Cette dernière remarque est importante, car elle montre qu'une fonction de compression faible, résistant mal aux mises en collision, ne met pas en péril le secret de la valeur d'authentification : elle est seulement propice aux fausses signatures (ce qui est quand même très grave !).

Note : La factorisation du nombre n est inconnue de tous, sauf de l'autorité qui, par principe, bénéficie de la confiance des usagers et des vérificateurs. Donc des fonctions de compression anti-collision à trappe basées sur le secret de la factorisation du nombre n présentent à un coût marginal dans la carte une sécurité en harmonie avec le système de base de génération des valeurs d'authentification.

Basée sur l'identité par le biais de la valeur d'authentification, ce procédé de signature possède une nouvelle propriété très intéressante : il ne peut être détourné de son usage, "l'intégrité", pour permettre un autre usage, "la confidentialité". **L'usager n'a pas de clé** ; il ne peut pas utiliser son secret (la valeur d'authentification) pour assurer une mise à la clé.

Ce schéma basé sur l'identité résoud certains problèmes politiques posés par une utilisation commerciale de la cryptologie. En effet, un bon schéma d'intégrité (authentification et signature) ne doit pas faire d'hypothèse sur l'intégrité (moralité et civisme) des usagers potentiels du schéma.

Ce souci politique semble peu préoccuper certains collègues étrangers : Que penser des contributions britanniques à la CEPT dans le cadre de la mise au point du radiotéléphone cellulaire Européen ? Authentification et confidentialité y sont allègrement mêlées. Les dernières propositions que j'ai vues consistent à mettre à la clé au début de la communication grâce à un algorithme à clé secrète et à utiliser ensuite cette clé pour assurer simultanément authentification et confidentialité. Ceci entrainera des discussions sans fin sur les spécifications des algorithmes à mettre en œuvre. Il me semble pourtant que la confidentialité sur ce type de service doit seulement être comparable à ce qui existe sur le réseau téléphonique en général : ne soyons pas plus royaliste que le roi !

Il serait bon également qu'un tel schéma soit introduit dans le projet d'avis X509 du CCITT simultanément normalisé par l'ISO sous l'appellation DIS 9594. Ces documents décrivent une structure d'identification et d'authentification grâce à un annuaire où chaque usager dispose de sa propre clé RSA. Notre schéma basé sur l'identité compléterait heureusement cette proposition : au moins ces usagers n'auraient pas de clé.

En fait, le problème primordial pour les opérateurs de réseaux est d'assurer correctement l'identification et donc la fiabilité et la sécurité des processus de taxation. Le deuxième problème est d'assurer les économies d'échelles en simplifiant les mécanismes de gestion autant que faire se peut.

D'une part, une carte émise par une autorité doit après des accords croisés pouvoir être utilisée facilement sur un autre réseau.

D'autre part, une valeur d'authentification pour un réseau doit pouvoir être facilement insérée sur une carte émise par une autre autorité, évidemment toujours après des accords croisés.

43. Les cartes avec ou sans émetteur ??

Demain, on peut imaginer des cartes sans émetteur, mises en vente dans les quincailleries. L'usager se rendrait chez les fournisseurs de services (banque, actel, société de transport, ...) et, après signature de contrats, obtiendrait l'insertion de valeurs d'authentification dans sa carte sans émetteur. A l'aide de chaque valeur d'authentification, l'usager pourrait alors émettre des signatures pour accéder au service correspondant : achat de marchandise par émission d'un chèque électronique, radiotéléphone, accès à un réseau avec identification réciproque, authentification réciproque des échanges, ...

Cependant, ces cartes sans émetteur, si elles présentent des avantages évidents, ne couvrent pas tous les cas. Il est alors impossible de faire participer la carte ou le dispositif de sécurité à une économie de gestion du système, car la puce d'une carte sans émetteur est réellement la propriété de son porteur. Il est impossible d'y gérer des droits et des clés. Ceci est très fâcheux dans le cas de l'accès conditionnel à des services radiodiffusés ou transmis par cable. Dans le cas du téléphone, il est impossible d'y gérer une zone porte-jetons. Il est pourtant peu économique de transmettre un ticket de taxation à chaque appel.

Aussi, bien que des cartes sans émetteur voient sans doute assez rapidement le jour, les cartes à émetteur subsisteront certainement. Il s'agit donc d'apprécier les évolutions du marché. Cette appréciation est très importante dans l'évolution des services.

Prenons l'exemple du radiotéléphone ; il est probable que l'on verra des radiotéléphones montés en série dans les automobiles. Il est aussi probable que l'on verra des "portaphones" vendus dans les quincailleries et donnant accès à des bornes publiques situés dans les trottoirs et dans les parkings.

Une première solution consiste à accréditer ces appareils dans les Actels, avec introduction d'une valeur d'authentification après signature d'un contrat avec le client. Cette solution, acceptable pour des portaphones agréés, semble moins réaliste pour des véhicules automobiles.

Une deuxième solution consiste à donner au client une carte à mémoire insérable dans un tel appareil. Cette carte serait émise par les Actels et participerait à l'économie du système en réduisant les émissions de tickets de taxation. Ce serait une évolution de la carte "publiphone".

Une troisième solution consiste à accréditer des cartes bancaires, en laissant les banquiers faire leur métier et gérer les flux monétaires. Ce serait une évolution des cartes bancaires.

Il est vraisemblable que la réalité sera multiple et que le meilleur choix des techniciens aujourd'hui est de maintenir possible la coexistence de ces diverses solutions, car elles sont en fait complémentaires.

A mon avis, les algorithmes à clé publique et les algorithmes à clé secrète coexisteront donc dans les futures cartes à puce ; et les domaines d'utilisation de ces deux techniques se différencieront nettement :

Les clés, les droits d'accès et les porte-jetons (porte-monnaie et tire lire) resteront gérés par des techniques cryptographiques qu'il me semble préférable de maintenir secrètes et de ne pas chercher à normaliser, afin de conserver toute la souplesse nécessaire aux changements d'algorithmes. L'étanchéité entre les réseaux me semble bonne sur ce point.

Au contraire, les mécanismes d'intégrité (authentification et signature) méritent la normalisation afin d'assurer sur ce plan une grande intéraction entre les différents réseaux.

Note : Il est cependant utile de diversifier au plus tôt les mécanismes de signature, c'est à dire de travailler d'autres problèmes complexes que la théorie des nombres afin de se prémunir contre les conséquences catastrophiques d'une progression inattendue en arithmétique, particulièrement en factorisation des grands nombres entiers.

Tout problème complexe donne lieu à des procédures à apport nul de connaissance.