HACKATHON 2021

by L'ÉQUIPE

Qui sommes-nous ?

Une team Made in Epita Paris

L’équipe, c’est:

- deux data scientists:

Maxime Leherle Kévin Guillet

- trois responsables sécurité:

Simon Guiot Sébastien Lanyi
Vianney Delaboudinière

Challenge 1 - Identifier

Intercepter des signaux sur les capteurs d'énergies:

- ➔ Une séquence de touches clavier a été enregistré
- ➔ Cette séquence contient des identifiants de connexion

Le but est de retrouver les identifiants encodés sous forme de signaux électrique.

<u>Problématique retenue :</u> Comment extraire le login et le mot de passe depuis les données du rayonnement électrique ?

Travail effectué en amont:

- Pose d'un appareil de mesure électrique (eavesdropping)
- Récupération des signaux dans une plage de fréquences donnée.

Notre solution

Une ressemblance forte à un OCR ...

Tout part de la base de données ...

- On charge les trames associées aux lettres en utilisant le script fourni.

- Ces trames sont chargées et nous en tirons les informations nécessaires afin d'en créer une base de données csv.

- Ce fichier csv contient l'ensemble des trames, avec les touches associées aux 17 pics (1 par colonne), ce qui donne un peu plus de 350.000 lignes.

- Lors de l'entraînement du modèle, on sépare la base de données aléatoirement avec 70% pour l'entraînement et 30% pour la validation.

Sébastien - Vianney

… pour construire un modèle de prédiction …

Le problème pouvant se rapprocher d'un OCR, nous avons choisi une solution similaire:

- Nous avons utilisé un modèle d'apprentissage supervisé type classifieur linéaire, le **SVM**.
- Cependant nous avons observés des résultats parfois très proche entre touches.
- Nous avons donc testé une approche avec les probabilités de chaque touche sur chaque trame pour tester les cas de détection imprécise, cela n'a rien donné. Nous avons également optimisé les prédictions en les parallélisant.

Maxime - Kévin

… pour retrouver des identifiants.

- On lance la prédiction sur la totalité de la capture découpées en trames de 17 pics.
- La séquence ‘CTRL+ALT+SHIFT’ est interprétée comme une séquence commençant par des ‘CTRL’ et finissant par des ‘NOKEY’ (*l’utilisateur relâche les touches*).
- Puis nous analysons la capture en considérant les séquences d’une même prédiction comme une seule touche. Les prédictions sans répétition sont considérées comme du bruit et sont ignorées.

- Résultats finaux -

Prédiction des identifiants:

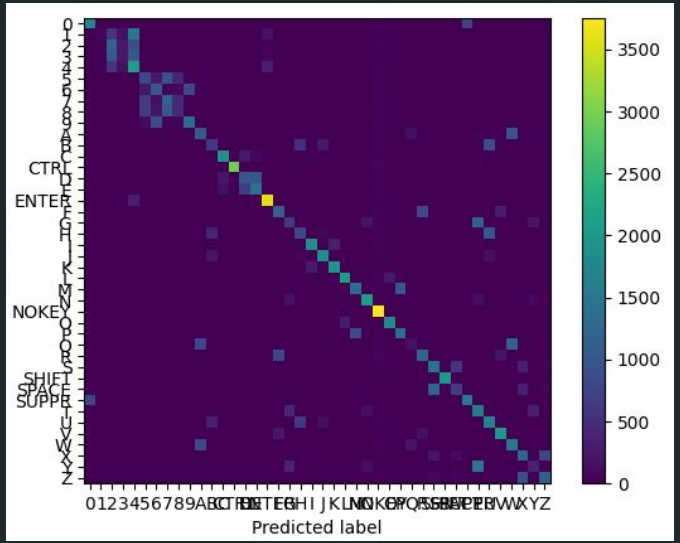
- après traitement, notre algorithme renvoie cette série de touche :

  SHIFTISHIFTUAWAWCKAWTGON202244

- nous pouvons supposer que les identifiants sont :

login: Hackathon
mdp: 2022

Entrainement du modèle:



score de validation: 61.05%

Ouverture

Comment gérer le bruit de façon automatisé ?

- Avoir une base de données qui contient les signaux associés à l’appui de SHIFT+touche et CTRL+ALT+SHIFT.

- Remplacer notre déduction d’identifiants par une IA qui :

  ➔ À partir d’une grande quantité de captures dont on connaît les identifiants associés.

  ➔ sur lesquels on lance notre modèle de prédiction d’appui de touches.

  ➔ détermine les identifiants à partir de la séquence de touches prédites.