Information theory

M. Filiol

Antoine Puissant

25 septembre 2015

Résumé

Table des matières

1 Plan

- Majeure
  - Information theory (Filiol)
  - Programmation sécurisée (Ray) (C & Web)
  - Error correcting code (Code correcteurs et détecteurs d'erreur) (Filiol)
  - Réseaux (Ray)
  - Mathematics for security (Filiol)
    - Théorie des graphs
    - Fonctions booléennes
  - 2 semaines de projet
- Mineure technique
  - Cryptographie (Filiol)
  - Réseaux (Ray)
  - Stéganographie
  - Forensic
  - 2 semaines de projet
- Mineure managériale
  - Methodology
  - Ethique, droit et réglementation
  - OSINT (Renseignement en sources ouvertes)
  - 2 semaines de projet

Partiel = 75% des questions des tests du moodle + 25% de questions plus ouvertes.

2 Théorie de l'information

2.1 Introduction

Historiquement, on l'appelle la théorie de Shannon. Deux textes originaux publiés en 1948 et 1949.
Qu'est-ce que l'information ?
L'incertitude = l'effort qu'un attaquant doit fournir pour trouver une clé secrète.
Il faut définir ce qu'est l'information et un système de mesure.
L'information n'existe que quand elle passe sur un support (!= d'idée, concept).
Le message passe ensuite par un transmetteur. On a ensuite le canal de communication (support). Dès qu'on a un support, la nature va pouvoir l'altérer.
Il faut luter contre l'altération du support, donc de l'information.
Pour la partie réception, il faut s'assurer que l'information reçue soit proche de celle émise, malgré le bruit incorporé par la nature.
Ici, on se place dans le domaine de la sûreté : pas d'actions malveillantes.
On parle de sécurité lorsque qu'il y a des actions malveillantes. On a 3 pb : Comment caractériser une chose Comment la transmettre Comment se prémunir des attaques malveillantes

- L'information doit être représentée
- Le concept d'incertitude est essentiel. On ne transmet pas qqch d'évident. L'information est alors incertaine.
- Seulement une partie de l'information n'est pas prédictible

Pour passer de la tête au support, il nous faut des outils. Quand on veut représenter un message, on utilise des signaux et des symboles. Mais comment bien choisir ces symboles de manière économique (taille de l'info -¿ taille du disque, bande passante, etc...). Selon les besoins, on ne va pas utiliser les même méthodes de stockage de l'information (binaire, hexadécimale, base dix, etc...).
On va retrouver deux types d'information : l'information utile et l'information parasite (le bruit).

2.2 L'incertitude

2.2.1 Intoduction :

L'incertitude d'une variable aléatoire qui va prendre des valeurs X_i avec des probabilités p_i sera une fonction qui va définir uniquement des probabilités p_i.
On note cette fonction $H(p_1, ..., p_n)$.
On veut une incertitude max quand toutes les probas sont égales 'distribution équiprobable).
Pour toutes permutation, l'entropie reste la même.
Dès qu'on à un espace plus grand, l'incertitude est plus grande.
Continue et concave : toute variation infime ne provoquera pas de grande variation sur la fonction.
On peut « sommer » les événements.
Dans le théorème de H, λ est souvent égal à 1.

Vérifier que H valide les axiomes et trouver l'entropie de l'AES → res = 11
Touver un couple $(p_1, p_2) \in [0; 1]^2$ ayant une entropie de 0.75. Dessiner la courbe d'entropie. Soit, A et B deux événements indépendants de

distribution P_1 et P_2. Soit $P_3 = P_1 * P_2$ le produit direct de P_1 et P_2. $P_3 = P_i * P_j / P_i \in P_1; P_j \in P_2$.
Montrer que $H(P_3) = H(P_1) * H(P_2)$. Dire si les point A, B et C (slide 2) sont vrais ou non avec preuves.

2.3 L'entropie et ses propriétés

On prend ici un vecteur de variables aléatoires (finies et dénombrables).
$H(X)$ est l'entropie conjointe.
Le premier théorème donne une borne supérieure au l'entropie.
Démontrer ça (dérivée au dessus du log, fonction concave)
Le second théorème donne que l'entropie conjointe est inférieure ou égale au produit des entropies. Dans le cas de variables indépendantes, la probabilité conjointe est égale à a somme des entropies.

2.4 L'entropie conditionnelle

Entropie conditionnelle rejoint la probabilité conditionnelle.
L'entropie conditionnelle est alors l'entropie de X sachant que A c'est déjà réalisé est de la forme : $H(X|A) = -\sum_{i=1}^{n} p_i.log(p_i)$
Quelle est l'incertitude résiduelle de ce message capturé : DFDJ FTU VO NFTT BHF TFDSFU
$H(X|Y)$ est l'incertitude de X après avoir observé Y.
Démontrer que $H(X|Y) = H(X)$ lorsque X et Y sont indépendants
Théorème « chain rule » : $H(X|Y) = H(X) + H(X|Y) = H(Y) + H(Y|X)$

2.5 Mesure de l'information

Montrer que pour tout X, on a $H(X^2|X) = 0$ mais donner un exemple pour montrer que $H(X|X^2)$ ne vaut pas 0.

2.6 Conclusion

3 Qu'est-ce qu'une source d'informations ?

Une source d'informations est une suite de symboles d'un même alphabet.
Source sans mémoire (zero-memory) : chaque symbole est vu comme une VA indépendante des autres. Les suivantes ne sont pas influencées par ce qui a été vu avant.
Le problème est : comment représenter l'information de la meilleure manière possible ?
Toute chaîne de codage a au plus, une image. On a ici des injection et pas des bijections. **On a une source S.**
Soient les mots ω_1**,** ω_2**,** ω_3 **et leurs probabilités respectives** $p_1 = 0.9$**,** $p_2 = 0.05$**,** $p_3 = 0.025$**,** $p_4 = 0.025$**.**
Comparer les codages suivants :
Voir feuille manu
$f1 et f2$ **instantanés ?**

« *A code f is instantaneous or a prefix code if...* »
Un code instantanée est uniquement déchiffrable (l'inverse n'est pas vrai). Il peut être déchiffré à la volé.
Exemples :
$f(\omega_1) = 0$
$f(\omega_2) = 10$
$f(\omega_3) = 110$
$f(\omega_4) = 1110$
Soit un mot reçu suivant : 0|110|10|0|10|10|0|10
Soit, une fois déchiffré : $\omega_1|\omega_3[\omega_2|\omega_1|\omega_2|\omega_1|\omega_2$
Décodage par l'arrière = mémoire qui stocke tout le message puis décodage une fois que tout est reçu. Donc pas intéressant (long, consomme beaucoup de ressources). **Montrer que quelque soit** $n \in N^*$**, il existe un code instantané sur [0, 1] qui a des mots de toutes les longueurs possibles dans l'ensemble** $\{1, ..., n\}$
L'inégalité de Kraft ne le fait que pour les codes instantanés. McMillan le fait pour les codes uniquement déchiffrables.
preuves pour Kraft et McMillan dans les articles donnés.

4 Construction de codes compacts

Algorithme d'Huffman = construction récursive.

5 Communication through noisy channels

Peut-on lutter contre les altération de la nature et si oui, comment ?
Un canal de communication est une boîte noire. Cela peut être un espace mémoire (HDD, RAM) sur laquelle on va lire et écrire.
Les canaux sans mémoire (un symbole ne dépend pas de ceux qui le précèdent) discrets (un nombre fini).
Un canal est défini par une matrice de canal/distribution.
La proba p est la proba d'erreur.

5.1 Codage

La probabilité que le symbole binaire soit altéré est de p.
La probabilité qu'un symbole ne soit pas changé est $q = 1 - p$.
Sur n mots émis, la probabilité que les mots soient émis correctement est de :

$$((1-p)^3)^n$$

soit,

$$(1-p)^{3n}$$

Doubler le message ne fait que de le détection d'erreurs.
Pour de la correction d'erreurs, on va pouvoir faire de la « k-répétition » :

1 0 1 1

est le message de base. Celui envoyé est le suivant :

111 000 111 111

On reçoit alors :

101 110 111 101

On va alors faire un « vote majoritaire » et avoir le message suivant :

1 1 1 1

On a donc ici un canal qui bruite plus que prévu. On va donc faire une répétition de 5 :

11111 00000 11111 11111

Et on reçoit :

11011 01100 11111 11011

Le vote majoritaire donne alors le résultat suivant :

1 0 1 1

5.2 Décodage

Une règle de décodage doit permettre de produire une partition.
L'objectif est de minimiser le taux d'erreur résiduelle.
L'observateur ne connaît pas des probabilité de distribution de ce qui lui arrive.
Lorsque les mots de codes ont la même probabilité, les deux règles sont équivalentes.

6 Capacité

7 Second théorème de Shannon

8 Conclusion

Moyennant un mécanisme de redondance, on pourra toujours s'assurer que Bob et Alice aient la même information.
On est ici dans la sûreté, la nature n'est pas mal icieuse (contrairement à l'attaquant qui va changer en permanence p de bruitage).

9 General information sources

Modèle zero-memoire pas très adapté aux langages naturels.
Source est une obj qui emet des signaux aléatoires.
On considère une source ayant de la mémoire (dépendance avec les carac précédents)
Source stationnaire : si on prend une séquence $s_1, s_2, \ldots, s_n$, l'apparition de la séquence ne dépend pas du temps.
Règle des trois σ : La proba que $[\mu - 3\sigma < X < \mu + 3\sigma] = 0.9$
Une source est ergodique si elle est stationnaire et si pour tout pattern, le proba de réalisation d'un pattern tend vers 1.

9.1 Chaînes de Markov

Markov source with ireductible matrix $\rightarrow$ langages naturels
Distribution limite = distribution stable.
Le premier théorème de Shannon s'applique aux sources générales.
Le théorème 2 de Shannon s'applique aussi.

9.2 Langage

Chumsky
On a un pb : On a une séquence et une grammaire. Est-ce que le mot $\in$ la grammaire.
On peut classer les langages en 4 classes :

- Classe 3 : grammaires/langages régulier (on a besoin d'un automate deterministe pour répondre au pb).
- Classe 2 : contexte dépendant (pour résoudre le pb, il faut des automate non déterministes). Comprend les langages informatiques.
- Classe 1 : contexte-free (pb compliqué à résoudre). Comprend les langages naturels.
- Classe 0 : récursivement énumérables (insolvable)

9.3 Conclusion

Ce sue l'on a vu dans le cas de sources simple (sans mémoire) fonctionne, de manière beaucoup plus complexe, avec des sources générales.

10 Projet

Propriété AOP :
Faire un petit rapport, par 2 max, rendre dans un mois (08/09/2015)