apprentissage automatique préservant la confidentialité basé sur le chiffrement homomorphique

Serigne Saliou Gueye

March 28, 2021

R E M E R C I E M E N T S

Acknowledgement goes here

RESUMÉ

INTODUCTION GÉNÉRALE

ABSTRACT

CONTENTS

PART I

CONTEXTE ET INTRODUCTION

CHAPTER 1

NOTION PRÉLIMINAIRES DE LA CRYPTOGRAPHIE

Avant de passer à la définition du chiffrement homomorphe, nous allons revenir sur quelques notions cryptographiques à savoir les différents systèmes cryptographique (systèmes à clés secrètes et à clés privés) et aussi parler de la sécurité prouvée en cryptographie. Pour des traitements beaucoup plus approfondis de ces concepts nous renvoyons le lecteur sur [références](Jonathan Katz and Yehuda Lindell. Introduction to Modern Cryptography (Chapman Hall/Crc Cryptography and Network Security Series). Chapman Hall/CRC, 2007.)

1.1 Introduction

La cryptographie est une discipline des maths incluant des principes et méthodes permettant de garantir les services de sécurité de l'information .Sa principale mission est de garantir la sécurité des communications c'est-à-dire de permettre à des entités qui ne se font pas confiance en général de communiquer en toute sécurité en présence de potentiels adversaires (susceptibles entre autres d'accéder à des secrets en violant la confidentialité, d'intercepter et de modifier les informations échangées ou d'usurper des identités lors d'une communication).
La cryptographie ne prend en charge que les 4 premiers sur les 5 services de

sécurité fondamentaux que sont la Confidentialité, l'intégrité, l'authentification, la Non Répudiation et la Disponibilité.Elle est composée des systèmes à clé secrète et des systèmes à clés publiques.
Dans un système à clés secrètes les entités se partagent la même clé pour le chiffrement et le déchiffrement (on parle de cryptographie symétrique)

Dans le cas du système à clé publique, on utilise deux clés dont l'une soit k' est difficile à déduire de l'autre soit k (on parle aussi de cryptographie asymétrique). La clé k est appelée clé publique et est utilisée pour le chiffrement ou la vérification de signature selon le système ; la clé k' est appelée clé privée et est utilisée pour le déchiffrement ou la signature selon le système.

Les algorithmes de chifffrements et les protocoles cryptographique sont impliqués dans une variété d'applications critiques: : les banques en ligne, le vote électronique, la vente aux enchères électroniques... En tant que tels, leurs securités doivent donc être formellement vérifiées et prouvées

1.2 Sécurité inconditionelle

1.2.1 Notion de sécurité parfaite

Un cryptosytéme est parfois définis par trois algorithmes :Algorithme de génération des clés ,de chiffrement et de déchiffrement ainsi qu'une specification d'un espace de message M avec #M > 1
L'algorithme de génération des clés que nous notons GEN est un algorithme probabiliste qui produit une clé k choisit selon une certaine distribution. Nous désignons par K l'espace des clés c'est à dire l'ensemble de toutes les clés possibles qui peuvent être sortis par GEN
L'algorithme de chiffrement que nous allons noté par ENC prend en entré une clé $\kappa \in$ K et un messsage m $\in$ M produit un chiffré c .Nous considérons que l'algorithme de chiffrement est probabiliste donc ENC(k,m)peut alors générer des chiffrés différents lorsque le même message est utilisé.On note c $\mapsto ENC_{\kappa}(m)$.On note C l'ensemble de tous les chiffrés possibles
L'algorithme de déchiffrement noté par DEC prend en entré une clé κ et un chiffré c et produit le message initial m. Contrairement à ENC l'algorithme de déchiffrement est déterministe puisque $\mathrm{DEC}_{\kappa}(c)$donne la même sortie à chaque exécution on notera donc m:=$\mathrm{DEC}_{\kappa}(c)$pour designer le caractére deterministe

Il est évident que si l'attaquant détient la clé alors il pourra déchiffrer tous les messages échangés par les entités. Alors qu'en est-il de ENC ? N'est-il pas mieux de garder tous les deux (l'algorithme et la clé) secret?
Auguste Kerckhoffs a soutenu le contraire à la fin du 19e siècle lors de l'élucidation de plusieurs principes de conception de chiffrements militaires. L'un des plus important d'entre eux, désormais simplement connu sous le nom de principe de Kerckhoff.

1.2.2 principe de Kerckhoff

La méthode de chiffrement ne doit pas être tenue secrète, et elle doit pouvoir tomber entre les mains de l'attaquant sans inconvénient.
Pour pouvoir bien comprendre ce principe et voir l'importance de garder la clé secrète nous pouvons considérer l'exemple suivant
Soit m un message $\in \{0,1\}^n$. On definit notre algorithme de chiffrement (ENC) en faisant la somme directe entre m et $\kappa \in$ K c = m $\oplus$ κ .On peut voir dans cet algorithme de chiffrement que si κ est fixé alors ce cryptosystéme ne sera pas sur et l'attaquant obtenir le massage initial à partir du chiffré en faisant m = c $\oplus$ κ.
Par contre si κ est choisit de façon uniformément aléatoire dans l'espace des clés $\{0,1\}^n$ et gardé secret pour l'attaquant alors le résultat du chiffré sera uniformément distribué dans l'espace des chiffrés $\{0,1\}^n$ et la sécurité sera atteinte

1.2.3 Principe de Shannon

Un chiffré doit apporter de la confusion et de la diffusion, c'est-à-dire :

Confusion
Il n'y a pas de relation algébrique simple entre le clair et le chiffré. En particulier, connaître un certain nombre de couples clair-chiffré ne permet pas d'interpoler la fonction de chiffrement pour les autres messages.
Diffusion
La modification d'une lettre du clair doit modifier l'ensemble du chiffré. On ne peut pas casser le chiffre morceau par morceau.

1.2.4 Théoréme du secret parfait

Avant d'énoncer ce théoréme,nous allons considérer les deux exemples ci-dessous

exemple 1

Considérons l'exemple simple du chiffrement de césar suivant
Soit $\kappa \in \{0,.....25\}$ avec $\Pr[K=\kappa]$ la probabilité que la clé soit k (on considère que les clés sont équiprobable) donc $\Pr[K=\kappa]=\frac{1}{26}$. Supposons que nous avons la distribution suivante sur M:$\Pr[M=a]=0{,}7$ et $\Pr[M=z]=0{,}3$. ($\Pr[M=a]$ probabilité que le message soit a et $\Pr[M=z]$ probabilité que le message soit z).Ainsi on cherche alors la probabilité que le message chiffré soit B
On peut voir qu'il n'y a deux possibilités :soit M=a et K=1 ,soit m=z et k= 2 .Par indépendance de M et K nous avons $\Pr[M=a\wedge K=1]$ =$\Pr[M=a]$*$\Pr[K=1]$,de même $\Pr[M=z\wedge K=2]$ =$\Pr[M=z]$*$\Pr[K=2]$.
Par conséquent,
$\Pr[C=B]=\Pr[M=a\wedge K=1]+\Pr[M=z\wedge K=2]=0{,}7\ \frac{1}{26}+0{,}3\ \frac{1}{26}=\frac{1}{26}$ donc
$\Pr[C=B]=\frac{1}{26}$
Nous pouvons également calculer la probabilité conditionnelle par exemple calculer la probabilité que le chiffré B soit issu de a ,$\Pr[a \mid B]$.En utilisant le théoréme de Bayes nous obtenons :

$$Pr[M=a \mid C=B]=\frac{Pr[C=B \mid M=a].Pr[M=a]}{Pr[C=B]}$$

Notons que $Pr[C=B \mid M=a]=\frac{1}{26}$, puisque si M = a alors la seule voie C = B peut se produire si κ = 1 ce qui se produit avec une probabilité de $\frac{1}{26}$

$$\text{Donc } Pr[M=a \mid C=B]=\frac{Pr[C=B \mid M=a].Pr[M=a]}{Pr[C=B]}=\frac{\frac{1}{26}.O,7}{\frac{1}{26}}$$

$$Pr[M=a \mid C=B]=0{,}7$$

Conclusion
on a $\Pr[M=a]=Pr[M=a \mid C=B]$ .ce résultat montre que la probabilité d'avoir une information claire ne varie pas même si on connait son chiffré