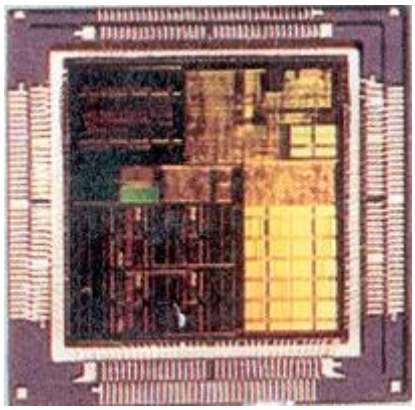
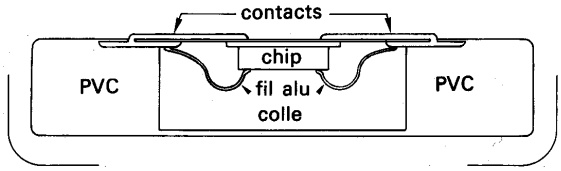
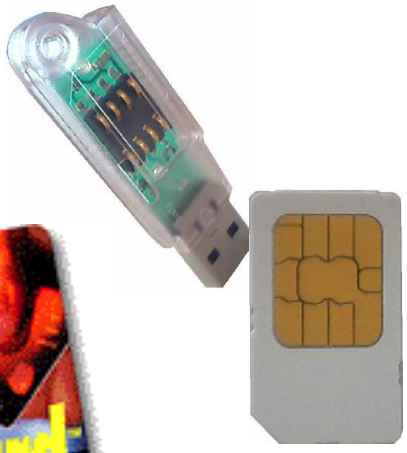
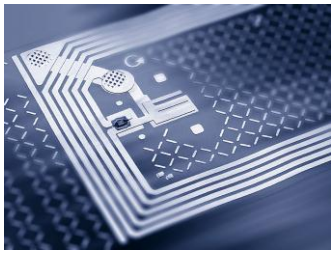


CARTES A PUCE







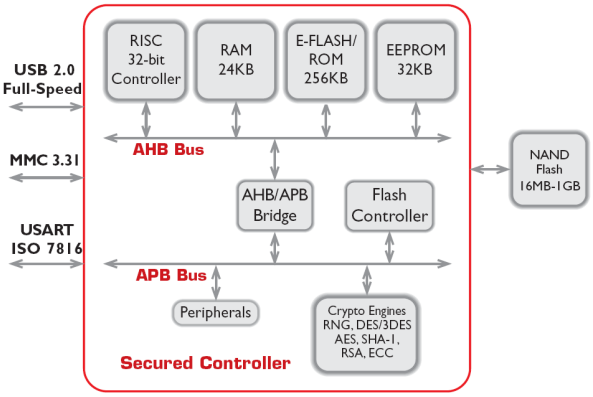


Table des matières

I- Aperçu de la carte à puce.

Historique

C'est en 1950 que la compagnie américaine *Diners'Club* lance la première carte de voyage et de loisirs. Elle se présente sous la forme d'un petit carnet (en carton) qui contient une liste d'adresses (des hôtels restaurants qui acceptent cette carte) et dont la page de garde comporte la signature du titulaire et diverses informations.

L'année 1958 voit la naissance des cartes *American Express* (sur support plastique), émises par les héritiers de la célèbre compagnie de diligences *Wells & Fargo*.

En France le groupe carte Bleue apparaît en 1967 afin d'offrir un moyen de paiement concurrent des cartes américaines. L'objectif est de réduire le nombre de chèques en circulation qui représente déjà 40 % des opérations de paiement et atteindra 90% en 1980.

Les premier distributeurs automatiques de billets (DAB) voient le jour en 1971 et utilisent des cartes bleues munies de pistes magnétiques (une piste magnétiques est constituée de deux zones de lectures d'une capacité de 200 octets).

En 1974 Roland Moreno dépose un premier brevet sur un objet *portable à mémoire*. Il décrit un ensemble (l'ancêtre des cartes à mémoires) constitué d'une mémoire électronique (E^2PROM) collé sur un support (une bague par exemple) et un lecteur réalisant l'alimentation (par couplage électromagnétique) et l'échange de donnés (par liaison optique). Il réalise la démonstration de son système à plusieurs banques. Il fonde la compagnie Innovatron.

Grâce au ministère de l'industrie il est mis en relation avec la compagnie Honeywell Bull qui travaille sur une technologie (TAB Transfert Automatique sur Bande) réalisant le montage de circuits intégrés (puces) sur un ruban de 35 mm, à des fins de test.

En 1977 Michel Ugon (Bull) dépose un premier brevet qui décrit un système à deux puces un microprocesseur et une mémoire programmable. La compagnie BULL CP8 (*Cartes des Années 80*) est crée. La première carte à deux composants est assemblée en 1979. Le Microprocesseur Auto-programmable Monolithique (MAM, en anglais Self Programmable One chip Microprocessor – SPOM)) voit le jour en 1981, c'est en fait le composant qui va équiper toutes les cartes à puces.

Marc Lassus (futur fondateur de Gemplus) supervise la réalisation des premières puces (microprocesseur et mémoire puis du MAM) encartées par CP8 (un nouveau process de fabrication est mis au point pour obtenir des épaisseurs inférieures à un mm).

Le Gie carte à mémoire est crée en 1980 et comprend des industriels (CP8, Schlumberger, Philips), le secrétariat d'Etat des P&T, et plusieurs banques.

En 1982 plusieurs prototypes de publiphones utilisant des cartes à mémoires (les télécartes) sont commandés par la DGT (Délégation Générale des Télécommunications) à plusieurs industriels. Les télécartes vont constituer par la suite un marché très important pour les cartes à puces.

En 1984 la technologie CP8 (MAM) est retenue par les banques françaises, le système d'exploitation B0 va devenir le standard des cartes bancaires françaises. Le groupement des cartes bancaires (CB) émet une première commande de 12,4 millions de cartes.

Les standards de base des cartes à puces (ISO 7816) sont introduits à partir de 1988.

A partir de 1987, la norme des réseaux mobiles de 2°génération (GSM) introduit la notion de module de sécurité (une carte à puce SIM –Subscriber Identity Module). En raison du succès de la téléphonie mobile, les télécommunications deviennent le premier marché de la carte à puce.

A partir des années 96, l'apparition des cartes java marque l'entrée des systèmes cartes à puce dans le monde des systèmes ouverts. Il devient en effet possible de développer des applications dans un langage largement diffusé.
Depuis 2005, certains composants intègrent des machines virtuelles .NET, ils sont usuellement dénommés *dotnet card*.

Les marchés.

La plus importante entreprise de cartes à puce GEMALTO, détient 80% de part de marché (?) et emploie environ 10,000 personnes. En 2009 son chiffre d'affaire était de 1,650 millions d'euros, dont 900 M€ pour la téléphonie mobile et 450M€ pour le bancaire.

Les tableaux suivants illustrent quelques aspects du marché de la carte à puce.

2009	Global shipment in Million of Units (Mu)	
	Memory	Microprocessor
Telecoms	300	3 400
Financial services – Retail – Loyalty	30	750
Government – Healthcare	170	160
Transport	160	40
Pay TV	-	100
Others (including corporate ID)	80	70
TOTAL	740	4 520

Marché de la carte à puce, chiffres 2009, source eurosmart.com

Year	Telecom	Banking	Government healthcare	Total	% Banking	% SIM
1999	200	108		398	27,1	50,3
2000	370	120		551	21,8	67,2
2001	390	140		599	23,4	65,1
2002	430	175		791	22,1	54,4
2003	670	205		979	20,9	68,4
2004	1050	280		1469	19,1	71,5
2005	1220	330		1727	19,1	70,7
2006	2150	480	140	2895	16,6	74,2
2007	2650	510	105	3445	14,8	76,9
2008	3200	650	140	4185	15,5	76,4
2009	3400	750	160	4520	16,6	75,2

Evolution du marché de la carte à puce depuis 1999

En 2002 Marc Lassus estimait que la puissance installée (en MegaDhrystone) du parc informatique était de 4K pour les mainframes, 20K pour les ordinateurs personnels et 34K pour les cartes à puce à microprocesseur.

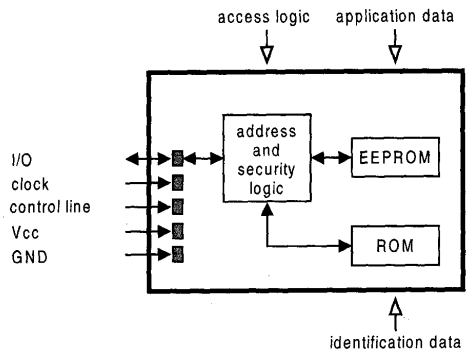
Les principales caractéristiques d'une carte à puce sont les suivantes,

- C'est un objet portable, qui loge des données et des procédures.
- C'est un objet sécurisé
  - Il est difficile de lire les données stockées dans les mémoires de la puce.
  - Le code est exécuté dans un espace de confiance, il n'est pas possible d'obtenir les clés associées à des algorithmes cryptographiques.
- C'est un objet de faible prix (1-5$ pour les SPOM, 0,1-0,5$ pour les cartes magnétiques), mais personnalisable pour des centaines de millions d'exemplaires.
- Une puce ne peut fonctionner seule, elle nécessite un CAD (*Card Acceptance Device*) qui lui délivre de l'énergie, une horloge (base de temps), et un lien de communication. Un CAD usuel est un lecteur de cartes.

La technologie des cartes à puce

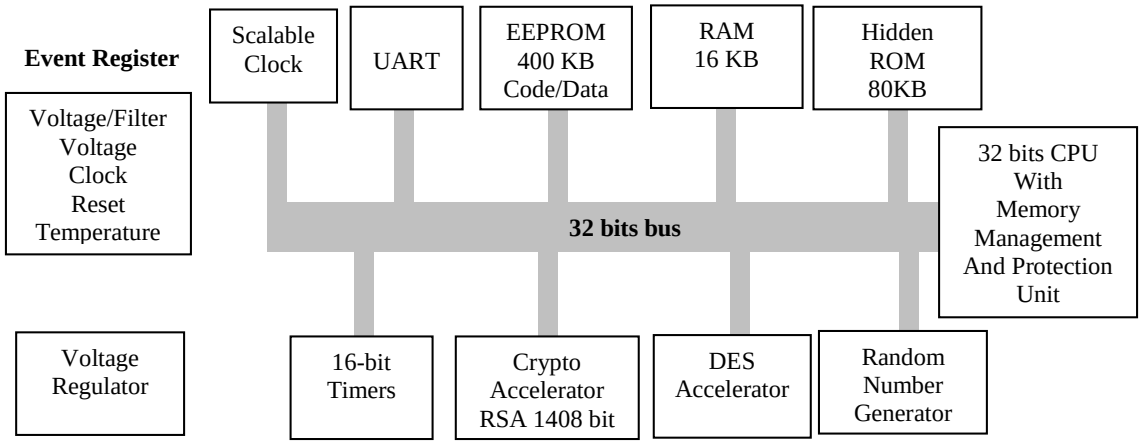
Les cartes à mémoire.

Elles comportent un bloc de sécurité (optionnel) qui contrôle les accès à des mémoires de type ROM ou E^2PROM. Les télécartes de 1ière génération (ou TG1) n'étaient pas sécurisées; les



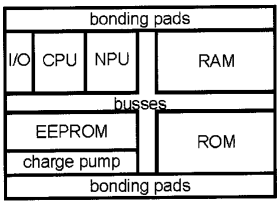
télécartes de 2ième génération (ou TG2) comportent un bloc de sécurité.

Les cartes à microprocesseurs.

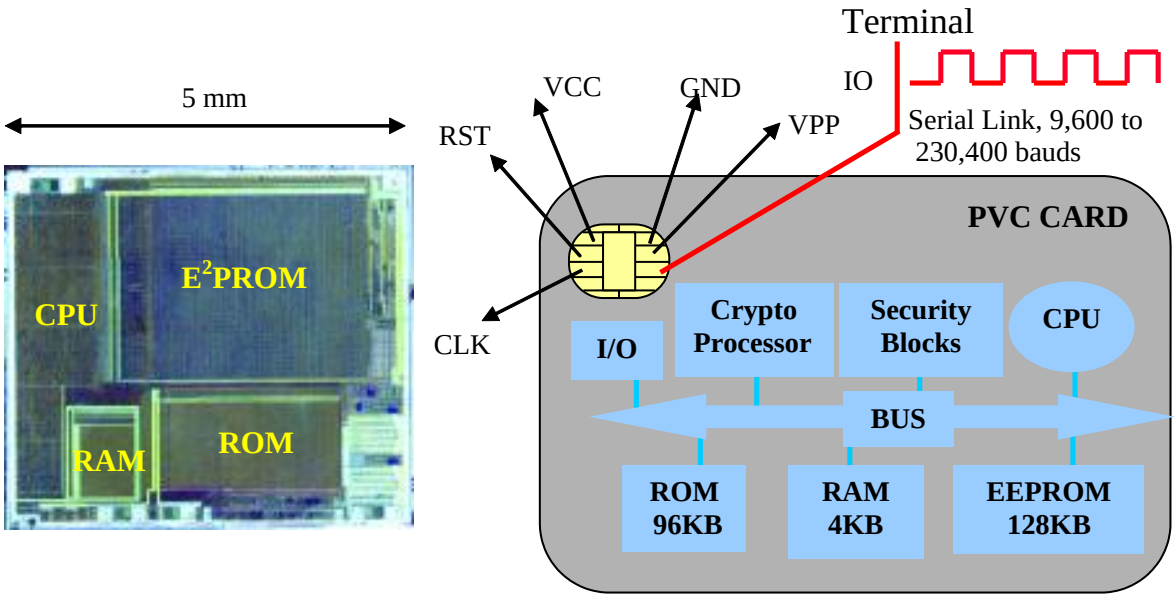


Le microcontrôleur 88CFX4000P

Un microcontrôleur se présente typiquement sous la forme d'un rectangle de silicium dont la surface est inférieure à 25 mm^2. D'une part cette taille est imposée par les contraintes de flexion induite par le support en PVC, et d'autre part cette dimension limitée réalise un compromis entre sécurité physique et complexité du composant.



Les capacités mémoires sont comprises entre 128 et 256 Ko pour la ROM (surface relative 1), 64 et 128 ko pour l'E^2PROM (surface relative 4), 4 et 8 Ko pour la RAM (surface relative 16). En raison de ces contraintes technologiques la taille de RAM est modeste; l'E^2PROM occupe une portion importante du CHIP. Les écritures en E^2PROM sont relativement lentes (de l'ordre de 1 ms par mot mémoire de 32 à 64 octets), et le nombre de ces opérations est limité (de 10^4 à 10^6). L'introduction des mémoires FeRAM devrait amoindrir ces contraintes (10^9 opérations d'écriture, capacités mémoire de l'ordre du Mo, temps d'écriture inférieur à 200ns).



Les classiques processeurs 8 bits ont des puissances de traitements comprises entre 1 et 3 MIPS, ce paramètre est supérieur à 33 MIPS pour les nouvelles architectures à bases de processeurs RISC 32 bits.
En termes de puissance de calcul cryptographique, les systèmes 32 bits réalisent un algorithme DES à un Mbit/s et un calcul RSA 1024 bits (clé privé) en moins de 300mS.
A l'horizon 2004 l'introduction des technologies de type mémoires FLASH devrait conduire à des capacités de l'ordre de 1 Mo. Les puissances de calculs estimées sont de l'ordre de 100 à 200 mips.

Couches de communications ISO 7816.

Requête.	Réponse.
CLA INS P1 P2 Lc [Lc octets]	sw1 sw2
CLA INS P1 P2 Le	[Le octets] sw1 sw2
CLA INS P1 P2 Lc [Lc octets] CLA C0 00 00 Le	61 Le [Le octets] sw1 sw2

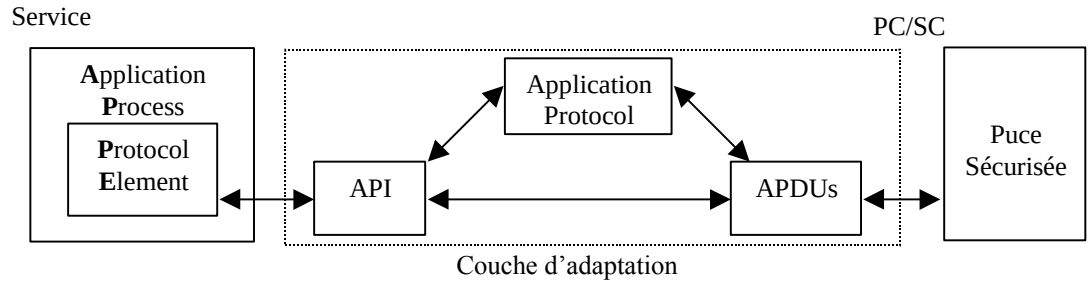
Commandes (APDUs) définis par la norme ISO 7816.

La norme ISO 7816 décrit l'interface de communication entre la puce et son terminal associé. Ce dernier fournit l'alimentation en énergie et une horloge dont la fréquence est typiquement 3.5 Mhz. L'échange de données entre puce et terminal est assuré par une liaison série dont le débit est compris entre 9600 et 230,400 bauds. La norme 7812-12 définir cependant une interface USB à 12 Mbit/s. Le terminal produit une requête (APDU) qui comporte conformément au protocole de transport T=0 au moins 5 octets (CLA INS P1 P2 P3) et des octets optionnels (dont la longueur *Lc* est précisée par la valeur de l'octet P3). La carte délivre un message de réponse qui comprend des octets d'information (dont la longueur *Le* est spécifiée par l'octet P3) et un mot de status (sw1 sw2, 9000 notifiant le succès d'une opération) large de deux octets. Lorsque la longueur de la réponse n'est pas connue à priori un mot de status «61 Le» indique la longueur du message de réponse. Une fois ce paramètre connu le terminal obtient l'information au moyen de la commande *GET RESPONSE* (CLA C0 00 00 Le).

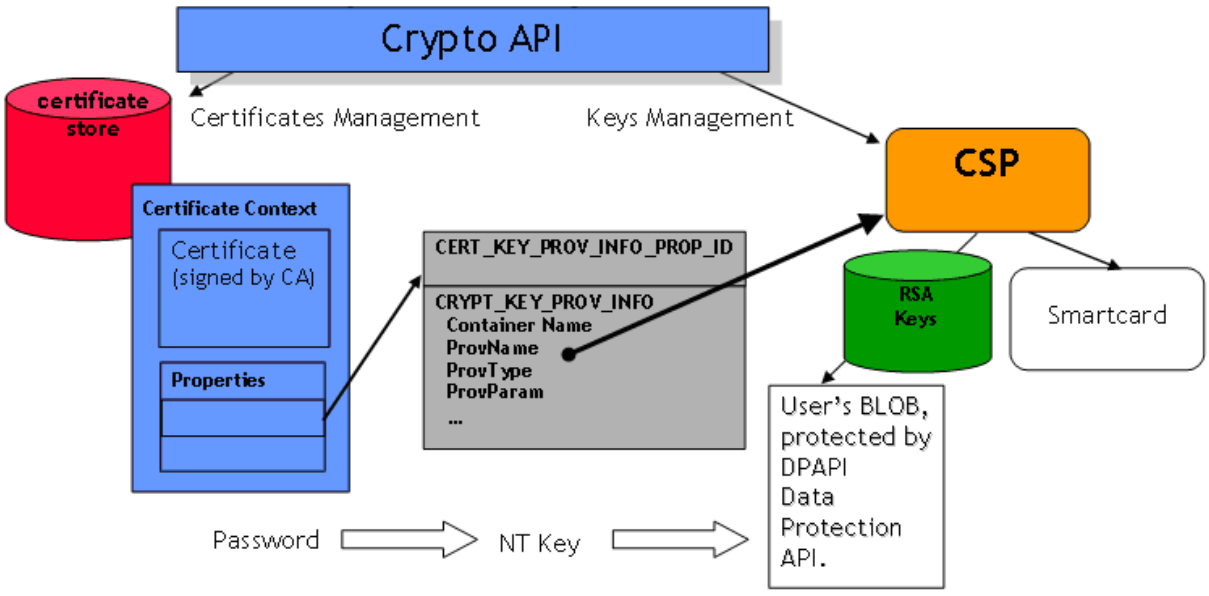
Les opérations de lecture et d'écriture, l'invocation des fonctions cryptographiques sont associées à des APDUs spécifiques. L'information stockée dans la puce sécurisée est organisée selon un système de fichiers qui comporte un répertoire racine (MF Master File), des sous répertoires (DF Dedicated File) et des fichiers (EF Elementary File). Chaque composant est identifié par un nombre de **deux octets**; la navigation à travers ce système s'effectue à l'aide d'APDUs particulières (SELECT FILE, READ BINARY, WRITE BINARY). La sécurité est assurée par des protocoles de simple ou mutuelle authentification (transportés par des APDUs), qui en cas de succès autorisent l'accès aux fichiers.
La mise en œuvre d'une carte utilise donc un paradigme d'appel de procédure, transporté par des APDUs (généralement définies pour un système d'exploitation spécifique); l'information embarquée est connue à priori et classée par un système de fichier 7816.

Intégration des cartes à puce aux technologies de l'information.

L'intégration des puces sécurisées aux technologies de l'information implique l'adaptation des logiciels applicatifs de telle sorte qu'ils génèrent les APDUs nécessaires à l'utilisation des ressources embarquées. Schématiquement le middleware classique consiste à définir les éléments protocolaires (PE) requis par un service (*Application Process*) et exécutés dans la puce; chaque élément est associé à une suite d'APDUs (*Application Protocol*) variable selon le type de carte utilisée. L'application localisée sur le terminal utilise la puce aux moyens d'APIs (Application Programmatic Interface) plus ou moins normalisées (par exemple PC/SC pour les environnements win32), qui offre une interface de niveau APDUs ou plus élevé (PE).



Middleware classique d'une carte à puce.



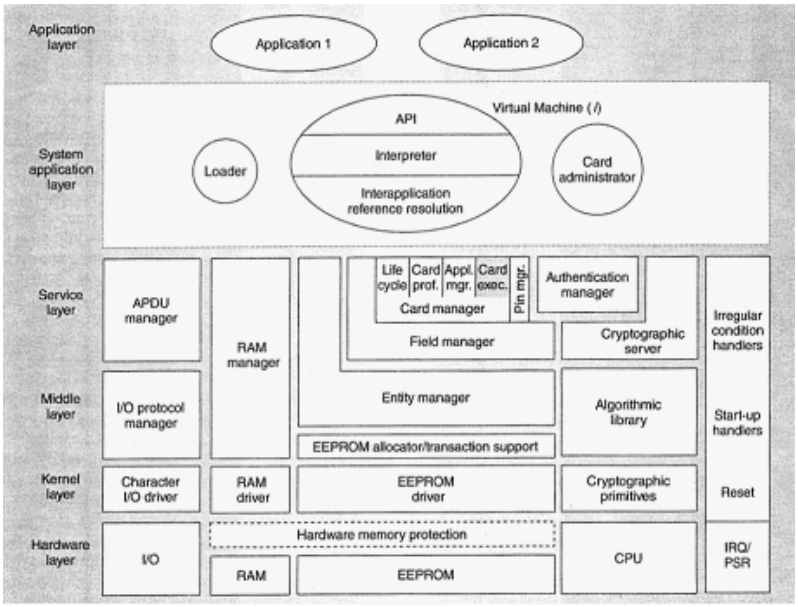
Crypto Service Provider Microsoft

Système d'exploitation d'une carte à puce.

Schématiquement un système d'exploitation d'une carte à puce comporte les éléments suivants,

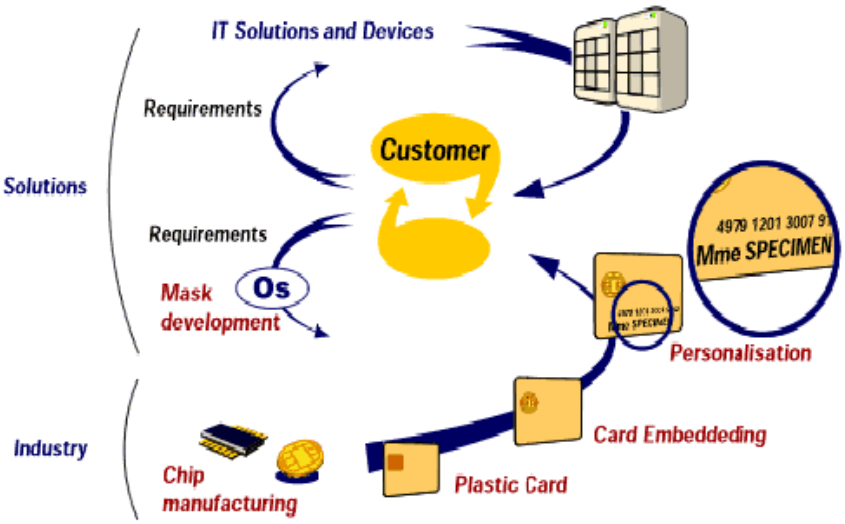
- Un bloc de gestion des ordres (APDUs) transportés par la liaison série.
- Une bibliothèque de fonctions cryptographiques, dont le code est réalisé de telle manière qu'il soit résistant aux attaques logiques connues (Timing attack, DPA, SPA, …).
- Un module de gestion de la RAM
- Un module de gestion de la mémoire non volatile (E^2PROM…), qui stocke les clés des algorithmes cryptographiques et les secrets partagés).
- Un module de gestion de la RAM, qui est une ressource critique en raison de sa faible quantité et de son partage entre procédures et applications.
- Un bloc de gestion d'un système de fichiers localisé dans la mémoire non volatile.
- Un module de gestion des événements indiquant une attaque probable de la puce sécurisée comme par exemple,
  - Une variation anormale de la tension d'alimentation (*glitch*)
  - Une variation anormale de l'horloge externe de la puce sécurisée.
  - Une variation anormale de température.
  - La détection d'une perte d'intégrité physique du système. La surface d'une puce est généralement recouverte par un treillis métallique qui réalise une sorte de couvercle dont le système teste la présence.

Le système d'exploitation est contenu dans la ROM dont le contenu n'est pas chiffré. La connaissance de son code, bien que difficile ne doit pas rendre possible des attaques autorisant la lecture de la mémoire non volatile.



Un exemple de système d'exploitation

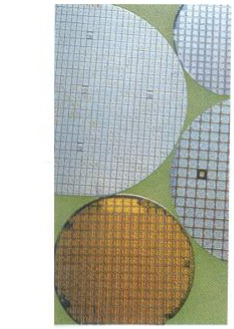
Cycle de vie d'une carte à puce.

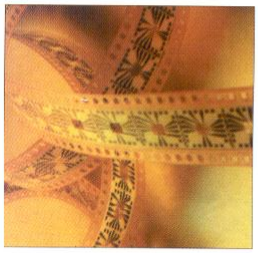


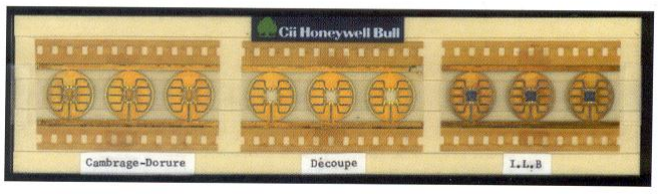
Cycle de vie d'une carte à puce

Un système d'exploitation est réalisé par une entreprise spécialisée. Ce logiciel étant ajusté pour un composant électronique particulier, il est appelé ***masque***.

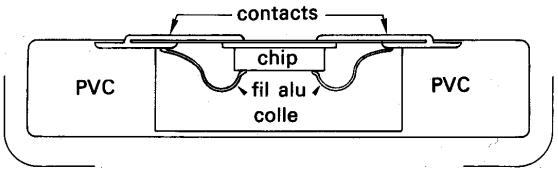
Le masque est stocké dans la ROM du composant lors du processus de **fabrication**. Au terme de cette phase le fondeur de silicium insère dans la puce une clé dite clé de fabrication et écrit dans la mémoire de cette dernière des informations telles que numéro de série du produit, date de fabrication etc.







Le wafer (plaque de silicium circulaire qui comporte un ensemble de puces) est alors envoyé à l'**encarteur** qui réalise sa découpe, colle les puces sur des ***micromodules*** et en réalise le micro câblage. L'ensemble est alors protégé par une substance isolante. Il est ensuite collé sur un support en plastique (PVC) dans lequel on a préalablement usiné une cavité (le *bouton*).



L'encarteur, qui connaît les clés de fabrication, inscrit de nouvelles informations dans la puce et active un verrou de fabrication qui annule la clé de fabrication. Une nouvelle clé est inscrite dans la puce permettant de contrôler les opérations ultérieures. Cette opération est encore dénommée **personnalisation**.
Les cartes sont par la suite transférées vers **l'émetteur** de la carte qui peut inscrire de nouvelles informations.
La vie de vie d'une carte consiste à poser à verrou d'invalidation (IV) qui rend non fonctionnel le système d'exploitation.

Systèmes fermés et systèmes ouverts.

On peut distinguer deux types de systèmes d'exploitation de cartes à puce :

- Les systèmes fermés, généralement mono application, dédiés à un usage unique par exemple cartes bancaires (masque CP8 M4 B0'), les cartes santé (VITALE), les cartes pour la téléphonie mobile (modules SIM).
- Les systèmes ouverts, tels que les javacard par exemple, qui ne sont pas destinés à une application particulière, et pour lesquels il est possible de ***charger*** des logiciels (applets) après la réalisation du masque et l'encartage.

Quelques exemples de systèmes fermés.

La carte bancaire BO'

Les cartes bancaires B0' sont dérivées du masque CP8 M4, conçu date du milieu des années 80. La mémoire E^2PROM est associée à des adresses comprises entre 0200h et 09F3 (les adresses référencent des ½ octets) soit environ 2 kilo octets. Elle se divise en sept zones,

- La zone secrète qui stocke les clés émetteurs primaires & secondaires, un jeu de clés secrètes, les codes PIN (Personal Identity Number) du porteur.
- La zone d'accès qui mémorise le nombre de présentations erronées de clés.
- La zone confidentielle, dont le contenu est défini en phase de personnalisation.
- La zone de transaction, qui mémorise les opérations les plus récentes.
- La zone de lecture, qui loge des données en accès libre.
- La zone de fabrication, qui réalise la description de la carte et comporte des informations sur sa fabrication.
- La zone des verrous, qui mémorise l'état de la carte (en fabrication, en service, annulée).

Les clés associées aux opérations de lecture ou d'écriture sont déterminées par le système d'exploitation.

Les cartes TB.

Cette carte dite d'usage général (*General Purpose*) était commercialisée dans le courant des années 90 par la société CP8. Elle intégrait des algorithmes cryptographiques DES, RSA ainsi que des mécanismes d'authentification par PIN code et blocage après trois présentations erronées. Son principe de fonctionnement est basé sur le contrôle des accès de fichiers élémentaires (lecture / écriture) à l'aide d'opérations de présentation de clés. Il existe deux types de procédures d'authentification,

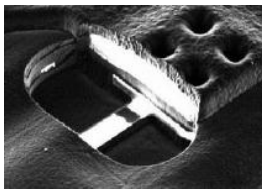
- Authentification par PIN code, avec blocage du répertoire (après trois échecs).
- Authentification par clé cryptographique. Un premier ordre demande au système d'exploitation de produire un nombre aléatoire. Un deuxième ordre présente au système la valeur chiffrée de la valeur précédemment fournie.
- Chaque répertoire dédié (DF) comporte trois types de fichiers
  - Des fichiers secrets, qui abritent les clés.
  - Des fichiers de contrôle d'accès, qui mémorisent le nombre d'échecs des opérations d'authentification.
  - Des fichiers dont les accès sont plus ou moins conditionnés à la présentation de clés.
- Il existe divers types de clés, associés à plusieurs algorithmes cryptographiques.
  - Clé de fabrication.
  - Clé de personnalisation.
  - Clé d'émetteur.
  - PIN codes
  - Clé d'authentification.

- Un répertoire est un bloc mémoire de taille fixe qui possède un en tête des fichiers et des sous répertoires.
  - Les opérations de création de fichiers et de sous répertoires peuvent impliquer la présentation de clés.
  - Les clés d'authentification sont définies lors de la création des fichiers ou sous répertoires.

Quelques attaques contre les cartes à puces.

Attaques matérielles (intrusives).

Pose de microsondes à la surface du circuit. L'attaquant désire obtenir les secrets de la mémoire non volatile, par exemple en l'isolant du reste de la puce sécurisé et en produisant les signaux électriques nécessaires à sa lecture.



Réactivation du mode test, via un plot de connexion, dans le but de lire la mémoire. En phase de fabrication une procédure de test, réalisé par le système d'exploitation permet de vérifier le bon fonctionnement du système et d'éliminer les composants défectueux. Un fusible désactive ce mode. L'attaquant essaye de rétablir cette connexion.

Reverse engineering, reconstruction du *layout* de la puce, visualisation du code ROM.

Injection de fautes; grâce à des interactions physiques (injection de lumière, etc.) on perturbe le fonctionnement normal du microcontrôleur, afin de produire des erreurs de calculs permettant de déduire la clé d'un algorithme cryptographique.

Attaques logiques (non intrusives).

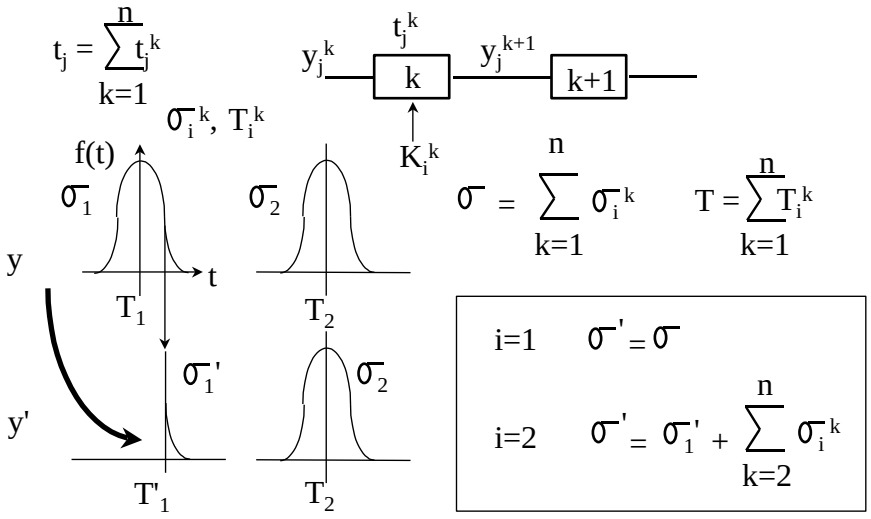
Attaques temporelles (moyenne, écart type). Certaines réalisations logicielles d'algorithmes peuvent présenter des temps de calculs différents en fonction des valeurs calculées et de la clé utilisée.

Attaques par corrélation statistique, telles que *Simple Power Attack* (SPA) ou *Differential Power Analysis* (DPA). Un processeur réalise un algorithme à l'aide d'une suite d'instructions nécessairement différentes en fonction de la clé.

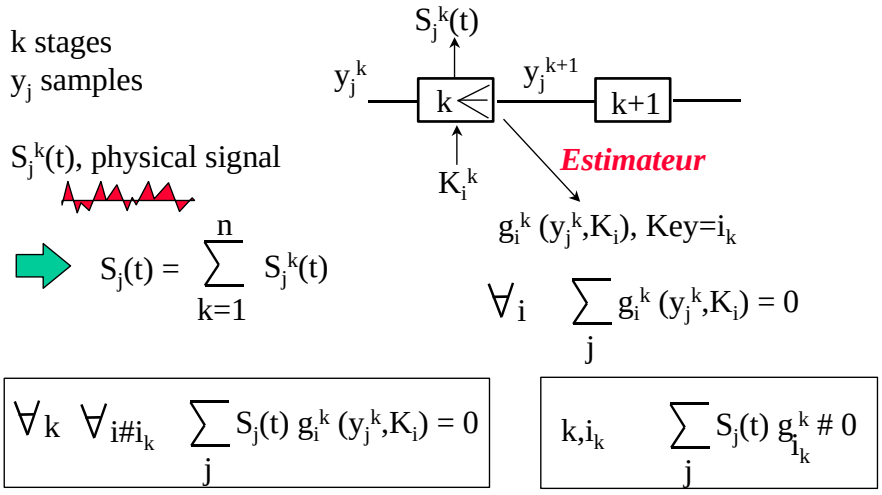
Ainsi un algorithme utilisant une clé parmi $n=2^p$ possible, utilise p instructions différentes pour une clé particulière. Il produit donc des signaux électriques (par exemple puissance consommée) ou radioélectriques qui sont corrélés à la clé opérationnelle.

Défauts de conception.

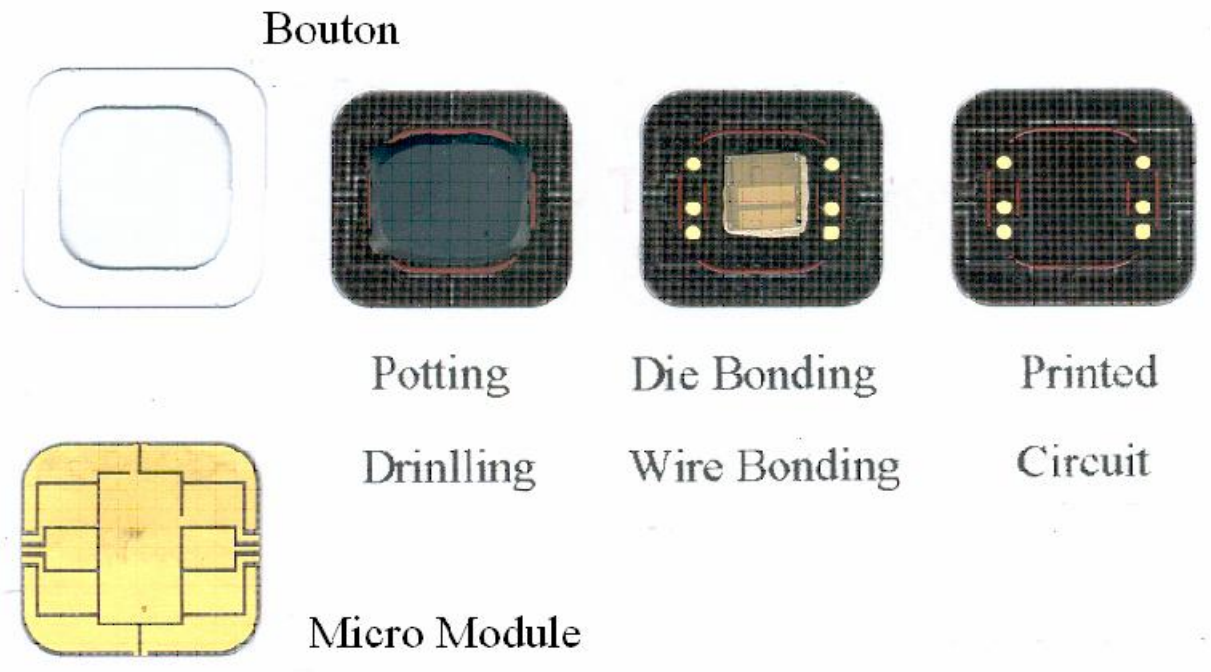
Coupure d'alimentation intempestive, parasite d'horloge, remise à zéro abusive, attaque par éclairement. L'attaquant cherche à créer un défaut dans le déroulement du programme. Il espère par exemple exécuter le code permettant de lire le contenu de la mémoire non volatile E^2PROM ou FLASH.



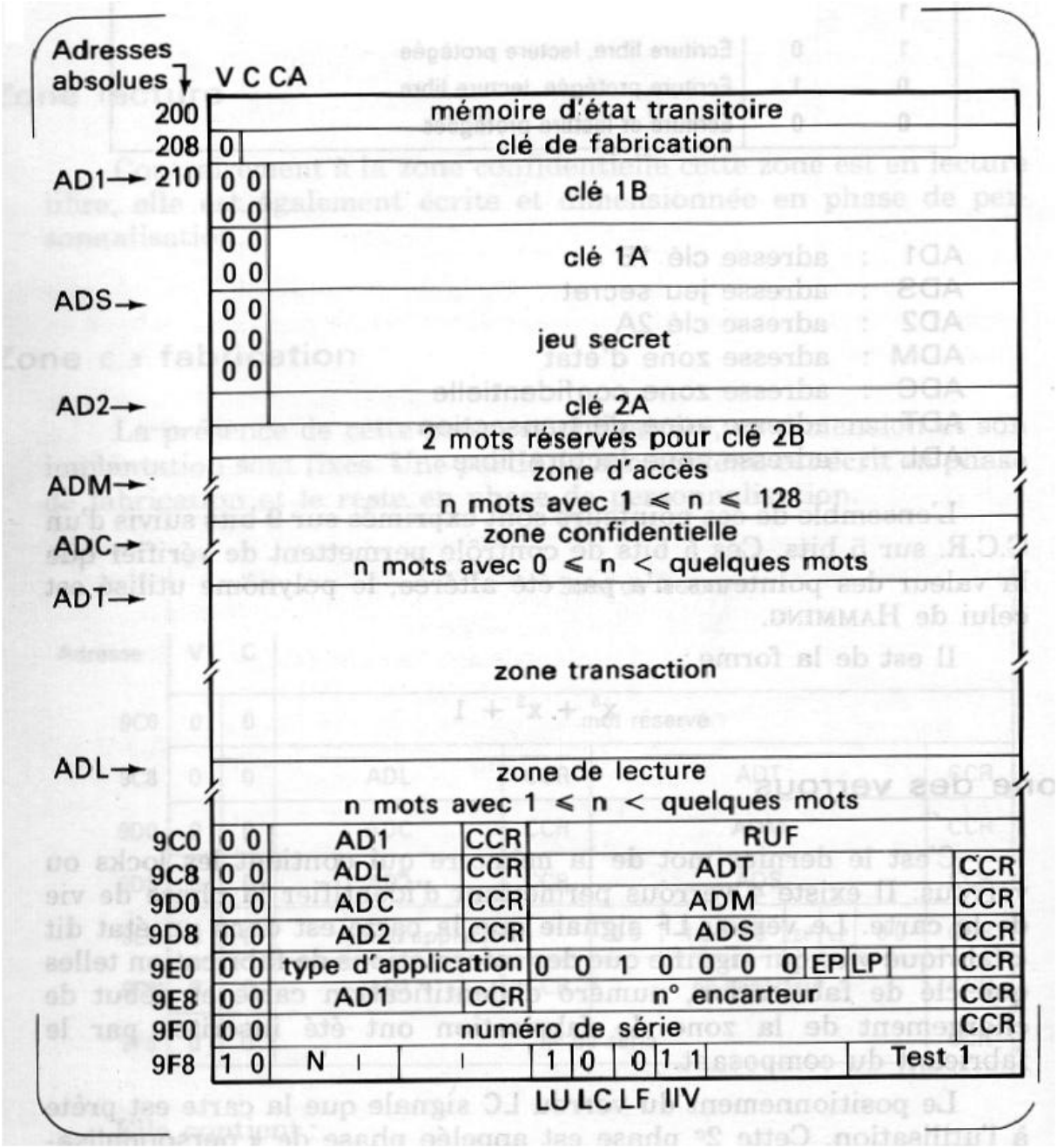
Attaque par écart type.



Attaque par corrélation.



Etapes de fabrication d'une carte à puce



Organisation de la mémoire d'une carte bancaire B0'