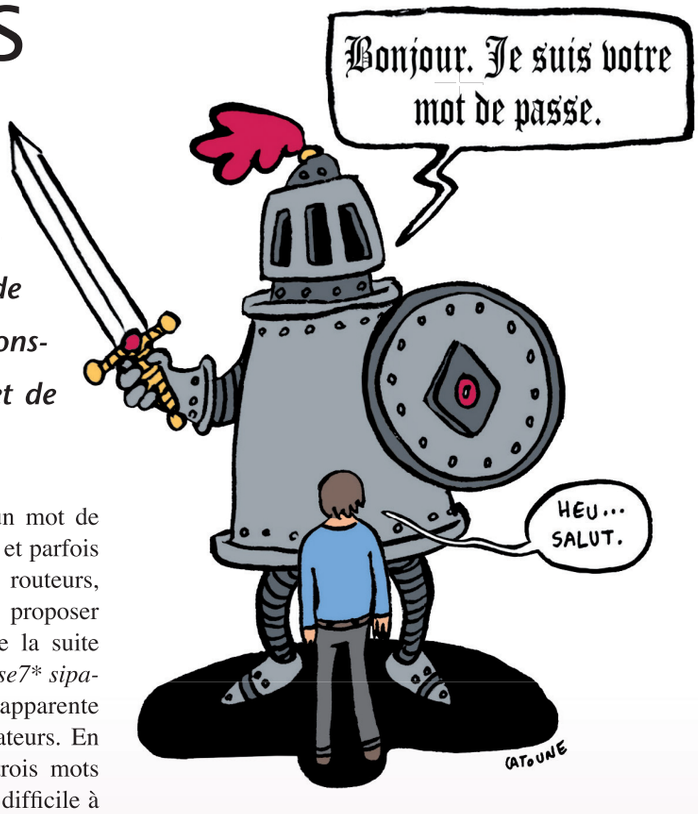
par Nicolas K. Blanchard

Des mots de passe... pas très secrets

De Dupont00 à 2r67W@SBZX2!r9, nous utilisons tous des mots de passe, et leur complexité apparente est le plus souvent trompeuse. Si nos mots de passe actuels sont si peu efficaces, comment pouvons-nous donc garantir la sécurité de nos données et de notre vie privée ?



Qui n'est pas déjà tombé sur un mot de passe impossible à mémoriser, et parfois impossible à changer ? Les routeurs, notamment les Freebox, adorent en proposer des particulièrement délicieux, comme la suite suivante (authentique) : *libitu& domuisse7* siparetur?!*. Pourtant, cette complexité apparente ne sert souvent qu'à frustrer les utilisateurs. En effet, voici une petite énigme : des trois mots de passe qui suivent, lequel est le plus difficile à « cracker » ?

1R&ducteur1998
milieu-jeu-calebassier-tatar-palais
FYkLh39pdR

Rendons à Shannon...

Il est en fait très difficile de répondre en toute généralité, car cela dépend des informations accessibles à une hackeuse potentielle. L'habitude est donc de considérer des *modèles d'attaque*, et le plus courant vient de Claude Shannon, père de la théorie de l'information, qui énonça le principe suivant : *« L'ennemi connaît le système. »* Quels sont donc les systèmes en question ?
Le premier correspond à la majorité des mots de passe utilisés aujourd'hui : un mot aléatoire, auquel on apporte quelques modifications (ajout d'une majuscule, d'un caractère spécial ou d'une date, presque toujours à la fin).
Le deuxième est simplement une suite de cinq mots pris au hasard dans un dictionnaire.
Le troisième système est une suite de dix chiffres et lettres générés par une machine.

Pour comparer ces systèmes très différents, un outil s'impose : l'entropie, qu'il s'agit donc de calculer indépendamment pour chacun (voir encadré). On peut alors se rendre compte que la première méthode (correspondant au premier mot de passe proposé) est la plus vulnérable, prenant à peine une minute à « cracker » sur une machine quelconque, alors que la troisième prend près de deux mois. Enfin, la deuxième méthode nécessiterait un million de machines, ou bien plusieurs jours d'accès au supercalculateur français le plus performant.

Malgré la supériorité apparente de cette solution, le choix n'est pas si simple. Déjà, un mot de passe ne doit pas simplement être dur à deviner, il doit aussi répondre à des exigences d'usage : personne ne veut passer une demi-heure à écrire son mot de passe avant d'accéder à ses courriers électroniques.
En outre, un mot de passe est avant tout de l'information à l'état brut. Contrairement à des caractéristiques personnelles difficiles à falsifier (base de l'identité biométrique), il a l'avantage d'être

échangeable. Il est fréquent que l'on transfère ainsi une partie de son pouvoir d'authentification à un tiers contre une aide (ainsi, un paraplégique peut utiliser sa carte de crédit en demandant à un proche de taper le code). Si la plupart des mots de passe sont faits pour être faciles à mémoriser, certains peuvent être construits avec l'objectif inverse, pour éviter qu'un tiers puisse nous nuire en se souvenant trop longtemps de nos secrets. On a affaire à un problème d'optimisation sophistiqué, avec plusieurs contraintes parfois contradictoires ; une solution unique ne saurait donc suffire.

Plusieurs poids, plusieurs mesures

Une étude récente se pencha sur une expérience de vote en ligne, où les utilisateurs avaient un fort taux d'abstention. La cause majeure ? Des mots de passe alphanumériques difficiles à saisir qui donnaient lieu à des erreurs (comme mélanger « g » et « q »). Ainsi, les structures alphanumériques complexes sont loin d'être optimales.

Les mots de passe prononçables (formés de syllabes) sont non seulement faciles à taper, ils sont aussi beaucoup plus rapides, compensant ainsi leur entropie plus faible par caractère. Pourquoi ne pas directement utiliser des mots entiers ? Parce que, dans un contexte de vote électronique, avoir un taux d'erreur (et donc d'abstention) différent causé par une maîtrise inégale de la langue serait discriminatoire…

La piste des méthodes biométriques semble salvatrice, mais on déchante vite : jusqu'à présent, toute méthode d'authentification biométrique a été piratée en moins d'un an. Et une fois piratée, les conséquences peuvent vous suivre toute une vie – une alternative prometteuse de sécurité biométrique décentralisée évitant ce risque vient cependant d'apparaître, savamment nommée *protocole horcruxe*.

Enfin, si la police dans certains pays peut déverrouiller votre téléphone malgré vous grâce à vos empreintes digitales, elle ne peut rien contre un mot de passe présent dans votre mémoire. Ceux-ci risquent donc de faire partie de notre vie pour les décennies à venir.

Au-delà des solutions technologiques potentielles, il ne faut pas oublier que ce perpétuel jeu de chat et de souris est parfois hors du contrôle des utilisateurs eux-mêmes. Ainsi, une des attaques les plus dangereuses aujourd'hui consiste à pirater un site ayant peu d'importance stratégique (par exemple un réseau de rencontres) et à récupérer les listes reliant mot de passe et coordonnées électroniques de l'utilisateur. Le pirate peut alors tester tous les sites plus importants (*email*, banque en ligne…) et espérer que ses cibles aient la même combinaison. Pour plus de 60 % des utilisateurs, ce sera le cas (et les 18–25 ans sont les plus vulnérables !). Certains sites évitent cette faiblesse (en chiffrant les mots de passe intelligemment), mais cette pratique est hélas loin d'être universelle. La meilleure méthode n'est donc pas d'avoir un mot de passe sécurisé pour toutes les utilisations, mais un bon algorithme qui en génère rapidement un nouveau – de tête si possible.

Méditez cette sage maxime sur le sujet : les mots de passe sont comme les sous-vêtements. Il faut éviter de trop les réutiliser, et en changer fréquemment ; il ne faut pas les montrer à tout le monde ou les laisser traîner ; et si possible il convient de garder une part de mystère.

N.K.B.

Calculez l'entropie de votre mot de passe

Dans le contexte de la génération d'un mot de passe, l'entropie correspond à la quantité d'information contenue dans le système, qui évolue en fonction du nombre de possibilités. Par exemple, avec mot de passe de six chiffres, on a un million de possibilités. Une hackeuse essayant de deviner le mot de passe sans information supplémentaire aura besoin de cinq cent mille essais en moyenne avant d'y arriver. Cet outil, bien qu'imparfait, rend possible une comparaison immédiate des différentes méthodes proposées. Calculons l'entropie de la dernière : chaque caractère peut être un chiffre (10 en tout), ou une lettre majuscule (26) ou minuscule (26), pour un total de 62 possibilités. Chaque caractère supplémentaire multipliant le nombre de possibilités par 62, on a 62^{10}, soit environ $8{,}4 \times 10^{17}$, mots de passe de dix caractères.

Passons au deuxième mot de passe. Les dictionnaires standards français comptent environ soixante mille mots. Si l'on en prend cinq au hasard indépendamment, on a donc $60\,000^5$, soit environ $7{,}8 \times 10^{23}$ possibilités.

La structure du premier mot de passe est la plus complexe. Il nous faut non seulement le nombre de mots en français, mais aussi leur longueur moyenne, qui se situe entre 10 et 11 (parmi les mots du dictionnaire). Chaque lettre pouvant, en moyenne, être transformée en une autre d'une seule manière (par exemple, A devient 4, O devient 0), on se retrouve avec $2^{11} = 2\,048$ possibilités pour ces modifications. On peut multiplier par trois cents possibilités pour les numéros à la fin (certaines dates peuvent s'écrire de plusieurs façon, comme 1998 et 98), et par une dizaine pour un chiffre initial. On se retrouve donc avec $60\,000 \times 2\,048 \times 300 \times 10$, soit environ $3{,}7 \times 10^{11}$ mots de passe différents. Cette méthode est de loin la plus mauvaise des trois !