

RÉCUPÉRATION DE DONNÉES ET INTELLIGENCE ARTIFICIELLE

Que peut-on réellement attendre ?

JFIN 2023 - CLUSIS

- TRISTAN PINCEAUX
- HEAD OF CERT CWATCH

Almond



Agenda

- Stockage et format de données
- Les ransomwares et leurs techniques de chiffrement
- Les opportunités pour retrouver ses données
- Et l'IA dans tout ça ?

01.

> WHOAMI

01.



Tristan Pinceaux

Incident handler / Forensic expert



Problem solver



> 12 ans d'expérience DFIR



Manager CERT

Ma vision : La sécurité ne doit pas coûter la liberté

Ma mission : Arrêter les cybercriminels avant qu'ils ne vous arrêtent



- Point de vue d'expert **DFIR**
- Enthousiaste mais pas cryptologue ni data scientist
- L'IA est un domaine qui évolue à très grande vitesse



Quand j'étais plus jeune, j'ai trouvé dans le grenier de mes grands-parents un vieux livre.

→ Ce livre était dans un sale état

→ Comment pouvoir lire le livre en l'état ?

→ **Comment récupérer la donnée ?**

02.

STOCKAGE ET FORMAT DE DONNÉES

02.

Rappel : C'est quoi une donnée ?

Définition du dictionnaire Le Robert

donnée

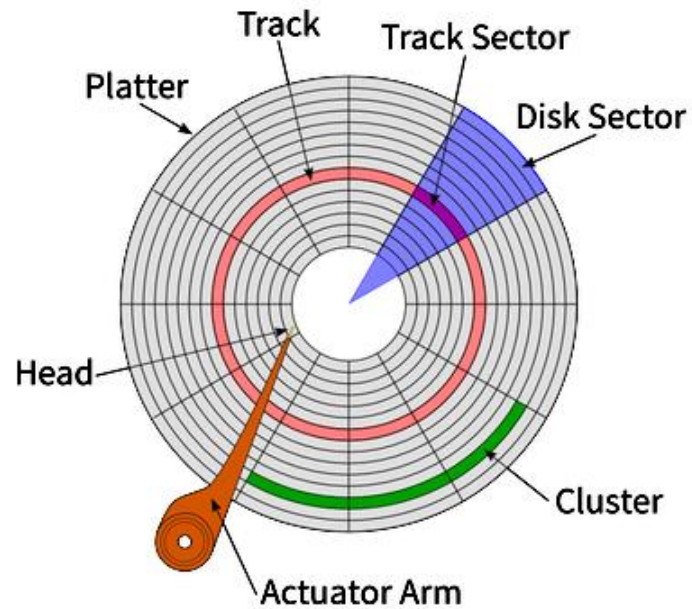
nom féminin

Représentation conventionnelle d'une information permettant d'en faire le traitement automatique.

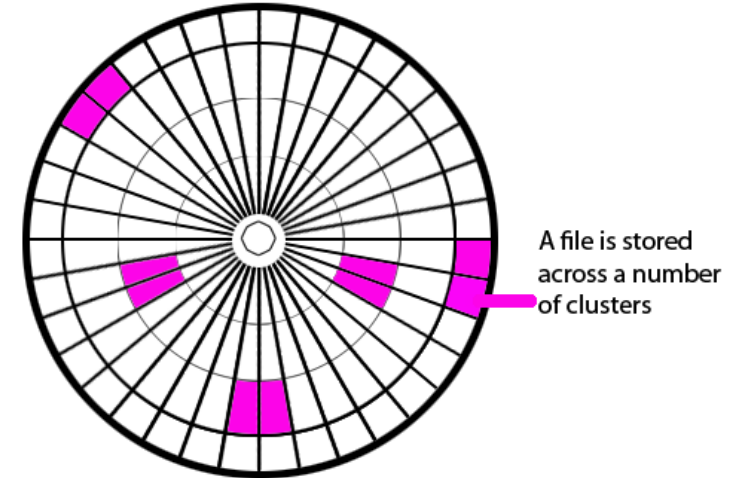


Où sont stockées nos données ?

Structure de données dans un disque NTFS



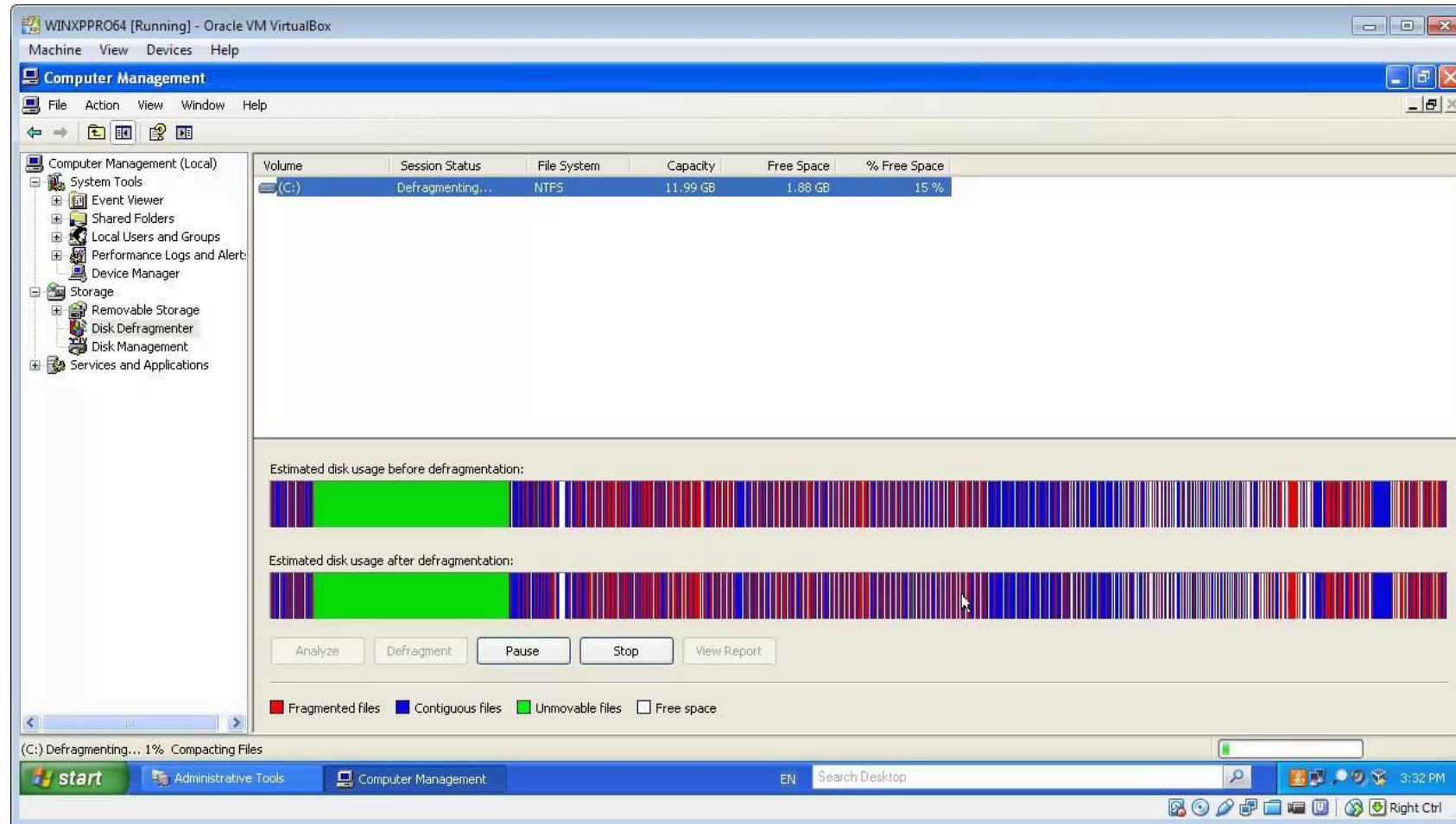
Le grenier de mes grands-parents



Les feuilles du livre,
éparpillées dans le grenier

Rappel : la défragmentation

Pour les moins jeunes...



Rappel : la défragmentation

Pour les encore moins jeunes...

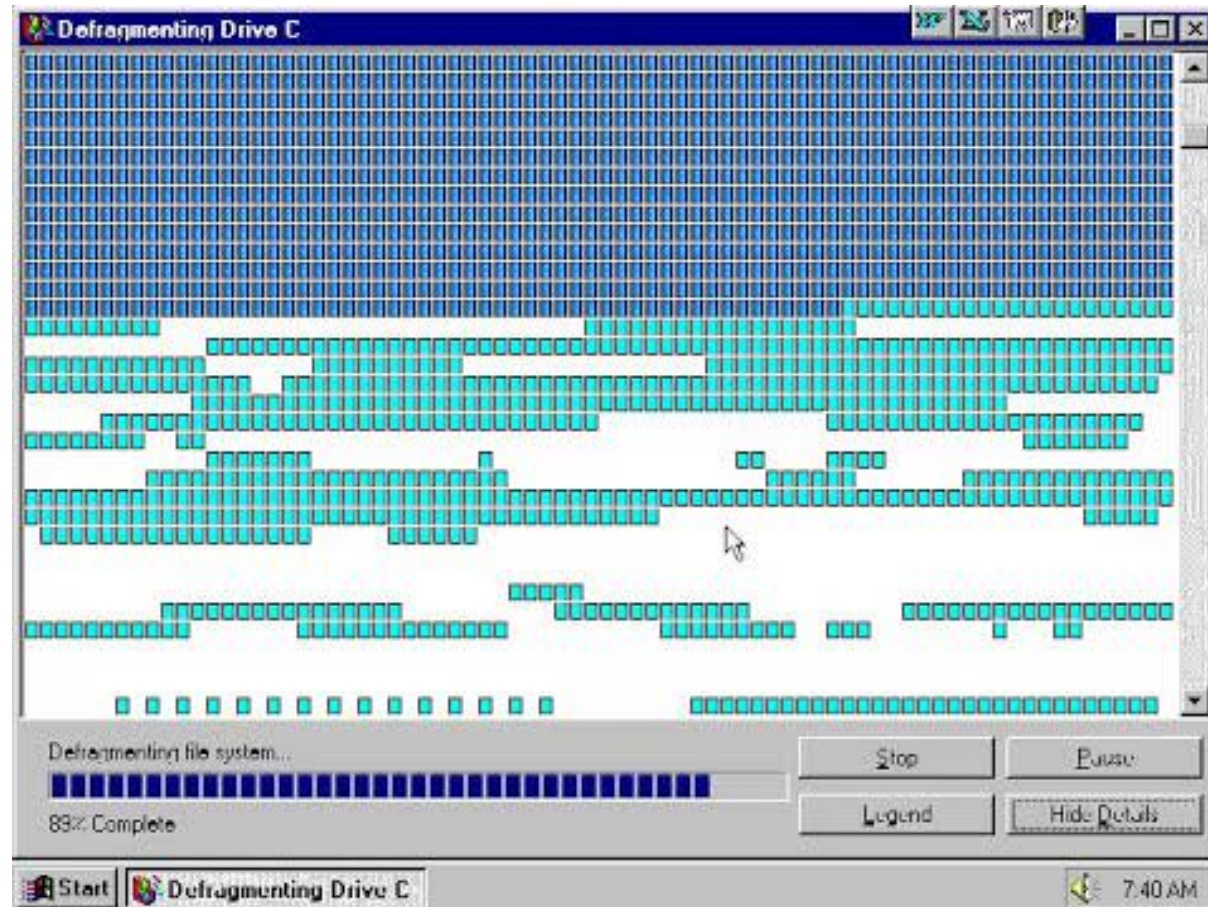



Table des matières (\$I30) et Index /annuaire (\$MFT)



	PAGE
1. JEROME AND THE LION	1
2. THE SAINT WHO STOOD ON A PILLAR	7
3. THE SCOURGE OF GOD	14
4. THE VANDAL HORDE	22
5. KING CLOVIS BECOMES A CHRISTIAN	30
6. A CAMEL DRIVER BECOMES A PROPHET	37
7. WHAT IT MEANT TO BE A KNIGHT	47
8. BERTHA WITH THE BIG FOOT	56
9. STORIES OF CHARLEMAGNE	66
10. CHARLEMAGNE AND THE MAGIC RING	76
11. CHARLEMAGNE AND THE ROBBER	82
12. ROLAND BECOMES A KNIGHT	89
13. THE DEATH OF ROLAND	97
14. HOW NORMANDY CAME BY ITS NAME	105
15. OLAF, THE BOY VIKING OF NORWAY	116
16. THE CID WINS HIS NAME	126
17. THE LAST DAYS OF THE CID	135
18. THE LORELEI	143
19. THE MOUSE TOWER	149
20. THE DEVIL'S LADDER	155
21. GERDA'S RIDE TO HER WEDDING	163
22. PETER THE HERMIT	170
23. THE WIVES OF WEINSBERG	178
24. THE MEETING OF RICHARD AND SALADIN	182

v

Sangria Recipes

Index

Absinthe, 27, 39, 106	lemon, 33, 35, 40, 42, 43, 47, 51, 61, 65, 75, 78, 83, 84, 87, 118, 128
apple, 21, 22, 24, 34, 37, 38, 41, 51, 79, 88, 104, 123, 125, 127	lime, 35, 38, 46, 47, 51, 61, 69, 71, 78, 82, 83, 84, 88, 113, 118
applejack, 124	lychee liqueur, 24, 66, 76, 102
apricot, 114	mango, 37, 78
apricot brandy, 30	melon, 34, 81, 87
artichoke liqueur, 22, 125	Midori, 84, 108
banana liqueur, 126	nectarine, 87, 114
blackberries, 26, 28, 29	orange, 21, 23, 33, 34, 35, 40, 41, 42, 44, 47, 51, 52, 61, 65, 67, 69, 78, 82, 83, 84, 87, 89, 105, 109, 113, 118, 123, 127
blackberry liqueur, 27	orange liqueur, 126
blueberries, 26, 28, 30, 31, 32, 49, 73, 74, 91, 97, 112	peach, 33, 36, 37, 43, 44, 51, 53, 56, 58, 66, 76, 86, 87, 92, 93, 100, 107, 124, 126
brandy, 21, 37, 42, 51, 52, 53, 65, 87, 89, 91, 92, 105, 109, 114, 115, 128	peach liqueur, 54, 102
cantaloupe, 114, 119	pear, 37, 94, 123
Chambord, 32, 49, 77, 103, 116	pear liqueur, 100
Chastreuse, 79, 85, 106	pineapple, 95, 96
cherries, 41, 52, 61, 67, 127	plum, 68, 97, 98, 108
cherry liqueur, 124	pluot, 99
Cognac, 41, 75	pomegranate liqueur, 36, 79, 86
Countreau, 78, 82	raspberries, 26, 39, 48, 49, 70, 85, 90, 99, 101, 103, 106, 121, 122
cranberries, 45	raspberry liqueur, 104
Crème de Cassis, 28, 101	rum, 23, 25, 29, 37, 63, 78
cucumber liqueur, 48, 70, 122	strawberries, 26, 27, 37, 42, 49, 69, 73, 80, 91, 115, 116, 117, 118, 119, 120, 121, 123
elderflower liqueur, 99, 107	strawberry liqueur, 90
Frangelico, 85	tangerine, 37
gin, 33	tangerine liqueur, 45, 73, 74, 98
ginger liqueur, 56, 58, 86, 90, 100, 101	
Goldschlager, 39, 58, 122	
Grand Marnier, 61, 69, 118	
grapes, 21, 30, 33, 52, 53, 62, 63, 64, 76, 91, 107, 109	
Irish whiskey, 68, 108	
kiwi, 72	
kumquat, 73, 74	

141

MFT = index / annuaire du disque



Le système de fichier va enregistrer les endroits physiques où sont stockés chaque fichier, sur le device de stockage

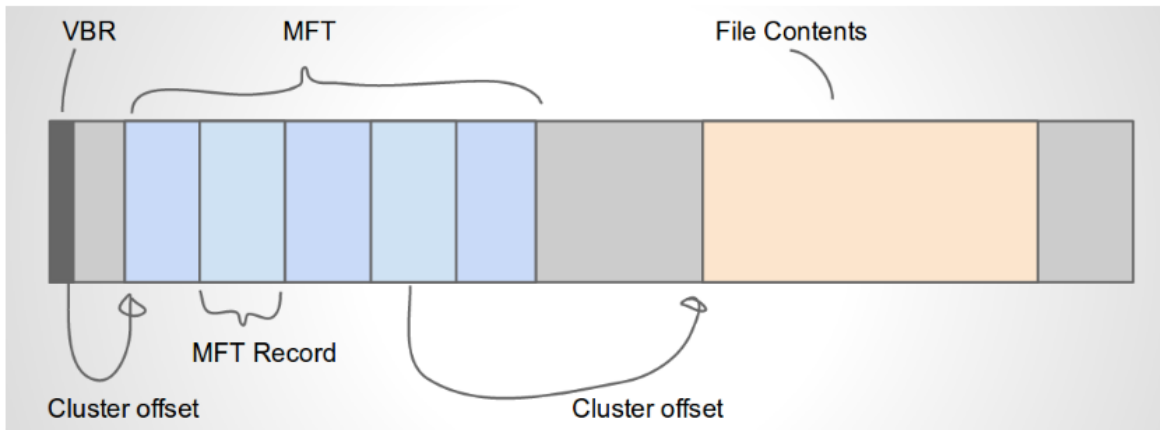


Figure 1: NTFS Volume Layout Showing the \$MFT

→ Pour NTFS : La **Master File Table** (\$MFT) contient des entrées pour chaque fichier

→ Chacune composée de métadonnées :

- Noms de fichier
- Taille,
- Date de création / modification,
- ACLs et permissions
- **Liste des clusters contenant la donnée**
- ...

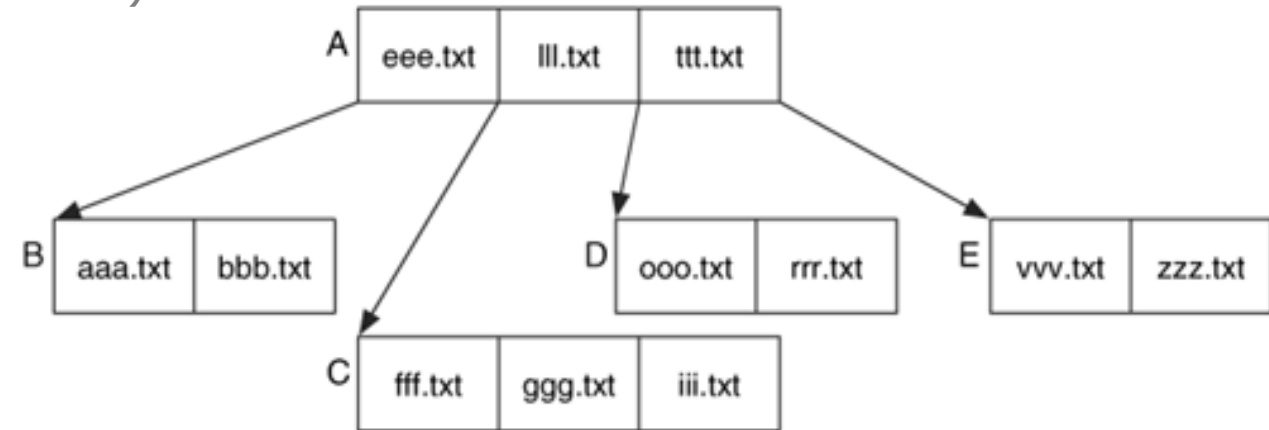
→ Ou de données lorsque le fichier est plus petit que la taille de l'entrée

- Fichier "Résident"

Ils donnent le contenu du livre : les chapitres

→ Les \$I30 sont des attributs d'index NTFS

- Sous forme de B-tree
- Index du contenu du dossier (fichiers et dossiers)
- Présent dans chaque dossier
- Leur taille varie => Slackspace



- **\$MFT = index**, contient les endroits du disque où sont stockés les fichiers qu'on cherche.
- L'index **\$I30 = table de matières**, contient les informations du contenu d'un dossier sur un NTFS
- Il n'y a plus qu'à aller au numéro de page indiqué pour retrouver le contenu.

La signature des fichiers

```
Applications ▾ Places ▾ Terminal ▾  
File Edit View Search Terminal Help  
brucewayne@brucewayne:~/Desktop/g4g$ xxd image.png | head  
00000000: 8950 4e47 0d0a 1a0a 0000 000d 4948 4452 .PNG.. ....IHDR  
00000010: 0000 00d6 0000 00eb 0803 0000 0022 d998 ....."  
00000020: 6f00 0001 4d50 4c54 45ff ffff ea00 01f4 o...MPLTE.....  
00000030: 0000 e100 00e0 e0e0 e800 00fe 0000 dbdb .....  
00000040: dbf2 0000 d9d9 d9e2 e2e2 dede ded7 d7d7 .....  
00000050: a100 00e9 0000 a400 00ac 0000 e8e8 e8bb .....  
00000060: 0000 8200 0099 0000 b400 00d1 0000 af00 .....  
00000070: 00f5 f5f5 d700 009d 0000 c100 0094 0000 .....  
00000080: 8900 00dd 0000 0000 b200 00cc 0000 efef ef8e .....  
00000090: 0000 7e00 00e4 eeee edf4 f4e8 e4e4 dbd1 ..~.....  
brucewayne@brucewayne:~/Desktop/g4g$
```

```
File Edit View Search Terminal Help  
brucewayne@brucewayne:~/Desktop/g4g$ xxd test.jpg | head  
00000000: ffd8 ffe0 0010 4a46 4946 0001 0100 0001 .....JFIF.....  
00000010: 0001 0000 ffdb 0084 0009 0607 1010 0f0f .....  
00000020: 100f 1210 100f 100f 0f0f 1010 0f0f 100f .....  
00000030: 0f0f 0f15 1116 1615 1515 1518 1d28 2018 .....(  
00000040: 1a25 1d15 1521 3121 2529 2b2e 2e2e 171f .%...!1!%)+....  
00000050: 3338 332d 3728 2d2e 2b01 0a0a 0a0e 0d0e 383-7(-.+.....  
00000060: 1510 101a 2d1d 1d1d 2b2d 2d2d 2d2d 2d2d ....~...+-----  
00000070: 2d2d 2b2d 2d2d 2d2d 2d2d 2d2d 2d2d 2d2d --+-----  
00000080: 2d2d 2d2d 2d2d 2d2d 2d36 2d2d 2d2d 2d2b -----6-----+  
00000090: 372d 2b2d 2b2d 372d 2d37 ffc0 0011 0800 7-+-+7--7.....  
brucewayne@brucewayne:~/Desktop/g4g$
```

```
Applications ▾ Places ▾ Terminal ▾  
File Edit View Search Terminal Help  
brucewayne@brucewayne:~/Desktop$ xxd test.zip | head  
00000000: 504b 0304 0a03 0000 0000 19aa 934a 15ed PK.....J..  
00000010: ac18 cebc 0200 cebc 0200 0e00 0000 6372 .....cr  
00000020: 6f70 7065 646e 6577 2e70 6e67 8950 4e47 oppednew.png.PNG  
00000030: 0d0a 1a0a 0000 000d 4948 4452 0000 012c .....IHDR...  
00000040: 0000 01b3 0806 0000 0095 4d2a e300 0000 .....M*....  
00000050: 0970 4859 7300 000b 1300 000b 1301 009a .pHYs.....  
00000060: 9c18 0000 0a4d 6943 4350 5068 6f74 6f73 ....MiCCPPhotos  
00000070: 686f 7020 4943 4320 7072 6f66 696c 6500 hop ICC profile.  
00000080: 0078 da9d 5377 5893 f716 3edf f765 0f56 .x..SwX...>..e.V  
00000090: 42d8 f0b1 976c 8100 2223 ac08 c810 59a2 B...l..">#...Y.  
brucewayne@brucewayne:~/Desktop$
```

- L'**extension** (dans le filerecord) est utilisée par l'OS pour le choix de l'application
- Ce **Magic Number** est utilisé par l'application pour reconnaître le type de fichier et son format de données



DATA LOSS

→ Suppression de données

- Malencontreuse
- **Volontaire**

→ Pannes logicielles

- Corruption du partitionnement (MBR / GPT)
- Corruption de la MFT (NTFS) / Inode Table (EXT)

→ Pannes matérielles

- Secousses
- Choc (appareil qui tombe)

→ Chiffrement

- Clé de chiffrement perdue
- **Ransomware**



**GPT = GUID Partition Table et non pas
Generative Pre-trained Transformer**

Ransomware : ce que le cybercriminel va faire

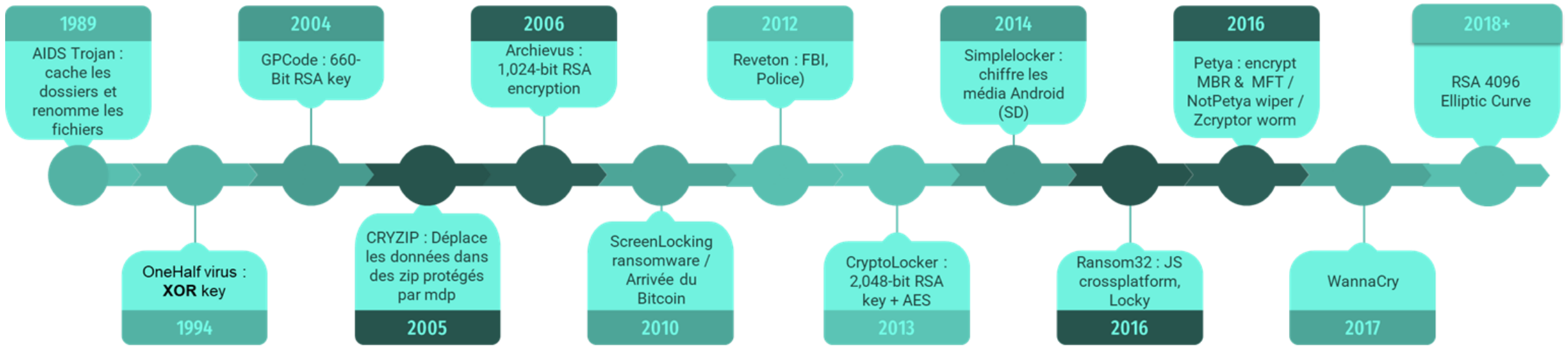
- **Définition** : « Forme d'extorsion imposée par un code malveillant sur un utilisateur du système »
- **Objectif** : Obtenir de la victime le paiement d'une rançon
- **Moyens** : Empêcher l'utilisateur d'accéder à ses données
- Les **cryptoransomwares** vont chiffrer les données
- 2 contraintes :
 1. Ne pas se faire détecter ou bloquer avant le chiffrement
 2. Donc aller vite



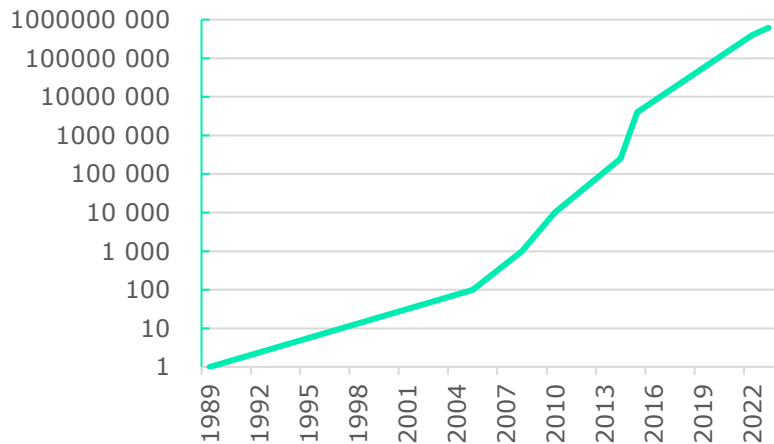
3 secondes avant de chiffrer
4 minutes pour chiffrer 50Go



L'histoire du Ransomware

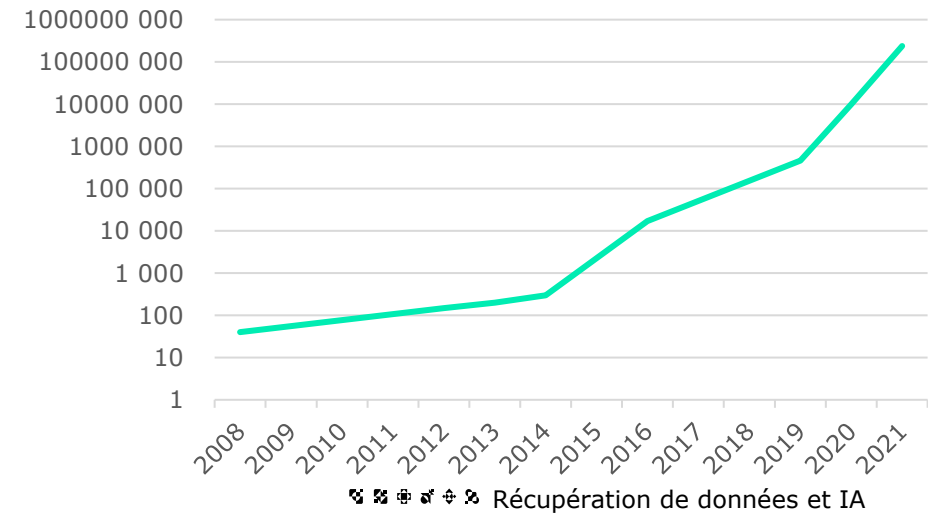


Nombre de Samples



- 623M attaques en 2021
- 500M en 2022
- 20 chaque seconde

Montant de la rançon (\$)



🔍 🛡️ 📊 📈 📉 📌 Récupération de données et IA

Quoi ? Que chiffrer ?

→ Sélectif

- uniquement les fichiers utilisateurs (.docx, .pdf, .xlsx, .jpg, ...)

→ Filtrant

- Tout sauf les fichiers critiques de l'OS (.dll, .sys, .exe, ...)

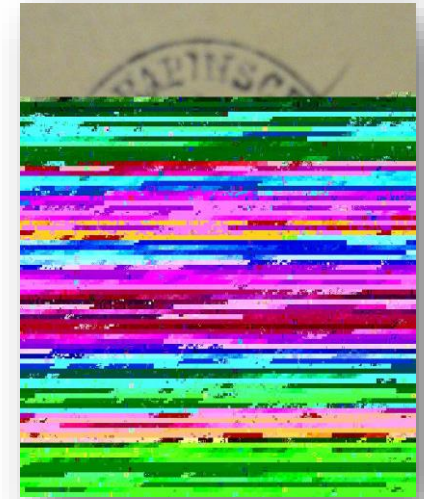
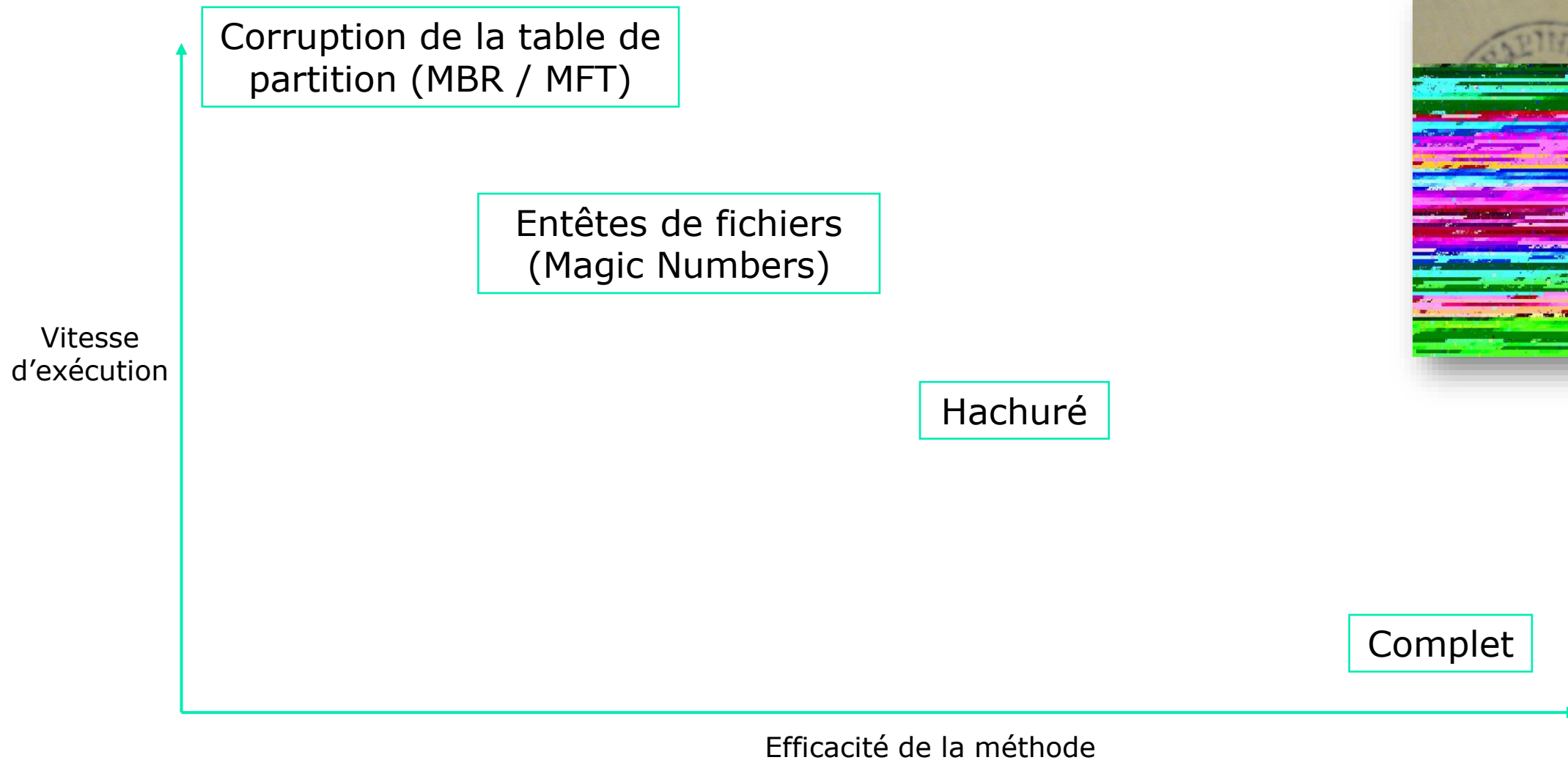
→ Conteneurs

- .vmdk (ESXI), .vbk (Veeam), ...

→ Total

En parallèle, l'attaquant va supprimer les autres opportunités de récupération telles que les VSS ou les backups.

Comment ?



Avec quoi ?

Type de cryptographie	Cryptographie symétrique	Cryptographie asymétrique	Cryptographie hybride
<i>Vitesse de chiffrement</i>	Très Rapide	Lent	Rapide
<i>Stockage de la clé</i>	En clair dans la RAM	Clé pour déchiffrer que côté attaquant (secure)	1 clé (sym) par fichier, chiffrée par une clé publique
<i>Utilisation de clés</i>	Même clé pour chiffrer et déchiffrer (secret partagé)	Différentes pour chiffrer et déchiffrer	Utilise RSA (asym) pour obtenir une clé AES (sym)
<i>Exemple d'algorithme</i>	AES	RSA, Diffie Hellman, ECDSA/ECDH	RSA puis AES
<i>Taille de clé</i>	256 bits	4096 bits (RSA)	
<i>Exemple de protocole</i>	Zip chiffré par mdp	PKE, EMV-CA	PGP/GPG, HTTPS (TLS)



- ❑ Pousser ses backups en production (quand on en a)
- ~~❑ Payer la rançon (NON!)~~
- ❑ Vérifier sur NoMoreRansom s'il n'existe pas un Decryptor (avoir de la chance)
- ❑ Espérer les clés de déchiffrement (être patient)
 - Attendre une saisie du FBI ou le génie d'un chercheur en cybersécurité
- ❑ Et sinon ...?



/!\ La seule préoccupation adressée ici est la récupération des données (**disponibilité**). Pas l'intégrité, la confidentialité (exfiltration) ni la traçabilité.

03.

RETROUVER SES DONNÉES

03.

Retrouver l'accès à ses données

3 options s'offrent à vous

01.

Le déchiffrement

02.

La restauration

03.

La récupération

1. Déchiffrement des données

Les opportunités de retrouver la clé

- Clé(s) stockée(s) dans un fichier sur le disque
 - Même temporairement
- Chiffrement en cours / Machine non éteinte
 - Possibilité de récupérer la clé (en RAM)
 - KO si cryptographie asymétrique
- Même clé utilisée pour plusieurs chiffrements
 - Théorème des restes chinois
 - *Ex: Harasom ransomware*
- Les déchiffreurs (Decryptors)
 - No More Ransom <https://www.nomoreransom.org/en/decryption-tools.html>
 - Utilise une faiblesse dans l'algorithme ou l'implémentation du ransomware



2. Restauration des données



- Les Points de restauration (OS)
 - Fonctionnalité **Backup and Restore / System Recovery**
 - Fonctionnalité **File History** (Windows File Versions)



- Les instantanés de machines virtuelles (VM snapshot)
 - Disques virtuels et état de la RAM à un instant t



- Les sauvegardes (backups)
 - **Sauvegarde hors site ou hors ligne :**
 - Stockée dans le cloud,
 - Bande LTO immuable
 - Stockage hors ligne (type disque dur externe) non accessible au moment de l'attaque
 - **Sauvegarde sur site**
 - Souvent supprimées par l'attaquant, ou chiffrées par le ransomware

→ **Volume Shadow Copy (VSS)**

- Fonctionnalité du FileSystem
- Mécanisme mis en place par Windows pour journaliser les modifications de fichiers
- « copy on write » snapshot
- Souvent effacées par les ransomwares
- Limites sur Windows 11 : seuls les fichiers Système critiques sont sauvegardés



Un Backup ne dispense pas d'une analyse forensique pour comprendre la chaine d'attaque

3. Récupération des données

Des opportunités

1. Clusters libérés mais donnée non effacée
2. Contenu pas entièrement chiffré par le ransomware
3. Entêtes reconnaissables (Magic Number) dans les espaces non-alloués et slackspace
4. ...

Le tout pour le tout

→ File Carving

« processus de réassemblage de fichiers à partir de fragments en l'absence de métadonnées du système de fichiers »

→ Possible sous condition

- Fragments récupérés pas toujours exploitables
- Le ransomware a-t-il été sélectif ou hachuré ?
- L'**espace non-alloué** a-t-il été réécrit ?
- La MBR ou la MFT sont-elles « décorrumpibles » ?
- Le **slackspace** est-il disponible et exploitable ?
- ...

3. Récupération des données

Le tout pour le tout

→ Fonctionne bien sur les containers

- .vmdk, .vbk, .vhdx

→ Plus facile si peu de hachures

- Fichiers écrits de manière linéaire donc impactés si > XMo
- Fichiers fragmentés (ex: .vhdx) : possède des tables d'information
- si compression invalide => c'est chiffré donc on « marque » les fragments touchés

→ Etude du format de données et de stockage nécessaire

- Base des blocs de données, versioning, déduplication, ...
- R&D Databack et partenariat éditeur (HP, DataCore, ...)

→ Reconstruction

1. Réassembler les fragments dans le bon ordre
 - Algorithme glouton, élagage $\alpha\beta$, Analyse séquentielle
 - Nombre élevé de permutations à essayer
2. Reconstruire les entêtes
3. Vérifier la cohérence

→ Les outils

- TestDisk / PhotoRec
- Scalpel (TSK)
- Bulk Extractor
- Custom tools (Databack)
- ...



Récapitulatif

01.

Le déchiffrement

02.

La restauration

03.

La récupération

 Donnée intègre et fraîche

Données pré-attaque
Systèmes sains

Dernière roue du carrosse

 **Possibilités faibles**
Problèmes de sécurité

Perte de quelques jours

Efficacité partielle (80-100%)
Prends du temps



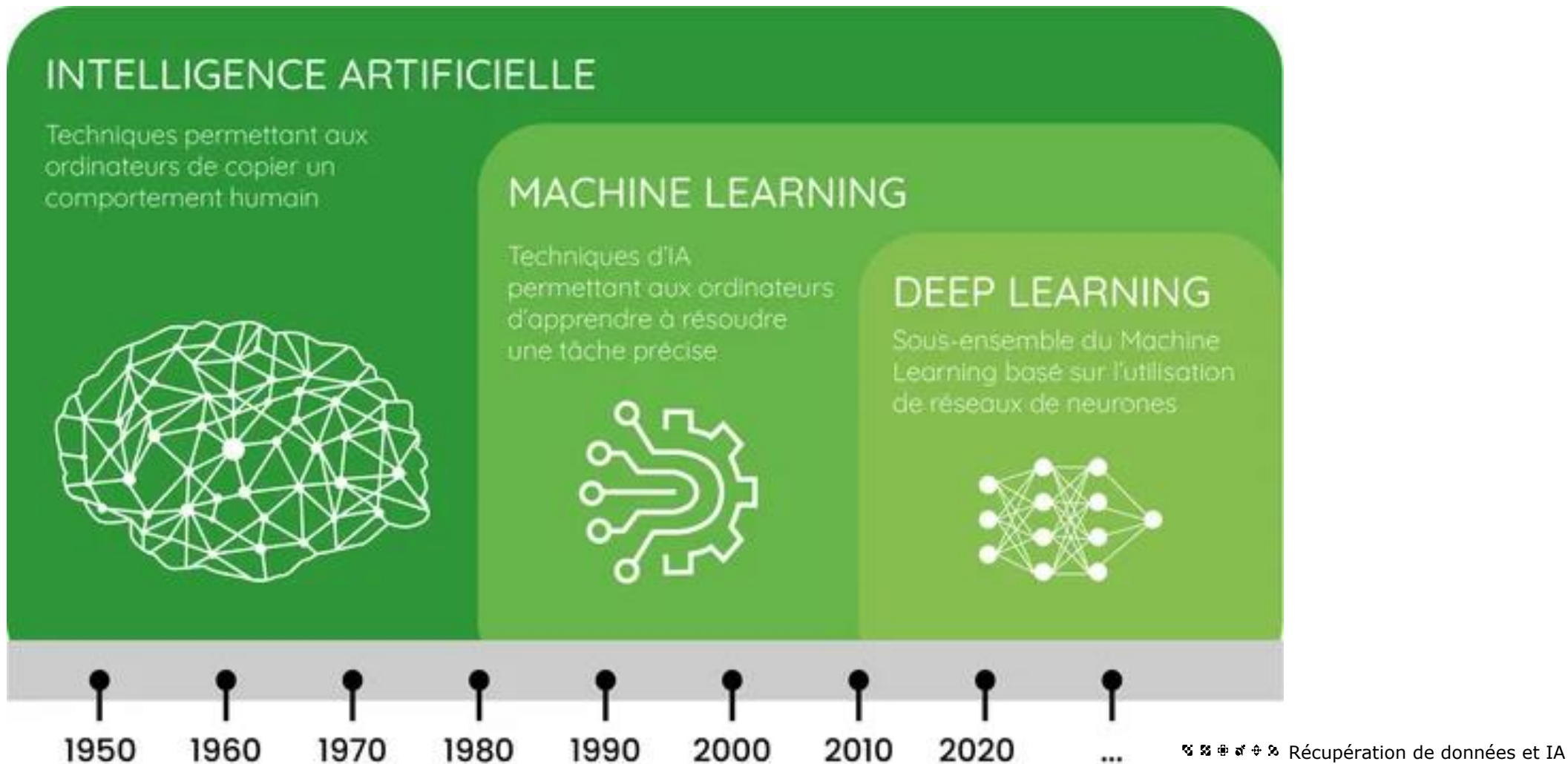
1. L'auteur des faits n'a pas laissé de traces de comment il a abimé le livre et où il a laissé la couverture
 1. Je n'avais pas les **clés de chiffrement**
2. Je n'avais pas de copie du livre même d'un autre éditeur
 1. Je n'avais pas de **backup**
3. Je n'avais pas les sous pour acheter un second exemplaire
 1. je n'ai pas payé **la rançon**
4. La valeur sentimentale du livre m'a poussé à investir de l'énergie (que vaut la donnée ?)
5. J'ai retrouvé les parties manquantes en fouillant dans le bazar du grenier
 1. J'ai fait du **carving**
6. J'ai pu remettre les pages dans le bon ordre en analysant leur contenu
 1. malgré l'absence des numéros (les Magic Numbers)
 2. J'ai pu **RECUPERER la donnée**
7. Enfin j'ai pu apprécier la lecture ! Mais ça m'a **pris du temps**

4.

L'IA PEUT-ELLE
NOUS AIDER ?

04.

1. Intelligence Artificielle : « Ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. »



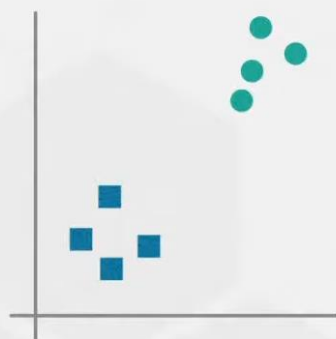
Apprentissage supervisé

Résultats rapides

Economique (compute)

Minimise l'erreur de prédiction

=> Se veut « exact »



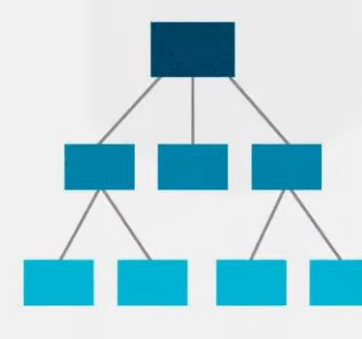
Binary
classification



Multi-class
classification



Regression
modeling



Ensembling

Apprentissage non-supervisé

Moins complexe

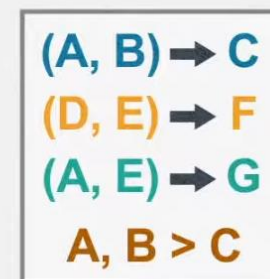
mais moins « exact » / précis



Clustering



Anomaly
detection



Association
mining



Dimensionality
reduction

Fonctionnalité

Crée du nouveau contenu

Analyse les données

Utilise l'apprentissage automatique

Simule les résultats

Industries utilisées dans

IA générative



Musique, mode, art

IA prédictive



Finances, commercialisation,
recherche, santé

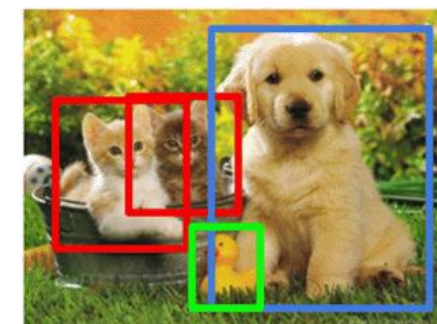


Classification



CAT

Object Detection



CAT, DOG, DUCK

L'IA dans la restauration de contenu



Original Photo

Restored Photo



photo restore AI

Images

Vidéos

Free

Actualités

Environ 59 900 000 résultats (0,34 secondes)



→ Détection et Classification

- Assister le carving : Analyse du disque et des structures / formats de données (partiellement chiffrés)
 - => Gain de temps R&D
- Faciliter la reconstruction d'entêtes : Reconnaissance du type de fichier / contenu
- Accélérer le traitement : Processus d'analyse très rapide (et rapide à entraîner => few-shot learning)

→ Classification du malware

- Signature, similarité (variants)
- Type d'actions sur le SI et le réseau (TTP) (Random Forest)
- Méthode et type de chiffrement des fichiers

→ Réorganisation

- Construction et/ou choix de l'algorithme de carving complexe et des heuristiques
 - Aide au développement d'outils pour chaque FS ou format de fichier
 - SmartCarving, BelkaCarving, fonctions de proximité
- Résolution de « puzzle » plus efficace qu'un algorithme classique
 - Livre sans table des matières et feuilles dans le mauvais ordre
- Réduction des erreurs de propagation (Parallel Unique Path)
- Extrapolation

→ Scalabilité

- Analyse de grandes quantités de données

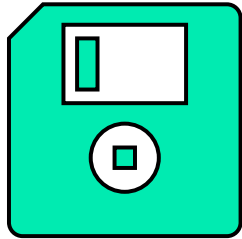
1. Aucune garantie de succès
2. Limité à de la donnée sémantique
3. Si totalement chiffré => Pas de magie
4. Espérer qu'une IA « casse » un algorithme de cryptographie est illusoire
 1. Casser RSA 4096 bit même avec l'IA et le quantique semble impossible (à date)

5.

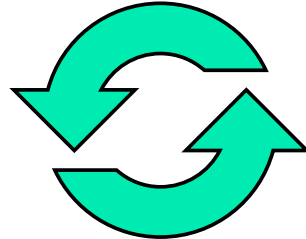
CONCLUSION

05.

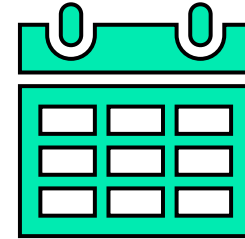
Appelez-nous



1. Priorité aux backups



2. Patcher, patcher, patcher



3. Garder la donnée chiffrée pour demain

Règle 3-2-1 :

- Au moins **3** copies de la donnée
- **2** lieux différents
- Dont **1** copie hors ligne ou immuable

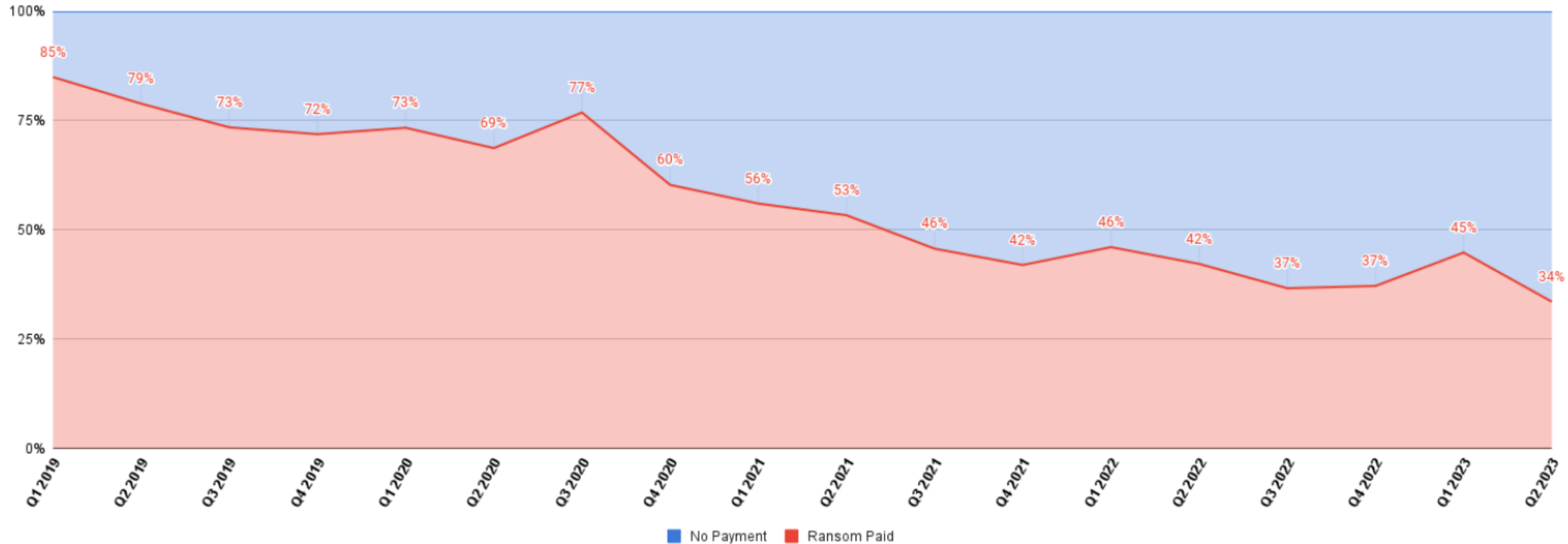


Ce qui est impossible aujourd'hui peut être trivial demain ;)

Takeaway : Ne payez pas la rançon !

Rejoignez les 66% d'entreprises malines

All Ransomware Payment Resolution Rates



- L'IA est un excellent outil, déjà indispensable
- Pirates toujours en avance et sans éthique ...
- Toujours un humain pour valider la « décision » finale
- Complexifie le monde mais peut nous aider dans beaucoup de domaines

« L'intelligence artificielle a piraté le système d'exploitation de notre civilisation. »

Yuval Noah Harari

DATABACK
RÉCUPÉRATION DE DONNÉES

inetum.
Positive digital flow

newsbridge 

 **AMOSSYS**

Almond

«L'intelligence ce n'est pas ce que l'on sait, mais ce que l'on fait quand on ne sait pas.»

Jean PIAGET



+33 (0)1 83 75 36 94



alerte@cwatch.almond.eu

Almond

MARCI

Tristan PINCEAUX

Head of CERT

tpinceaux@almond.eu

+33 6 99 20 49 73



+33 (0)1 83 75 36 94



alerte@cwatch.almond.eu

MOVE
FORWARD
WE'LL
WATCH
YOUR
BACK