

RÉSUMÉ THÉORIQUE - FILIÈRE INFRASTRUCTURE DIGITALE

M205 – Appréhender les méthodologies d'investigation numérique



85 heures

Concepteur : M. EL ANBAL

PARTIE 1 :

Maîtriser la technologie SIEM (SEM / SIM)

M205 – Appréhender les méthodes d'investigation numérique

Partie 1 : Maîtriser la technologie SIEM

- **Chapitre 1 : Connaître le fonctionnement de SIEM**
 - Définition de SIEM
 - Solutions du marché
- **Chapitre 2 : Concevoir les règles de corrélation**
 - Utilisation de SIEM pour la détection d'intrusion (Elastic)

Plan du cours

Durée : 40h

Partie 1 : Maîtriser la technologie SIEM

- **Chapitre 1 : Connaître le fonctionnement de SIEM**
 - Définition de SIEM
 - Solutions du marché
- **Chapitre 2 : Concevoir les règles de corrélation**
 - Utilisation de SIEM pour la détection d'intrusion (Elastic)

Plan du cours

Durée : 10h

Chapitre 1 : Connaître le fonctionnement de SIEM

Définition de SIEM

les SOC (Security Operations Center), une solution des entreprises pour lutter contre les attaques

- Pour identifier et investiguer des incidents de sécurité, de nombreuses entreprises comptent sur un ***Security Operations Center (SOC)***.
- Le SOC est une équipe dédiée à la **supervision de la sécurité du système d'information**.
- Pour ce faire, elle utilise des outils de monitoring, mais aussi des outils de collecte, d'intervention à distance et de corrélation d'événements.
- Elle recherche les signes d'un incident ou d'une compromission, par exemple des signaux faibles ou des comportements anormaux, afin de protéger le SI. Cette surveillance aide à la détection des événements de sécurité : intrusion, exécution de code non autorisé, exploits, élévation de privilèges, tentative d'accès à un compte admin, etc.
- Le SOC est donc un élément capital pour la sécurité des données de l'entreprise.

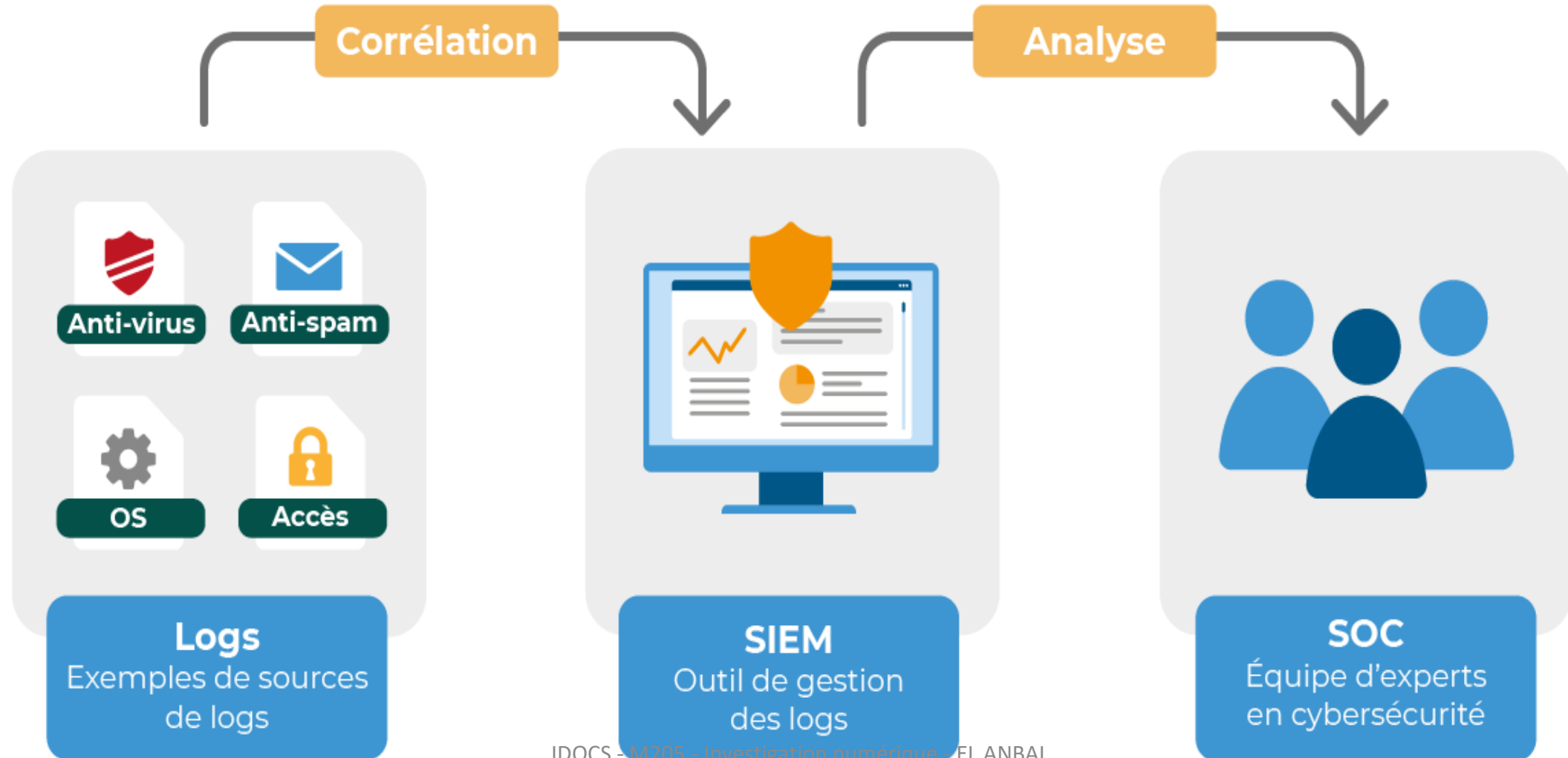
les SOC (Security Operations Center), une solution des entreprises pour lutter contre les attaques

- Pour gérer les alertes et détecter des intrusions, les équipes SOC utilisent un *Security Information Event Management (SIEM)*. C'est un des outils centraux pour monitorer la sécurité que nous verrons plus en détail dans les prochaines parties de ce cours.
- La mission principale d'un SOC est d'**identifier, analyser et remédier aux incidents de cybersécurité**. Cela, grâce au monitoring des différents équipements, mais aussi grâce aux méthodes d'analyse et de veille.

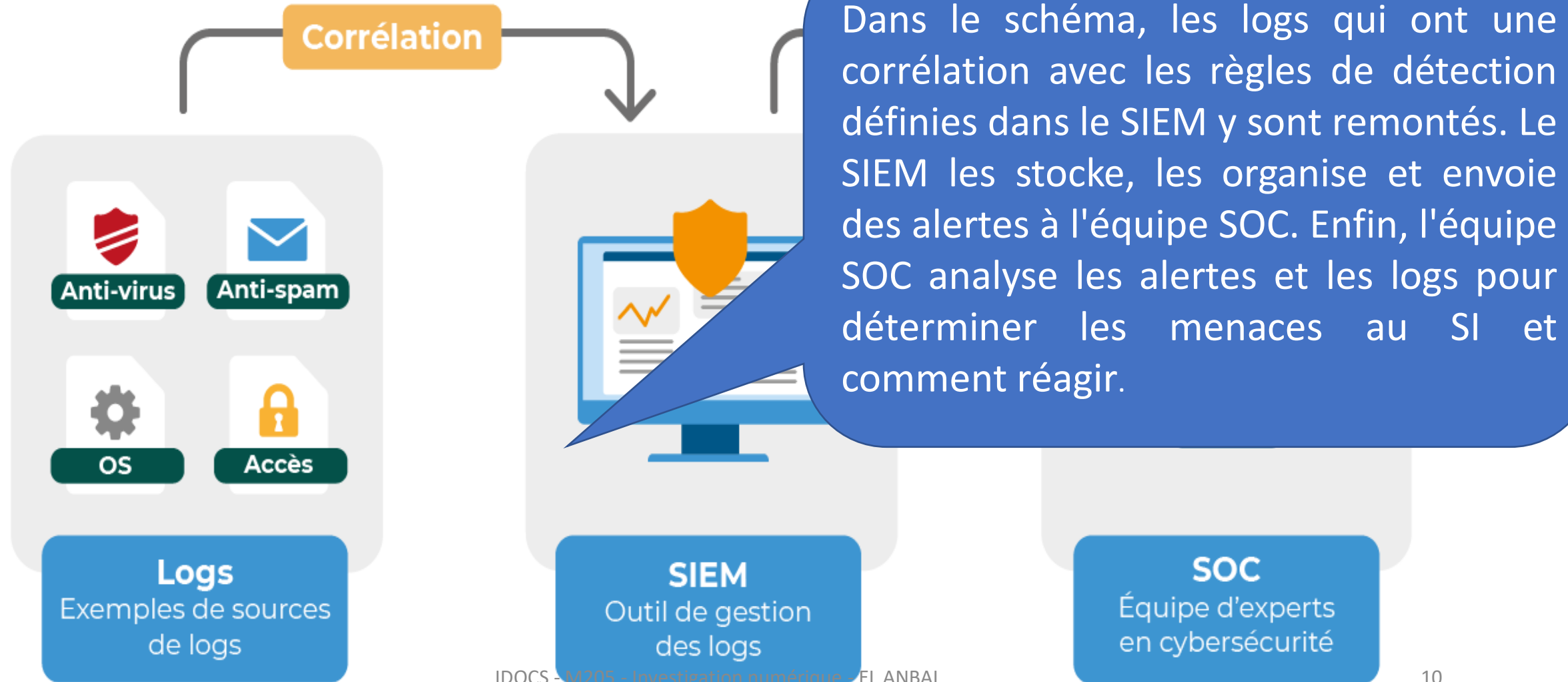
Comment installer un SOC ?

- La mise en place d'un SOC peut être compliquée et coûteuse. Toutefois, c'est un investissement à ne pas négliger pour être capable de **protéger les données** de l'entreprise, mais aussi pour **répondre rapidement** en cas de compromission. En effet, la centralisation des logs permettra des investigations plus organisées, dans le but de trouver les sources et les vecteurs de l'attaque en cas d'incident.
- La surveillance du SI en continu 24h/24 et 7j/7 permet donc de monitorer les activités réseaux, les machines, les serveurs, bref, tout élément connecté au SI.

Le schéma ci-dessous présente une architecture simplifiée type, avec comme point central le SIEM.



Le schéma ci-dessous présente une architecture simplifiée type, avec comme point central le SIEM



Récapitulation

- Une cyberattaque est composée de 7 étapes : *recon, weaponize, deliver, exploit, control, execute, maintain*.
- Les SOC veillent et administrent la sécurité du système d'information pour identifier, analyser les incidents de sécurité, et y remédier.
- Les bénéfices du SOC incluent la protection des données sensibles et la conformité aux règles de l'industrie.
- **Et convient aussi à déployer et mettre en place les solutions** permettant la détection et l'analyse, comme des serveurs de collecte de logs, des *Intrusion Prevention Systems (IPS)/Intrusion Detection Systems (IDS)*, des *Endpoint Detection and Response (EDR)*, un SIEM.

Maîtriser la technologie SIEM

- Le **SIEM (*Security Information and Event Management*)** est une approche du management de la sécurité. Le SIEM donne aux professionnels de la sécurité un aperçu et un historique des activités au sein de leur environnement informatique.
- Il génère des alertes basées sur l'analyse et la corrélation de plusieurs sources d'événements de sécurité.
- La technologie SIEM existe depuis plus d'une décennie. Elle a évolué initialement de la discipline de gestion des logs. Elle combine :
 - **la gestion des événements de sécurité (SEM)**, qui analyse les données des journaux et des événements en temps réel pour fournir une surveillance des menaces, la corrélation des événements et la réponse aux incidents ;
 - avec **la gestion des informations de sécurité** qui collecte, analyse et établit des rapports sur les données des journaux.

Fonctionnement d'un SIEM

- Le SIEM **collecte et agrège les données de journaux** générées dans tout votre système d'information, des applications aux périphériques réseau et de sécurité, tels que les pare-feux et les détections antimalware.
- Il identifie et catégorise ensuite les incidents et événements, et les analyse. Le logiciel répond à **deux objectifs principaux** :
 - **Fournir des rapports** sur les incidents et événements liés à la sécurité : les connexions réussies et échouées, l'activité des logiciels malveillants et d'autres activités malveillantes possibles.
 - **Envoyer des alertes** si l'analyse montre qu'une activité s'exécute sur des ensembles de règles prédéterminées, comme par exemple l'exécution d'un logiciel malveillant, et indique ainsi un problème de sécurité potentiel.

Chapitre 1 : Connaître le fonctionnement de SIEM

Solutions SIEM du marché

<https://gandalsmart.com/telecharger-installer-et-configurer-un-serveur-siem-pour-recueillir-et-stocker-les-logs-pour-le-monitoring-et-alertes-de-votre-entreprise/>

Comparaison des solutions SIEM du marché

- Vous avez le choix d'une multitude de SIEM sur le marché, des solutions libres ou open source à des solutions plus avancées (incluant des mécanismes de machine learning qui détectent les événements de sécurité, par exemple).
- Cette année, le Gartner a proposé [un tableau de comparaison des meilleurs SIEM du marché](#) (en anglais) pour l'aide à décision – je vous invite à y jeter un œil !

Le SIEM que nous allons utilisé : Elastic

- Dans ce cours, nous utiliserons la **solution open source ELK** (Elasticsearch, Logstash, Kibana) qui permet le monitoring de logs et la détection d'alerte. ELK est très intéressant, car il exploite une multitude de sources de données et les visualise de manière graphique.
- La suite ELK est composée de 4 outils, qui vous permettra de gérer vos logs dans l'ordre suivant :
 - **Beats**, pour l'envoi de logs en fonction des systèmes utilisés : Windows, Linux, logs réseau ;
 - **Logstash**, permettant de modifier les données envoyées dans ElasticSearch ;
 - **Elasticsearch**, la base de données principale pour le stockage des données ;
 - et enfin **Kibana**, l'interface graphique permettant de visualiser les données remontées au moyen de dashboards et l'envoi des alertes...

Remonte les logs sources



Permet l'agrégation des données



Centralise les logs, stockage et indexation



Recherche, visualisation et envoi des alertes



Récap

- nous venons de voir le fonctionnement d'un SIEM, qui a pour objectifs :
 - de fournir des rapports sur les incidents et événements liés à la sécurité ;
 - d'envoyer des alertes pour de probables problèmes de sécurité.
- Nous avons également vu le SIEM que nous déploierons dans la suite du cours : ELK. Il comprend Elasticsearch, Logstash, Kibana et Beats.

Partie 1 : Maîtriser la technologie SIEM

- **Chapitre 1 : Connaître le fonctionnement de SIEM**
 - Définition de SIEM
 - Solutions du marché
- **Chapitre 2 : Concevoir les règles de corrélation**
 - Découverte et manipulations manuelles des logs
 - Utilisation de SIEM pour la détection d'intrusion (Elastic)

Plan du cours

Durée : 15h

Chapitre 2 : Concevoir les règles de corrélation

Découvrir et manipuler les logs

Découvrir les logs

- Lorsque vous monitoriez le réseau d'une entreprise, il est nécessaire d'effectuer une collecte de logs pour **comprendre et identifier des événements sur le réseau**. Cette centralisation est primordiale dans le maintien opérationnel du système d'information.
- En informatique, un **log** est un fichier qui enregistre des événements qui se produisent sur un système d'exploitation ou tout autre équipement informatique, routeur, switch, serveur. La collecte de logs vous donnera des informations sur ce qui se passe sur votre S'.



Comprendre les logs avant de voir Elastic



- Les logs peuvent contenir des messages entre différents utilisateurs d'un logiciel de communication, par exemple, mais également des connexions réseau, des actions de création ou de suppression de fichier, des tentatives d'accès, etc.

Types de logs

Événements (Logs)	
Événements générés par les systèmes d'exploitation, les services et les applications	Les événements (Logs) des systèmes d'exploitation, des services et des applications (en particulier les événements liés à la sécurité) sont souvent d'une grande valeur pour la détection et le traitement d'un incident, telles que l'enregistrement des événements relatifs à l'accès aux comptes et les actions qui ont été réalisées. Les entités doivent mettre en place des références exigeant l'activation des logs sur tous les systèmes et surtout sur les systèmes critiques sans oublier les postes utilisateurs. Ces logs peuvent être analysés en utilisant des règles de corrélation. Une alerte peut être générée suite à cette analyse pour indiquer un incident.

Types de logs

Événements des équipements réseaux et FW	Les événements (Logs) générés par ces équipements identifient les connexions bloquées et aussi autorisées, même s'ils fournissent peu d'informations sur la nature de l'activité. Ils peuvent être utiles pour identifier les tendances du réseau et faire des analyses comportementales comme ils peuvent être corrélés avec d'autres événements détectés par d'autres sources.
NetFlow	Les routeurs, les switches et autres périphériques réseau peuvent fournir ces métadonnées relatives au protocole TCP/IP. Ces informations sur le flux réseau, peuvent être utilisées pour identifier des activités anormales provoquées par des logiciels malveillants, exfiltration de données, et d'autres actes de malveillance.

Emplacement	Contenu
/var/log/alternatives.log	Les logs d'update-alternatives.
/var/log/apache2/*	Les logs du serveur http apache2.
/var/log/apt/*	Les logs d'apt. Tous les paquets installés avec apt-get install, par exemple.
/var/log/aptitude	Les logs d'aptitude. Contient toutes les actions demandées, même les abandonnées.
/var/log/auth.log	Les informations d'autorisation de système. Y sont consignées toutes les connexions (réussies ou pas) et la méthode d'authentification utilisée.
/var/log/bind.log	Les logs du serveur de nom bind9, s'il sont activés.
/var/log/boot.log	Les informations enregistrées lors du démarrage du système. Ce fichier n'est pas activé par défaut.
/var/log/cron	Les informations sur les tâches cron. Enregistrement à chaque fois que le démon cron (ou anacron) commence une tâche.
/var/log/daemon.log	Les informations enregistrées par les différents daemons (processus) de fond qui fonctionnent sur le système.
/var/log/debug	Les logs de debugging.
/var/log/dmesg	Les messages du noyau Linux depuis le démarrage.
/var/log/dpkg.log	Les informations sur les paquets installés ou retirés en utilisant la commande dpkg.
/var/log/faillog	Les échecs de connexion. # faillog -u root.
/var/log/kern.log	Les informations enregistrées par le noyau. Utile pour déboguer un noyau personnalisé, par exemple.
/var/log/lastlog	Les informations de connexion récente de tous les utilisateurs. Ce n'est pas un fichier ascii. Vous devez utiliser la commande lastlog pour afficher le contenu de ce fichier.
/var/log/mail.*	Les informations du serveur de messagerie. Par exemple, sendmail enregistre des informations sur tous les éléments envoyés dans ces fichiers.
/var/log/messages	Les messages du système, y compris les messages qui sont enregistrés au démarrage. Beaucoup de choses sont enregistrées dans /var/log/ messages y compris le courrier, cron, daemon, kern, auth, etc.
/var/log/syslog	Tous les messages, hormis les connexions des utilisateurs. Plus complet que /var/log/messages.
/var/log/user.log	Les informations sur tous les journaux de niveau utilisateur.

Objectifs de gestion des logs



- La gestion de votre collecte de logs consistera à mettre en place la journalisation et la centralisation. La **journalisation** est la mise en place d'un système permettant la remontée automatique de logs. L'envoi de vos logs à un point central décrit la **centralisation**, qui vous donnera une vision globale des événements du système d'information.
- Cette gestion des logs a plusieurs objectifs :
 - **Obtenir un état général du SI** et identifier des événements anormaux.
 - **Détecter les intrusions**, par exemple au moyen de solution IPS/IDS ou de règle SIEM.
 - **Retracer l'historique** et les actions d'un attaquant dans le cadre d'une investigation forensic.
 - **Visualiser les actions du SI**, définir des statistiques et identifier les signaux faibles.
- Mais avant que vous définissiez une supervision et détectiez les événements de sécurité, il est nécessaire de savoir identifier les logs d'intérêt en fonction des systèmes. Cela va optimiser votre collecte de logs, et vous évitera la remontée de logs inutiles

Identifier les logs d'intérêt

- La première étape pour monitorer la sécurité est d'**identifier les logs d'intérêt qui permettent la détection d'événements suspects**. Par exemple, si vous recevez un log d'une connexion au serveur Active Directory en dehors des heures de travail, cela peut être **un comportement anormal auquel vous devez réagir** ! Le rôle du monitoring est justement d'identifier ce type d'événement.
- Les attaquants rivalisent d'ingéniosité pour éviter les détections. Il existe plusieurs techniques permettant de **contourner le logging**. Ils peuvent, par exemple, supprimer les journaux locaux, désactiver la remontée de logs, ou encore effectuer de l'injection de processus, c'est-à-dire camoufler un code malveillant dans un programme légitime. Soyez-en conscient lors de l'analyse des logs.
- Toutefois, vous n'êtes pas obligé de monitorer tout le réseau, car cela provoquerait une surconsommation des ressources du SI, et n'apporterait pas forcément de logs pertinents.



Quels sont les logs pertinents à monitorer ?

- La liste ci-dessous propose des **logs d'intérêt à monitorer** dans le cadre de la surveillance de sécurité du SI, également recommandée par l'ANSSI.
 - Authentification.
 - Gestion des comptes et des droits.
 - Accès aux ressources.
 - Modification des stratégies de sécurité.
 - Activité des processus.
 - Activité des systèmes.
- Cette liste est un bon point de départ. Ainsi, vous éviterez un nombre ingérable de logs dans votre collecte – mais n'hésitez pas à éventuellement ajouter des logs qui sont pertinents à votre SI, même s'ils ne figurent pas dans cette liste.
- Pour aller plus loin, l'ANSSI a mis à disposition un [guide de bonnes pratiques pour la mise en place de système de monitoring de logs](#).
- En cas d'incident de sécurité ou simplement dans le cadre de la surveillance journalière, il peut être intéressant d'explorer les logs manuellement, ce qu'on appelle également le **parsing de logs**. (voir TP 1 et TP2)



Chapitre 2 : Concevoir les règles de corrélation

Utilisation de SIEM pour détection d'intrusion (Elastic ; splunk)

La stack ELK

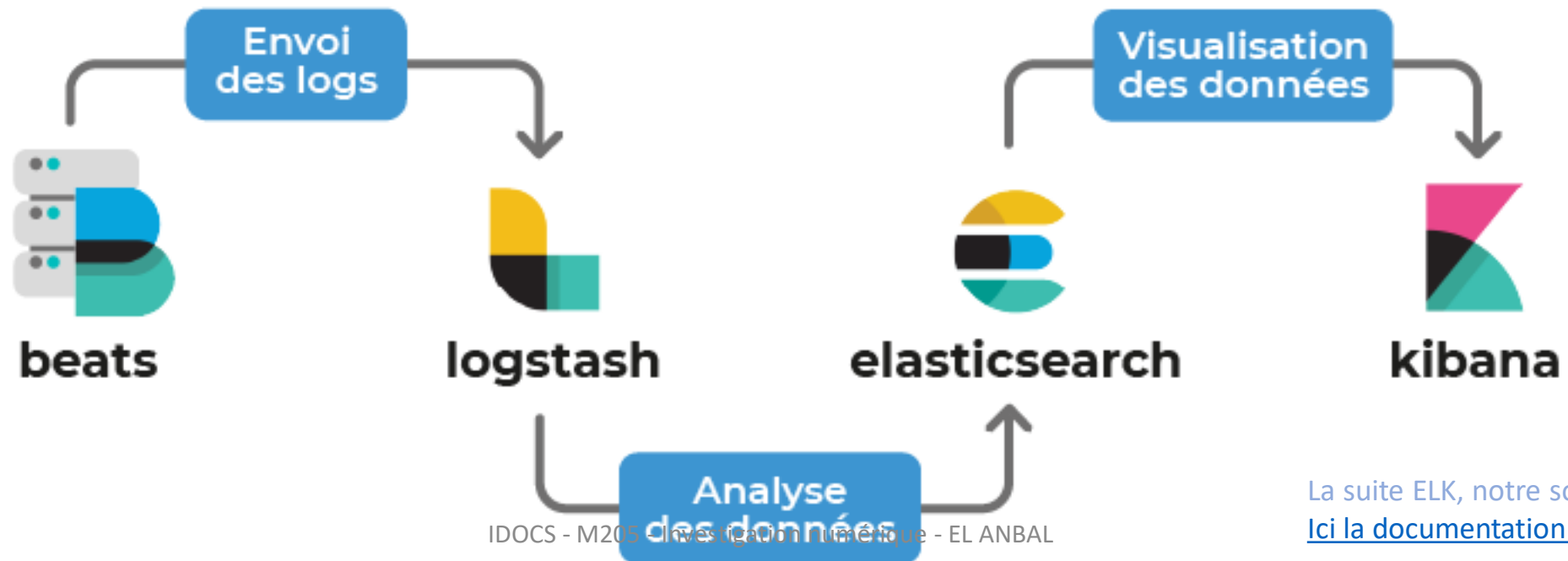
ELK est une suite open source comprenant 3 composants principaux : Elasticsearch, Logstash et Kibana. Beats a ensuite été ajouté pour former la stack ELK.

ELK permet l'indexation et l'analyse de données. Vous pourrez par exemple charger différents types de données, comme vos logs, et les visualiser sous forme de diagrammes personnalisés comme dans le screenshot ci-contre :



La stack ELK

- **Beats** : peut être installé sur les machines à monitorer en agent pour vous remonter les logs.
- **Logstash** : permet l'agrégation des données dans Elasticsearch.
- **Elasticsearch** : le composant principal, qui centralise les informations et y accède via une [API RESTful](#).
- **Kibana** : permet la création de dashboards et la visualisation des données dans ElasticSearch.



La suite ELK, notre solution SIEM
[Ici la documentation officielle.](#)

Elastic Stack

User Interface



Kibana

Store, Index & Analyze



Elasticsearch

Ingest



Logstash



Beats

les composants de la stack ELK

- **Remontez vos logs avec Beats**
- Beats est un ensemble d'outils permettant l'envoi de logs. Ces outils devront être installés sur les machines que vous souhaitez monitorer. Ils agiront comme des agents qui collectent les journaux d'événement et logs :
 - Filebeat : ingestion de fichiers de logs.
 - Packetbeat : ingestion de fichiers de capture réseau.
 - Auditbeat : ingestion de fichiers audit.
 - Heartbeat : vérification si un service est disponible ou non.
 - Functionbeat : monitoring des environnements cloud.
 - Journalbeat : ingestion des logs systemd.
 - Metricbeat : collection des métriques de différents systèmes.
 - Winlogbeat : collection de logs Windows.
- Pour installer les différents outils Beat, je vous invite à vous référer à [la documentation](#).



les composants de la stack ELK

- **Agrégez les données avec Logstash**
- Logstash permet l'agrégation et le formatage de données pour l'envoi dans Elasticsearch. Logstash vous permettra donc d'envoyer différents types de données autres que des logs.
- Pour en savoir plus et installer Logstash, vous pouvez vous référer à [la documentation officielle](#).



Les composants de la stack ELK

- **Centralisez vos données avec Elasticsearch**

- Elasticsearch est le moteur principal de la stack ELK : c'est lui qui va stocker les données et les rendre accessibles.
- Il peut être utilisé pour plusieurs objectifs, mais sa fonctionnalité principale est l'indexation de flux de données. Dans notre cas, cette fonctionnalité permet le stockage centralisé des logs, tels que des journaux ou des paquets réseau, par exemple. Il est donc très utile pour notre monitoring de la sécurité.
- ***Comment Elasticsearch stocke les données remontées ?***
- Elasticsearch stocke les données au format JSON. Ces données sont contenues dans des index qui sont des bases de données. Je vous conseille de créer un index par type de données ingérées (index1, index2...).

Les composants de la stack ELK

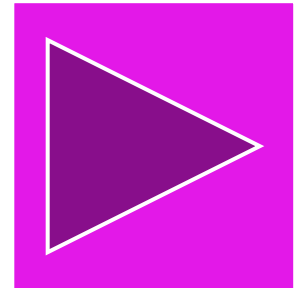
- **Centralisez vos données avec Elasticsearch**
 - *Et comment sont organisés ces index ?*
 - Les index contiennent des documents dans lesquels les données sont organisées dans des “fields”, c'est-à-dire des champs. Dans le screenshot à droite, nous pouvons voir que les fields sont définis dans la colonne de gauche.
 - Pour récupérer les données, Elasticsearch fonctionne comme une API RESTful.

Table	JSON
@timestamp	Dec 18, 2020 @ 17:46:17.102
_id	ME6-dnYBmsDm0F9auUWI
_index	winlogbeat-7.10.0-2020.12.15-000002
_score	-
_type	_doc
agent.ephemeral_id	d3146df2-d2ea-4818-979d-653c83ff1308
agent.hostname	DESKTOP-UDKLLH0
agent.id	587a526c-33d1-4da1-8518-6e4a1a432c5c
agent.name	DESKTOP-UDKLLH0
agent.type	winlogbeat
agent.version	7.10.0
ecs.version	1.5.0
event.action	Process Create (rule: ProcessCreate)
event.category	process
event.code	1
event.created	Dec 18, 2020 @ 17:46:18.464
event.kind	event
event.module	sysmon
event.provider	Microsoft-Windows-Sysmon
event.type	start, process_start
hash.imphash	30ad68b9dc9737d8c720dd9284051add
hash.md5	f5801470145fe1b446e98e7709311271

Les composants de la stack ELK

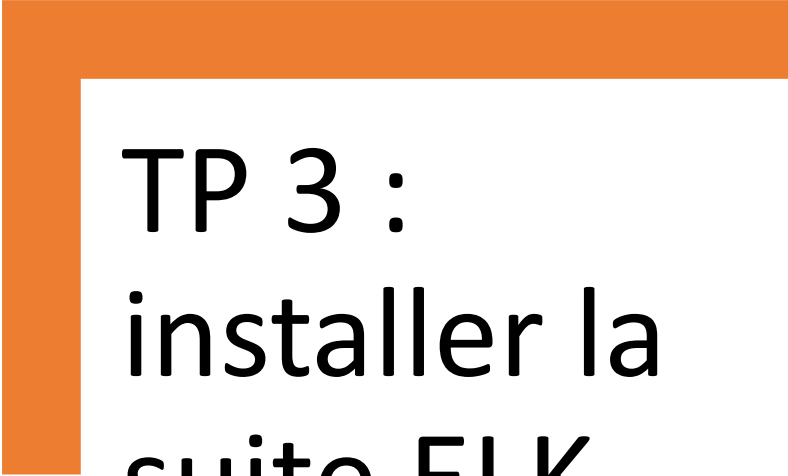


- **Visualisez vos données avec Kibana**
- Kibana va permettre de visualiser les données d'Elasticsearch en temps réel. Cet outil vous propose des dashboards préconfigurés pour analyser les logs qui vous sont remontés. Il est également possible de visualiser et d'explorer vos données dans la section Discover. Ceci est très utile pour visualiser vos logs et comprendre ce qu'ils contiennent. Cela vous permettra de mettre en place vos règles de SIEM plus facilement.
- Dans le screencast ci-dessous, découvrons cette fonctionnalité et la stack ELK :



En résumé : la suite ELK est composée de 4 outils :

- Beats, installé sur les machines à monitorer pour remonter les logs ;
- Logstash, qui reçoit et agrège les logs dans Elasticsearch ;
- Elasticsearch, le composant principal qui centralise les données ;
- Kibana, où l'on peut visualiser les données avec des dashboards.



TP 3 : installer la suite ELK

<https://www.digitalocean.com/community/tutorials/how-to-install-elasticsearch-logstash-and-kibana-elastic-stack-on-ubuntu-20-04-fr>

Liens vers des démonstrations

- Comparatif des Solutions SIEM : <https://www.gartner.com/reviews/market/security-information-event-management>
- Démo de **Splunk** en Ar : https://www.youtube.com/watch?v=xmo9ZZnLRIM&ab_channel=DalalAlharthi%D8%AF%D9%84%D8%A7%D9%84%D8%A7%D9%84%D8%AD%D8%A7%D8%B1%D8%AB%D9%8A
- Démo de **Splunk** en Fr : <https://www.youtube.com/watch?v=mmTPupSUInM>
- Alien Vault Ossim : <https://gandalsmart.com/telecharger-installer-et-configurer-un-serveur-siem-pour-recueillir-et-stocker-les-logs-pour-le-monitoring-et-alertes-de-votre-entreprise/>
- Quick start : <https://www.elastic.co/guide/en/kibana/8.6/get-started.html>

PARTIE 2 :

Appréhender les Méthodologies d'investigation réseaux et systèmes

M205 – Appréhender les méthodes d'investigation numérique
MH : 90h

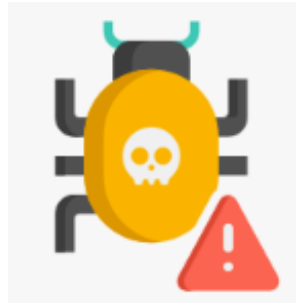
Formatrice : M. EL ANBAL

Partie 2 : Appréhender les Méthodologies d'investigation réseaux et systèmes

- **Préambule :**
 - Méthodes d'infection par un malware
 - Signes d'infection directs et indirects
- **Chapitre 1 : Processus d'investigation numérique**
 - Définition / vocabulaire
 - Etapes du processus
 - Activités d'investigation
 - Avantages et inconvénients
 - Cas d'utilisation et types d'investigation
- **Chapitre 2 : Répertoirer les indices de compromission**

Plan du cours

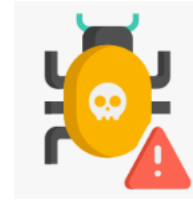
Durée : 25h



Préambule

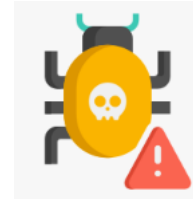
Introduction au Digital forensic

Méthodes d'infection



Avant d'entamer les concepts de l'investigation numériques ,nous allons d'abord faire un rappel sur quelques méthodes utilisées par les attaquants permettant d'infecter une machine.

Méthodes d'infection



La faille informatique

- c'est une vulnérabilité au sein d'un logiciel ou équipement informatique permettant à un pirate d'infecter une machine en essayant d'installer un malware.
- Les vulnérabilités les plus dangereuses sont les vulnérabilités zero-day car le fournisseur de l'outil ou du logiciel infecté n'a pas encore le correctif de la faille.



La faille humaine

- fait partie de la catégorie des attaques par l'ingénierie sociale qui exploite le manque d'expérience des personnes pour réussir à infecter une machine (exemple: envoi d'un lien malicieux par mail, le faux antivirus, etc.). C'est l'une des méthodes la plus efficace permettant à un attaquant de réussir un piratage de la machine cible. On peut parler aussi de la navigation web à travers le téléchargement d'un malware de façon non intentionnelle.



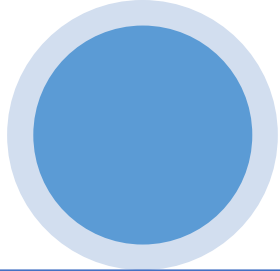
Les clés USB

- survient si nous branchons une clé USB infectée. Les attaquants peuvent même jeter des clés USB infectées près des entreprises cibles. Si un employeur trouve une de ces clés et la branche sur son ordinateur professionnel, le malware peut même se propager sur le réseau de l'entreprise. Certaines entreprises bloquent les ports physique des clés USB.



Le tiers de confiance (Supply Chain Attack)

- l'attaquant dans ce cas ne va pas attaquer l'entreprise cible directement mais en passant par une autre entreprise ou un client de confiance. Exemple: attaquer une entreprise "A" qui vend un outil à l'entreprise "B" cible. Une fois l'entreprise "A" lance des mises à jour logiciel de l'outil infecté par un malware sans correctif, tous les clients de l'entreprise "A" vont télécharger le malware à leur insu.



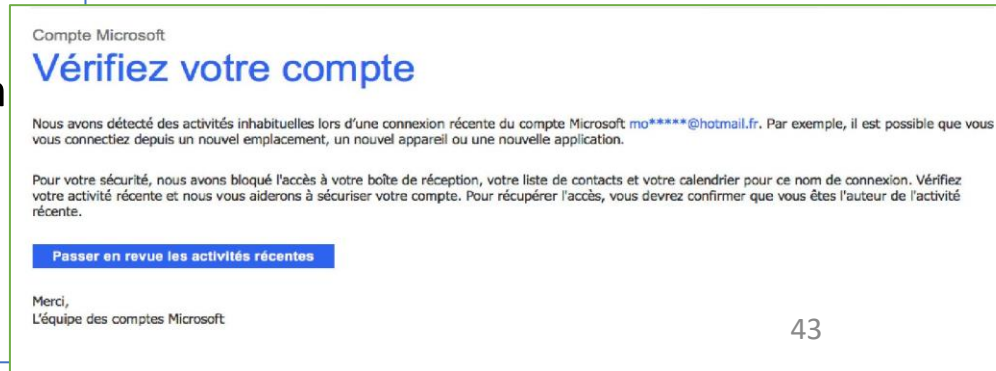
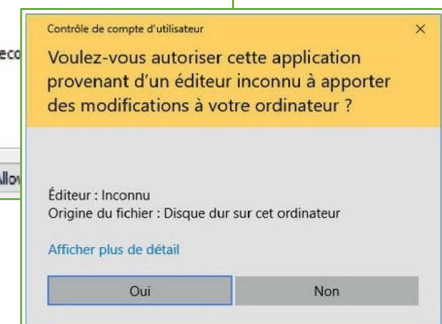
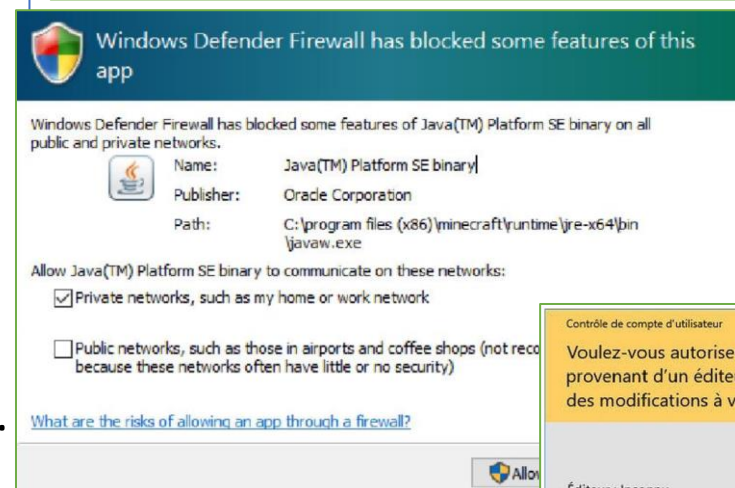
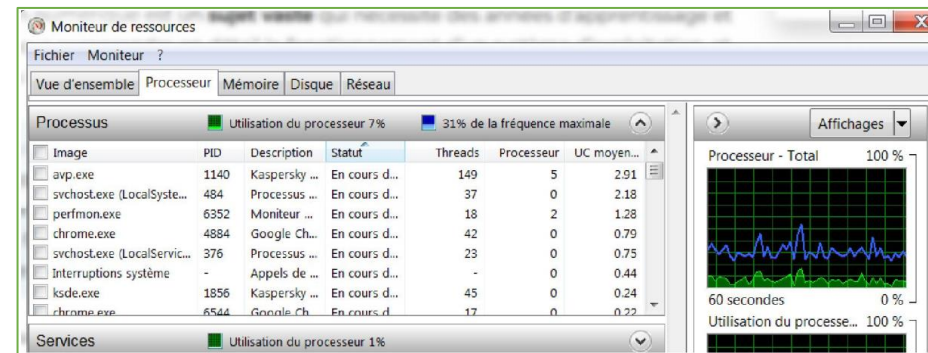
Signes d'une infection par un malware :

Les signes indirects

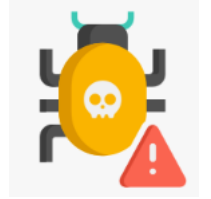
- Ralentissement des performances de la machine
- Messages d'avertissement
- Messages d'alerte
- La fenêtre noire
- La publicité sur votre machine

Les signes Directs

- L'impossibilité d'accéder à l'un ou à plusieurs de vos comptes en ligne.
- Le curseur de votre ordinateur qui se déplace sans aucune action de votre part.
- Chiffrement de vos fichiers (dans le cas d'une attaque de type Ransomwares).
- Utilisation de votre carte bancaire pour effectuer des achats en ligne (en utilisant par exemple un logiciel de type Spyware).
- Modification de vos dossiers et/ou vos fichiers.



Introduire l'analyse forensic pour trouver le malware



- Un incident de sécurité peut survenir sur **n'importe quel type de système** : Windows, Linux, Mac, Android, iOS... Il existe des malwares pour toutes ces plateformes.
- Il existe toutefois une différence notoire : un malware destiné à infecter un système Windows **ne pourra pas infecter un système MacOS**, et vice-versa.
- Afin de trouver le programme malveillant, une analyse forensic peut être effectuée sur une machine infectée.



Processus d'investigation numérique

Chapitre 1

Etat de l'art de l'investigation numérique (Vocabulaire)

À propos de l'investigation numérique ...

Le terme anglais "Forensic" = "Légal" en Fr ;

-> Informatique légale, investigation numérique légale / informatique judiciaire ou criminalistique.

Application de **méthodes, techniques et protocoles** d'investigation numérique respectant des **procédures légales** et destinés à apporter des **preuves** numériques.

En informatique, on parle de "Digital Forensic" ou "Investigation numérique",





Définition

L'analyse forensic ou investigation numérique consiste à analyser une machine infectée afin de comprendre ce qui s'est passé dans cette machine et d'en préparer un rapport avec des conclusions de l'analyse et des recommandations.

L'analyse forensic se base sur un ensemble de techniques et d'outils de recherche, de collecte et d'analyse de données permettant de trouver et d'analyser des informations (traces) laissées sur une machine.

L'objectif principal est de comprendre d'où vient une attaque ainsi que le mode de fonctionnement de cette dernière.

L'analyse forensic



Elle se base sur 2 points principaux :

La recherche des traces sur Internet: dans ce cas il faut étudier un site malveillant en essayant de retrouver le propriétaire du site ou des traces comme son adresse mail, sa position géographique, etc.

La recherche des traces en locale: consiste à faire une investigation numérique sur une machine en cherchant sur les fichiers logs Windows, par exemple: trouver la trace des fichiers supprimés. On essaie de reproduire le "timeline" de l'utilisation de la machine concernée.

Elle peut aussi être utilisée dans plusieurs activités:

L'analyse forensic judiciaire.

L'analyse forensic en réponse à un incident de sécurité.

L'analyse forensic scientifique ou ingénierie inversée.



Objectifs de l'investigation numérique

- ❶ Elle permet de récupérer, d'analyser et de préserver les ordinateurs et les matériels connexes de manière à aider l'agence d'investigation à les présenter comme preuves devant un tribunal.
- ❷ Elle aide à déterminer le motif du crime et l'identité du principal coupable.
- ❸ Conception de procédures sur une scène de crime présumée qui vous aide à garantir que les preuves numériques obtenues ne sont pas corrompues.
- ❹ Acquisition et duplication de données : Récupération de fichiers et de partitions supprimés sur des supports numériques afin d'en extraire les preuves et de les valider.
- ❺ Cela vous aide à identifier rapidement les preuves et vous permet également d'estimer l'impact potentiel de l'activité malveillante sur la victime.
- ❻ Produire un rapport d'expertise informatique qui offre un rapport complet sur le processus d'investigation.
- ❼ Préserver les preuves en suivant la chaîne de possession.



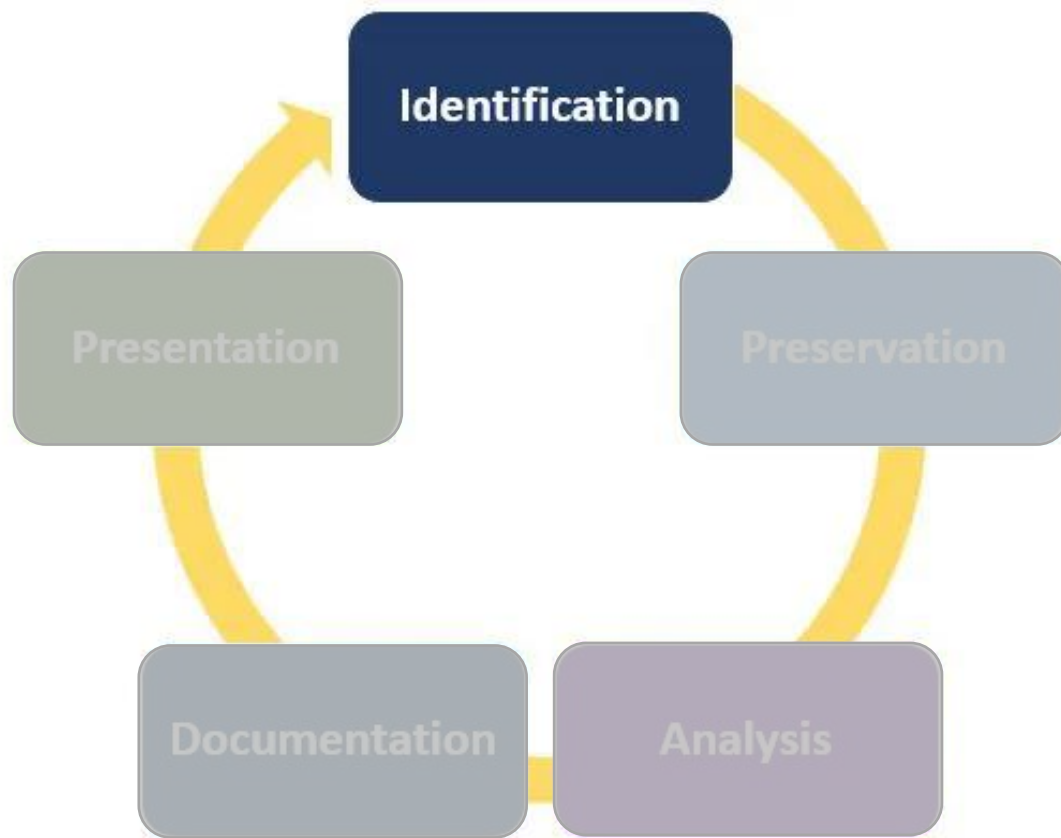
Processus de l'investigation numérique

- La criminalistique numérique comprend les étapes suivantes :





Processus du digital forensic



• Identification

- Il s'agit de la première étape du processus forensic.
- Le processus d'identification comprend principalement des éléments tels que les preuves présentes, l'endroit où elles sont stockées et, enfin, la manière dont elles sont stockées (dans quel format).
- Les supports de stockage électroniques peuvent être des ordinateurs personnels, des téléphones portables, des tablettes, etc.



Processus du digital forensic

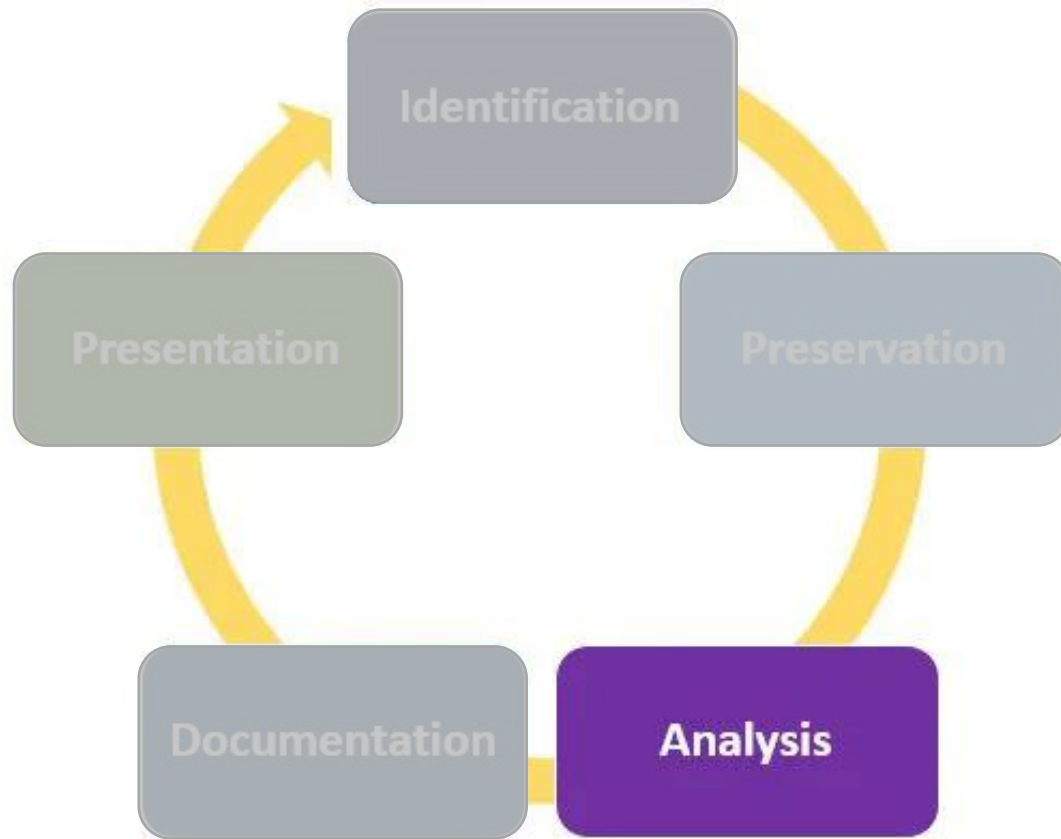


- **Préservation**

- Dans cette phase, les données sont isolées, sécurisées et préservées.
- Il s'agit notamment d'empêcher les personnes d'utiliser le dispositif numérique afin que les preuves numériques ne soient pas altérées.



Processus du digital forensic

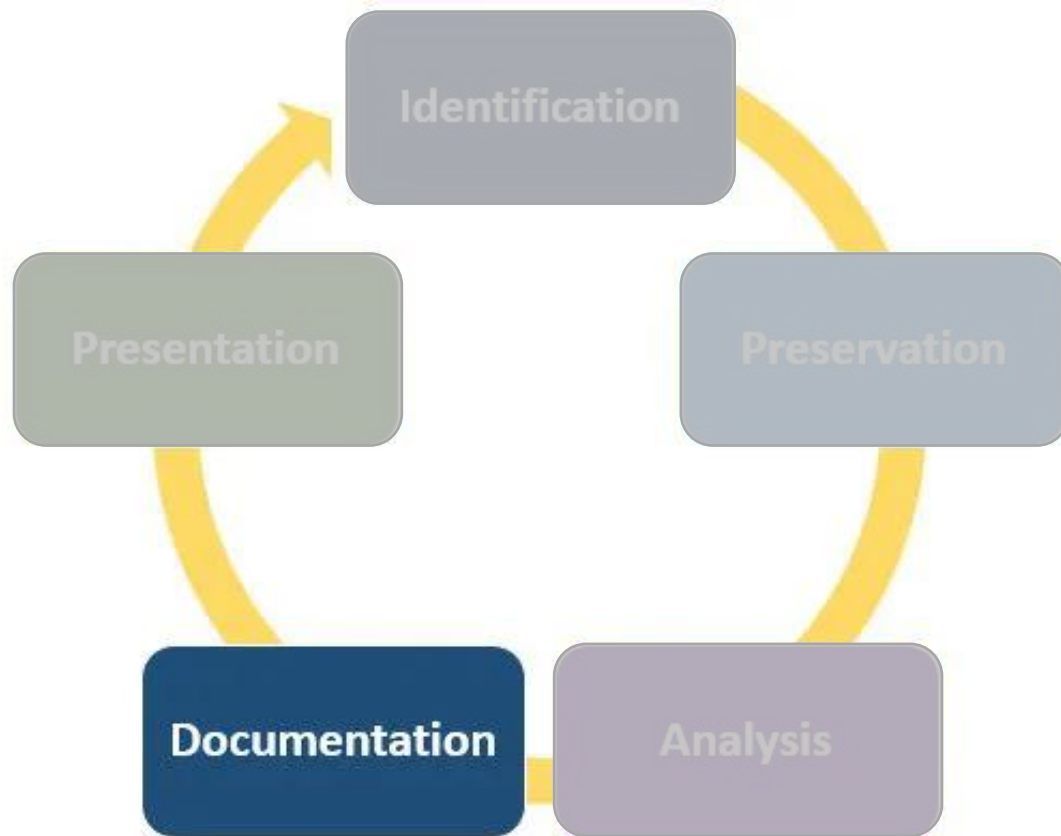


- **Analyse**

- Au cours de cette étape, les agents d'enquête reconstituent des fragments de données et tirent des conclusions sur la base des preuves trouvées.
- Toutefois, de nombreuses itérations d'examen peuvent être nécessaires pour étayer une théorie criminelle spécifique.



Processus du digital forensic



- **Documentation**

- Au cours de ce processus, un enregistrement de toutes les données visibles doit être créé. Il permet de recréer la scène de crime et de l'examiner.
- Elle implique une documentation appropriée de la scène de crime, ainsi que la prise de photographies, la réalisation de croquis et la cartographie de la scène de crime.



Processus du digital forensic



- **Présentation**

- Cette dernière étape consiste à résumer et à expliquer les conclusions.
- Cependant, elle doit être rédigée en termes simples en utilisant des terminologies abstraites.
- Toutes les terminologies abstraites doivent faire référence aux détails spécifiques.

Récapitulation



Identification

- Identify the purpose of investigation
- Identify the resources required

Preservation

- Data is isolate, secure and preserve

Analysis

- Identify tool and techniques to use
- Process data
- Interpret analysis results

Documentation

- Documentation of the crime scene along with photographing, sketching, and crime-scene mapping

Presentation

- Process of summarization and explanation of conclusions is done with the help to gather facts.



Exemples d'utilisation (incidents)

- Ces derniers temps, les organisations commerciales ont eu recours à l'investigation numérique dans les cas suivants :

Vol de propriété
intellectuelle

Espionnage
industriel

Conflits de travail

Enquêtes sur les
fraudes

Utilisation inappropriée
de l'Internet et du
courrier électronique sur
le lieu de travail

Enquêtes sur les
faillites

Questions relatives
à la conformité
réglementaire



Avantages de la criminalistique numérique

les avantages
et les
bénéfices de
l'investigation
numérique

- Assurer l'intégrité du système informatique.
- Produire des preuves au tribunal, ce qui peut conduire à la punition du coupable.
- Elle aide les entreprises à saisir des informations importantes si leurs systèmes ou réseaux informatiques sont compromis.
- Elle permet de traquer efficacement les cybercriminels, où qu'ils se trouvent dans le monde.
- Aide à protéger l'argent et le temps précieux de l'entreprise.
- Permet d'extraire, de traiter et d'interpréter les preuves factuelles, afin de prouver les actions des cybercriminels devant les tribunaux.

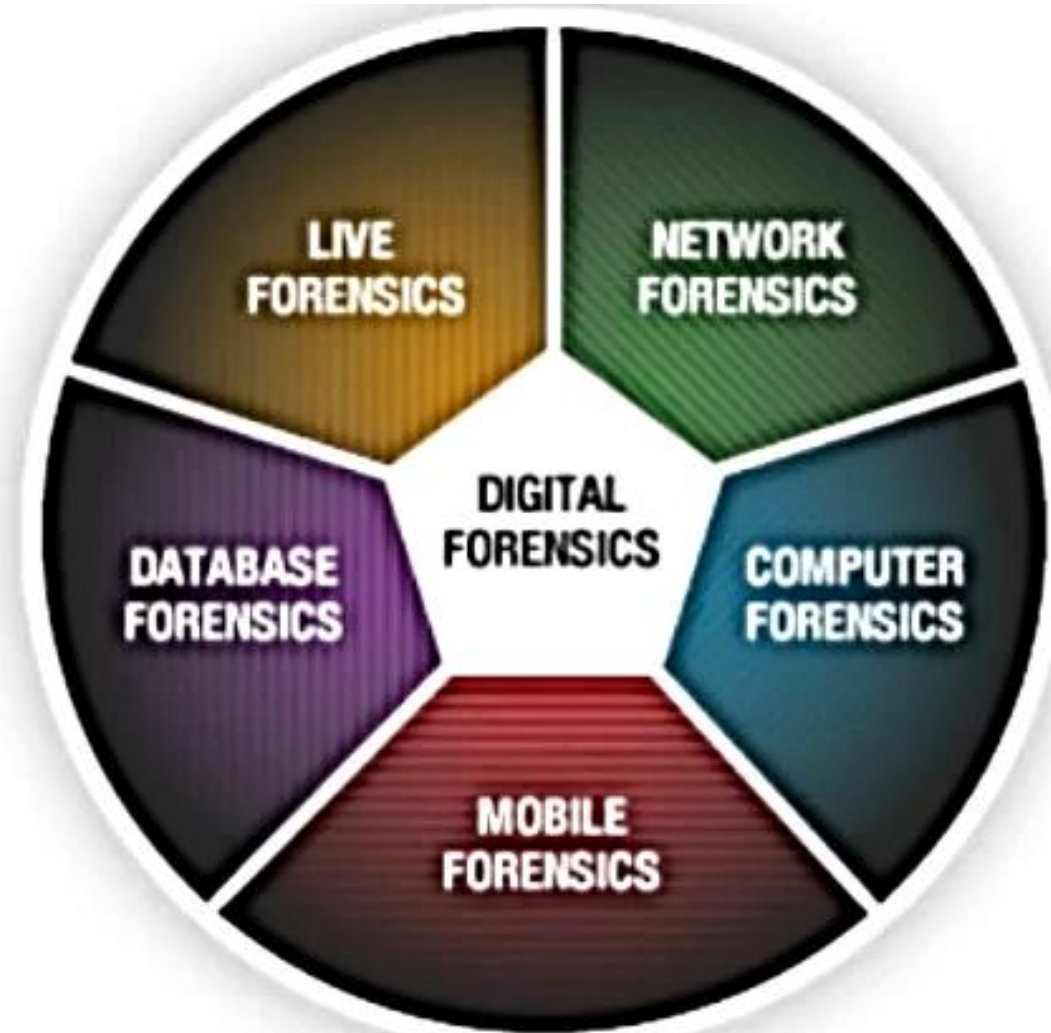
Inconvénients de la criminalistique numérique



Voici les principaux
coûts/inconvénients
de l'utilisation de la
criminalistique
numérique

- Les preuves numériques sont acceptées par les tribunaux. Cependant, il faut prouver qu'il n'y a pas eu de falsification.
- La production de documents électroniques et leur stockage sont extrêmement coûteux.
- Les praticiens du droit doivent avoir des connaissances approfondies en informatique.
- Il faut produire des preuves authentiques et convaincantes.
- Si l'outil utilisé pour la criminalistique numérique n'est pas conforme aux normes spécifiées, la preuve peut être désapprouvée par la justice.
- Le manque de connaissances techniques de l'enquêteur peut ne pas donner le résultat escompté.

Les disciplines/types d'investigation





Types de criminalistique numérique

Il existe quelques types de criminalistique numérique, dont les suivants :



Disk Forensics :

- Il s'agit d'obtenir des preuves à partir de supports de stockage numériques tels que les dispositifs USB, les DVD, les CD, etc., en rassemblant les fichiers actifs ou en modifiant ou supprimant les fichiers.



Network Forensics :

- Il s'agit généralement d'une sous-partie de la criminalistique numérique relative à la surveillance et à la détection du trafic réseau du système afin d'extraire des données cruciales pour toutes les preuves légales à présenter au tribunal.



Wireless Forensics :

- Il s'agit d'une partie de la criminalistique des réseaux qui vise à fournir les outils nécessaires à la collecte et à l'extraction de preuves à partir du trafic réseau sans fil.



Database forensics:

- Il s'agit d'un type de criminalistique numérique qui concerne l'étude et la collecte de bases de données et de leurs métadonnées pertinentes. Elle suit des techniques d'investigation pour interroger la base de données et recueillir des preuves.



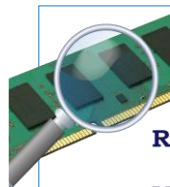
Malware Forensics :

- Cette branche de la criminalistique s'occupe de l'identification des codes malveillants et de l'étude de leurs problèmes liés à leur charge de travail, aux chevaux de Troie, aux virus, etc.



Email Forensics :

- Cette branche de l'informatique légale s'occupe de la récupération des données effacées et analyse le contenu des courriels, y compris les courriels qui ont été supprimés ou le calendrier ou les contacts dans le courriel.



Analyse de la mémoire :

- Il s'agit d'une analyse médico-légale qui recueille les données de la mémoire cache ou de la mémoire vive de l'ordinateur, puis rassemble les preuves à partir de cette mémoire.

Partie 2 :
Appréhender
les
Méthodologies
d'investigation
réseaux et
systèmes

- **Chapitre 2 :Répertoirer les indices de compromission**
 - ***Utilisation de Mitre Att&ck***
 - Les types d'IoC
 - Répertoirer les IoC
 - Les tactiques de mitre Att&ck
 - ***Analyses de données numériques***
 - Méthodologie d'analyse
 - Données volatiles et non volatiles
 - Analyse forensic du dump mémoire
 - Analyse forensic du disque dur

Plan du cours

Durée : 45h



Répertoirer les indicateurs de compromission

Chapitre 2

1- Utilisation de Mitre ATT&ck

Les indicateurs de compromission

- ou *IoC* sont des **éléments découverts pendant l'analyse** qui permettront de caractériser l'incident et d'identifier les menaces.
- Les indicateurs de compromission peuvent être de **plusieurs types** ; découvrons-les ensemble afin de mieux les répertorier.

Les types IoC

- Les **IoC réseaux** peuvent faire référence à une **activité réseau malveillante**, telle que la connexion à un serveur de commande et de contrôle. La connaissance de ces indicateurs permettra très rapidement de bloquer ces connexions pour contenir une menace sur tout un réseau. En analysant les trames réseau générées par un logiciel malveillant, il sera également possible de créer des **règles IDS pour détecter de manière proactive** une intrusion ou une connexion malveillante.
- Les **noms de domaines et URL** font également partie de ces IoC.

Les types IoC

- Les **hash de fichiers**
 - Les hash sont des **empreintes uniques** qui identifient des fichiers. Ces hash sont des IoC qui permettront d'identifier les fichiers malveillants sur une machine. Il peut également être utile d'identifier les répertoires d'écriture de ces fichiers sur le système.
- Les **adresses emails**
 - Les **adresses emails délivrant du spam, du phishing** ou autre sont également des indicateurs de compromission qui pourront être bloqués sur les passerelles de messagerie.
- Les **actions sur le système**
 - Les **actions menées sur le système** par un logiciel malveillant peuvent également être des IoC. Par exemple, si le malware crée un service Windows en particulier, ou une clé registre.

Répertoire les IoC

- La société Mandiant a créé un format **OpenIOC** basé sur le format XML, pour stocker et gérer ses indicateurs de compromission. Il existe un outil permettant la génération de fichiers IoC, appelé [IoC-Editor, disponible ici](#).
- L'utilitaire permet de créer son propre fichier d'IoC. Il suffit de faire un clic droit puis d'ajouter un item supplémentaire. Il est ensuite possible d'ajouter un IoC en fonction de ses besoins. Ici, nous ajoutons un IoC en rapport avec l'adresse email.



Liste des IoC

[illegible]

Découvrez la matrice ATT&CK

- La matrice ATT&CK est une initiative du MITRE qui permet d'identifier les tactiques et les techniques utilisées par un attaquant. Elle fournit un ensemble de classifications permettant d'identifier les différentes phases d'une attaque, ainsi que les techniques utilisées pour chaque phase.
- Les objectifs d'utilisation de la matrice pour votre SI sont donc multiples :
 - effectuer une analyse de vos moyens de détection et identifier les faiblesses en amont ;
 - améliorer vos détections des menaces et votre méthodologie d'investigation ;
 - tester vos règles de détection pour vous assurer que vous êtes alerté comme prévu.

APT28

APT28 is a threat group that has been attributed to Russia's General Staff Main Intelligence Directorate (GRU) 85th Main Special Service Center (GTsSS) military unit 26165.^[1] This group has been active since at least 2004.^{[2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12]}

APT28 reportedly compromised the Hillary Clinton campaign, the Democratic National Committee, and the Democratic Congressional Campaign Committee in 2016 in an attempt to interfere with the U.S. presidential election.^[4] In 2018, the US indicted five GRU Unit 26165 officers associated with APT28 for cyber operations (including close-access operations) conducted between 2014 and 2018 against the World Anti-Doping Agency (WADA), the US Anti-Doping Agency, a US nuclear facility, the Organization for the Prohibition of Chemical Weapons (OPCW), the Spiez Swiss Chemicals Laboratory, and other organizations.^[13] Some of these were conducted with the assistance of GRU Unit 74455, which is also referred to as Sandworm Team.

ID: G0007

Associated Groups: SNAKEMACKEREL, Swallowtail, Group 74, Sednit, Sofacy, Pawn Storm, Fancy Bear, STRONTIUM, Tsar Team, Threat Group-4127, TG-4127

Contributors: Sébastien Ruel, CGI; Drew Church, Splunk; Emily Ratliff, IBM; Richard Gold, Digital Shadows

Version: 3.0

Created: 31 May 2017

Last Modified: 06 October 2020

[Version Permalink](#)

Associated Group Descriptions

Name	Description
SNAKEMACKEREL	^[14]
Swallowtail	^[11]
Group 74	^[15]
Sednit	This designation has been used in reporting both to refer to the threat group and its associated malware.
Sofacy	This designation has been used in reporting both to refer to the threat group and its associated malware.

- En sélectionnant un des groupes, vous pourrez accéder à plus d'informations. Dans l'exemple ci-dessous, nous pouvons retrouver plus d'informations concernant le groupe d'attaquants APT28 :
 - la description ;
 - les secteurs ciblés ;
 - les alias (noms utilisés dans les rapports de threat intelligence) ;
 - les techniques et tactiques associées.
- Ces informations sont utiles car elles permettent d'identifier des comportements suspects sur notre SI. Elles sont également utiles pour simuler des attaques.

Découvrez la matrice ATT&CK

Elle est composée de 12 tactiques qui contiennent différentes techniques et décrivent les phases classiques d'une attaque.

ATT&CK Matrix for Enterprise

[layouts ▾](#)
[show sub-techniques](#)
[hide sub-techniques](#)

Reconnaissance 10 techniques	Resource Development 6 techniques	Initial Access 9 techniques	Execution 10 techniques	Persistence 18 techniques	Privilege Escalation 12 techniques	Defense Evasion 27 techniques	Credential Access 14 techniques	Discovery 25 techniques	Lateral Movement 8 techniques	Collection 17 techniques	Command and Control 16 techniques	Exfiltration 9 techniques	Impact 13 techniques
Active Scanning (2)	Acquire Infrastructure (2)	Drive-by Compromise	Command and Scripting Interpreter (2)	Account Manipulation (2)	Abuse Elevation Control Mechanism (2)	Abuse Elevation Control Mechanism (2)	Brute Force (2)	Account Discovery (2)	Exploitation of Remote Services	Archive Collected Data (2)	Application Layer Protocol (2)	Automated Exfiltration (1)	Account Access Removal
Gather Victim Host Information (2)	Compromise Accounts (2)	Exploit Public-Facing Application	Exploitation for Client Execution	RTPS Jobs	Access Token Manipulation (2)	Access Token Manipulation (2)	Credentials from Password Stores (2)	Application Window Discovery	Internal Spearphishing	Audio Capture	Communication Through Removable Media	Data Transfer Size Limits	Data Destruction
Gather Victim Identity Information (2)	Compromise Infrastructure (2)	External Remote Services	Inter-Process Communication (2)	Boot or Login Assistant Execution (1)	Boot or Login Assistant Execution (2)	Boot or Login Assistant Execution (2)	Exploitation for Credential Access	Browser Bookmark Discovery	Lateral Tool Transfer	Automated Collection			Data Encrypted for Impact
Gather Victim Network Information (2)	Devised Capabilities (2)	Hardware Additions	Native API	Boot or Login Initialization Scripts (2)	Boot or Login Initialization Scripts (2)	Boot or Login Initialization Scripts (2)	Forced Authentication	Cloud Infrastructure Discovery	Remote Service Session Hijacking (2)	Clipboard Data	Data Encoding (2)	Exfiltration Over Alternative Protocol (2)	Data Manipulation (2)
Gather Victim Org Information (2)	Establish Accounts (2)	Phishing (2)	Scheduled Task/Job (2)	Browser Extensions	Create or Modify System Process (2)	Create or Modify System Process (2)	Input Capture (2)	Cloud Service Dashboard	Remote Services (2)	Data from Cloud Storage Object	Data Obfuscation (2)	Exfiltration Over C2 Channel	Data Removal (2)
Phishing for Information (2)	Obtain Capabilities (2)	Replication Through Removable Media	Shared Modules	Compromise Client Software Binary	Event Triggered Execution (2)	Event Triggered Execution (2)	Man-in-the-Middle (2)	Cloud Service Discovery	Application Through Removable Media	Data from Configuration Repository (2)	Dynamic Resolution (2)	Exfiltration Over Other Network Medium (2)	Endpoint Denial of Service (2)
Search Closed Sources (2)		Supply Chain Compromise (2)	Software Deployment Tools	Create Account (2)	File and Directory Permissions Modification (2)	File and Directory Permissions Modification (2)	Modify Authentication Process (2)	Domain Trust Discovery	Software Deployment Tools	Data from Information Repositories (2)	Encrypted Channel (2)	Exfiltration Over Physical Medium (2)	Firmware Corruption
Search Open Technical Databases (2)		Trusted Relationship	System Services (2)	Create or Modify System Process (2)	Group Policy Modification	Group Policy Modification	Network Sniffing	File and Directory Discovery	Taint Shared Content	Data from Local System	Failback Channels	Exfiltration Over Web Service (2)	Inhibit System Recovery
Search Open Websites/Domains (2)		Valid Accounts (2)	User Execution (2)	Event Triggered Execution (2)	Hide Artifacts (2)	Hide Artifacts (2)	OS Credential Dumping (2)	Network Service Scanning	Use Alternate Authentication Material (2)	Ingress Tool Transfer	Multi-Stage Channels	Exfiltration Over Web Service (2)	Network Denial of Service (2)
Search Victim-Owned Websites			Windows Management Instrumentation	External Remote Services	Hijack Execution Flow (2)	Hijack Execution Flow (2)	Steal Application Access Token	Network Share Discovery		Data from Removable Media	Non-Application Layer Protocol	Scheduled Transfer	Resource Hijacking
				Hijack Execution Flow (2)	Indicator Removal on Host (2)	Indicator Removal on Host (2)	Steal or Forge Kerberos Tickets (2)	Peripheral Device Discovery		Data Staged (2)	Non-Standard Port	Transfer Data to Cloud Account	Service Stop
				Process Injection (2)	Scheduled Task/Job (2)	Scheduled Task/Job (2)	Steal Web Session Cookie	Permission Groups Discovery (2)		Email Collection (2)	Protocol Tunneling		System Shutdown/Reboot
				Implant Container Image	Valid Accounts (2)	Valid Accounts (2)	Modify Authentication Process (2)	Process Discovery		Input Capture (2)	Proxy (2)		
				Office Application Start/Stop (2)			Modify Cloud Compute Infrastructure (2)	Query Registry		Man-in-the-Browser	Remote Access Software		
				Pre-OS Boot (2)			Modify Registry	Remote System Discovery		Web Service (2)			
				Scheduled Task/Job (2)			Modify System Image (2)	Software Discovery (2)					
				Server Software Component (2)			Network Boundary Bridging (2)	System Information Discovery					
				Traffic Signaling (2)			Obfuscated Files or Information (2)	System Network Configuration Discovery					
				Valid Accounts (2)			Pre-OS Boot (2)	System Network Connections Discovery					
								System Owner/User Discovery					
								System Service Discovery					
								System Time Discovery					

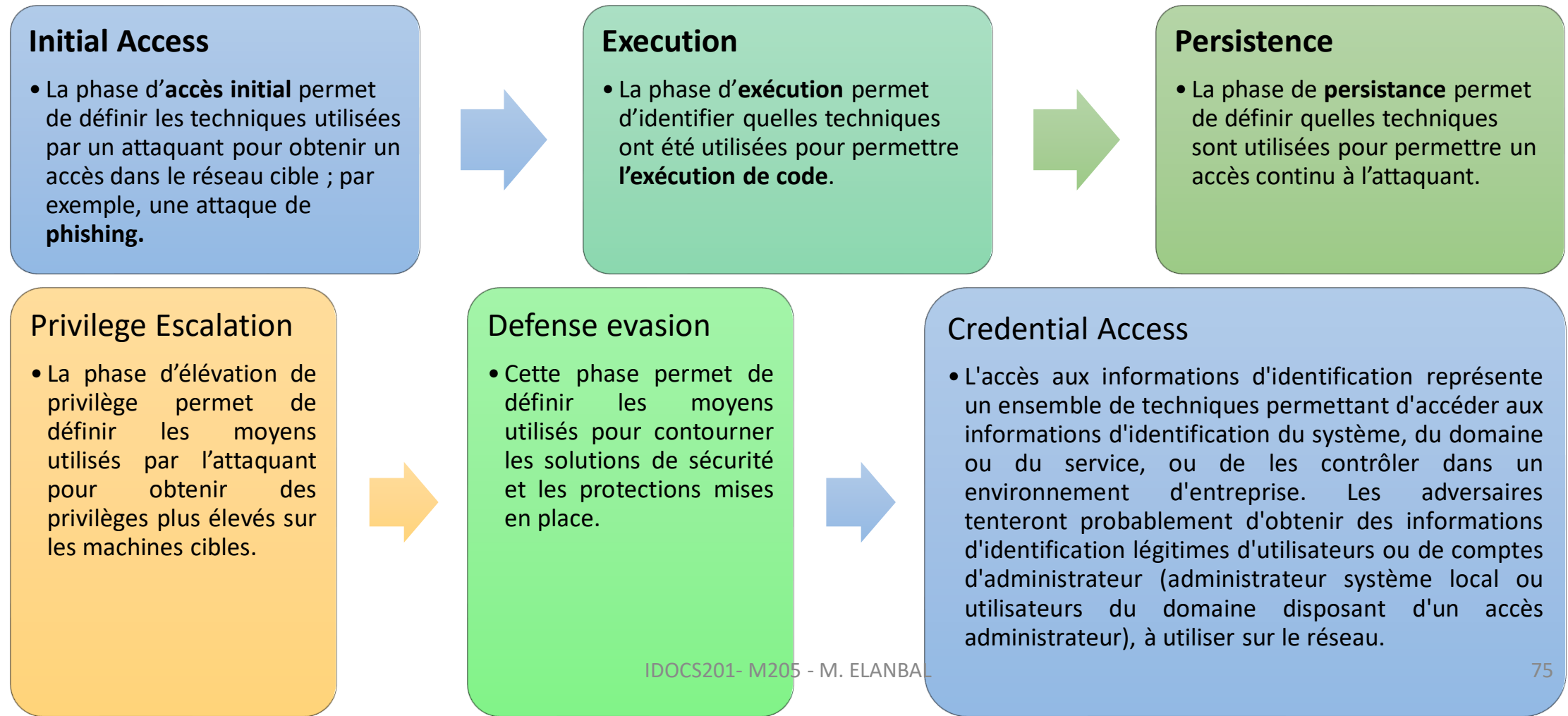
Tactiques

- L'objectif stratégique d'un attaquant peut être d'extorquer de l'argent au moyen d'une rançon, de voler des informations, ou tout simplement de détruire l'environnement informatique d'une organisation. Pour atteindre ces objectifs, les attaquants mettent en place toute **une série d'actions progressives – des tactiques**.
- On va généralement retrouver les différentes étapes d'une attaque, en commençant par l'accès initial (Id TA0001 sur la matrice), en passant par des phases intermédiaires d'exécution (TA0002), d'élévation de privilèges (TA0004) ou de mouvement latéraux (TA0008), pour finalement aboutir aux phases de collection (TA0009) et d'exfiltration (TA0010).
- Il est important de comprendre que les tactiques sont une classification et une description des modes opératoires. Les tactiques décrivent ce que l'attaquant essaie de faire à **n'importe quelle étape** de l'attaque. Elles ne suivent pas forcément un ordre défini.

Techniques

- Les tactiques spécifient ce que l'attaquant tente de faire, les techniques décrivent les **différentes méthodes développées par les attaquants pour mener une tactique**.
- Par exemple, si les attaquants maintiennent leur accès sur le SI compromis, c'est la tactique de persistance. Cette tactique peut être menée à bien de **plusieurs manières répertoriées en techniques**. Il est possible de créer une clé de registre RUN (T1547.007), ou encore de créer un service malveillant qui permettra l'exécution du logiciel malveillant à chaque reboot ou à un instant T (T1543.002).
- Les techniques ayant le même objectif sont regroupées sous la même tactique.
- Pour chaque technique, la matrice fournit des exemples de cas connus où la technique est :
 - utilisée par un groupe d'attaquants ;
 - ou implémentée par un logiciel malveillant.
- Toutes ces informations vous permettent d'identifier de manière plus précise les techniques utilisées, afin d'ajuster en profondeur vos mécanismes de supervision et de détection.

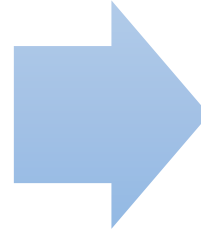
Les tactiques de Mitre Att&ck



Les tactiques de Mitre Att&ck

Discovery

- La **découverte** consiste en des techniques permettant à l'adversaire d'**acquérir des connaissances sur le système et le réseau** interne. Lorsqu'ils ont accès à un nouveau système, les attaquants doivent découvrir ce qu'il est et comment il fonctionne.



Lateral Movement

- Le **mouvement latéral** consiste en des techniques permettant à un attaquant d'accéder à des systèmes distants sur un réseau et de les contrôler. Il peut inclure l'exécution d'outils sur ces systèmes distants. Les techniques de déplacement latéral pourraient permettre à un adversaire de collecter des informations d'un système sans avoir besoin d'outils supplémentaires, tels qu'un outil d'accès à distance.

Les tactiques de Mitre Att&ck

Collection

- La collection consiste en des techniques permettant d'identifier et de rassembler des informations, telles que des fichiers sensibles, à partir d'un réseau cible avant l'exfiltration. Cette catégorie couvre également les emplacements sur un système ou un réseau où l'adversaire peut rechercher des informations pour exfiltrer des données.



Command and Control

- La tactique de commande et de contrôle montre comment les adversaires communiquent avec les systèmes sous leur contrôle au sein d'un réseau cible. Un adversaire peut établir le commandement et le contrôle de différentes manières, selon la configuration du système et la topologie du réseau. En raison du large degré de variation disponible pour l'adversaire au niveau du réseau, seuls les facteurs les plus courants ont été utilisés pour décrire les différences de commandement et de contrôle. Les méthodes documentées contiennent encore de nombreuses techniques spécifiques, dues en grande partie à la facilité avec laquelle il est possible de définir de nouveaux protocoles et d'utiliser des protocoles et des services réseau légitimes existant pour la communication.

Exfiltration

- L'exfiltration consiste à voler des données, les altérer ou les supprimer. L'attaquant pourra utiliser des techniques pour masquer ces actions, vous ne saurez donc peut-être pas s'il l'a réellement fait ou pas ! L'attaquant a eu accès ici à tous les fichiers de la machine compromise, il a potentiellement supprimé, altéré ou volé des informations !



Impact

- La tactique **Impact** représente des techniques dont l'objectif principal **réduit directement la disponibilité ou l'intégrité d'un système**, d'un service ou d'un réseau. y compris la manipulation de données, pour impacter une activité ou un processus opérationnel. Ces techniques peuvent représenter l'objectif final de l'adversaire ou fournir une couverture en cas de violation de la confidentialité.



2- Analyse des données numériques

Les étapes à suivre pour bien mener son analyse forensic

La phase d'identification de contexte:

- cette étape consiste à rencontrer les personnes travaillant en entreprise (administrateurs, responsable de la machine infectée, etc.) afin de collecter le plus d'informations possible sur la machine infectée et sur l'incident de sécurité de manière générale.

La phase de collecte des informations sur la/les machine(s) infectée(s):

- cette étape consiste à copier les données pour pouvoir les analyser sans compromettre les données originales (copie de la mémoire vive et du disque dur). Vous calculez les hashes des informations collectées afin de garantir l'intégrité des fichiers analysés.
- Cette phase est très importante, car elle nécessite de respecter une procédure stricte qui **n'altérera pas les données stockées**. Les données collectées seront stockées dans un container que l'on appelle **image**. Une image est un **dump brut extrait d'un support numérique**. Il existe plusieurs formats d'image.

Formats des données collectées

Les images RAW

- Les images brutes, au format RAW, ne sont pas un format en soi, mais un **bloc de données brutes** reproduites à partir d'une image. Les images brutes ne contiennent aucune métadonnée supplémentaire en dehors des informations sur le fichier image lui-même (nom, taille, horodatage et autres informations).

Les formats de forensic

- Plusieurs problèmes avec les images brutes ont conduit à la création de formats de fichiers pour le forensic. Les formats de forensic comportent des éléments supplémentaire tels que l'horodatage, les hash des images et d'autres **métadonnées**. Par ailleurs, il peut être nécessaire de compresser ou chiffrer une image acquise. Les formats de forensic facilitent la mise en œuvre de ces fonctionnalités. Vous retrouverez entre autres :
 - **EnCase EWF**, développé par Guidance Software, l'une des plus anciennes **entreprises de logiciels de forensic**. Il utilise le format EWF (*Expert Witness Format*) qui prend en charge les métadonnées, la compression, le chiffrement, le hachage, etc. ;
 - **FTK Smart**, par AccessData, est un concurrent direct d'EnCase EWF. Ce **format propriétaire** inclut également les métadonnées, la compression, le chiffrement, le hachage, etc. ;
 - **AFF**, pour *Advanced Forensic Format*, a été créé par Simson Garfinkel en tant que **format ouvert**. Il comprend toutes les fonctionnalités attendues et inclut également des fonctionnalités de chiffrement et de signature utilisant des certificats X.509 standard.

Les étapes à suivre pour bien mener son analyse forensic (suite)

La phase d'analyse des informations collectées:

- C'est la phase technique de l'investigation numérique où vous allez essayer de remonter au moment exact où l'incident de sécurité s'est produit.
- Dans cette phase, vous serez amené à analyser le dump mémoire (la RAM), le disque dur ainsi que les fichiers douteux.
- Afin de réaliser une analyse de la mémoire volatile (la RAM) il faut dans un premier temps réaliser ce qu'on appelle un dump mémoire. Le dump de la RAM consiste à faire une copie vers un support externe de la RAM avant l'extinction de la machine. Cela permet de récupérer des informations très utiles pour l'investigation tels que: les identifiants, les clés de chiffrement, les processus actifs, etc.
- Tandis que, pour l'analyse du disque dur, une copie bit à bit du disque dur de la machine infectée est à réaliser (**copie fidèle de chaque bit du disque**). Dans ce cas, il vous faut un disque dur externe pour la copie, et le temps nécessaire pour faire la copie dépend de la taille du disque dur de la machine infectée.

Analyse des données collectées

- Du coup, Les données numériques peuvent être volatiles (on parle aussi de données dynamiques) et non-volatiles (on parle alors de données statiques). Les enquêteurs peuvent recueillir deux types de preuves numériques :

Données volatiles :

- Les données volatiles sont des informations numériques stockées sur un support temporaire. Ces données sont perdues lorsque l'appareil est mis hors tension. La mémoire vive (RAM) est la donnée volatile la plus courante dans le cadre d'une enquête sur les preuves numériques. D'autres exemples sont les connexions réseau, les fichiers ouverts, les processus en cours d'exécution et les sessions actives. En général, on peut recueillir des données résiduelles à partir de ces sources.

Données non volatiles :

- Les données non volatiles sont des informations numériques stockées sur des supports permanents, tels que les disques durs. Les données ne sont pas perdues même lorsque le périphérique est éteint. Les données non volatiles comprennent les fichiers système, les journaux d'événements, les fichiers de vidage, les fichiers de configuration et les informations de compte. Ces données sont moins difficiles à récupérer à des fins de preuve que les données volatiles.

Analyse des données collectées

Les données volatiles sont par nature fragiles et sont modifiées en permanence. Par exemple, éteindre un ordinateur détruit certaines de ces données, brancher une clé USB ajoute de nouvelles données dans le registre, ... Ces données se trouvent dans le registre, le cache, la RAM (mémoire vive).



Par contre, les données non-volatiles, c'est-à-dire les données inscrites sur le disque dur. Ces données sont stockées et ne dépendent pas de l'alimentation électrique du système.

Analyse des données collectées

L'image bit-à-bit est différente d'un backup

- Le principe de l'image bit-à-bit est de prendre une image complète du disque dur, un peu comme un miroir. Concrètement, si le disque dur a une capacité de 500 Go, le fichier image qui en résulte aura lui aussi une taille de 500 Go. On va ainsi collecter les fichiers cachés, les données résiduelles comme le slack-space, les espaces inutilisés (mais sur lesquels il restent peut-être des anciennes données ?), etc ...
- *C'est comme un Backup alors ?* 🤔
- Non ! Faire un Backup consiste à créer une archive des données, pour les restaurer ensuite ultérieurement, en cas de problème sur le système, par exemple. Ainsi, les espaces inutilisés ne sont pas collectés.

Analyse des données collectées

Analyse forensic du dump mémoire

- Durant cette phase l'analyste va essayer d'extraire le plus d'information possible de la RAM tels que: **connexions réseaux**, des **clés de registre**, des **mots de passe** ou encore des **processus** en cours d'exécution.
- Récupération de la liste des processus: Lors de l'exécution d'un programme, un processus est créé pour ce dernier ainsi qu'un ID spécifique appelé PID (Process ID). Le code de ce programme est chargé dans la RAM pour son exécution ainsi que les bibliothèques partagées, les données numériques et la pile d'exécution. Avoir la liste des processus dans la RAM permet de savoir qu'elle programme tournait sur la machine au moment du dump mémoire permettant de récupérer et d'analyser les données stockées dans l'espace d'adressage (qui contient le code du programme) dans l'optique de trouver le programme potentiellement malveillant et de comprendre son origine.
- Cependant, cette phase porte ses fruits uniquement si le programme est exécuté pendant le dump mémoire et que son fonctionnement n'est pas dissimulé derrière un processus légitime.
- Parmi les informations les plus importantes à récupérer, nous trouvons: l'adresse mémoire du processus (offset), le nom du processus en cours d'exécution (name), le numéro d'identification du processus (PID), le PID du processus parent (PPID), la date et l'heure de lancement du processus (start).

Analyse des données collectées

Analyse forensic du dump mémoire

- Récupération des DLL d'un processus: Les DLL (Dynamic Link Library) sont des librairies disponibles et utilisées dans Windows. Ces DLLs peuvent être utilisées pour manipuler des données, créer des connexions réseau ou même d'écrire sur des fichiers ou les créer. Il est également important de récupérer les DLLs utilisées par un programme pour déduire son fonctionnement.
- L'analyse du registre: Le registre de Windows contient plusieurs informations pertinentes. Lors de l'exécution d'un programme il est accédé en permanence ce qui signifie que le système d'exploitation charge une partie des fichiers du registre dans la RAM. Il peut contenir les programmes récemment exécutés ainsi que les valeurs introduites par un programme malveillant.
- L'analyse des connexions réseaux: l'objectif d'un programme malveillant est de communiquer à distance avec son propriétaire afin de recevoir les ordres. Pour réussir son attaque, le programme malveillant cherche à ouvrir des ports.
- L'analyse mémoire du réseau permet de voir si des communications ont été effectuées vers des IPs distantes malveillantes (exemple: IPs des serveurs de Command and control), l'utilisation des ports suspects (exemple: port 4444) ou même l'échange de données.

Analyse des données collectées

Analyse forensic du disque dur

- Lors de l'analyse du dump mémoire nous avons pu analyser les processus en cours d'exécution avant l'extinction du système (c'est-à-dire les données volatiles). Dans l'analyse du disque dur nous allons analyser le comportement et la chronologie des événements sur le système afin de confirmer notre analyse du dump mémoire.
- Afin de comprendre cette partie, il faut avoir des connaissances en terme de système d'exploitation.
- Les données dans un système d'exploitation Windows sont stockées dans un système de fichiers appelé **NTFS (New Technology File System)** permettant de spécifier la manière dont les fichiers et dossiers sont nommés, stockés et organisés.
- Le système **NTFS** permet de:
 - Gérer les ADS (Alternate Data Stream): permettant de localiser un fichier par son nom ou son auteur. Cependant, il peut être utilisé par les attaquants afin de dissimuler des flux de données.
 - Améliorer la sécurité en utilisant les ACL (Access Control List) permettant de gérer les droits d'accès aux fichiers.
 - Chaque fichier stocké contient un timestamp, ainsi que des informations sur la création, la modification, l'accès.
 - NTFS utilise une structure de données appelée MFT (Master File Table) qui stocke les données sur le système avec des informations de timestamp

Analyse des données collectées

Analyse forensic du disque dur

- Lors de l'analyse du dump mémoire nous avons pu analyser les processus en cours d'exécution avant l'extinction du système (c'est-à-dire les données volatiles). Dans l'analyse du disque dur nous allons analyser le comportement et la chronologie des événements sur le système afin de confirmer notre analyse du dump mémoire.
- Afin de comprendre cette partie, il faut avoir des connaissances en terme de système d'exploitation.
- Les données dans un système d'exploitation Windows sont stockées dans un système de fichiers appelé **NTFS (New Technology File System)** permettant de spécifier la manière dont les fichiers et dossiers sont nommés, stockés et organisés.
- Le système **NTFS** permet de:
 - Gérer les ADS (Alternate Data Stream): permettant de localiser un fichier par son nom ou son auteur. Cependant, il peut être utilisé par les attaquants afin de dissimuler des flux de données.
 - Améliorer la sécurité en utilisant les ACL (Access Control List) permettant de gérer les droits d'accès aux fichiers.
 - Chaque fichier stocké contient un timestamp, ainsi que des informations sur la création, la modification, l'accès.
 - NTFS utilise une structure de données appelée **MFT (Master File Table)** qui stocke les données sur le système avec des informations de timestamp

Analyse des données collectées

Analyse forensic du disque dur

- La MFT enregistre toutes les données relatives à un fichier, tels que: son nom, sa taille, son horodatage, ses droits d'accès, et parfois ses données.
- Ci après quelques exemples d'attributs qu'on peut trouver dans les entrées MFT:

Valeur	Description
\$STANDARD_INFORMATION 0x10	Attributs de fichier (Read-Only ou Archive), horodatage et nombre de liens physiques.
\$ATTRIBUTE_LIST 0x20	Une liste d'attributs qui constituent le fichier et la référence de fichier du fichier MFT dans lequel chaque attribut est situé.
\$FILE_NAME 0x30	Le nom du fichier en Unicode
\$VOLUME_NAME 0x60	Le nom du volume.
\$VOLUME_INFORMATION 0x70	Information sur le volume
\$DATA 0x80	Le contenu du fichier

Pour exploiter les données contenues dans une MFT, il existe des outils open source permettant de convertir le fichier MFT en un fichier de format CSV. Plusieurs fichiers .csv peuvent être créés suite à la conversion, le plus important c'est celui qui va contenir toutes les entrées permettant de retrouver toutes les créations de fichiers ordonnées par date et heure (timestamp).

Cette information est cruciale lors d'une investigation forensic car elle va nous fournir une chronologie des évènements qui ont eu lieu avant l'attaque

Analyse des données collectées

Analyse forensic du disque dur

- D'autres investigations peuvent être effectuées lors d'une analyse de disque dur:
 - **Windows event logs:** permettant de retracer toutes les actions de chaque application sur le système.
 - **Les services:** concerne les applications qui se lancent discrètement au démarrage de votre ordinateur jusqu'à l'extinction de ce dernier (exemple d'outils: autorun tool et services.msc). Cette fonctionnalité peut être utilisée par les attaquants pour que le malware se ré-exécute même après le redémarrage de la machine (on parle de malware persistant).
 - **Les prefetch:** sont des fichiers créés par le système d'exploitation lorsqu'une application s'exécute pour la première fois. Il est possible de déterminer les programmes qui se sont exécutés sur la machine car les prefetch stocke les 128 derniers programmes exécutés sur le système en les classant par ordre chronologique.
 - *Prefetch (et [Prefetcher](#)) est un **système de cache** inclus par défaut dans Windows. Il a pour but d'**optimiser et accélérer le lancement des applications**. Cela se présente sous la forme d'un service Windows qui crée des petits **fichiers .pf** (ex CHROME.EXE-CCF9F3FC.pf) dans le dossier **C:\Windows\Prefetch***

Processus et règles de base

- Il y a trois grandes règles à avoir en tête avant de commencer.
- Ne jamais travailler sur la machine à étudier car on va alors altérer les données.
- Prendre deux copies : on travaille sur l'une, l'autre doit être conservée « propre ».
- Calculer l'empreinte des données à l'aide d'un algorithme de hachage (hash). Le hash permet de s'assurer de l'intégrité des données (on peut utiliser les algorithmes MD5, SHA1, etc...).

PARTIE 3 :

Maitriser les outils d'investigation

M205 – Appréhender les méthodes d'investigation numérique
MH : 90h

Formatrice : M. EL ANBAL



Identifier les outils d'investigation du marché

Chapitre 1



Les outils d'analyse forensic

- Pour réaliser une analyse forensic de support numérique, il existe sur le marché des outils reconnus.



- La société Guidance Software propose une suite d'utilitaires appelée [Encase](#) dédiée à l'analyse forensic, en passant de **l'analyse du disque et au tri des données** jusqu'à **l'analyse des fichiers et le déchiffrement des volumes analysés**. Les licences sont payantes et restent relativement chères. Toutefois, ce genre d'outil est largement utilisé par les experts judiciaires ou encore les organismes de police.

Les outils d'analyse forensic

- Il existe également des **outils hardware permettant la collecte de supports numériques sans altération des données**, tels que des [bloqueurs en écriture](#) que nous avons évoqués plus haut.
- Les bloqueurs en écriture sont donc des dispositifs qui permettent de faire **l'acquisition d'une image d'un disque dur en bloquant le mécanisme d'écriture**, mais pas de lecture, dans le but de préserver le contenu. L'utilisation de ce type de dispositif permet de protéger le contenu du disque et **garantit ainsi son intégrité**.
- Un bloqueur en écriture matériel **s'interpose entre le disque dur (la preuve) et le PC** qui servira à l'acquisition de l'image. Il existe également des bloqueurs en écriture **logiciels**.



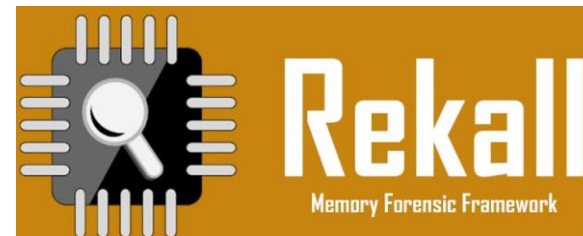
Les outils d'analyse forensic



- Le framework [The Sleuth Kit](#) permet de réaliser une analyse forensic en passant de la **génération d'une timeline au triage des données et à l'analyse des artefacts Windows (registre, email, historique...)**, jusqu'à la génération d'un rapport. Il comporte une interface graphique appelée **Autopsy**.
- Les utilitaires [NirSoft](#) ou encore la suite [Sysinternal](#) de Microsoft sont également utilisés dans l'analyse forensic. Google a également développé son propre framework d'analyse, [Google Rapid Response](#) (GRR), qui permet de faire de l'analyse forensic à distance de postes compromis.

Les outils d'analyse forensic

- Certains projets proposent également des **OS Linux destinés à l'analyse forensic** et embarquant des outils préinstallés tels que [SIFT](#), [Tsurugi](#), [CAINE](#) ou encore [DEFT](#).
- Pour l'analyse de la **mémoire vive**, il existe également des frameworks tels que [Volatility](#) et [Rekall](#).
- Pour la **copie des disques durs et de la RAM**, il est possible d'utiliser le freeware [FTK Imager Lite](#).



TP n°3



Activité 1

Activité 2

Rapport d'investigation forensic :

- Une fois votre analyse forensic faite, un rapport d'investigation doit être effectué et communiqué à l'entreprise. Ce rapport doit comprendre les éléments suivants:

Les indicateurs de compromission réseaux:

- consiste à identifier les adresses IP, les URLs et les domaines qui ont été utilisés par l'attaquant et découverts à l'étape d'investigation. Ces informations appelées IoC sont à inclure dans votre rapport d'investigation final afin de bloquer ces communications sur les autres machines de votre réseau et ainsi contenir la menace. Ces IoCs peuvent aussi être utilisés pour créer des règles de sécurité au niveau de FW, Proxy, IPS/IDS.

Les indicateurs de compromission fichiers:

- dans ce cas nous parlons des hash de fichiers malveillants découverts à l'étape d'investigation. Vu que chaque fichier a son propre hash qui est unique, cette information est cruciale afin de bloquer tout fichier avec ce(s) hash(s) au niveau des anti-virus ou même sur les outils de gestion de messagerie permettant de bloquer les emails avec une pièce jointe dont le hash est malveillant.

Les indicateurs de compromission adresses email:

- si la compromission de la machine a été effectuée à travers la réception d'un email malveillant (exemple: phishing, URL malicieuse, etc.), l'adresse email émettrice de ce mail constitue aussi un IoC. Cette information permet de bloquer tout email en provenance de cette dernière.

Les indicateurs de compromission système

- si le logiciel malveillant mène des actions sur le système, tel que la création d'un service Windows spécifique, le service créé constitue un IoC. Un scan à distance peut être effectué sur toutes les machines de l'entreprise pour identifier les machines compromises.

La structure du rapport d'investigation

Page de garde.

Sommaire.

Résumé des investigations, rappel du contexte et des conclusions.

Détail des investigations :

- phase de collecte et hash (copie d'écran) ;
- analyse réalisée et les découvertes (copie d'écran).

Hypothèse de l'analyse.

Recommandations.

Liste des indicateurs de compromission.

Activité 4 :
Appliquez cette
structure sur
l'activité
précédente



Fin du module

Merci pour votre attention

Etat de l'art de l'investigation numérique

Technopédia définit l'informatique légale comme :

- le processus de **découvrir** et **d'interpréter** des données électroniques ;
- l'objectif principal de ce processus est de **préserver** toutes les preuves sous leur forme la plus **originale** tout en effectuant une enquête structurée en :
 - recueillant
 - identifiant et
 - validant les informations numériques afin de reconstruire des événements passés.

En d'autres mots :

- l'investigation numérique est une branche des sciences légales ;

À l'époque, de telles preuves auraient été :

- le journal intime d'une personne ;
- empreinte digitale sur un verre.

ADF – Anti Digital forensic

- En français : Anti-criminalistique
- Définition du Dr Marc Rogers (université de Purdue) :

"Tentatives d'affecter négativement l'existence, la quantité et / ou la qualité des preuves provenant d'une scène de crime, ou de rendre l'analyse et l'examen des preuves difficiles ou impossibles à réaliser."

ADF

- Obfuscation des données :
 - Chiffrement
 - Stéganographie
 - Caches avancées (secteurs défectueux, espaces libres, partitions cachées ...)
- Suppression des artefacts :
 - Nettoyage de disque (nécessite beaucoup de temps)
 - Nettoyage de fichiers (rapide)
 - Démagnétisation (machine) / destruction (incinération)
- Brouiller les pistes
 - Suppression / modifications des journaux de logs
 - Usurpation d'identité
 - Compte zombifié
 - Commandes par cheval de troie

Timestomp (metasploit) : modifier les métadonnées d'un fichier (création / modification)

Transmogrify (metasploit) : modifier l'en-tête des fichier (Jpeg -> Doc)

- Attaque contre l'investigation
 - exploitation des failles des outils médico-légales
 - modification des hash (contestatation possible de l'enquête)
 - falsification des résultats obtenus