Formalisme Algébrique

Introduction

Tous les chercheurs avancés ont, à un moment ou à un autre, été confrontés à la difficulté d'informer leurs congénères sur l'état d'avancement de leurs recherches. J'ai souvent pu relever que, découragés par cette difficulté, beaucoup se limitaient à des déclarations plutot vagues du genre "j'ai essayé beaucoup de trucs dans telle direction" sans que le lecteur puisse avoir une idée précises des pistes effectivement envisagées Par ce post je vous propose une approche qui va considérablement faciliter les échanges de ce genre en leur donnant un cadre formel. Le système que je présente ici est puissant mais pas sans limite (certains cryptos faisant appel à des transformations spatiales complexes échappent à ce mode de représentation) mais il n'en demeure pas moins souvent fort utile.

Eléments de base

Une poignée de symboles suffisent à la description :

+Mt+	Cette expression signifie cryptage par la méthode Mt
-Mt-	" " " décryptage par la méthode Mt
M	Représente le message en clair
K	Clé (Key) Note : parfois il n'y a pas de clé et parfois il y en a plusieurs (cfr illustrations et relations fondamentales ci-après).
C	Message crypté
F()	Application de la fonction F au contenu de la parenthèse. Exemples : - Romain(x) : transformation du nombre x en chiffres romains. - Tranposée(y) : transposée de y. y peut être une chaine de car. ABC => CBA, une matrice, etc...
F(G(x))	Composées de fonctions, image par le fonction F de l'image de x par la fonction G. "Imbriquables" à l'infini.
"val"	Valeur littérale (à utiliser telle quelle)
(var)	Variable (représentant une élément quelconque de la scène du crypto).
a..z	Désigne toutes les valeurs discrètes successives de a à z (nombres, lettres ou autres). Cette notation fait implicitement référence à un processus itératif.
a-z	Désigne la séquence a-z considérée dans son ensemble.
=	Egalité
!=	Egalité unique/destructive (voir "Relations fondamentales" plus loin).

Illustrations

Quelques illustrations pour bien cerner les choses :

> M +Mt+ K = C
Le message en clair M, crypté par la méthode Mt en utilisant la clé K donne le message crypté C.

> M +Mt+ = C
Le message en clair M, crypté sans clé par la méthode Mt donne le message crypté C.
C'est le cas notamment du cryptage en dent de scie.

> "MESURE" +Adeba+ "MBPSLA" = "SAMDALMAS"
Le litteral MESURE crypté par la méthode Adeba avec la clé MBPSLA donne le texte SAMDALMAS.

> (Montant chèque en lettres) -Cesar- (1..25)
Décryptage du montant en lettres du chèque par la méthode César en utilisant successivement les 25 offsets possibles (voir également note "x mod 26 = 0" dans la partie "Extensions").

> (Montant chèque en lettres) +Vigénère+ Romain(Somme(Toutes les pièces))
Désigne le cryptage par la méthode Vigénère du montant en lettres du chèque en utilisant la clé constituée par la somme de toutes les pièces transformée en chiffres romains.
Note : pour ceux qui cherchent encore, ce n'est pas la solution de T12.

> M +Périple+ LMouvSatur(Start(PolyReg(2,8),B2))
Je vous laisse deviner... J'ai volontairement introduit une erreur pour qu'on ne puisse pas me reprocher de donner des soluces.

> M +Atb+ = C
Une première représentation de l'atbash. Ici on considère la clé est intrinsèque au système.
M +Atb+ Transposée(A-Z) = C
Deuxième approche, avec clé extrinsèque cette fois.
La question de savoir laquelle de ces 2 représentations est la meilleure procède du débat philosophique.

> CHR(B18convB10(LMovSatur(Start(PolyIrreg(Page droite),C5)))-x)
Celle-ci est un clin d'oeil à Yael59, c'est la représentation de mon hypothèse B18 pour T1-D (qui je l'admets était très tirée par les cheveux !)

Relations fondamentales

Quelques relations fondamentales :

M +Mt+ K = C (C'est la base même de la crypto, sans elle point de salut)

et aussi :

C -Mt- K = M (Et ça, son corrolaire direct et aussi la porte de sortie)

et donc :

(M +Mt+ K) -Mt- K = M

Note :
certains cryptos (militaires notamment) sont de la forme (M +Mt+ K) -Mt- K != M
(le décryptage est destructif et ne peut s'opérer qu'une seule fois).

Dans les cryptos industriels ou militaires on trouve aussi souvent :

M +Mt+ Kpu = C
C -Mt- Kpr = M

Note :
cryptage asymétrique, les clés de cryptage et de décryptage sont différentes.
Exemple :

(M +RSA+ Kpu) -RSA- Kpr = M

Dans l'algorithme Rivest-Shamir-Adleman la clé de cryptage est connue de tous - clé publique = Kpu - alors que la clé de décryptage n'est connue que des seules personnes autorisées à décrypter - clé privée = Kpr.

M +Mt+ K = K +Mt+ M Commutativité.

Ca vous paraît exotique ? Détrompez-vous !
Essayez donc un Vigénère dont M et K ont les mêmes longueurs...

M +Mt+ N = M Neutre, le crypto est non codant.

Etc...

Extensions

On le voit ci-dessus le formalisme algébrique permet d'aller beaucoup plus loin dans l'étude des propriétés des systèmes cryptographiques en leur fournissant un cadre simple de représentation. Grace à lui il est par exemple possible de facilement modéliser des propriétés telles que la commutativité, l'associativité, la réversivité, l'existence d'un élément neutre (exemple : pour le cryptage César l'élément neutre est le décalage x tel que x mod 26 = 0), etc...
On l'aura compris poussé à fond ce formalisme peut conduire à une théorie descriptive complète de la crypto, incluant des notions proches des concepts mathématiques de corps, de groupes, d'anneaux, etc... qui dépasse de loin le cadre des énigmes de Sam.

Application à T12

Pour en revenir à T12 et pour récompenser ceux qui ont eu le courage de poursuivre leur lecture jusqu'ici, je vous livre quelques petits amuse-gueule tous non pertinents (ne gaspillez donc plus votre temps, j'ai déjà perdu le mien pour vous):

> (Lettres des billets) -Vigénère- ("MBPSLA")
J'ai essayé avec et sans reset de clé sur chaque billet

> (Lettres des billets) -Cesar- (1..25)
Rien à voir, circulez !

Celui-ci n'est pas tout à fait non pertinent, mais personnellement je n'y crois pas trop :
> ("MBPSLA") -Cesar- (1..25)
Deux résultats sortent du lot : J... et F...

Celle-ci je l'ai testée un jour de pluie, il n'y avait rien à la télé :
> (Lettres des billets) -Vigénère- [("MBPSLA") -Cesar- (1..25)]
Résultat : des clous.

Etc, etc...

Clés isomorphes et relation d'équivalence

En me perdant dans les méandres labyrintiques de la crypto T12-G je suis tombé sur une intéressante équivalence dont je vous livre ici l'essence : l'utilisation de clés isomorphes Kn (clés constituées de n fois la répétition du caractère K) fait naître une - somme toute naturelle - relation d'équivalence entre le crytposystème Vigénère et le cryptosystème Soustractif utilisé par Sam dans T12-D. Cette équivalence s'exprime de la manière suivante :

M -Vigénère- (An..Zn) = Transposée (M -Soustractif- (An..Zn))

Note : vous pouvez remarquer qu'il ne s'agit plus de représenter des cas particuliers de décryptages, mais bien de comparer des méthodes entre-elles sur des sets entiers de clés.

Cette relation d'équivalence est moins triviale qu'il n'y paraît car elle signifie que dans une attaque en force brutale à clés isomorphes, seule une des deux méthodes est pertinente, l'autre n'est que la transposée de la première (note: si vous essayez veillez à utiliser le même offset pour A pour les 2 méthodes, sinon vous allez avoir des surprises).

Périodicité d'une clé

Dans les systèmes monosubstitutifs standard (Vigénère, Soustractif, etc... - que pour la facilité je noterai "mono-standard" dans la suite de ce post) la notion de longueur de clé est tout à fait intuitive : la longueur de la clé est égale au nombres de caractères qui la constituent.
La périodicité d'une clé se définit alors aussi simplement :

$$P = \frac{1}{l}$$

Evidemment plus la clé est longue plus la périodicité est faible (autrement dit plus longtemps il faudra attendre pour observer un reset de clé au niveau des caractères du message crypté).
Plus précisément dans un message codé mono-standard un reset de clé sera observé tous les :

$$i = (l * \sum_{0 \to len(C)} n) + 1$$

Par exemple pour une clé de longueur 7, dans le message crypté les indices des caractères suivants correspondant à des reset de clé :
1, 8, 15, 22, 29, ...

Note importante : dans les années 40, Shannon démontra que, pour être totalement surs, les systèmes à clés symétriques privées (auxquels appartiennent les mono-standard) doivent utiliser des clés de longueur au moins égales à celle du message à chiffrer.
Sam n'a - heureusement pour nous - pas suivi ce conseil, ce qui comme nous allons le voir plus loin a conduit à certains problèmes de sécurité...

Toute méthode est une fonction

Jusqu'à présent dans le formalisme j'ai toujours tracé une frontière claire entre d'une part les méthodes de cryptage et d'autre part les fonctions de transformation des éléments primaires :

- une méthode correspond in fine toujours à un algorithme discret dont toutes les étapes sont parfaitement définies et s'enchainent selon une séquence prédéterminée (exception faite du crypto faisant intervenir l'opérateur !=, qui est un cas vraiment à part).
- une fonction correspond quant à elle à une mise en forme préliminaire (ou éventuellement intermédiaire) des éléments de la scène crypto avant qu'interviennent les méthodes cryptographies elles-mêmes.

Par exemple dans l'expression :

(Lettres billets) +Vigénère+ Romain(Pièces orphelines)

Vigénère est la méthode et Romain() la fonction de mise en forme; on voit clairement la préséance des deux.

Maintenant disons le tout net : la séparation entre méthode et fonction n'est qu'une vue de l'esprit, elle n'a pas d'existence réelle; toute méthode n'étant rien d'autre qu'une composée plus ou moins complexes de fonctions elles-mêmes plus ou moins complexes (en fait j'avais déjà cette unification en tête lorsque j'ai introduit l'opérateur x..y).

Une méthode est toujours décomposable en un algorithme articulé autour des 3 structures suivantes (propres aux langages procéduraux) :
- la séquence
- la répétitive
- l'alternative

La séquence est restituée par l'ordre naturel d'évaluation des fonctions imbriquées qui s'opère toujours du paramètre le plus encapsulé vers celui qui l'est le moins.

Par exemple :

F(G(H(I(x)))))

s'interprète selon la séquence : x => I => H => G => F
La séquence s'en trouve donc modélisée.
La question des séquences de paramètres indépendants ne se pose pas vraiment, elle peut être rendue par l'usage d'une fonction identité (G(x)=x) ou par une fonction de séquençage pur (rappellez-vous que vous avez toute latitude dans la définition de vos fonctions).

La répétitive est le domaine de l'opérateur X..Y, pas besoin d'en dire d'avantage.

A l'alternative correspond la fonction ternaire Si(Cond, V, F) : si la condition est vérifiée, la fonction renvoie V, sinon elle renvoie F.

La conclusion logique de la transposabilité de ces 3 structures de bases est que toute méthode est effectivement bien une composée de fonctions.

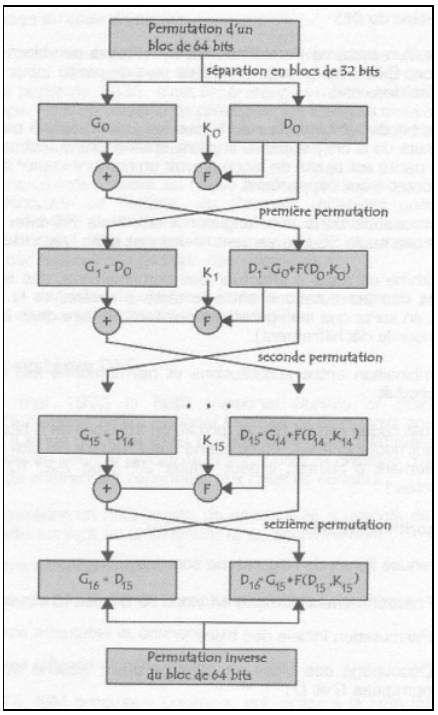
L'expression de départ :

(Lettres billets) +Vigénère+ Romain(Pièces orphelines)

est donc équivalente à :

Fn(...Fn-1(...Fn-2(...(lettres billets),...)...,(Pièces orphelines)...)...)

Autre exemple avec l'algorithme du cryptosystème DES :



A y regarder de plus près il ne contient rien d'autres que des fonctions encapsulées...

Bien qu'aucun doute ne soit plus permis en ce qui concerne la subjectivité de la frontière méthode-fonction, je vais continuer à l'utiliser pour des raisons pratiques : elle est plus intuitive et conduit à des notations plus synthétiques.

Propriétés des cryptosystèmes

Vigénère et inversibilité - décryptage par indiscrètion

Il convient d'être très prudent lors l'utilisation de cryptosystèmes mono-standard dans le cadre d'un concours faisant intervenir un grand nombre de participants (comme la chasse de Sam par exemple).

Ces systèmes possèdent le gigantesque talon d'achille suivant :

M +Vigénère+ K = C (phase de cryptage)

C -Vigénère- K = M (phase de décryptage normal)

Et ici le piège se referme (décryptage par indiscrétion) :

C -Vigénère- Mid(M,p,q) = Mid(K,r,q)

Note : Mid(Chaine,a,b) = extraction de la sous-chaine de Chaine, commençant à l'indice a et de longueur b.

Si on applique l'alorithme de Vigénère au texte crypté en utilisant une partie du texte en clair, on retrouve

Dans les cryptosystèmes mono-standard, le non respect de la condition de Shannon sur les clés de cryptage (voir plus haut) établit une relation bijective directe et inversible entre l'ensemble des caractères du message en clair et ceux de la clé [(1)]. Autrement dit en cas d'indiscrètion sur un seul mot du message en clair c'est quasiment la totalité de la protection du message crypté qui tombe.

En fait plus (l * n) + 1 << len(M), plus la situation est explosive.

(1)Pour être complet il faut signaler que cette relation existe aussi si la condition de Shannon est respectée, mais dans ce cas, elle est totalement inexploitable.

Equivalence Alpha-linéaire - Vigénère

Pour l'alpha-linéaire :

C +AphaLineaire+ K = C +Vigénère+ (A-Z)

Et pour l'alpha-linéaire inverse :

C +AphaLineaireInv+ K = C +Vigénère+ (Z-A)

Commutativité du Vigénère

Lorsque M et K sont de même longueur on observe une équivalence entre crypter le message avec la clé et crypter la clé avec le message :

M +Vigénère+ K = K +Vigénère+ M

Transposition Théosophale

M +TransTheosophale+ = M mod 9

Fin de la version publique