IMPOSSIBLE HASARD PARFAIT

1 **Les méthodes algorithmiques de production de suites aléatoires ne donnent pas des suites aléatoires au sens fort défini par Per Martin-Löf en 1965. En effet, « être produit par un algorithme » entraîne que celui qui détient l'algorithme peut anticiper ce qui sera produit. De plus, ce qui est produit par un algorithme peut être compressé en un fichier de taille au plus égale à la taille du programme de l'algorithme ; or ce qui est authentiquement aléatoire est incompressible. Les décimales de π, en particulier, sont produites par des algorithmes tirés de séries numériques classiques et ne constituent donc pas une suite aléatoire au sens fort.**

Les méthodes physiques (loterie, bruit thermique, mécanique quantique, etc.) ne produisent des suites aléatoires au sens fort que (1) sous l'hypothèse que les théories associées (par exemple la mécanique quantique) sont les bonnes théories qui régissent la physique du mécanisme producteur, ce qui est impossible à prouver, et (2) si le dispositif physique producteur a été correctement monté et ne subit aucune manipulation, et en particulier n'est victime d'aucune porte dérobée. Il se peut donc que les méthodes quantiques produisent de véritables suites aléatoires, mais l'affirmer est impossible car cela reviendrait à affirmer que la mécanique quantique est une théorie qui ne sera jamais remise en cause, et que l'on a bien modélisé le dispositif physique utilisé, deux affirmations risquées, comme le montre l'histoire de la physique.

Il en va de même plus généralement pour tous les sites internet qui proposent des jeux où le hasard intervient, et cela concerne aussi les loteries nationales, lotos et autres jeux d'argent organisés à grande échelle. Tous devraient proposer des procédures de tirage non plus fondées sur des dispositifs physiques locaux que certains ont su manipuler, mais sur des balises aléatoires sécurisées dont l'utilisation serait rendue parfaitement contrôlable *a posteriori* grâce aux données enregistrées.

L'utilisation de balises aléatoires concerne d'autres domaines et devrait y être obligatoire. Si une entreprise qui produit des médicaments assure qu'elle teste au hasard les lots mis en circulation, il est souhaitable que le système de tirage au sort qu'elle utilise pour ces contrôles soit fondé sur un tirage aléatoire extérieur, car c'est la seule façon d'être certain que l'entreprise ne choisisse pas les cobayes à sa convenance.

Lorsqu'un tribunal désigne au hasard des jurés ou qu'une procédure politique choisit aléatoirement des représentants, il serait souhaitable, pour garantir l'absence de biais et de manipulations, que la procédure elle aussi se fonde sur une balise aléatoire indépendante publique et autorisant des contrôles *a posteriori*.

D'autres exemples concernent les applications décentralisées, par exemple liées aux blockchains et aux monnaies cryptographiques. Ils confirment l'importance de bonnes balises aléatoires et, depuis quelques années, le sujet préoccupe les informaticiens et les cryptologues qui ont, nous le verrons, élaboré des idées subtiles afin que personne ne triche.

Explorons pas à pas le problème de la création de balises aléatoires sûres.

LA MÉTÉO, UNE BALISE ALÉATOIRE ?

Notre premier exemple de balise élémentaire donnant un aléa public pour le pile ou face par téléphone s'appuie sur la météo. La balise serait déduite des publications sur internet de Météo-France. Elle produirait un bit par jour en se fondant sur la température à midi relevée par Météo-France à la station météorologique de Paris-Montsouris. Cette température serait arrondie au dixième de degré inférieur et un 0 (correspondant à *pile*) serait produit si le premier chiffre après la virgule de la température est pair et un 1 (correspondant à *face*) serait produit si le premier chiffre est impair.

Un site internet donne cette température « officielle » et peut donc, sans intermédiaire, servir de balise aléatoire : www.infoclimat.fr/observations-meteo/temps-reel/paris-montsouris/07156.html. La température à midi y est chaque jour définitivement inscrite sur le site et reste ensuite accessible. On trouve par exemple les données suivantes : 11,8 °C pour le 28 novembre 2019 à midi, qui donnerait un 0, donc *pile* ; 9,5 °C pour le 29 novembre 2019 à midi, qui donnerait un 1, donc *face* ; 7,8 °C pour >

> le 30 novembre 2019 à midi, qui donnerait un 0, donc *pile*; 3,5 °C pour le 1er décembre 2019 à midi, qui donnerait un 1, donc *face*; 5,2 °C pour le 2 décembre 2019 à midi, qui donnerait un 0, donc *pile*. La suite des tirages est donc, à partir du 28 novembre: *pile-face-pile-face-pile*. On peut opérer un tirage à pile ou face par téléphone tous les jours à midi.

Cette balise est très probablement aléatoire: il n'y a aucune raison de penser que les nombres pairs (ou impairs) soient favorisés, ni qu'aucune séquence particulière de *pile* ou *face* soit plus fréquente qu'une autre de même longueur. Elle est bien imprévisible, car les prévisions météorologiques pour Paris ne sont pas assez précises pour qu'on puisse savoir à l'avance ce qui va venir comme premier chiffre après la virgule de la température à midi. La balise est vérifiable *a posteriori*: tout le monde peut aller consulter le site mentionné pour voir ce qui y est indiqué. On peut raisonnablement faire confiance au sérieux des opérateurs du site pour qu'ils ne changent pas les données inscrites et qu'ils les maintiennent accessibles assez longtemps.

Trois problèmes persistent cependant avec cet aléa public.

(1) Prévoir le résultat du tirage devient possible quand on s'approche de midi, car la température varie continûment. Il faudrait donc convenir au moins quelques heures avant midi que l'on va jouer à pile ou face et attendre midi pour que s'effectue le tirage.

(2) Le procédé n'est pas rapide. En considérant d'autres lieux de relevé, d'autres heures, d'autres données (pression, hygrométrie, etc.) on pourrait faire mieux, mais on resterait limité à quelques bits par minute au plus.

(3) Peut-on faire totalement confiance à ce qui est publié par le site de Météo-France? Est-ce qu'il ne se pourrait pas qu'un employé, par négligence, erreur de manipulation ou soudoyé par l'un des joueurs, modifie le relevé publié pour produire des températures autres que celles vraiment relevées? L'un des joueurs ne pourrait-il pas s'introduire sur le site et en modifier les chiffres publiés? Nous verrons par la suite que nous pouvons pallier cette incertitude sur la stabilité et la sécurité de ce qui est inscrit.

Les cours de bourses, les taux de change, les résultats sportifs, le contenu de ce qui est publié par les médias (journaux, radio, télévisions, pages web), sont d'autres sources publiques d'aléas et vérifiables *a posteriori*. Certaines sont assez rapides, mais aucune ne donne une assurance de stabilité supérieure aux données publiées par Météo-France. Avec elles, on aura donc en définitive un aléa public ayant les mêmes imperfections qu'avec Météo-France.

AUTRES SOURCES PARTAGEABLES D'ALÉAS

Conscients de l'intérêt de disposer de meilleures sources publiques d'aléas, certains sites internet en proposent. Ainsi, le site random.org rend disponible des nombres aléatoires de toutes sortes depuis 1998.

Ce site, conçu et réalisé par Mads Haahr, du Trinity College, à Dublin, propose des nombres aléatoires qu'il qualifie de «véritables», car

2 LES FONCTIONS DE HACHAGE

Une fonction de hachage est une fonction qui, à un x donné (une suite de caractères ou un nombre) de longueur quelconque, associe une valeur $f(x)$ de taille fixe, l'« empreinte de x ». Le passage de x à $f(x)$ se fait par une série d'opérations qui commencent en découpant x en morceaux (d'où le nom de hachage). Ces morceaux sont rassemblés et emmêlés de façon à obtenir un brouillage aussi parfait que possible de l'information présente dans x. Pour la fonction classique de hachage SHA-256, la longueur de $f(x)$ est de 256 bits.

Les fonctions de hachage à usage cryptographique, comme SHA-256, ont les propriétés suivantes.
(a) Elles sont « à sens unique » : le passage de x à $f(x)$ est un calcul rapide, mais pour un y donné, trouver x tel que $y = f(x)$ est impossible en pratique.
(b) Il est impossible en pratique de trouver x et x' différents avec $f(x) = f(x')$.
(c) Elles produisent des valeurs assimilables à des valeurs aléatoires uniformes et, en particulier, même si x et x' sont deux suites de caractères n'ayant qu'un seul caractère différent, les valeurs $f(x)$ et $f(x')$ sont sans aucun lien apparent.

Exemple simple. On transforme la donnée x en une suite de 0 et de 1 que l'on découpe en morceaux de longueur 8. Les morceaux sont ensuite additionnés comme des nombres en base 2. On ne garde du résultat que les quatre chiffres centraux :
x = 00010010100101010100100101
→ 00010010 + 10010101 + 00100101 = 11001100
→ $f(x)$ = 0011.
Cet exemple est bien sûr trop simple pour avoir les propriétés (a), (b) et (c).

Exemples d'utilisation.
– Authentifier un utilisateur sans garder son mot de passe *(voir l'illustration ci-contre)*.
– Assurer l'intégrité d'un fichier x. Si l'on dépose un gros fichier sur un serveur, on lui associera son empreinte $f(x)$, beaucoup plus courte. Celui qui téléchargera x s'assurera que x lui parvient sans erreur et en totalité en calculant l'empreinte du fichier qu'il aura reçu, laquelle doit être la même que celle publiée sur le site.
– Engagement d'une donnée sans la révéler. Vous choisissez un nombre que vous ne devez plus changer car, par exemple, vous allez participer à une enchère simultanée ; vous calculez son empreinte et la communiquez à l'organisateur. Le site d'enchère recueille les empreintes des autres participants. Les empreintes sont toutes rendues publiques et ceux qui les ont proposées ne peuvent donc plus rien changer. À cet instant, personne ne connaît les enchères proposées (car f est impossible à inverser), et donc, même si les soumissions ne sont pas simultanées, ceux qui participent ignorent les propositions des autres. La phase d'enchère terminée, chacun révèle son nombre. Le calcul des empreintes des nombres révélés permet de vérifier qu'il n'y a pas eu de changements. Tout s'est donc déroulé comme si chacun avait écrit sa proposition sur un papier et avait placé celui-ci dans une enveloppe cachetée qui n'a été ouverte qu'au dernier moment.

En cryptographie, les fonctions de hachage sont conçues avec soin et soumises à des tests variés. Des compétitions pour concevoir les nouvelles générations de fonctions de hachage sont organisées. Ainsi est née la fonction SHA-3, disponible depuis 2015 (voir https://en.wikipedia.org/wiki/SHA-3).

tirés du bruit atmosphérique capté par radio. On peut par exemple obtenir, à la demande, le résultat d'un lancer de deux dés, ou le mélange d'un paquet de 52 cartes, ou même une image aléatoire faite de pixels noirs et blancs.

Random.org affirme extraire 12000 bits par seconde du bruit atmosphérique, et c'est à partir d'eux que les autres services sont fournis. Si on le souhaite, les nombres que le site communique sont mémorisés, mais ce n'est pas très facile car le site n'est pas conçu pour jouer le rôle de balise aléatoire sécurisée émettant périodiquement. Par ailleurs, on ne sait pas vraiment comment le bruit atmosphérique est capté et transformé en données aléatoires. Utiliser random.org pour les applications n'exigeant pas une certitude d'honnêteté et de sécurité totale sera une assez bonne solution, mais, tel qu'il est conçu, ce site internet n'est pas en mesure de fournir assez de garanties pour des applications où par exemple des sommes importantes d'argent seraient en jeu.

LA BALISE DU NIST: UNE LISTE DE 512 BITS CHAQUE MINUTE

C'est sans doute pourquoi l'institut américain des normes et de la technologie, le NIST (National Institute of Standards and Technology), travaille sur ce problème depuis 2011 et propose une balise aléatoire, conçue cette fois comme un émetteur public et périodique d'aléas dont les données produites sont protégées et restent consultables sans limitation de temps. La balise du NIST publie chaque minute une liste de 512 bits aléatoires (*voir* https://beacon.nist.gov/home). Et le NIST encourage la création d'autres balises indépendantes.

Le NIST utilise les bits émis par un dispositif quantique. Il fournit un moyen de vérifier que les bits publiés ne sont pas modifiés une fois publiés et affirme que son site est protégé contre toute intervention extérieure, ce qui assure que personne ne manipule le mécanisme de production ou de publication.

Si l'on considère que le NIST est un tiers de confiance parfait, on peut utiliser ce qu'émet sa balise. Par exemple, si des sites de jeux qui utilisent des nombres aléatoires vont les y rechercher en indiquant comment, on pourra avoir confiance en eux et vérifier qu'il n'y a pas tricherie, car il est improbable que le NIST se laisse corrompre par eux. Malgré tout, comme le notent de nombreux spécialistes, le NIST ne donne aucun moyen pour contrôler que les bits qu'il publie sont produits par des phénomènes quantiques et qu'aucune manipulation n'est opérée par l'institut sur ce qu'il émet. Le NIST ayant déjà été soupçonné de tentatives de trucage liées à des outils de génération d'aléas (*voir* https://en.wikipedia.org/wiki/Dual_EC_DRBG), il est impossible de recommander d'utiliser sa balise pour les applications exigeant une sécurité absolue.

Remarquons que, sur sa page, le NIST indique en rouge qu'il ne faut pas utiliser sa balise pour choisir des mots de passe. La recommandation est tout à fait justifiée et s'applique à tous les générateurs publics d'aléas. En effet, les données publiées restent accessibles (c'est nécessaire pour mener des contrôles *a posteriori*), ce qui signifie que si votre mot de passe a été engendré par une balise publique, quelqu'un cherchant à le deviner a accès aux mêmes nombres aléatoires, ce qui facilite son travail. Pour vos mots de passe, utilisez des générateurs privés et si possible indépendants de votre machine!

D'autres sites se présentent comme des balises aléatoires sécurisées et réussissent à faire mieux que la balise du NIST.

Le Laboratoire de sécurité computationnelle et de cryptographie appliquée de l'université du Chili (https://random.uchile.cl/en/) propose une balise dont les bits dépendent à la fois d'une source quantique locale et de données extérieures (données sismiques, radio de l'université du Chili, blockchain Ethereum, Twitter). Ce site est l'une des meilleures balises actuelles: il utilise le principe décrit plus loin et dans l'encadré 3.

Le site brésilien d'Inmetro, l'Institut national de la métrologie et de la qualité (www4.inmetro.gov.br/) et le projet collaboratif *League of entropy* (www.cloudflare.com/leagueofentropy/) offrent divers outils de génération de hasard privé et public. Là encore, ils s'appuient >

AUTHENTIFICATION SANS CONSERVATION DES MOTS DE PASSE

Première connexion : enregistrement de l'identifiant de l'utilisateur et de l'empreinte du mot de passe

Jean-Paul choisit le mot de passe JP1234 et le communique au serveur

Le serveur calcule l'empreinte du mot de passe reçu, trouve X2e9qq92 et mémorise le couple Jean-Paul/X2e9qq92.

Connexion ultérieure : vérification par le serveur que celui qui se connecte est la bonne personne

Jean-Paul se connecte au serveur et indique son identifiant ainsi que son mot de passe

Le serveur calcule l'empreinte du mot de passe reçu et vérifie qu'elle est bien la même que celle du couple qu'il a enregistré Jean-Paul/X2e9qq92