Comment corriger efficacement les typos dans les mots de passe

Nikola K. Blanchard

Institut de Recherche en Informatique Fondamentale, Université Paris Diderot

Algotel 2019

Saint Laurent de la Cabrerisse, 7 Juin 2019

Petit cours de cuisine: comment faire une tapenade ?

Les mots de passe aujourd'hui

En moyenne:

- Une utilisatrice a $\sim$ 100 comptes
- Elle créé 50 MDP par an
- Pas de méthode générale à cause des contraintes

En conséquence:

- Fréquente réutilisation (75% des utilisateurs)
- Partage des MDP (40% des utilisateurs)
- Perte très fréquent (40% à 60% tous les trois mois)

Les mots de passe aujourd'hui

En moyenne:

- Une utilisatrice a $\sim$ 100 comptes
- Elle créé 50 MDP par an
- Pas de méthode générale à cause des contraintes

En conséquence:

- Fréquente réutilisation (75% des utilisateurs)
- Partage des MDP (40% des utilisateurs)
- Perte très fréquent (40% à 60% tous les trois mois)

Sécurité des mots de passe

Attaquer le mot de passe:

- ”123456” est toujours le plus fréquent
- Les contraintes sont contre-productives
- La longueur bat la complexité

Attaquer le serveur:

- Les vulnérabilités principales viennent du phishing et de la réutilisation
- Rarement hachés ou salé
- Rarement avec une bonne fonction de hachage (pas SHA-256)
- Tout devrait avoir lieu côté client

Sécurité des mots de passe

Attaquer le mot de passe:

- ”123456” est toujours le plus fréquent
- Les contraintes sont contre-productives
- La longueur bat la complexité

Attaquer le serveur:

- Les vulnérabilités principales viennent du phishing et de la réutilisation
- Rarement hachés ou salé
- Rarement avec une bonne fonction de hachage (pas SHA-256)
- Tout devrait avoir lieu côté client

Pourquoi corriger les typos ?

Pourquoi corriger les typos ?

Elles gênent les utilisateurs

- Très frustrant
- Fréquent (3% des essais de login)
- Plus prévalent sur les longs mots de passe

Corriger ne réduit pas la sécurité

- Pas d'effet sur les attaques hors-ligne
- Les mots de passe courant sont loins les uns de autres
- Cela permet de limiter la fréquence de login

Pourquoi corriger les typos ?

Elles gênent les utilisateurs

- Très frustrant
- Fréquent (3% des essais de login)
- Plus prévalent sur les longs mots de passe

Corriger ne réduit pas la sécurité

- Pas d'effet sur les attaques hors-ligne
- Les mots de passe courant sont loins les uns de autres
- Cela permet de limiter la fréquence de login

Types de typos (recalculé depuis [Chatterjee *et al.*, 2016])

Catégorie de typo	Proportion de mauvais MDP
Substitution simple	29.7
AZERTY \ voisin numpad	14.0
Single shift	8.5
Suppression simple	19.4
Caps lock	14.7
Insertion simple	13.1
Espace	2.0
Lettre dupliquée	3.8
Transposition simple	3.9
Autres	19.0

Contraintes

Securité: ne pas introduire de nouvelles vulnérabilités

Faible coût:

- Compatible avec le hachage
- simple à implémenter
- Pas trop de calcul/stockage sur le serveur

Corriger autant de typos *légitimes* que possible (32% chez [Chatterjee *et al.*, 2016])

Contraintes

Securité: ne pas introduire de nouvelles vulnérabilités

Faible coût:

- Compatible avec le hachage
- simple à implémenter
- Pas trop de calcul/stockage sur le serveur

Corriger autant de typos *légitimes* que possible (32% chez [Chatterjee *et al.*, 2016])

Contraintes

Securité: ne pas introduire de nouvelles vulnérabilités

Faible coût:

- Compatible avec le hachage
- simple à implémenter
- Pas trop de calcul/stockage sur le serveur

Corriger autant de typos *légitimes* que possible (32% chez [Chatterjee *et al.*, 2016])

Méthodes triviales ?

Méthodes triviales

Stocker tout en clair... : problème de sécurité

Stocker toutes les typos possibles : problème de coût

Envoyer toutes les corrections de typos possibles : problème de communication

Utiliser du chiffrement asymétrique: problème de complexité+de sécurité

Méthodes triviales

Stocker tout en clair... : problème de sécurité

Stocker toutes les typos possibles : problème de coût

Envoyer toutes les corrections de typos possibles : problème de communication

Utiliser du chiffrement asymétrique: problème de complexité+de sécurité

Méthodes triviales

Stocker tout en clair... : problème de sécurité

Stocker toutes les typos possibles : problème de coût

Envoyer toutes les corrections de typos possibles : problème de communication

Utiliser du chiffrement asymétrique: problème de complexité+de sécurité

Méthodes triviales

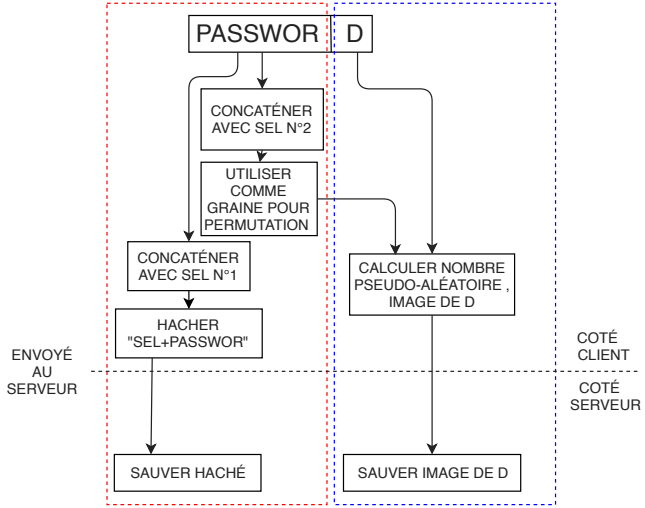
Stocker tout en clair... : problème de sécurité

Stocker toutes les typos possibles : problème de coût

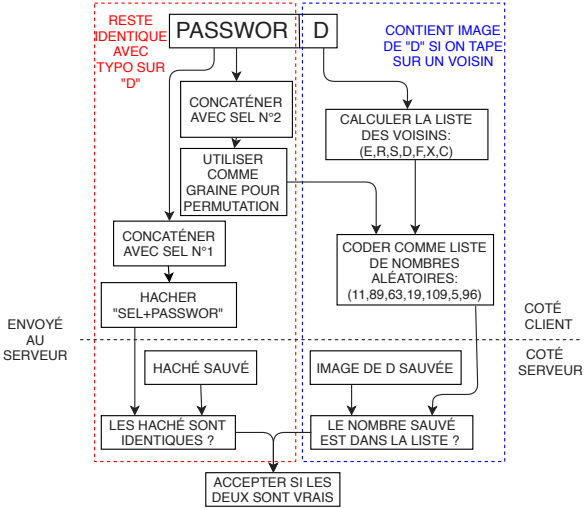
Envoyer toutes les corrections de typos possibles : problème de communication

Utiliser du chiffrement asymétrique: problème de complexité+de sécurité

Corriger les substitutions



Corriger les substitutions



Extensions aux autres typos

Transposition:

- Enlever deux lettres avant de hacher
- Encoder chaque lettre avec deux permutations différentes

Insertion:

- Combiner les deux méthodes précédentes
- Enlever deux lettres après en insérer une donne un haché de substitution

Extensions aux autres typos

Transposition:

- Enlever deux lettres avant de hacher
- Encoder chaque lettre avec deux permutations différentes

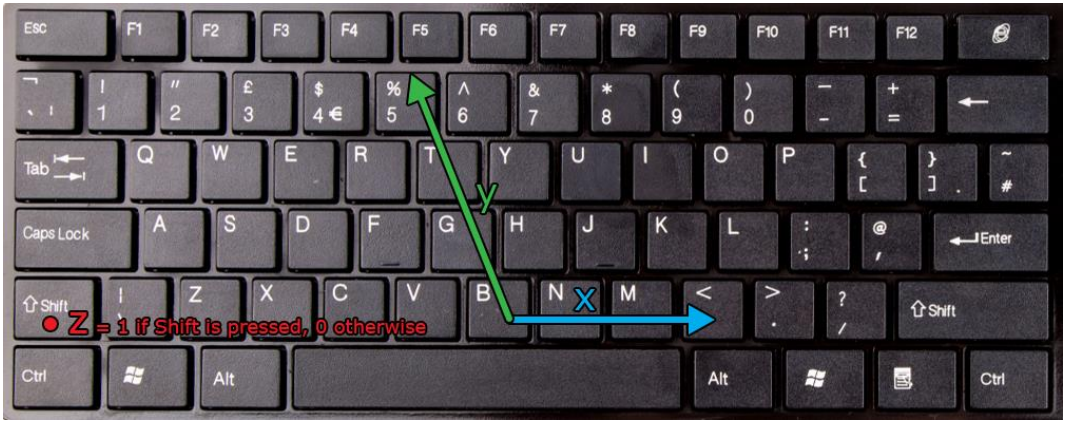
Insertion:

- Combiner les deux méthodes précédentes
- Enlever deux lettres après en insérer une donne un haché de substitution

Comparaison des méthodes

Algorithm	Substitution	Transposition	Insertion	Complet
Calcul,# de				
Permutations	n	$4n-4$	$4n-4$	$\max(4(n-1), 60)$
Hachés	$n+1$	n	n	$\max(n+1, 17)$
Nombres	$n \times k$	$(n-1) \times 4k$	$(n-1) \times 4k$	$\max(4(n-1)k, 60k)$
Stockage, # de				
Hachés	$n+1$	n	$2n$	$\max(2n+1, 33)$
Nombres	n	$4n$	$5n$	$\max(5n, 80)$
Typos corrigées				
Strict	24.2 %	28.4 %	34.5 %	50.2 %
Tolérant	24.2 %	28.4 %	42.2 %	57.7 %

Algorithme générique basé sur le logarithme disret



Système de coordonnées du clavier

Algorithme générique basé sur le logarithme disret

Pour de petits nombres premiers p_i, le mot de passe est codé comme

$$X(P) = \prod_{1 \leq i \leq n} p_i^{x_i} \times p_{i+n}^{y_i} \times p_{i+2n}^{z_i}$$

Envoyer $g^{X(P)}$ pour un g aléatoire dans un grand groupe.

Pour un MDP tapé avec une typo P':

Si $P' \approx P$: $\quad g^{X(P')} = (g^{X(P)})^{p_i} \quad$ OU $\quad (g^{X(P')})^{p_i} = g^{X(P)}$

Propriétés de l'algorithme basé sur le log discret

Positif

- Très haute sécurité
- Stockage et commmunication asymptotiquement optimaux

Négatif:

- Ne marche que sur les substitutions
- Coût très élevé (potentiellement plusieurs secondes)

Résumé du framework principal

Securisé:

- Même résistance resistance en ligne que [Chatterjee *et al.*, 2017]
- Speed-up < 1.5 hors ligne sur des données réelles.

Bas coût:

- Pas de calculs supplémentaires par le serveur en espérance
- Les communications tiennent dans un paquet standard
- Compatible with previous systems

Corrige 57% des typos soit 91% des typos *légitimes*.