Cryptographie

Quentin Deschamps

Automne 2020 *M2 SRI, ISFA*

Théorie de l'information

Généralités

- Question : quelle est la quantité d'information contenue dans un message ?

Généralités

- Question : quelle est la quantité d'information contenue dans un message ?
- Théorie de Shannon (1948), basée sur la théorie des probabilités.

Généralités

- Question : quelle est la quantité d'information contenue dans un message ?
- Théorie de Shannon (1948), basée sur la théorie des probabilités.
- Permet de définir rigouresement les notions de sécurité et de secret.

Définitions de probabilités

Variable aléatoire

Une variable aléatoire discrète **X**, consiste en un ensemble X, une distribution de probabilité discrète $(p_x)_{x \in X}$ sur X et de la donnée $Pr[X = x]$: la probabilité que X se réalise en x.

Exemples

Lancer de pièces, tirage de dés...

Définitions de probabilités

Probabilité mutuelle

Soient deux variables aléatoires **X** et **Y** définies sur des ensembles finis X et Y respectivement. La probabilité mutuelle $Pr[X = x, Y = y] = Pr[x, y]$ est la probabilité pour que X se réalise en x et Y se réalise en y.

Probabilité conditionnelle

Soient deux variables aléatoires **X** et **Y** définies sur des ensembles finis X et Y respectivement. La probabilité conditionnelle $Pr[X = x|Y = y] = Pr[x|y]$ est la probabilité que X se réalise en x sachant que Y s'est réalisé en y.

Définitions de probabilités

Variables indépendantes

Les variables aléatoires X et Y sont dites des variables aléatoires indépendantes si $Pr[x, y] = Pr[x|y]$ pour tout $x \in X$ et tout $y \in Y$

Définitions de probabilités

Variables indépendantes

Les variables aléatoires X et Y sont dites des variables aléatoires indépendantes si $Pr[x, y] = Pr[x|y]$ pour tout $x \in X$ et tout $y \in Y$

Idée : X représente le message, Y son code et on veut que X et Y soient indépendantes.

Définitions de probabilités

Variables indépendantes

Les variables aléatoires X et Y sont dites des variables aléatoires indépendantes si $Pr[x, y] = Pr[x|y]$ pour tout $x \in X$ et tout $y \in Y$

Idée : X représente le message, Y son code et on veut que X et Y soient indépendantes.

Confidentialité parfaite

Un système cryptographique probabiliste (P, C, K, E, D) assure une confidentialité parfaite si $Pr[x|y] = Pr[x]$ pour tout $x \in P$ et tout $y \in C$

Exemple : Chiffrement de César

- Chiffrement par décalage : la clef indique le décalage.
- Exemple : Cesar décalé de 3 donne fhvdu.

Exemple : Chiffrement de César

- Chiffrement par décalage : la clef indique le décalage.
- Exemple : Cesar décalé de 3 donne fhvdu.

Théorème

Le chiffrement ce César assure une confidentialité parfaite si et seulement si chaque clef est choisie avec une probabilité $\frac{1}{26}$

Exemple : Chiffrement de César

- Chiffrement par décalage : la clef indique le décalage.
- Exemple : Cesar décalé de 3 donne fhvdu.

Théorème

Le chiffrement ce César assure une confidentialité parfaite si et seulement si chaque clef est choisie avec une probabilité $\frac{1}{26}$

Confidentialité parfaite $\neq$ Sécurité

Entropie

- Idée : Plus un évènement est probable, moins il apporte d'information.
- Exemple : Pour connaître une carte au tarot, savoir que c'est un roi donne plus d'information que de savoir que c'est un pique.

Entropie

- Idée : Plus un évènement est probable, moins il apporte d'information.
- Exemple : Pour connaître une carte au tarot, savoir que c'est un roi donne plus d'information que de savoir que c'est un pique.

Definition Entropie

L'entropie de la variable aléatoire **X** est définie comme étant la quantité

$$H(X) = -\sum_{x \in X} Pr[x] \log_2(Pr[x])$$

A propos de la définition

Definition Entropie

L'entropie de la variable aléatoire **X** est définie comme étant la quantité

$$H(X) = -\sum_{x \in X} Pr[x] \log_2(Pr[x])$$

- Notion venant de la thermodynamique.

A propos de la définition

Definition Entropie

L'entropie de la variable aléatoire **X** est définie comme étant la quantité

$$H(X) = -\sum_{x \in X} Pr[x] \log_2(Pr[x])$$

- Notion venant de la thermodynamique.
- Le signe $-$ permet d'avoir une entropie toujours positive.

A propos de la définition

Definition Entropie

L'entropie de la variable aléatoire **X** est définie comme étant la quantité

$$H(X) = -\sum_{x \in X} Pr[x] \log_2(Pr[x])$$

- Notion venant de la thermodynamique.
- Le signe $-$ permet d'avoir une entropie toujours positive.
- Exercice : quand est-ce qu'une entropie est nulle ?

A propos de la définition

Definition Entropie

L'entropie de la variable aléatoire **X** est définie comme étant la quantité

$$H(X) = -\sum_{x \in X} Pr[x] \log_2(Pr[x])$$

- Notion venant de la thermodynamique.
- Le signe $-$ permet d'avoir une entropie toujours positive.
- Exercice : quand est-ce qu'une entropie est nulle ? Exercice : Calculer l'entropie d'une pièce équilibrée. Comment varie l'entropie dans le cas d'une pièce non équilibrée ?

A propos de la définition

Definition Entropie

L'entropie de la variable aléatoire **X** est définie comme étant la quantité

$$H(X) = -\sum_{x \in X} Pr[x] \log_2(Pr[x])$$

- Notion venant de la thermodynamique.
- Le signe $-$ permet d'avoir une entropie toujours positive.
- Exercice : quand est-ce qu'une entropie est nulle ? Exercice : Calculer l'entropie d'une pièce équilibrée. Comment varie l'entropie dans le cas d'une pièce non équilibrée ?
- $H(X) = p \log(p) + (1-p) \log(1-p)$

Propriétés de l'entropie

Propriété

L'entropie d'une variable aléatoire **X** correspond au nombre de bits à utiliser pour encoder sans perte d'information.

Propriétés de l'entropie

Propriété

L'entropie d'une variable aléatoire **X** correspond au nombre de bits à utiliser pour encoder sans perte d'information.

- L'entropie de la clef secrète doit être suffisamment élevée.
- Exercice : quelle est l'entropie maximale de la distribution de la clef pour un code de César ?

Rappel

- Pour prouver qu'un protocole est sécurisé on prouve que si on arrive à casser le protocole on arrive à résoudre un problème réputé difficile.
- Pour cela on suppose que l'on fait casser le protocole et on l'utilise pour résoudre le problème. Réduction

Qu'est-ce qu'un problème difficile

Définiton

Un problème est calculatoirement difficile si il n'existe pas d'algorithme le résolvant en un temps polynomial en la taille de l'entrée

Qu'est-ce qu'un problème difficile

Définiton

Un problème est calculatoirement difficile si il n'existe pas d'algorithme le résolvant en un temps polynomial en la taille de l'entrée

Les résultats d'impossibilité sont difficiles à prouver, on se contente souvent de problème dont on ne connaît pas d'algorithme polynomial.

Problèmes difficiles classiques

- RSA : factoriser le produit de deux grands nombres premiers
- Logarithme discret : Dans un groupe discret connaissant un générateur g, trouver x tel que $h = g^x$.
- LWE : différencier une distribution aléatoire d'un message bruité par une gaussienne discrète.

Un exemple : RSA flexible

$\mathcal{C}$		$\mathcal{R}$		$\mathcal{F}$
$(N, e, d) \leftarrow RSA\,\mathrm{KeyGen}(k)$				
$y \xleftarrow{\$} \mathbb{Z}/N\mathbb{Z}$				
	$(N, y) \rightarrow$			
		$m' \xleftarrow{\$} \mathcal{M}$		
		$E \leftarrow \prod_{m \in \mathcal{M}, m \neq m'} h(m)$		
		$u \leftarrow y^E[N]$		
			$pk^* = (u, N) \rightarrow$	
			loop	
			$\leftarrow$	m_i
		si $m_i = m'$ **le programme s'arrête**		
		sinon $\sigma = y^{E/h(m)}$		
			$\rightarrow$	
			$\leftarrow$	(m^*, σ^*)

Si $m^* = m'$ on peut déduire un couple (x, e) solution pour RSA flexible

Remarques

- Il faut faire attention à être sur une instance difficile du problème : factoriser un nombre n'est pas toujours difficile.
- Il n'est pas toujours possible d'établir une réduction, parfois il n'y a qu'un seul sens qui est valide.

Problème

Comment échanger une clef secrète sans aucune communication préalable ?

Exemple historique : Puzzles de Merkle

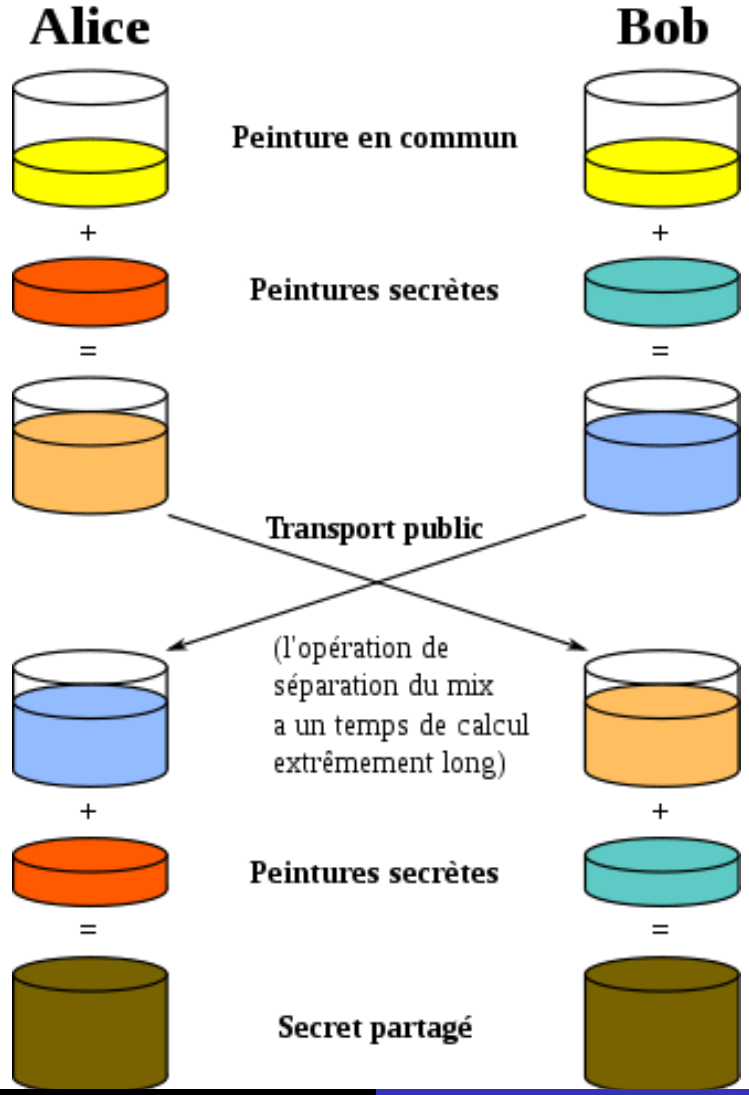
- Alice envoie un grand nombre de puzzles de complexité moyenne à Bob. Chaque puzzle contient un identifiant et une clef.
- Bob résout un puzzle au hasard et renvoie l'identifiant à Alice.
- Alice et Bob connaissent tous les deux la clef.
- Un attaquant doit résoudre tous les puzzles jusqu'à trouver celui avec le bon identifiant pour connaître la clef.

Exemple historique : Puzzles de Merkle

- Alice envoie un grand nombre de puzzles de complexité moyenne à Bob. Chaque puzzle contient un identifiant et une clef.
- Bob résout un puzzle au hasard et renvoie l'identifiant à Alice.
- Alice et Bob connaissent tous les deux la clef.
- Un attaquant doit résoudre tous les puzzles jusqu'à trouver celui avec le bon identifiant pour connaître la clef.

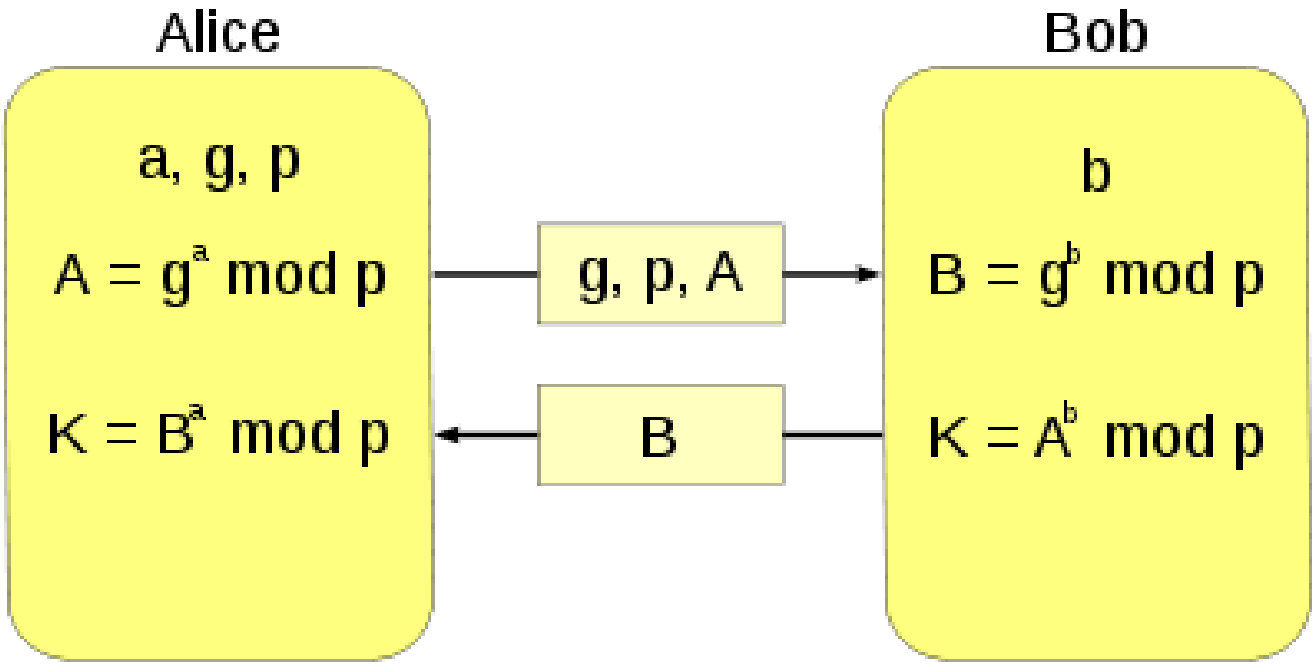
Le protocole n'atteint pas un niveau de sécurité suffisant pour les standards actuels.

Le protocole Diffie-Hellman en peinture



Le protocole Diffie-Hellman en mathématiques

On considère les éléments d'un groupe G dans lequel le logarithme discret est difficile



$K = A^b \bmod p = (g^a \bmod p)^b \bmod p = g^{ab} \bmod p = (g^b \bmod p)^a \bmod p = B^a \bmod p$

Masque de Véran

- Codage d'un message m de longueur l.

Masque de Véran

- Codage d'un message m de longueur l.
- Générer une chaîne aléatoire de caractère de longueur l : $r \xleftarrow{\$} \{0,1\}^r$

Masque de Véran

- Codage d'un message m de longueur l.
- Générer une chaîne aléatoire de caractère de longueur l : $r \xleftarrow{\$} \{0,1\}^r$
- Sommer (ou exclusif les deux chaînes)

Masque de Véran

- Codage d'un message m de longueur l.
- Générer une chaîne aléatoire de caractère de longueur l : $r \xleftarrow{\$} \{0,1\}^r$
- Sommer (ou exclusif les deux chaînes)
- Clef secrète : r, message codé $c = r \oplus m$

Masque de Véran

- Codage d'un message m de longueur l.
- Générer une chaîne aléatoire de caractère de longueur l : $r \xleftarrow{\$} \{0,1\}^r$
- Sommer (ou exclusif les deux chaînes)
- Clef secrète : r, message codé $c = r \oplus m$
- Déchiffrement $m = c \oplus r$

Masque de Véran

- Codage d'un message m de longueur l.
- Générer une chaîne aléatoire de caractère de longueur l : $r \xleftarrow{\$} \{0,1\}^r$
- Sommer (ou exclusif les deux chaînes)
- Clef secrète : r, message codé $c = r \oplus m$
- Déchiffrement $m = c \oplus r$

Théorème

Le masque de Véran achève une sécurité parfaite si r est parfaitement aléatoire.

Limitations

- Clef secrète de même longueur que le message.
- Impossible d'utiliser deux fois la même clef.
- Aléatoire parfait : difficile à créer.

Limitations

- Clef secrète de même longueur que le message.
- Impossible d'utiliser deux fois la même clef.
- Aléatoire parfait : difficile à créer.

C'est un idéal que l'on cherche à approcher.

Deux pistes de solution

- Extension de clef
- Réseaux de substitution-permutation

Principe

Agrandir la clef secrète de façon à garder une clef "ressemblant à de l'aléatoire".

Principe

Agrandir la clef secrète de façon à garder une clef "ressemblant à de l'aléatoire".

Fonction pseudo-aléatoire

Une fonction f est dite pseudo-aléatoire si un attaquant polynomial distingue $Im(f)$ de la distribution aléatoire avec probabilité négligeable.

Vision cryptographique de la définition

On cherche à attaquer $f : E \to F$.

$\mathcal{C}$		$\mathcal{A}$
$b \xleftarrow{\$} \{0,1\}$		
	loop	
	$\leftarrow$	$e \in E$ choisi par $\mathcal{A}$
Si $b = 0$ alors $r \xleftarrow{\$} F$		
Sinon $r \leftarrow f(e)$		
r		
	$\rightarrow$	
	end loop	
		renvoyer $s \in \{1\}$.

La fonction f est pseudo-aléatoire si pour tout attaquant $Pr[s = b] - \frac{1}{2}$ est négligeable.

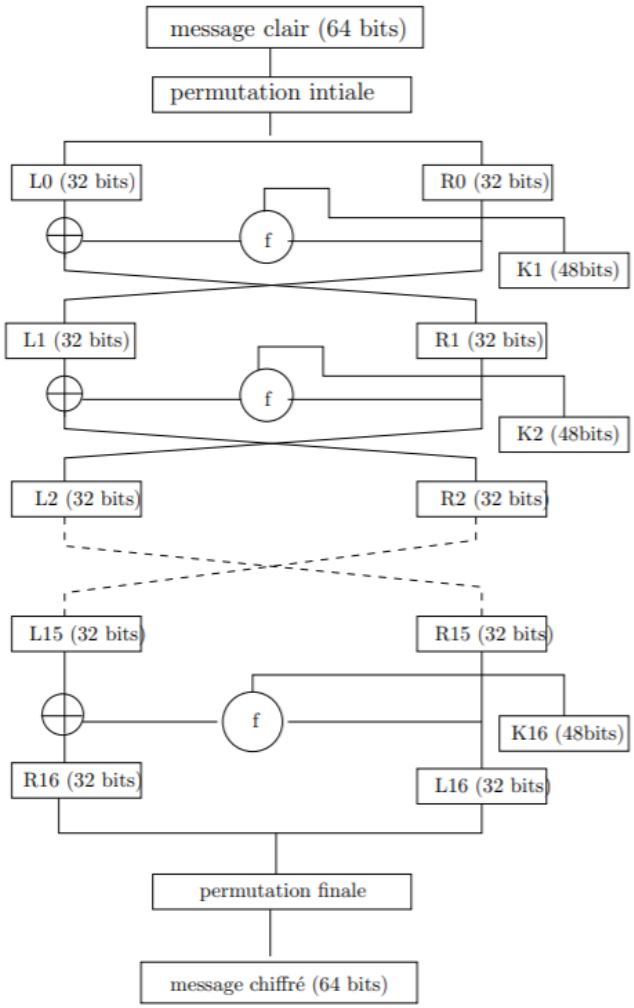
Principe

- Coder un morceaux de texte, effectuer une permutation, recoder un bout de texte, etc.

Principe

- Coder un morceaux de texte, effectuer une permutation, recoder un bout de texte, etc.
- Facile à mettre en place
- Difficile à analyser

Exemple : AES


message clair (64 bits)
permutation intiale
L0 (32 bits)
R0 (32 bits)
f
K1 (48bits)
L1 (32 bits)
R1 (32 bits)
f
K2 (48bits)
L2 (32 bits)
R2 (32 bits)
L15 (32 bits)
R15 (32 bits)
f
K16 (48bits)
R16 (32 bits)
L16 (32 bits)
permutation finale
message chiffré (64 bits)

Méthode de cryptage

- On somme la clef et le message.
- Exemple message :CRYPTANALYSE, clef secète : cle message encodé : FDDSFFQMQBEJ

Méthode de cryptage

- On somme la clef et le message.
- Exemple message :CRYPTANALYSE, clef secète : cle message encodé : FDDSFFQMQBEJ
- Exercice : décoder (collectivement) :

Enigme

Qygb Curxlut gvlzwa xk zh pozsy ictgk g'pf y'sjbgdwuoh. Pfygl goh h ggfjbkf lh zcbnk vhnk rhhy zlm ivhgvg, wlkbhhz zlm ivlgobz yz zlm ysunosym wip mk dyyysungwlhz. Ws yxfh uobzc zcbnk zh gghphks. Pf khhcz su jxcpy g iuy lcbfk rl mkbzuzwvhy bvobssfkg.

Analyse fréquentielle

- Calculer le nombre d'itération de chaque lettre en prenant une lettre sur N pour différente valeur de N.
- Calculer $\delta = \sum_{lettre}(frequence - \frac{1}{26})^2$ pour chaque valeur de N. La valeur maximale correspond à la longueur de la clef.

N	1	2	3	4	5	6
δ	0,017	0,028	0,039	0,06	0,031	0,047

On suppose que la clef a taille 4

Détermination de la clef

On calcule le nombre d'occurrence de chaque lettre en comptant une lettre sur 4 et on essaie de déterminer la clef. En partant de la première lettre :
2, 3, 1, 3, 1, 0, 0, 0, 0, **9**, 0, 2, 0, **9**, 0, 0, 0, 6, 1, 0, 3, 1, 4, 3, 2, 0
Les deux "9" espacé de 4 pourraient correspondre à "a" et "e" soit un décalage de 8 correspondant à la lettre "h".

Détermination de la clef

On calcule le nombre d'occurrence de chaque lettre en comptant une lettre sur 4 et on essaie de déterminer la clef. En partant de la première lettre :
2, 3, 1, 3, 1, 0, 0, 0, 0, **9**, 0, 2, 0, **9**, 0, 0, 0, 6, 1, 0, 3, 1, 4, 3, 2, 0
Les deux "9" espacé de 4 pourraient correspondre à "a" et "e" soit un décalage de 8 correspondant à la lettre "h".
Le même raisonnement en commençant à la lettre 2,3 et 4 donne la clef "hugo"

Une bonne référence

https://www.math.univ-paris13.fr/ boyer/enseignement/PolyCrypto2010.pdf