Crypted or Uncrypted ?

FOUCHÉ Stanislas 22007315
PRIOU Antoine 22008452

Introduction

Sommaire

Partie 1

Dataset

Description du dataset

Partie 2

Expérimentations

Analyse des différentes méthodes/applications

Partie 3

Résultats

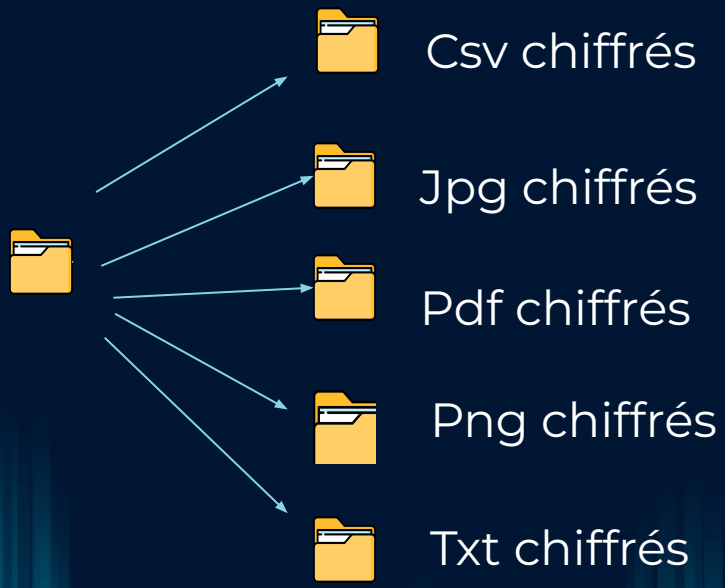
Interprétation des résultats

Résumé du premier semestre

- Production d’un état de l’art
- Explications des méthodes de chiffrement
- Choix et descriptions des logiciels

Dataset

init_chiffrement .py



Algorithmes de chiffrement :

- AES 256 CBC
- AES 128 CBC
- AES 256 ECB
- des-ede3-cbc
- des-ede3-ecb
- Bf-cbc
- Camellia-256-cbc
- Cast5-cbc
- Rc4

MAGNET
Encrypted Disk Detector

&&

Elcomsoft
Encrypted Disk Hunter

2 outils pour WINDOWS, spécialisés dans la détection de volumes chiffrés sur un disque

- Détection approfondie et automatique
- Rapide et Efficace
- Large compatibilité
- Analyse forensic
- Recherche approfondie Bitlocker

Cipher.exe

Outil intégré à Windows qui permet de gérer le chiffrement des fichiers et des dossiers via
(EFS - Encrypting File System)

Avantages :

- Rapide et Simple
- Identifie très bien les fichiers cipher

Inconvénient :

- Identifie seulement les fichiers cipher

Commandes :

- cipher /u /n
- cipher /s:<directory>

TCHunt

Outil de détection de fichiers chiffrées, mais aussi de volumes chiffrés

Il se base sur plusieurs critères :

- La taille du fichier
- L'entropie du fichier
- L'en-tête du fichier

Avantages:

- Rapide et Efficace
- Spécifie la nature des données (clé, données)
- Spécifie les algorithmes détectés

Inconvénients :

- Ne détecte pas tous les types de fichiers
- Détecte des faux positifs (entropie élevée)

Recherche de fichiers par extension :

- Algorithme simple par recherche d’extension
- On se base sur une liste d’extensions connues
- Mettre à jour régulièrement

SVM

Prétraitement

- Récupération de tous les datasets à comparer
- Calcul d'entropie de chaque fichier
- Structuration des données en un set

- Fusion aléatoire du dataset non chiffré/ avec chaque dataset sélectionné ratio 70/30 avec sklearn.model_selection.train_test_split

- Au total toujours 195 fichiers par dataset dont :
  - 70% de training data
  - 30% de test data

Modélisation

- Implémentation d'un svm sklearn avec un noyau linéaire

- Entraînement du modèle sur training data
- Comparaison

Évaluation

- Comparaison prédictions/ labels test data

- Concaténation des résultats

- Affichage pour chaque dataset des résultats en utilisant matplotlib

- Sauvegarde des résultats dans le dossier results/

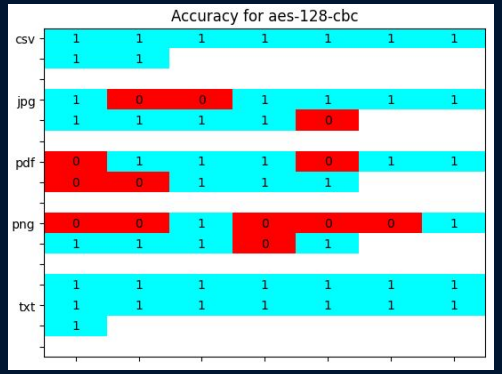
Comparaison Outils

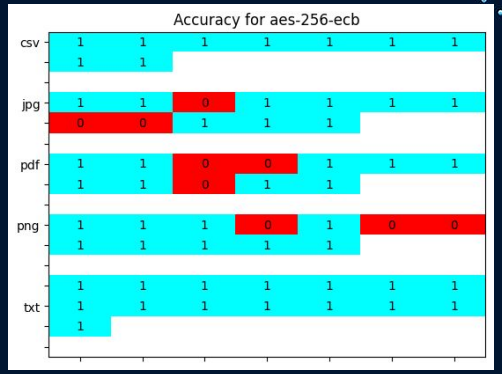
	Temps d'execution	Conteneur	Partition/Volume
MAGNET	30s	✓	✓
ELCOMSOFT	30s	✓	✓

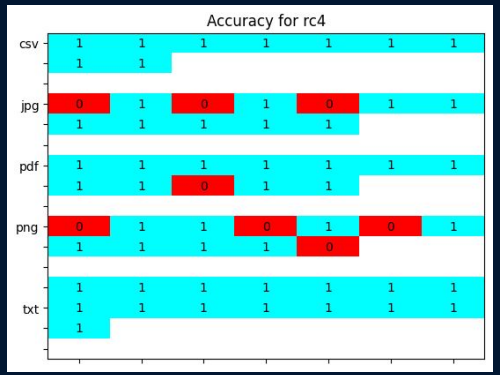
Comparaison Outils

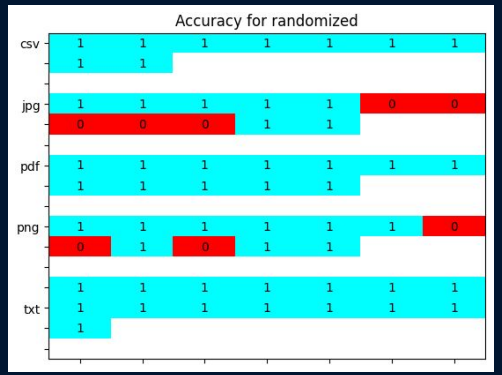
	Temps d'execution	Base recherche	Conteneur
Cipher	O(1)	fichiers cipher	
Recherche extension	1min	Extensions connus	
TCHunt	1min	- Taille fichier - Entropie - Analyse en-tête	

Résultat du SVM









Interprétation & Résultats

On obtient les chiffres suivant, chaque numéro représente le nombre d'erreurs de prédictions

	Chiffrement	TOTAL	CSV	JPG	PNG	PDF	TXT
1.	DES EDE3 ECB	: 6	(0/9 + 1/12 + 2/12 + 3/12 + 0/15)				
2.	HEADLESS	: 7	(0/9 + 4/12 + 1/12 + 2/12 + 0/15)				
3.	Randomized	: 8	(0/9 + 5/12 + 0/12 + 3/12 + 0/15)				
4.	HALFHEADED	: 8	(0/9 + 1/12 + 4/12 + 3/12 + 0/15)				
5.	RC4	: 8	(0/9 + 3/12 + 1/12 + 4/12 + 0/15)				
6.	BF-CBC	: 8	(0/9 + 2/12 + 3/12 + 3/12 + 0/15)				
7.	AES 256 ECB	: 9	(0/9 + 3/12 + 3/12 + 3/12 + 0/15)				
8.	GPG	: 9	(0/9 + 3/12 + 4/12 + 2/12 + 0/15)				
9.	CAST5 SBC	: 11	(0/9 + 1/12 + 4/12 + 6/12 + 0/15)				
10.	DES EDE3 CBC	: 12	(0/9 + 3/12 + 5/12 + 4/12 + 0/15)				
11.	CAMELLIA 256 CBC	: 12	(0/9 + 4/12 + 4/12 + 4/12 + 0/15)				
12.	AES 128 CBC	: 13	(0/9 + 3/12 + 4/12 + 6/12 + 0/15)				
13.	AES 256 CBC	: 16	(0/9 + 4/12 + 6/12 + 6/12 + 0/15)				

Répartition des tâches

- Dataset
- Expérimentation
- Rapport
- Analyse résultats

Antoine :

— Script pour ordonner le dataset non chiffré

— Réalisation d'un SVM basé sur l'entropie

— Expérimentation et affichage des résultats obtenus

— Expérimentation des applications

— Rédaction du rapport

Stanislas

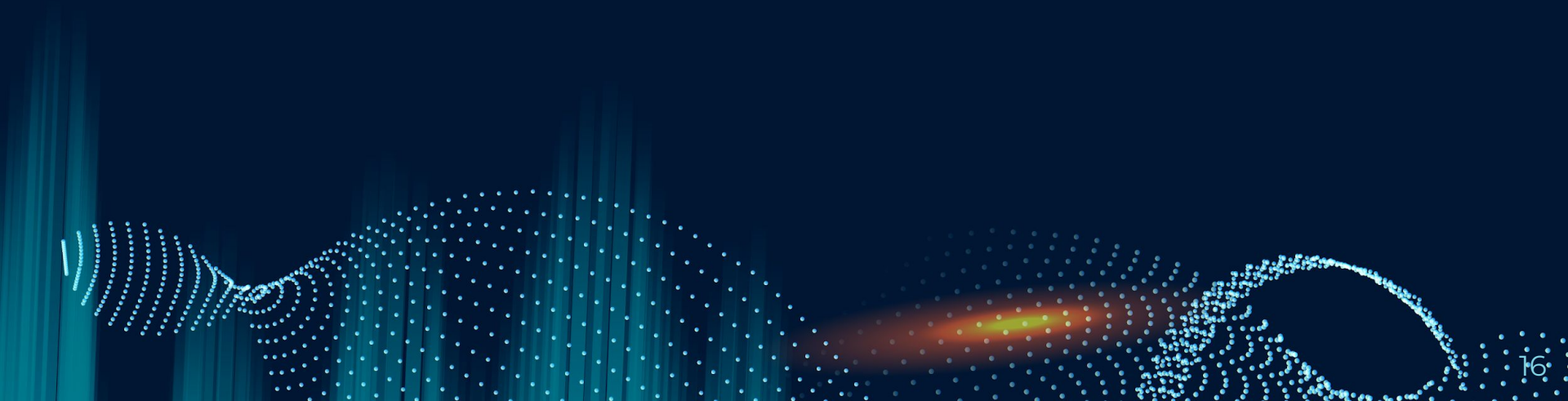
— Constitution du dataset chiffré

— Script de chiffrement pour obtenir l'ensemble des datasets chiffrés

— Expérimentation des applications

— Rédaction du rapport

Conclusion



THANKS!

Code disponible sur le GitLab Unicaen :

https://git.unicaen.fr/22007315/crypted_or_uncrypted.git