Mots de passe : le choix humain plus sécurisé que la génération aléatoire

Nicolas Blanchard, Clément Malaingre, Ted Selker

Mots de passe : le choix humain plus sécurisé que la génération aléatoire

Nicolas K. Blanchard[1] et Clément Malaingre[2] et Ted Selker[3]

[1] *IRIF, Université Paris Diderot,* [2] *Teads France,* [3] *University of California, Berkeley*

Les mots de passe formés d'une suite de mots d'un langage donné sont une manière simple d'obtenir beaucoup d'entropie sans perdre en capacité de mémorisation. La génération automatique de ces phrases est cependant problématique car elle doit se faire sur un vocabulaire réduit pour garantir que les mots sont familiers, mais laisser les utilisateurs choisir eux-même leurs phrases mène aussi à une faible entropie.

Nous étudions la possibilité d'allier les deux en présentant une liste de mots aléatoires à partir de laquelle les utilisateurs doivent créer leur phrase. Grâce à une étude utilisateur sur des listes de 20 et 100 mots, nous montrons que cette méthode conserve entre 98% et 99% de l'entropie maximale. De plus, malgré un élément de distraction visant à les faire oublier leur phrase, plus de trois quarts des utilisateurs se souvenaient de tous les mots de leur phrase à la fin de l'expérience.

Mots-clefs : Mots de passe, Etude utilisateur, Sécurité accessible

1 Introduction

La nécessité d'avoir une forte entropie dans ses mots de passe et la découverte progressive que les contraintes de complexité sont contre-productives [MCTK10, KSK+11] sont en train de transformer les usages en recommandant des mots de passe plus longs et moins complexes [SKD+14]. Une alternative prometteuse passe par l'utilisation de phrases entières au lieu de simples mots de passe, mais la création de celles-ci laisse à désirer. En effet, lorsque les humains sont laissés à eux-mêmes, ils utilisent souvent des citations ou des passages de poèmes et chansons, ce qui compromet gravement la sécurité [DMR10, Bon12].

Suivant la génération aléatoire de mots de passe, la génération aléatoire de phrases semble naturelle mais pose plusieurs problèmes. Tout d'abord, imposer un tel mot de passe à un utilisateur rend sa mémorisation difficile. Ensuite, le dictionnaire d'où sont tirés ces mots compte pour beaucoup, et la taille de ce dernier est limitée si l'on veut que les utilisateurs connaissent tous les mots (ce qui améliore la mémorisation).

Nous considérâmes donc la possibilité de présenter une liste de mots aux utilisatrices à partir de laquelle elles devaient créer leur propre phrase. Pour analyser les comportements et vérifier la viabilité d'une telle méthode, nous recrutâmes 99 utilisatrices † et un groupe contrôle de 26 personnes, au départ via un site indexant les expériences en ligne [Kra98], et ensuite par les réseaux sociaux, l'expérience étant partagé par les utilisatrices. L'expérience eut intégralement lieu en anglais avec 51 personnes indiquant que l'anglais était leur langue principale, suivie 28 francophones, et 14 hébréophones, d'âges allant de 16 à 69 ans.

Durant cette expérience, une liste de mots de longueur 20 ou 100 (en test A/B équiprobable avec la même proportion d'anglophones dans chaque groupe) répartis en cinq colonnes fut présentée à chaque utilisatrice. Les mots étaient tirés uniformément dans un dictionnaire des 87691 mots anglais les plus fréquents. Les utilisatrices étaient censées choisir 6 mots et les écrire dans 6 cases situées sous la liste, étant encouragées à faire des phrases ou des combinaisons faciles à retenir. Des séquences aléatoires de six mots étaient assignées au groupe de contrôle. Il était alors demandé de mémoriser leur séquence de mots, après quoi une question leur demandant les deux premières lettres de chaque mot servait de renforcement. Dans une deuxième phase, elles devaient deviner une ou plusieurs fois les mots qu'une autre utilisatrice avait choisis, ce qui les faisait passer plusieurs minutes sur d'autres listes de mots dans le but de réduire leur capacité à se souvenir de leur phrase originale. Enfin, elles devaient rentrer leur phrase, d'abord sans indice puis à partir de la liste de mots montrée au départ si leur premier essai était incorrect.

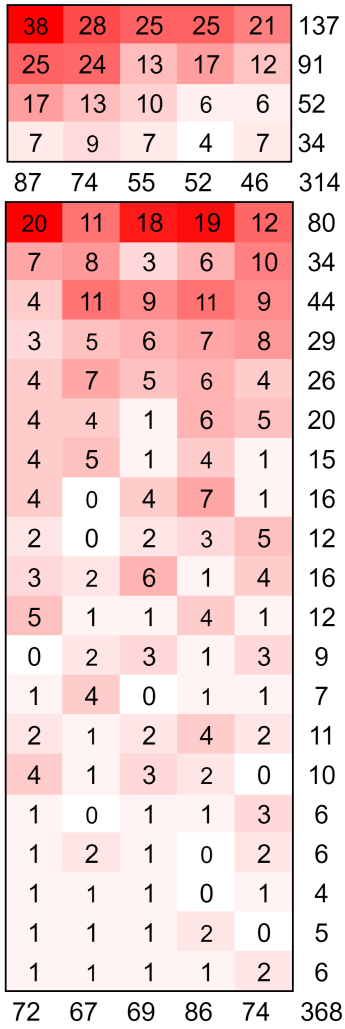
†. Nous arrêtâmes l'expérience à 100, avant de repérer et de supprimer la deuxième performance d'une utilisatrice.

2 Résultats

2.1 Choix des mots

Comme pressenti, les choix des utilisatrices étaient fortement influencés à la fois par la position relative des mots et par leur familiarité avec ces derniers. La Figure 1 ci-contre montre la fréquence de choix de chaque mot en fonction de sa position dans la liste (les nombres à côté de la carte représentent les sommes par ligne et colonne). Deux phénomènes sont visibles : tout d'abord, la préférence pour les colonnes de gauche qui est très présente dans la liste de 20 mots ne l'est plus dans la liste de 100 mots. Ensuite, la préférence pour les lignes supérieures est très présente chez les deux, mais ne s'accompagne pas d'un sursaut de fréquence pour les dernières lignes sur la liste longue, malgré la proximité de celles-ci aux zones de texte.

FIGURE 1: Carte thermique des fréquences relatives sur 20 mots et 100 mots.



Le choix dépendait aussi fortement de la fréquence relative des mots présentés (un mot plus fréquemment utilisé dans le langage courant étant plus souvent choisi). Cette tendance à choisir les mots plus fréquents est légèrement plus marquée chez les personnes ayant à priori un vocabulaire plus faible (celles n'ayant pas indiqué l'anglais comme langue principale). L'impact de ce choix est étudié dans la partie suivante.

2.2 Mémoire

Durant la dernière phase, 48% des utilisateurs parvinrent à taper leur phrase initiale sans erreur, et 32% supplémentaires ne firent qu'une faute de frappe[‡]. Le tableau ci-dessous récapitule les différentes erreurs par essai et par groupe. *Variante* correspond ici à l'écriture d'un mot très similaire (généralement l'addition ou la suppression d'un 's').

Essai	Correct	Manquant	Faux	Typo	Variante	Ordre
1 :20	18/47	26	5	6	8	6
1 :100	26/51	16	4	10	5	3
Contrôle	6/26	31	12	11	11	10
2 :20	14/29	0	3	1	2	8
2 :100	15/26	1	4	4	2	3

Au final, 81% des utilisatrices du groupe ayant 20 mots se souvinrent d'au moins cinq mots de leur phrase, contre 94% dans l'autre groupe. Il semble donc y avoir une facilité de mémorisation plus élevée quand les utilisatrices choisissent parmi une liste plus longue.

Le temps passé à choisir et à écrire sa phrase n'avait pas d'impact significatif. La langue principale n'avait pas non plus un impact sur le taux d'erreur global, mais les personnes ayant indiqué l'anglais comme langue principale oubliaient plus souvent leurs mots (29 oublis au total contre 13).

3 Modèles

Afin d'estimer les performances des utilisatrices, nous modélisâmes leurs comportements par plusieurs stratégies possibles. Chacune suppose (comme vérifié expérimentalement) que la fréquence de choix des mots est une fonction croissante de leur fréquence d'utilisation dans le langage écrit habituel, suivant par exemple une loi de Zipf (où la fréquence d'un mot est directement proportionnelle à l'inverse de son rang dans l'ordre des mots les plus fréquents $f(n) = \frac{c}{n}$)[FG11].

‡. Une utilisatrice passa directement au deuxième essai à cause d'un bug.

On peut voir ci-dessous l'explication de chaque stratégie, avec une table indiquant leurs entropies respectives (calculées selon la formule $E = \sum_{i=1}^{n} p_i \log(p_i)$, où p_i est la probabilité de choisir le mot i).

— *Plus Fréquents(n)* : correspond à prendre les 6 mots les plus fréquents (de rang le plus bas) présents dans les n mots affichés.

— *Corpus(n)* : correspond à choisir ses mots avec le même biais que le langage général, parmi une liste de taille n. Un mot de rang r_k est donc pris dans une liste de n mots avec probabilité

$$\frac{\frac{1}{r_k}}{\sum_{i=1}^{n} \frac{1}{r_i}}$$

— *Uniforme(n)* : correspond à prendre les mots de manière uniforme parmi un dictionnaire de taille n (le tirage étant uniforme sans remplacement, la taille de la liste n'a aucune importance).

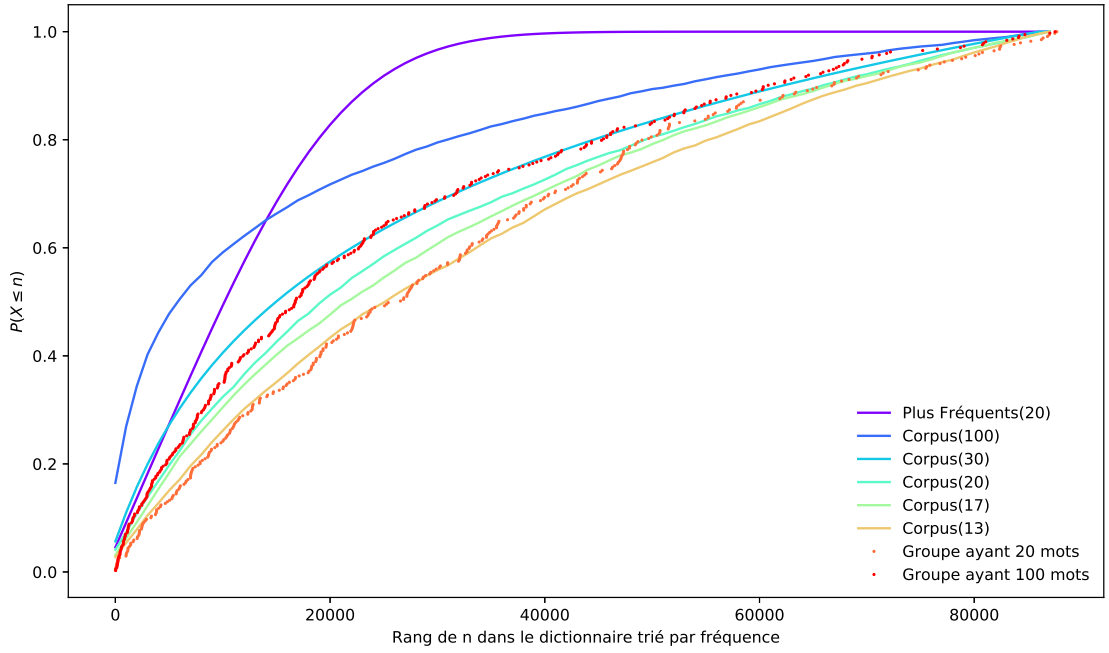
Entropie par mot des différentes stratégies

Stratégie	Entropie (bits)
Uniforme(dictionnaire de 87691 mots)	16.42
Corpus(13)	16.25
Corpus(17)	16.15
Corpus(20)	16.10
Corpus(30)	15.91
Corpus(100)	15.32
Plus Fréquents(20)	12.55
Uniforme(5000)	12.29
Plus Fréquents(100)	10.69
Corpus(langue anglaise : 276663 mots)	8.94
Corpus(dictionnaire de 87691 mots)	8.20

Si les utilisatrices avaient suivi un stratégie du *Plus Fréquent*, ou une qui s'en rapprochait, la méthode proposée aurait eu très peu d'intérêt. Par contre, une stratégie comme *Corpus* est supérieure à un choix uniforme, même sur un dictionnaire plus petit que celui utilisé.

Les résultats expérimentaux sur ce point excédèrent nos attentes, comme l'indique la Figure 2 qui montre la fonction de répartition (probabilité de prendre un mot de rang $\geq n$ en fonction de n) des différents modèles et des deux groupes expérimentaux. Pour les modèles, nous effectuâmes 10^9 simulations avec chaque stratégie afin d'obtenir une haute précision sur les distributions et l'entropie.

FIGURE 2: Fonction de répartition des mots selon différents modèles



Comme on peut le voir, le modèle *Plus fréquents* n'est pas du tout adapté, et le modèle *Corpus* correspond beaucoup mieux aux données expérimentales. Cependant, pour une liste de n mots, il semble que le modèle le plus satisfaisant ne soit pas *Corpus*(n) mais *Corpus*(k) avec $k < n$. Il existe une explication simple pour ce phénomène : la position des mots dans la liste. Ainsi, le groupe choisissant dans une liste de 20 mots prit très rarement les mots en bas à droite (fréquence réelle de 7.3% au lieu des 20% attendus). La stratégie correspondante s'apparente plutôt à *Corpus*(k) avec $13 \leq k \leq 17$ qu'à *Corpus*(20). De même, pour le groupe des 100 qui choisit 63.3% des mots parmi les six premières lignes, la stratégie est similaire à *Corpus*(30).

L'entropie étant concave et les événements étant de surcroît en ordre de probabilité décroissante, un gain sur la fin de la courbe est plus que compensé par un gain au début de la courbe. L'entropie expérimentale du groupe devant choisir parmi 20 mots est donc comprise entre celle de *Corpus*(13) et de *Corpus*(17), soit entre 98.4% et 99.0% du maximum théorique de 16.42 bits par mot sur notre dictionnaire.

Afin de vérifier qu'il n'existe pas une autre stratégie gagnante non mentionnée, nous analysâmes en sus les performances des utilisatrices qui devaient deviner les mots qu'avait choisis une autre utilisatrice. Celles-ci eurent un taux de succès supérieur à un choix uniforme, mais inférieur à celui d'une stratégie intelligente (comme *Corpus*). Aucune ne parvint à deviner l'ensemble des mots (sans même tenir compte de l'ordre).

4 Conclusion

Comme nous l'avons vu, le choix de mots dans une liste mène à une entropie proche de l'optimale, supérieure à celle atteinte par un générateur uniforme sur un dictionnaire de taille raisonnable. Par rapport au choix humain sans contrainte, cette méthode permet de presque doubler l'entropie par mot ; les phrases personnalisées sont faciles à mémoriser sans pour autant se retrouver dans un corpus standard. Elle est même valable en anglais pour des étrangers maîtrisant moins bien la langue. Une version étendue de ce papier est en libre accès sur HAL [BMS18].

Plusieurs questions subsistent cependant :

— Ces résultats sont-ils indépendant de la langue, et sont-ils directement applicables en français ?

— Choisir parmi 100 mots mène-t-il à des phrases plus mémorables parce qu'elles sont plus personnalisées, ou simplement parce que les mots choisis sont plus familiers ?

— Est-il possible d'utiliser le biais positionnel pour contrebalancer le biais linguistique et rendre le choix des mots encore plus uniforme ?

— Est-ce que la fréquence d'un mot et sa position sur la liste sont les seuls facteurs déterminants affectant son choix potentiel ? Un modèle mathématique intégrant cela et collant plus précisément aux données serait intéressant.

Références

[BMS18] Nicolas Blanchard, Clément Malaingre, and Ted Selker. Improving security and usability of passphrases with guided word choice. https ://hal.archives-ouvertes.fr/hal-01781233, April 2018.

[Bon12] J. Bonneau. The science of guessing : Analyzing an anonymized corpus of 70 million passwords. In *2012 IEEE Symposium on Security and Privacy*, pages 538–552, May 2012.

[DMR10] Matteo Dell'Amico, Pietro Michiardi, and Yves Roudier. Password strength : An empirical analysis. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

[FG11] Stephen Fagan and Ramazan Gençay. An introduction to textual econometrics. *Handbook of empirical economics and finance*, pages 133–154, 2011.

[Kra98] JH Krantz. Psychological research on the net. *WWW document*, 1998.

[KSK+11] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. Of passwords and people : Measuring the effect of password-composition policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2595–2604, New York, NY, USA, 2011. ACM.

[MCTK10] W. Ma, J. Campbell, D. Tran, and D. Kleeman. Password entropy and password quality. In *2010 Fourth International Conference on Network and System Security*, pages 583–587, Sept 2010.

[SKD+14] Richard Shay, Saranga Komanduri, Adam L. Durity, Phillip (Seyoung) Huh, Michelle L. Mazurek, Sean M. Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Can long passwords be secure and usable ? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2927–2936, New York, NY, USA, 2014. ACM.