Des codes secrets dans la carte bleue

François Dubois [1]

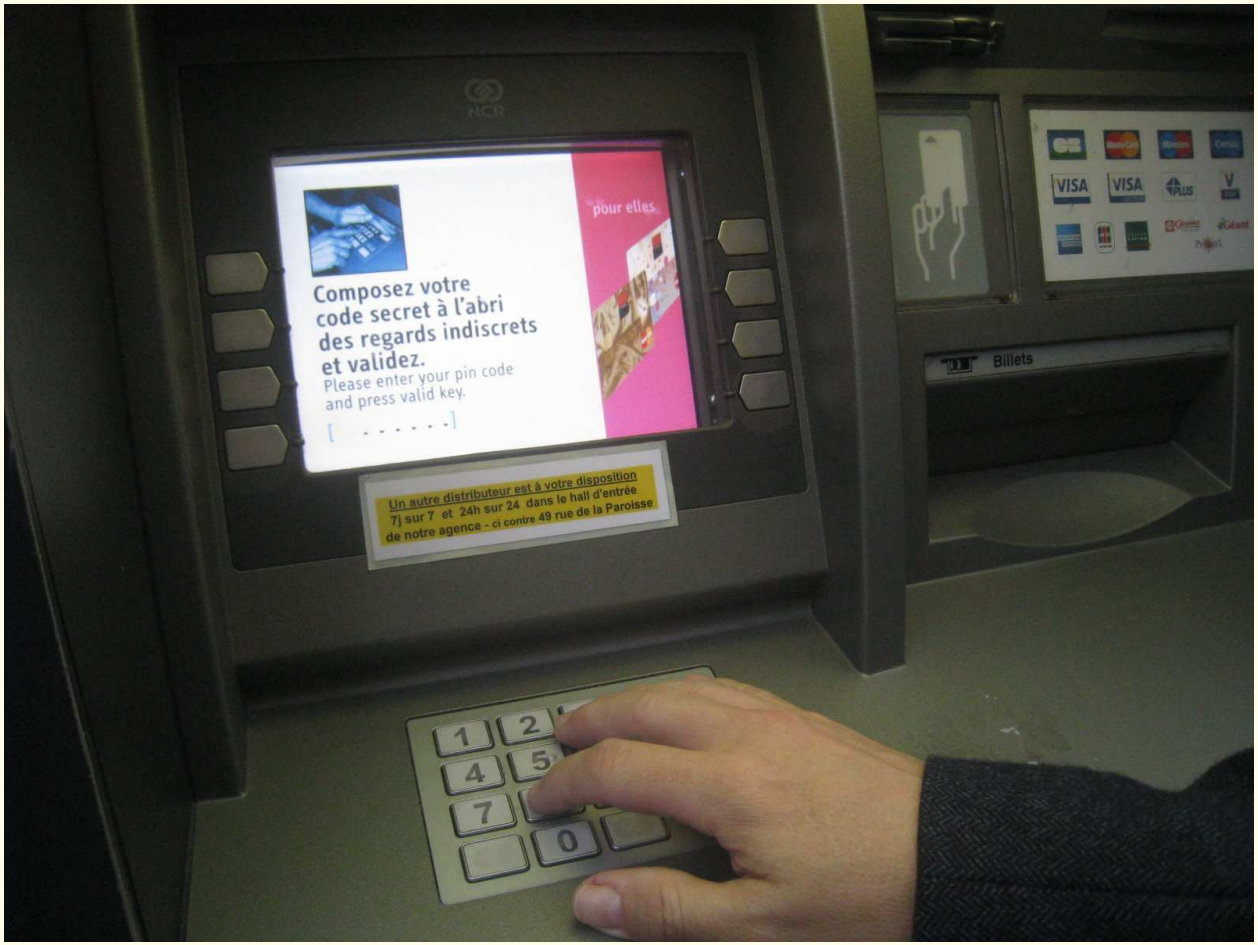
Kafemath
“Le Mouton Noir”, Paris 11ième
jeudi 25 juin 2009

[1] animateur du Kafemath, café mathématique à Paris.

Carte bleue

Un geste du quotidien...

Carte bleue (ii)



Que se passe-t-il dans la machine ?

Carte bleue (iii)

Protocole d'intercommunication à trois :

Terminal, Carte, Utilisateur

Premier temps : la carte s'authentifie

("vous débiterez le compte de M. Martin")

Second temps : le porteur de la carte s'identifie

(c'est bien M. Martin qui manipule la carte)

Troisième temps : expliciter le montant du retrait

Codage public

La règle est connue de tous :

remplacer une suite de lettre par une suite de chiffres

Exemple : code ascii

"American Standard Code for Information Interchange"

7 bits : $2^7 = 128$ caractès codables, plus un "bit de contrôle".

symboles spéciaux :	1 à 47
chiffres de 0 à 9 :	48 à 57
d'autres symboles spéciaux :	58 à 64
lettres majuscules :	65 à 90
encore des symboles spéciaux :	91 à 96
lettres minuscules :	97 à 122
derniers symboles spéciaux :	123 à 128

On remarque qu'il n'y a pas les lettres accentuées...

Codage public (ii)

Si m est un message de longueur "une lettre",
nous pouvons écrire formellement : $\widetilde{m} = f(ascii, m)$
Alors $\widetilde{m}$ est un nombre entre 0 et 128 codé en base 2.

Exemple : $m =$'a'. Alors $f(ascii,$ 'a'$) = 1100001$
$(64 + 32 + 1 = 97$ en base deux$)$

Réciproquement, si $\widetilde{m}$ est un nombre entre 0 et 128,
nous pouvons écrire formellement : $m = \widetilde{f}(ascii, \widetilde{m})$
Exemple : $\widetilde{m} = 97$. Alors $\widetilde{f}(ascii, 97) =$ 'a'

Les fonctions f et $\widetilde{f}$ sont "inverses" l'une de l'autre
$\widetilde{f}\big(ascii, f(ascii, m)\big) = m$ pour tout m
$f\big(ascii, \widetilde{f}(ascii, \widetilde{m})\big) = \widetilde{m}$ pour tout $\widetilde{m}$

La règle de codage "ascii" est connue de tout le monde !

Code secret

Pour des besoins militaires (par exemple !),
il peut être utile de garder la confidentialité d'un message

Stéganographie (art de cacher les message) :
classique chez les grecs
Echec par Démarate de l'offensive de Xerxès, roi des Perses,
contre la Grèce en 484 avant JC, :
ré-écriture sur une tablette d'argile.

Chiffre de César :
décalage de trois lettres vers la droite dans l'alphabet
message :
"Belle Marquise, vos beaux yeux me font mourir d'amour"
message codé : "EHOOH PDUTX LVHYR
VEHDX ABHXA PHIRQ WPRXU LUGDP RXU"

Codage par transformation "one to one" (bijection)

Code secret (ii)

Attention à la statistique d'emploi des lettres de l'alphabet !

lettre	occurence en %	lettre	occurence en %
e	10,5	n	7,8
a	8,7	r	7,7
i	8,5	s	7,2

etc. Vous savez tous jouer au Scrabble !

Exception : *La Disparition* de Georges Perec, Denoël, 1969.

Qui, d'abord, a l'air d'un roman jadis fait où il s'agissait
d'un individu qui dormait tout son saoul

Anton Voyl n'arrivait pas à dormir. Il alluma. Son Jaz marquait minuit vingt. Il poussa un profond soupir, s'assit dans son lit, s'appuyant sur son polochon. Il prit un roman, il l'ouvrit, il lut ; mais il n'y saisissait qu'un imbroglio confus, il butait à tout instant sur un mot dont il ignorait la signification.

Code secret (iii)

Blaise de Vigenère (1523-1596)

Traité des chiffres ou secrètes manières d'écrire, 1586

introduit la notion de clef

Le nombre de lettres qui sert au décalage

change à chaque lettre en suivant la clef !

message :

"Belle Marquise, vos beaux yeux me font mourir d'amour"

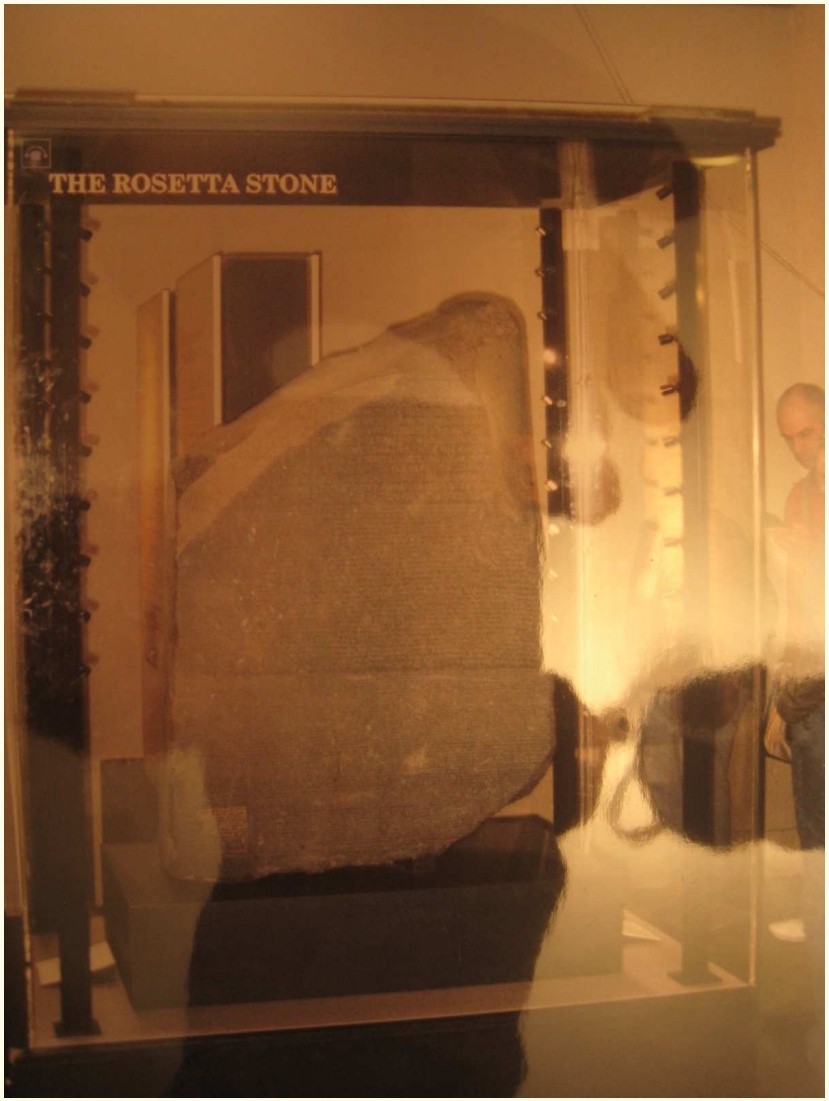
Clef : Molière (13, 15, 12, *etc.*)

message codé : "NSWTI DEDEF QWVZA GMMEL BKSFF QVJAB EUSLV UFOIQ FYD"

L'effet de la statistique d'emploi des lettres de l'alphabet

est largement amoindri.

Code secret (iv)



Jean-François Champollion déchiffre la Pierre de Rosette (1822)

Code secret (v)

Code symétrique :

les deux partenaires disposent de la clef "B" du code

Codage : $\widetilde{m} = f(B, m)$, décodage : $m = \widetilde{f}(B, \widetilde{m})$

Les deux fonctions sont inverses l'une de l'autre :

$$m = \widetilde{f}\big(B, f(B, m)\big), \quad \widetilde{m} = f\big(B, \widetilde{f}(B, \widetilde{m})\big)$$

Exemples :

- Code "Data Encryption Standard" (1975),
  le texte est tronqué par paquets de 64 bits
  et utilise une clef de longueur 56 bits ($2^{56} \approx 7,2\,10^{16}$)

  Abandonné en 1999 car la longueur de la clef était trop faible.
- Code "Advanced Encryption Standard" (1998),
  le texte est tronqué par paquets de 128 bits
  utilisation de clefs de longueurs 128 ($2^{128} \approx 3,4\,10^{38}$),
  192 ($2^{192} \approx 6,3\,10^{57}$) ou 256 bits ($2^{256} \approx 1,1\,10^{77}$)

Mais comment se transmettre la clef de façon fiable ?

Une clef publique !

Code **non** symétrique :

le propriétaire du code dispose de la clef $\widetilde{B}$

Il publie (!) une version dégradée, la clef publique B

Codage : $\widetilde{m} = f(B, m)$, décodage : $m = \widetilde{f}(\widetilde{B}, \widetilde{m})$

Les deux fonctions sont inverses l'une de l'autre :

$$m = \widetilde{f}\big(\widetilde{B}, f(B, m)\big), \quad \widetilde{m} = f\big(B, \widetilde{f}(\widetilde{B}, \widetilde{m})\big),$$

Tout le monde peut coder le mot m transformé en $\widetilde{m}$

mais seul le propriétaire du code peut décoder m à partir de $\widetilde{m}$

Authentification

Etape 1 : le propriétaire du code $\widetilde{B}$ émet le message codé

$$m = \widetilde{f}(\widetilde{B},\ \text{“je suis la banque”})$$

Etape 2 : à l'aide de la clef publique, nous décodons ce message comme si nous étions en train de coder un message en clair

$$\begin{aligned} f(B, m) &= f\left(B, \widetilde{f}(\widetilde{B},\ \text{“je suis la banque”})\right) \\ &= \text{“je suis la banque”} \end{aligned}$$

Alors l'auteur de ce message ne peut être que le propriétaire du code $\widetilde{B}$ puisque $\widetilde{B}$ est resté en la seule possession de son propriétaire

Authentification (ii)

Dans le cas de l'étape d'authentification d'une carte bancaire,
la Carte transmet au Terminal un double message du type
"je suis la carte de M. Martin",
$\widetilde{f}(\widetilde{B},$ "je suis la carte de M. Martin"$)$

La seconde partie de ce message a été émise
à l'aide du code secret $\widetilde{B}$ de la banque

Le Terminal dispose de la clef "publique" B de la banque,
il calcule $f\big(B, \widetilde{f}(\widetilde{B},$ "je suis la carte de M. Martin"$)\big)$
et le résultat est simplement "je suis la carte de M. Martin"

Par comparaison des deux messages,
l'authentification du compte à débiter est établie.

Authentification (iii)

Concernant le code secret qui identifie la personne (M. Martin) ?

Dans sa version de base,

le terminal lit le code tapé sur la clavier,
il le transmet à la carte,
laquelle dispose d'une copie (!) de ce code

La Carte répond simplement "oui, c'est le bon code"...
comme une "Yes-card", qui dit toujours "oui" !!

Authentification (iv)

Pour passer au travers des diverses sécurités, il "suffit" de

1) comprendre le processus d'authentification

2) lire le code "public" B diffusé par la banque dans le Terminal

3) "casser" ce code en calculant le code "privé" $\widetilde{B}$ de la banque

4) fabriquer une "Yes-card" qui comporte pour l'identification
un double message du type "je suis une fausse carte",
$\widetilde{f}(\widetilde{B},$ "je suis une fausse carte"$)$

La démonstration pratique a été effectuée en 1998,
par un "fraudeur" qui s'est fait connaître !!

La sécurité des cartes bancaires a été renforcée...

Le Terminal dispose de la clef publique A
d'une Autorité de certification
La carte dispose aussi de sa propre clef privée $\widetilde{C}$
La clef publique B de la banque
est codée par l'Autorité de certification : $\widehat{B} = \widetilde{f}(\widetilde{A}, B)$
La banque code avec sa clef privée $\widetilde{B}$
la clef publique C de la carte : $\widehat{C} = \widetilde{f}(\widetilde{B}, C)$
Outre les enregistrements précédents,
la carte porte les informations $\widehat{B}$ et $\widehat{C}$

Lors de l'authentification, le Terminal lit $\widehat{B}$ et $\widehat{C}$
il calcule la clef publique de la banque : $B = f(A, \widehat{B})$
avec celle-ci, il calcule la clef publique de la carte : $C = f(B, \widehat{C})$
il envoie un message aléatoire m
et la carte le code à l'aide de sa clef secrète : $\widetilde{m} = \widetilde{f}(\widetilde{C}, m)$
Le Terminal vérifie la relation $m = f(C, \widetilde{m})$.
Puis on s'intéresse enfin au compte de M. Martin !

Codage de Rivest, Shamir et Adleman (1977)

Comment disposer d'une clef publique B
qui ne permet pas (facilement !) de "calculer" la clef privée $\widetilde{B}$?

Grâce à la difficulté à décomposer certains nombres
en facteurs premiers !

Principe : le cryptage est public
et le décryptage reste privé (en principe !)
On considère p et q deux nombres premiers "de grande taille"
on calcule et on publie $n = p\,q = 145$
On se donne un nombre entier $\alpha = 33$ compris entre 0 et n
et premier à $(p-1)\,(q-1)$,
ce qui signifie que les nombres α et $(p-1)\,(q-1)$
n'ont **pas** de facteur premier en commun

L'auteur d'un code RSA **publie** le clef "publique" $n = 145$, $\alpha = 33$
et garde **secret** le nombre $(p-1)\,(q-1)$

Code RSA (ii)

Pour coder le nombre $m = 123$ ($1 \leq x \leq n$ par convention),
nous l'élevons à la puissance α **modulo** n
donc $\widetilde{m} = f_{RSA}(n = 145, \alpha = 33, m) = m^{\alpha}$ "modulo n"

En pratique m^{α} est un nombre de 212 (!) chiffres
et nous n'avons pas besoin de calculer car "modulo n" nous avons :

$$2\,m = 246 = 145 + 101, \text{ donc } \quad 2\,m = 101 \text{ modulo } 145$$
$$m^2 = 123 \times 123 = 104 \times 145 + 49, \text{ donc } \quad m^2 = 49 \text{ modulo } 145$$

De proche en proche :
$$m^4 = m^2\,m^2 = 49 \times 49 = 81 \quad \text{modulo } 145$$
$$m^8 = m^4\,m^4 = 81 \times 81 = 36 \quad \text{modulo } 145$$
$$m^{16} = m^8\,m^8 = 36 \times 36 = 136 \quad \text{modulo } 145$$
$$m^{32} = m^{16}\,m^{16} = 136 \times 136 = 81 \quad \text{modulo } 145$$
$$m^{33} = m^{32}\,m = 81 \times 123 = 103 \quad \text{modulo } 145$$

et $\widetilde{m} = f_{RSA}(n = 145, \alpha = 33, 123) = 103$

Code RSA (iii)

Pour décoder le message $\widetilde{m} = 103$ qui vient de nous parvenir, nous avons au préalable déterminé β de sorte que

$$\alpha\,\beta = 1 \quad \text{modulo } (p-1)(q-1).$$

On calcule ;

on a $\quad m = \widetilde{f}_{RSA}(n = 145, \beta = ??, \widetilde{m}) = \widetilde{m}^{\beta}$ "modulo n"

On admet dans un premier temps que $\beta = 17$

on vérifie ensuite facilement la propriété :

$$\begin{aligned}
\widetilde{m}^2 = 103 \times 103 = 73 \times 145 + 24, \text{ donc} \quad & \widetilde{m}^2 = 24 \text{ modulo } 145 \\
\widetilde{m}^4 = \widetilde{m}^2\,\widetilde{m}^2 = 24 \times 24 = 141 \quad & \text{modulo } 145 \\
\widetilde{m}^8 = \widetilde{m}^4\,\widetilde{m}^4 = 141 \times 141 = 16 \quad & \text{modulo } 145 \\
\widetilde{m}^{16} = \widetilde{m}^8\,\widetilde{m}^8 = 16 \times 16 = 111 \quad & \text{modulo } 145 \\
\widetilde{m}^{17} = \widetilde{m}^{16}\,\widetilde{m} = 111 \times 103 = 123 \quad & \text{modulo } 145
\end{aligned}$$

Nous retrouvons le "message" m que nous avions choisi au départ.
Nous avons vérifé (dans un cas particulier !) la relation

$$\widetilde{f}_{RSA}\big(n, \beta, f_{RSA}(n, \alpha, m)\big) = m$$

Code RSA (iv)

Est-il facile de “casser” le code RSA ?

- Oui si nous savons factoriser l'entier n
  en produit de deux facteurs premiers !!!
  Trouver p et q de sorte que $n = p\,q$

Pour $n = 145$, c'est (très) facile
car il est clairement divisible par 5, qui est un nombre premier !
Donc $p = 5, \quad q = 29 \quad (5 \times 29 = 145)$,
et le nombre **vraiment secret** $(p-1)\,(q-1)$ vaut ici $4 \times 28 = 112$

Pour $\alpha = 33$, son inverse modulo 112 vaut $\beta = 17$:
$17 \times 33 = 5 \times 112 + 1 = 1$ modulo 112

- Non si la recherche de p et q de sorte que $p\ q = n = 3\,464\,759$
  (pour fixer les idées) échoue.

S'authentifier à l'aide du code RSA

Nous avons par exemple reçu un message,
contenant outre l'information précédente $n = 3\,464\,759$,
les données $\alpha = 379$, $m = 7777$ et $x = 3\,286\,179$.

Remarquons qu'il est "facile"
(mais un peu long, c'est un véritable **défaut** du code RSA !)
de remarquer que le nombre x "n'a rien à voir" avec

$$\widetilde{m} = f_{RSA}(n = 3\,464\,759, \alpha = 379, m = 7777) = 2\,941\,631$$

Il est également facile de vérifier que

$$f_{RSA}(n = 3\,464\,759, \alpha = 379, x = 3\,286\,179) = 7777 = m$$

ce qui **démontre** la relation réciproque

$$\widetilde{f}_{RSA}(n = 3\,464\,759, \beta = ??, m = 7777) = 3\,286\,179 = x$$

Notre correspondant sait calculer $\widetilde{f}$
donc il connaît la clef β de ce code particulier ;
il s'est identifié de façon "authentique"

Les grands nombres premiers

Avec un peu de travail, il est “facile” de se rendre compte que l'exemple précédent est un modèle simple car

$$n = 3\,464\,759 = 2833 \times 1223$$

produit de deux nombres premiers de quatre chiffres... seulement

(on a $\beta = 566\,131$ pour les fanas !)

Il est “assez facile” de générer des nombres premiers

de “grande taille”

Les grands nombres premiers (ii)

Message de Herman te Riele (né en 1947, CWI Amsterdam)
du 26 août 1999 :
"On August 22, 1999, we found that the 512-bits number
RSA-155 =
109417386415705274218097073220403576120037329454492059909138421314763499842889347847179972578912673324976257528997818337970765372440271467435315933543338973
can be written as the product of two 78-digit primes:
102639592829741105772054196573991675900716567808038066803419335217907113077799 *
106603488380168454820927220360012878679207958575989291522270608237193062808643"

Une vingtaine d'informaticiens théoriciens des nombres,
7 mois de travail
un nombre "RSA-155" de 155 chiffres comporte environ 512 bits.

Les grands nombres premiers (iii)

- Plus récemment (2003, 2005) des "nombres RSA"
  de 193 et 200 chiffres (640 et 663 bits) ont été cassés
  par Jens Franke (né en 1964, Université de Bonn) et son équipe.

- Mai 2007.
  Un nombre de 307 chiffres cassé après onze mois de calculs
  Equipe de Arjen Lenstra (EPFL Lausanne)...

- Les codes RSA opérationnels comportent 1024 ou 2048 bits

- Lu sur le net :
  "Le cassage des clés RSA 1024 bits annonçé pour 2012"

Les grands nombres premiers (iv)

Actualité d'août 2008
Projet "Great Internet Mersenne Prime Search"
Le "nombre de Mersenne" $2^{43\,112\,609} - 1$ est premier
(un nombre de Mersenne est de la forme $2^p - 1$ avec p premier)
il s'écrit avec près de 13 millions de chiffres en base 10 (!!)

Edson Smith (Université de Californie, Los Angeles)
George Woltman (né en 1957, animateur du projet GIMPS)
Scott Kurowski (University of Central Missouri)
Marin Mersenne (Moine Français, 1588-1648)

Prix de l'Electronic Frontier Foundation

A venir :
150 000 dollars pour un nombre premier de 100 millions de chiffres,
250 000 dollars pour un nombre premier d'un milliard de chiffres.

Les gens

Auguste Kerckhoffs (1835-1903) : auteur d'un article en 1883.
L'efficacité d'un système cryptographique repose sur la clef, qui doit rester **secrète**.
Le protocole de codage doit lui être **public**.

Claude Shannon (1916-2001), ingénieur et mathématicien.
Père de la théorie mathématique de l'information ("entropie de Shannon").
"L'adversaire connaît le système", reformulation du principe de Kerckhoffs.

Alan Turing (1912-1954), mathématicien
Un des pères de l'informatique (calculabilité, "machine de Turing")
A dirigé l'équipe chargée de casser le chiffre allemand durand la seconde guerre mondiale (machine Enigma).

Les gens (ii)

Ronald Rivest (né en 1947),
Adi Shamir (né en 1952),
Leonard Adleman (né en 1945), cryptologues.

Auteurs du code "RSA" (1977),
système "à clef publique" le plus utilisé actuellement.

Xuejia Lai (né en 1970 ?) et James Massey (né en 1934).

Auteurs de l'algorithme de chiffrement "symétrique" (1991)
"International Data Encryption Algorithm".

Vincent Rijmen (né en 1970) et Joan Daemen (né en 1965)

Concepteurs de l'algorithme de chiffrement "symétrique"
"Advanced Encryption Standard" (1998).

Philip Zimmermann (né en 1954)

Auteur du protocole "Pretty Good Privacy" (1991),
algorithme de codage "en libre accès"

Les gens (iii)

Louis Claude Guillou, ingénieur crytologue

Un des concepteurs du système d'authentification des cartes bancaires (1983).

En a compris les limites (articles en 1988 et 1990).

Serge Humpich (né en 1963)

A compris (1998) le protocole de communication entre une carte bleue et un terminal.

A démontré que la fabrication d'une "Yes-card" est possible.

Condamné en 2000.

Lectures utiles

Thomas Genet.
"Le protocole cryptographique de paiement par carte bancaire",
http://interstices.info/jcms/c_33835, février 2008.

Jacques Patarin.
"La cryptographie des cartes bancaires",
Pour La Science, numéro spécial, juillet 2002.

Simon Singh.
Histoire des Codes Secrets,
Livre de Poche, numéro 15097, 1999.

Jacques Stern. *La science du secret*, Odile Jacob, 1997.

Bonus : pourquoi RSA marche-t-il ?

$n \equiv p\,q$, p et q premiers
$k = 1$ modulo $(p-1)\,(q-1)$. Alors $m^k = m$ modulo n.

On montre d'abord que $m^k = m$ modulo p
Si p divise m, alors $m = 0$ modulo p donc le résultat est vrai.
Si p ne divise pas m, le petit théorème de Fermat
entraîne que $m^{p-1} = 0$ modulo p.
Alors $m^k = m^{\ell\,(p-1)\,(q-1)+1} = \left(m^{(p-1)}\right)^{\ell\,(q-1)} m = m$ modulo p
De même avec le nombre q. D'où la conclusion *via* le lemme de Gauss car p et q sont premiers entre eux.

Soit α premier à $(p-1)\,(q-1)$
Soit β de sorte que $\alpha\,\beta = 1$ modulo $(p-1)\,(q-1)$
Si $\widetilde{m} = m^{\alpha}$ modulo n, alors $\widetilde{m}^{\beta} = m$ modulo n

On pose $k \equiv \alpha\,\beta$ qui est congru à 1 modulo $(p-1)\,(q-1)$
Alors $\widetilde{m}^{\beta} = \left(m^{\alpha}\right)^{\beta} = m^k$ qui est congru à m modulo n.