Chapitre 5
La cryptographie et la Blockchain

1. Introduction

Cette discipline mathématique est au centre du processus de validation et de vérification des informations échangées au sein d'un réseau, c'est cela qui permet d'assurer la confiance. Dans les sections qui vont suivre, nous allons décrire les concepts mathématiques sous-jacents et leur utilisation pour sécuriser et authentifier l'information.

Historiquement, la cryptographie est associée à des problématiques permettant de réaliser et d'analyser des schémas de cryptage (ce qui autorise l'échange d'informations secrètes à travers des canaux de communication non sécurisés). Toutefois, à partir des années 1970, la construction de signatures électroniques inviolables et le développement de protocoles de tolérance d'erreurs ont été inclus dans la discipline de la cryptographie. Dans ce cas, la cryptographie concerne tout système ayant besoin de faire face à des abus ou à des attaques malveillantes. Notre définition de la cryptographie rend essentielle et nécessaire l'utilisation de certains outils mathématiques, tels que les fonctions à sens unique, les générateurs pseudo-aléatoires et la preuve à divulgation nulle de connaissance, qui seront traités dans ce chapitre.

Il est important de souligner que la cryptographie est basée sur une hypothèse très importante qui est l'existence des fonctions à sens unique. Ces fonctions capturent les difficultés de calcul qui sont inhérentes à la cryptographie et qui capitalisent sur ces limitations de calcul. Sans limitations de calcul, cette approche de la cryptographie devient inopérante. Les difficultés générées par ces problèmes constituent une opportunité pour la cryptographie, mais une difficulté supplémentaire pour les algorithmes permettant de générer les fonctions à sens unique.

2. Les schémas de cryptage

L'échange d'informations secrètes sur des canaux de communication publics représente le plus basique et traditionnel problème de cryptographie. Cela consiste pour deux parties à communiquer de manière secrète sans qu'une tierce partie exploite ces informations. Typiquement, un schéma de cryptage est constitué d'une paire d'algorithmes. Le premier algorithme est l'algorithme de cryptage appliqué par l'expéditeur du message et le second est l'algorithme de décryptage utilisé par le destinataire afin de lire le message d'origine. Dans cette configuration, seul le texte chiffré transite dans le canal de communication. Pour garantir le caractère secret de ce type d'échange, les deux parties doivent partager une information qu'elles seules détiennent. Cette information supplémentaire ou « extra knowledge » peut prendre la forme d'un algorithme de décryptage ou d'un paramètre auxiliaire utilisé par l'algorithme de cryptage. Cette information supplémentaire est appelée la clé de décryptage. Nous pouvons noter que l'algorithme de décryptage peut être connu de tout le monde, car il est nécessaire d'avoir en sa possession la clé de décryptage afin de déchiffrer le contenu. Il est aussi important de souligner que la clé de décryptage doit toujours être gardée secrète.

L'évaluation du degré de sécurité de ce genre de schéma est une tâche qui n'est pas facile, et elle s'avère dans certains cas très compliquée ou même impossible. Il est important de définir la notion de sécurité et son périmètre. Deux approches sont possibles.

– La première dérive de la théorie de l'information, elle concerne les informations sur le texte source contenues dans le texte chiffré. Si de l'information sur le texte source est contenue dans le texte crypté, alors le schéma de cryptage est considéré comme non sécurisé. Il a été aussi démontré que, pour avoir un niveau de sécurité élevé, il fallait une clé de cryptage au moins aussi longue que le texte source. Cette condition drastique représente une limitation pour ce schéma, en particulier quand la quantité d'informations à crypter est élevée.
– La seconde approche est plus moderne et est basée sur le degré de complexité du calcul (théorie de la complexité). Dans ce cas, il n'est pas important de savoir si le texte chiffré contient ou non des informations concernant le texte source. La seule question à se poser est liée à la faisabilité d'extraire de l'information de ce texte. En d'autres termes, nous nous intéressons au degré de complexité pour déchiffrer les informations cryptées. Il s'avère que dans cette situation la longueur de la clé de cryptage n'est pas un élément-« clé » dans la sécurité du chiffrement. Par exemple, nous pourrions utiliser un générateur pseudo-aléatoire de clés de cryptage aussi longues que possible à partir de clés plus courtes.

La théorie de la complexité nous amène à introduire certains concepts tels que le schéma de cryptage par clé publique. Avant de décrire ce procédé, il faut souligner qu'il est possible d'adjoindre à un algorithme de cryptage, un paramètre auxiliaire qui est la clé de cryptage. Ce qui signifie que, pour chiffrer un message, nous devons aussi fournir la clé de cryptage à l'algorithme de cryptage. La plupart des schémas de chiffrement utilisent des clés de cryptage et de décryptage égales. Ceci soulève le problème de la distribution de clés. C'est pour cela qu'un nouveau schéma de cryptage basé sur la théorie de la complexité a été proposé, où la clé de cryptage (clé publique) peut être connue de tous et est différente de la clé de décryptage qui, elle, est gardée secrète. Il est impossible de remonter à la clé de décryptage à partir de la clé de cryptage. Ce schéma résout le problème de distribution de clés et permet de rendre publiques les clés de cryptage (identifiants).

Ce genre de schéma de cryptage est utilisé dans les systèmes distribués comme la blockchain, afin de contrôler les accès et la validité des informations circulant au sein du réseau.

3. Les générateurs pseudo-aléatoires

Les générateurs pseudo-aléatoires jouent un rôle central dans la construction des schémas de cryptage. Ils permettent en particulier de générer des clés de cryptage privées. Cependant, il est à noter que le terme « générateur pseudo-aléatoire » est également utilisé dans d'autres contextes, comme les procédures probabilistes. Par conséquent, il est primordial de donner une définition claire et précise du fait de leur importance en cryptographie.

Les générateurs pseudo-aléatoires sont des algorithmes déterministes qui permettent d'allonger une séquence de nombres aléatoires "graine", afin d'obtenir une séquence de bits plus longue et qui paraît aléatoire bien qu'elle ne le soit pas. Le pseudo-aléatoire et la théorie de la complexité sont liés de manière fondamentale, car ces générateurs sont construits sur l'hypothèse d'insolubilité (*intractability assumption*). L'hypothèse de l'existence du générateur pseudo-aléatoire est liée à l'existence de fonctions à sens unique, car il est construit à partir de fonctions à sens unique particulières.

4. Les fondamentaux de la théorie des probabilités

Dans cette section, nous allons définir les fonctions à sens unique en mettant l'accent sur leurs descriptions mathématiques. Nous proposerons également des exemples de ces fonctions et de leurs applications.

Avant de présenter ces fonctions à sens unique, nous allons faire quelques rappels sur les calculs de probabilités. Ces fondamentaux sont le socle des fonctions à sens unique et de la théorie de la complexité.

Les probabilités jouent un rôle déterminant en cryptographie. Elles sont en particulier essentielles pour traiter les informations ou le manque d'informations (le secret). Dans cette partie, nous abordons certains concepts et inégalités pertinents en cryptographie.

Dans ce qui va suivre, nous nous référerons uniquement aux distributions de probabilité discrètes. Dans notre cas, l'espace de probabilité consiste en une séquence de caractères d'une certaine longueur l et qui sont distribués de manière uniforme. C'est un échantillon d'une longueur totale de caractères de l bits, et chaque caractère est assigné d'une probabilité de 2^{-l}.

Les variables aléatoires sont des fonctions qui assignent une valeur de l'espace de l'échantillon vers l'espace réel. La variable aléatoire représente l'ensemble des résultats définis sur l'ensemble des éventualités.

Exemple

Soit X une variable aléatoire et $B(.,.)$ une expression booléenne qui dépend de la variable X, alors la probabilité $Pr[B(X,X)]$ signifie que $B(X,X)$ est vérifiée quand x est choisi avec la probabilité $Pr[X = x]$.

En d'autres termes :

$$Pr[B(X,X)] = \sum_x Pr[X = x] \cdot \chi(B(x,x))$$

où χ est la fonction indicatrice, ce qui implique que $\chi(B) = 1$ si l'événement B est réalisé, et 0 dans le cas contraire. Et pour toutes variables aléatoires X, nous avons : $Pr[X = X] = 1$.

Ce résultat peut se généraliser à deux variables aléatoires X, Y indépendantes et identiquement distribuées. Nous avons alors $Pr[B(X,Y)]$, la probabilité que $B(x,y)$ est réalisé pour la paire (x,y) choisie avec une probabilité $Pr[X = x].Pr[Y = y]$:

$$Pr[B(X,Y)] = \sum_{xy} Pr[X = x] \cdot Pr[Y = y].\chi(B(x,y))$$

À titre d'exemple, nous avons $Pr[X = Y] = 1$ uniquement si X et Y sont triviales (assigne toute la masse de probabilité à un seul caractère).

Une variable aléatoire très utile dans le contexte de la cryptographie est la variable aléatoire uniformément distribuée U_n, qui désigne une distribution d'une séquence de caractères sur une longueur n. La probabilité $Pr[U_n = \alpha] = 2^{-n}$ si $\alpha \in \{1;0\}^n$, et 0 sinon.

Trois inégalités fondamentales en théorie des probabilités vont être très utiles pour la suite, car elles permettent de donner les limites supérieures de variables aléatoires. Il s'agit de l'inégalité de Markov, de l'inégalité de Tchebychev et de l'inégalité de Hoeffding (Stein, 2005).

L'inégalité de Markov

Cette inégalité montre que pour une variable aléatoire avec une limite inférieure ou supérieure, il existe une relation entre la déviation de cette valeur de l'espérance de la variable aléatoire et la probabilité que cette variable aléatoire soit supérieure ou inférieure à cette valeur.

Soit X une variable aléatoire positive et v un nombre réel. Alors :

$$Pr[X \geq v] \leq \frac{E(X)}{v}$$

De manière équivalente : $Pr[X \geq r \,.\, E(X)\,] \leq \frac{1}{r}$.

La démonstration peut se faire de la manière suivante :

$$E(X) = \sum_x \Pr[X = x] \,.\, x$$

$$\geq \sum_{x<v} \Pr[X = x] \,.\, 0 + \sum_{x \geq v} \Pr[X = x] \,.\, v$$

$$= Pr[X \geq v] \,.\, v$$

L'inégalité de Markov est très utile dans les cas où nous ne connaissons que peu de paramètres sur la distribution de la variable aléatoire. Il suffit alors de connaître son espérance et au moins une des bornes ou limites de l'intervalle de définition de ses valeurs.