
OSINT Avancé
Investigation en source ouverte
ali.koudri@gmail.com

Organisation de la Formation

Module 1 : Maîtrise de la Collecte

1. Éthique et Légalité Avancées

- OSINT passif vs. Actif
- Légalité du scraping vs. CGU des sites
- Cas des "leaks" de données
- OPSEC Avancée

2. Deep Web & Analyse de Surface Approfondie

- Analyse de code source
- Analyse des Sitemaps et robots.txt
- Fuzzing & Monitoring

3. Dark Web Opérationnel

- Navigation sécurisée (Whonix, TAILS) vs. simple Tor Browser .
- Cartographie des forums et marketplaces (BreachForums, Dread, etc.)
- Techniques de recherche sur le Dark Web

4. Techniques d'Extraction

- Scraping statique vs dynamique
- Outils No-Code
- Outils Low-Code

5. Atelier Pratique

- Profiler une entreprise suspecte

Module 2 : De la Donnée à l'Intelligence

1. SOCMINT / HUMINT Avancé

- Analyse de réseaux sociaux
- Analyse comportementale
- Nouveaux terrains: Telegram / Discord

2. TECHINT Avancé

- Web Crawling

3. Maîtrise des Outils d'Analyse

- Import de données CSV, ...
- Transforms Avancées
- Création de Transforms locales avec Python
- Analyse de Graphes

4. IMINT & GEOINT Avancé

- Géolocalisation Manuelle
- VIDINT

5. BLOCKCHAIN-INT Avancé

- Analyse de Transaction
- Concepts avancés : Mixers et le Chain Hopping.
- Outils de visualisation

6. Travaux Pratiques

Module 3 : OSINT Opérationnel et Étude de Cas

1. Planification Stratégique Avancée

- Key Intelligence Questions
- Analysis of Competing Hypotheses
- Gestion de la Connaissance

2. OSINT Opérationnel

- Outils No-Code
- Plateformes CTI

3. Contre-OSINT

- Red Team : Comment l'OSINT est utilisé par les attaquants.
- Blue Team : Comment se défendre ? Auditer sa propre empreinte.

4. Étude de Cas Complète

- Collecte
- Analyse de Données
- Analyse Financière
- Analyse d'Infra
- LockShadow
- Analyse de Liens
- Production de rapport

5. Restitution, Bilan et Perspectives

Module 4 : L'Ère de l’Analyste Augmenté

1. Introduction et Changement de Paradigme

- De la Recherche à la Synthèse.
- Les 3 Rôles de l'IA en OSINT
- La Ligne Rouge (OPSEC & Éthique)

2. LLM au service du Renseignement

- TECHINT Augmenté (Code Assistant)
- SOCMINT Augmenté (Analyse Sémantique)
- Traduction & Contexte Culturel

3. OPSEC : L'IA en Local

- Pourquoi le Local ?
- Les Outils

4. Automatisation Intelligente

- Le Concept d'Agent
- n8n (L'Orchestrateur)

5. Computer Vision & Machine Learning "Accessible"

- GEOINT Augmenté
- Teachable Machine
- Transcription (Audio)

6. Travaux Pratiques

Module 1
Maîtrise de la Collecte et Sources Non-Conventionnelles

Rappels de la Partie 1

Qu'est-ce que l'OSINT ?

Définition Formelle : Le renseignement (Intelligence) collecté à partir de sources d'information ouvertes, publiques et accessibles.

"Ouvert" / "Public" ne veut PAS dire :

- **Gratuit** : Une source peut être publique mais payante (ex: registres d'entreprises, bases de données spécialisées).
- **Facile à trouver** : L'information peut être "cachée" dans le "Deep Web" (non indexé par Google), comme des bases de données ou des forums.
- **Numérique** : L'OSINT inclut aussi les médias traditionnels (TV, radio, journaux), les publications académiques, les conférences…

La Règle d'Or : Si vous devez hacker, forcer un accès ou violer un système pour obtenir l'information, ce n'est PAS de l'OSINT.

La Pyramide du Renseignement

Données

- Quoi : Faits bruts, non structurés, sans contexte.
- Répond à : Rien.
- Exemple : "Jean Dupont", "Paris", "Projet Y", "Entreprise X"

Information

- Quoi : Données contextualisées.
- Répond à : "Qui ?", "Quoi ?", "Où ?", "Quand ?"
- Exemple : "L'ingénieur Jean Dupont de l'entreprise X a publié une offre d'emploi pour un 'Spécialiste Projet Y' à Paris."

Renseignement

- Quoi : Information analysée, vérifiée et contextualisée.
- Répond à : "Pourquoi ?" et "Et alors ?" (Actionnable)
- Exemple : "L'entreprise X déplace ses ressources vers le Projet Y, ce qui indique un lancement de produit concurrent en Q3."

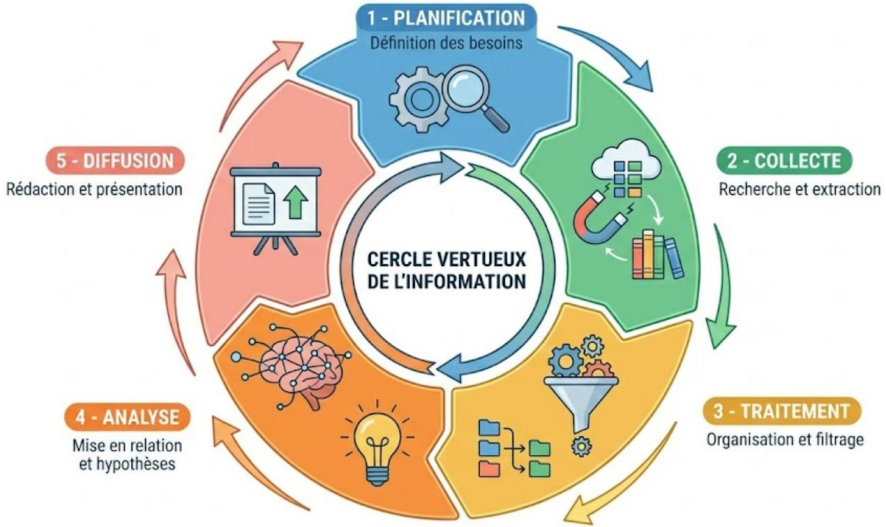
Le Facteur Humain : Le Maillon Faible de l'Information

Facteur Humain	Description	Exemple OSINT Clé
1. L'Habitude	La recherche de confort et la répétition des comportements.	**Réutilisation des Identifiants :** La cible utilise le même pseudonyme (username) sur toutes les plateformes (forum, Instagram, jeu vidéo), créant un fil d'Ariane.
2. La Fainéantise	La loi du moindre effort (préférer la facilité à la sécurité).	**Paramètres par Défaut :** Les paramètres de confidentialité par défaut sont rarement modifiés. L'utilisateur clique sur "Accepter" sans lire.
3. La Vanité (ou l'Égo)	Le besoin de reconnaissance sociale et de partage d'informations personnelles.	**Surcharge de Preuves :** Poster une photo d'une réussite (diplôme, récompense, lieu de travail) sans couper les éléments identifiables (badge, document en arrière-plan).
4. L'Oubli / La Négligence	L'incapacité à maintenir une vigilance constante sur tous ses anciens comptes et publications.	**Le Compte Abandonné :** Oublier un vieux compte MySpace ou Twitter avec des publications juvéniles, mais contenant un email personnel ou un numéro de téléphone vérifiable.

Le Cœur de la Méthode : Le Cycle du Renseignement

L'OSINT n'est pas une "recherche". C'est un cycle rigoureux. Omettre une étape compromet tout le résultat.

1. **Planification** (Définition des besoins)
2. **Collecte** (Recherche et extraction)
3. **Traitement** (Organisation et filtrage)
4. **Analyse** (Mise en relation et hypothèses)
5. **Diffusion** (Rédaction et présentation)



Itérations:

- Feedback : La diffusion du rapport génère presque toujours de nouvelles questions.
- Ces nouvelles questions alimentent une nouvelle phase de Planification.
- L'OSINT n'est pas un processus linéaire, mais une boucle d'amélioration continue.

CSI Linux

Qu'est-ce que c'est ?

- C'est une distribution Linux (basée sur Ubuntu) spécialisée pour l'investigation numérique, la cybersécurité et l'OSINT.
- Pensez-y comme la "boîte à outils" complète de l'enquêteur.

Pourquoi l'utiliser ?

- Gain de temps : Des centaines d'outils (dont la plupart utilisée dans cette formation) sont déjà pré-installés, pré-configurés et organisés par catégories.
- Efficacité : Il est conçu pour suivre le Cycle du Renseignement , avec des outils pour la Collecte , le Traitement et l'Analyse.

Comment l'utilise-t-on ?

- De manière sécurisée, comme vu dans notre pilier OPSEC.
- Nous l'installerons dans une Machine Virtuelle (comme VirtualBox) pour cloisonner notre environnement d'enquête de notre machine personnelle.

Exfiltration de Données Cloisonnée

Le Problème de l'Analyste : Vous avez téléchargé une preuve (un document leaké, une capture d'écran) dans votre VM sécurisée (Whonix/Tails).

Comment transférer ce fichier vers votre ordinateur principal pour rédiger le rapport ?

À NE PAS FAIRE : Se connecter à son Google Drive, s'envoyer un email, ou utiliser WeTransfer depuis la VM (Risque de corrélation d'identité et logs serveurs).

Exfiltration avec OnionShare

Qu'est-ce que c'est ? Un outil open-source (intégré à CSI Linux) qui permet de partager des fichiers de manière anonyme via le réseau Tor.

Architecture "Serverless" : Il ne stocke rien dans le Cloud. Il transforme temporairement votre VM en un serveur web sécurisé accessible via une adresse .onion.

Sécurité :

- **Anonymat** : Votre IP réelle est masquée (Tor).
- **Chiffrement** : Transfert chiffré de bout en bout (E2EE).
- **Éphémère** : Dès que vous fermez l'outil, le lien disparaît à jamais.

OnionShare : Cas d’usage 1

Scénario : Sortir un PDF suspect de Whonix vers votre Desktop pour analyse antivirus locale ou rapport.

1. Dans la VM (Source) : Ouvrez OnionShare, glissez le fichier et cliquez sur "Start Sharing".
2. Le Lien : L'outil génère une URL unique (ex: http://onionshare...onion) et un code d'accès.
3. Sur l'Hôte (Destination) : Ouvrez le navigateur Tor, collez le lien.
4. Transfert : Téléchargez le fichier directement depuis votre VM.
5. Fermeture : Arrêtez le partage. Le lien est mort.

OnionShare : Cas d’usage 2

Scénario : Une source (humaine) veut vous envoyer un document sans laisser de trace email.

1. Mode "Receive" : Lancez OnionShare en mode "Réception".
2. Partage : Envoyez l'URL .onion générée à votre source (via messagerie chiffrée).
3. Dépôt : La source upload le fichier via son navigateur Tor. Le fichier arrive directement sur votre disque dur sécurisé.

Note OPSEC : Le transfert passe par Tor. C'est lent. Idéal pour des documents (PDF, Images, Bases SQL légères), mais inadapté pour des dumps de 50 Go.

Travaux Pratiques

Vous avez trouvé un document PDF sensible sur le Dark Web (simulé par un fichier texte créé dans la VM).

Ce fichier pourrait contenir un malware.

Vous devez le transférer sur votre machine hôte pour l'analyser, sans jamais connecter votre VM à vos comptes personnels (Cloud/Email).

Note: Il faut au préalable que le timezone de votre VM soit le même que celui de la machine physique. *sudo timedatectl set-timezone Europe/Paris*

Éthique et Légalité Avancées

RGPD et Doxing

RGPD (Règlement Général sur la Protection des Données)

- **Le Principe** : S'applique à tout traitement de données personnelles de résidents EU.
- **Votre Impact** : Même si une donnée est publique (ex: post de forum), sa collecte, son stockage et son analyse systématiques constituent un "traitement".
- **Vos Obligations** :
  - Minimisation des données : Ne collecter que ce qui est strictement nécessaire (défini dans la Planification).
  - Limitation de la finalité : L'enquête doit avoir un but légitime (lutte anti-fraude, due diligence...).

Doxing vs. OSINT

- **OSINT** : Analyse factuelle pour un rapport privé destiné à un client.
- **Doxing** : Publication d'informations personnelles dans le but de nuire.

La "Zone Grise"

La formation initiale vous a appris ce qui est "blanc" (légal) et "noir" (illégal).

Une enquête avancée opère constamment dans le "gris".

Les 3 questions de l'analyste avancé :

- **Passif vs. Actif** : Ai-je le droit de toucher le serveur de la cible ?
- **Scraping vs. CGU** : Ai-je le droit d'aspirer massivement un site web ?
- **Leaks vs. Recel** : Ai-je le droit d'exploiter des données volées ?

Passif vs. Actif

OSINT Passif (100% Légal)

- **Définition** : Vous ne "touchez" jamais la cible. Vous consultez des sources tierces qui ont déjà collecté l'information.
- **Exemples** :
  - Consulter le Cache Google ou la Wayback Machine.
  - Analyser les résultats de Shodan ou Censys (ils scannent pour vous).
  - Utiliser DNS Dumpster ou SecurityTrails.
  - Consulter les profils via des interfaces alternatives (ex: Nitter).

OSINT Actif (La Zone Grise)

- **Définition** : Vous interagissez directement avec l'infrastructure de la cible. Votre IP (même masquée) apparaît dans leurs logs.
- **Exemples** :
  - Visiter le site web de la cible depuis votre VM/VPN.
  - Lancer des outils comme Sherlock ou WhatsMyName (qui contactent des centaines de sites).

"Regarder" vs. "Toucher"

Approche Passive (OSINT Pur) :

- **Outil** : Shodan.
- **Action** : Je cherche hostname:"serveur-cible.com" sur Shodan.
- **Résultat** : Shodan me dit que les ports 80, 443 et 22 sont ouverts.
- **Légalité** : **Parfaite**. Je consulte une base de données publique.

Approche Active (HORS OSINT) :

- **Outil** : Nmap (un scanner de ports).
- **Action** : Je lance *nmap -sV serveur-cible.com* depuis ma VM.
- **Résultat** : J'obtiens la même information.
- **Légalité** : **ILLÉGAL**. Un scan de ports est considéré par la loi (Art. 323-1 Code Pénal) comme une tentative d'accès ou de maintien frauduleux dans un système.
- C'est la différence entre "lire le rapport de Shodan" et "tenter d'ouvrir les portes soi-même".

Le Web Scraping

Définition

Collecte automatisée et massive de données sur un site web (vs. collecte manuelle).

Problématique

C'est extrêmement efficace, mais souvent "interdit" par les sites.

Le Web Scraping - Les blocages

Technique (robots.txt) :

- Un simple fichier texte qui "demande" aux robots (ex: Google) de ne pas indexer certaines pages.
- **Valeur légale** : Nulle. Le respecter est une question de "bonne conduite".

Contractuel (CGU / Conditions Générales d'Utilisation) :

- En visitant un site, vous acceptez implicitement ses CGU.
- 99% des CGU interdisent "l'extraction automatisée".
- Violer les CGU n'est pas en soi un crime, c'est une rupture de contrat civil.

Légal (La Loi) :

- Droit des producteurs de bases de données.
- RGPD (si vous scrapez des données personnelles).

Le Web Scraping - Risques

Alors, **je peux scraper ?** C'est un **flou juridique**.

- Cas LinkedIn vs. HiQ (US) : La justice a statué que le scraping de données publiques (profils LinkedIn) n'enfreignait pas la loi (CFAA).
- Cas RyanAir (EU) : La Cour de Justice de l'UE a statué que RyanAir pouvait interdire contractuellement (via ses CGU) le scraping de sa base de données (tarifs).

Conclusion pour l'analyste :

- Scraper des données publiques non-personnelles (articles, prix, offres d'emploi) : Faible risque (bannissement d'IP/compte).
- Scraper des données personnelles (profils, posts de forums) : Risque RGPD élevé (collecte non-minimisée, non-proportionnée).
- Scraper des données derrière un login : **Illégal**. C'est un contournement de mesure de sécurité.

Le Cas des "Leaks"

Rappel : Une brèche de données ("leak") est une base de données (utilisateurs, clients) volée lors d'un piratage et mise en ligne.

La Question Clé : Ai-je le droit de télécharger et d'exploiter une base de données issue d'un vol, même si elle est "publiquement" accessible sur un forum Tor ?

La Loi (France - Art. 323-3 C. Pénal) : "Le fait (...) de détenir, (...) des données (...) résultant d'une atteinte à un système (...) est puni..."

En clair : Télécharger le "leak" peut être assimilé à du recel de données volées.

Leaks : Pratiques “Safe”

HavelBeenPwned

- Ce que vous faites : Vous demandez "Est-ce que cible@email.com est dans le leak XYZ ?".
- Ce que vous ne faites pas : Vous ne téléchargez pas le leak.
- **Légalité** : Parfaite. HIBP est une source ouverte reconnue.

DeHashed / IntelX

- Ce que vous faites : Vous interrogez leur copie des leaks via leur moteur de recherche.
- **Légalité** : C'est une source ouverte payante. Vous payez pour un service d'accès à l'information, vous ne détenez pas la donnée volée.

Exemples : Utilisation d’un “sock puppet” sur X

Action : Consulter son profil via Nitter (interface alternative).

- **Verdict** : LÉGAL (Passif).

Action : Utiliser son pseudo sur Sherlock.

- **Verdict** : LÉGAL (ZONE GRISE - Actif léger). Sherlock "touche" 300 sites. Pas d'intrusion, mais génère du bruit.

Action : Trouver son email via un "leak" sur DeHashed.

- **Verdict** : LÉGAL (Passif - Utilisation d'un service tiers).

Action : Lancer nmap sur le serveur de son blog (trouvé via l'email).

- **Verdict** : ILLÉGAL (Hacking - Scan de port).

Action : Télécharger le "leak" complet où l'email a été trouvé.

- **Verdict** : ILLÉGAL (Recel de données).

La Responsabilité de l'Analyste

Ce n'est pas parce que c'est techniquement possible que c'est légal ou éthique.

Votre Éthique :

- **La Fin vs. les Moyens** : La fin (résultat de l'enquête) justifie-t-elle les moyens (intrusion dans la vie privée) ?
- **Intrusion** : Ne jamais interagir avec la cible, son entourage ou ses contacts.
- **Désinformation** : Votre rôle est de détecter la désinformation, non d'y contribuer.
- **Biais Cognitifs** : Votre pire ennemi reste votre propre biais de confirmation.

Votre Risque :

- Une erreur légale peut entraîner des poursuites judiciaires.
- Elle peut invalider toute votre enquête et détruire votre réputation.

Conclusion

L'OSINT Avancé vous rapproche de la "ligne jaune", mais vous ne devez jamais la franchir.

OSINT Passif (Toujours privilégier) :

- Utilisez Shodan, HIBP, Caches, Wayback Machine.

OSINT Actif (À utiliser avec conscience) :

- Se limite à des interactions minimales (ex: Sherlock).
- Tout scan de port (Nmap) ou Fuzzing (Dirbusting) est HORS OSINT.

Scraping :

- Toléré pour les données publiques, risqué pour les données personnelles (RGPD).

Leaks :

- Utilisez toujours des services tiers (DeHashed, IntelX).
- Ne téléchargez jamais la base de données brute.

QUIZZ

1. Quelle est la différence fondamentale entre l'OSINT passif et l'OSINT actif ?

A. L'OSINT passif concerne les données personnelles, tandis que l'actif concerne les entreprises.
B. L'OSINT passif est gratuit, tandis que l'OSINT actif nécessite des outils payants.
C. L'OSINT actif est toujours légal, contrairement à l'OSINT passif.
D. L'OSINT passif ne touche jamais l'infrastructure de la cible, tandis que l'actif interagit directement avec elle.

2. Selon le RGPD, quel principe s'applique si vous collectez des données personnelles accessibles publiquement sur un forum ?

A. Aucun, car les données ont été rendues publiques par l'utilisateur.
B. Le RGPD ne s'applique qu'aux données bancaires.
C. Vous devez obtenir le consentement écrit de chaque utilisateur avant la collecte.
D. Le principe de minimisation et de limitation de la finalité.

3. Quelle est la valeur juridique du fichier 'robots.txt' qui interdit l'indexation d'un site ?

A. Le non-respect du fichier est une infraction pénale (intrusion).
B. Il a la valeur d'un contrat civil contraignant.
C. Il n'a qu'une valeur technique de 'bonne conduite', mais aucune valeur légale directe.
D. Il est illégal de lire son contenu.

QUIZZ

4. Pourquoi télécharger une base de données issue d'un 'leak' (piratage) pour l'analyser en local est-il risqué juridiquement en France ?

A. C'est légal si la base a été trouvée sur le web de surface.
B. Cela peut être qualifié de recel de données volées (Art. 323-3 C. Pénal).
C. Ce n'est pas risqué tant que vous ne revendez pas les données.
D. C'est risqué uniquement si la base contient des virus.

3. Quelle est la valeur juridique du fichier 'robots.txt' qui interdit l'indexation d'un site ?

A. Le non-respect du fichier est une infraction pénale (intrusion).
B. Il a la valeur d'un contrat civil contraignant.
C. Il n'a qu'une valeur technique de 'bonne conduite', mais aucune valeur légale directe.
D. Il est illégal de lire son contenu.

OPSEC Avancée : Cycle de Vie d'un "Sock Puppet"

Phase de Création

Infrastructure (Fondations) :

- **VM Dédiée** : Le sock puppet ne "vit" que dans sa VM dédiée (ex: CSI Linux , ou une VM de travail dédiée). Il n'existe jamais sur votre machine hôte.
- **IP Dédiée** : L'IP de cette VM doit être neutre (VPN payé anonymement) ou anonyme (Tor/Whonix).

Identité (La "Personne") :

- **Email Anonyme** : Créer un email dédié (Proton.me, Tutanota). NE PAS utiliser Gmail.
- **Numéro de Téléphone** : L'étape la plus critique (nécessaire pour Telegram, Discord, etc.).
    - Option 1 (Virtuel) : Services de numéros virtuels (SMSPVA, etc.). Risqué, car souvent banni.
    - Option 2 (Physique - Gold Standard) : Acheter une carte SIM prépayée, en espèces, avec un téléphone "bête" (burner phone).
- **L’Identité** : Qui est-il ? (Nom, âge, localisation fictive, hobbies). Doit être cohérent.

Phase de Gestion (La "Crédibilisation")

Un compte créé hier est suspect. Un compte créé il y a 5 ans est une source.

Le "Chauffage" (Warming-up) :

- Un nouveau sock puppet ne doit pas enquêter tout de suite.
- Pendant des semaines/mois, il doit avoir une activité "normale" et non-controversée (ex: suivre des comptes de hobbies, retweeter des actualités, poster sur ses faux centres d'intérêt).

Cohérence de l’Identité :

- Si votre puppet est un "développeur", il suit des comptes tech. S'il est "journaliste", il suit l'actualité. Il doit rester dans son rôle.
- Patience : La crédibilité se construit sur la durée.

Phase de Maintenance (Les Erreurs Fatales à Éviter)

Le Risque : La Corrélation (L'ennemi qui relie votre puppet à vous).

Contamination Technique (L'Erreur n°1) :

- Utiliser le même mot de passe ou le même email de récupération que vos comptes réels.
- Utiliser la même photo (même si recadrée).
- Oublier d'activer le VPN/Tor (fuite d'IP).
- Poster une photo avec des métadonnées EXIF (coordonnées GPS de votre domicile) .

Contamination Comportementale (La Signature de l'Analyste) :

- **Analyse Temporelle** : Votre puppet (censé être à New York) poste-t-il toujours à 3h du matin, heure de NY (ce qui correspond à 9h, heure de Paris) ?
- **Analyse Linguistique** : Utilisez-vous le même jargon, les mêmes tics de langage ou les mêmes fautes d'orthographe que sur vos comptes réels ?
- **Analyse de Réseau** : Votre puppet suit-il les mêmes 10 comptes que votre profil réel ?

Sock Puppets 2.0 : L'Ère du Génératif

La Fin de "ThisPersonDoesNotExist"

Les outils historiques (TPDNE) basés sur les GANs sont devenus détectables :

- Yeux toujours centrés ("Alignement parfait").
- Artefacts visuels étranges en arrière-plan.
- Regard "vitreux" reconnaissable par les algorithmes anti-bot de LinkedIn/Facebook.

La Solution : Perchance AI

- Outil de génération de personnages par IA (Stable Diffusion) qui permet de contrôler le résultat via des prompts.
- https://perchance.org/ai-character-generator

Utiliser l'IA pour forger une légende cohérente et unique

Critère	Ancienne Méthode (TPDNE)	Nouvelle Méthode (Perchance)
Unicité (Reverse Search)	Détectable (Image dupliquée ou patterns connus)	**Unique au monde** (Aucun résultat sur Google Images)
Cohérence Légende	Visage aléatoire (On subit le résultat)	**Dirigée** (On demande un âge, un style, un métier précis)
Imperfection Humaine	Trop "parfait" ou artefacts bizarres	**Réaliste** (On peut demander "flou", "mauvais éclairage", "selfie miroir")

Astuce : Demandez toujours une qualité "imparfaite" (ex: "grainy photo", "low light", "casual selfie"). Les profils trop "propres" déclenchent la méfiance des cibles humaines et des algorithmes.

Travaux Pratiques (1 / 2)

Objectif : Créer une identité numérique cohérente et persistante, capable de résister à une vérification superficielle, en utilisant une méthodologie rigoureuse.

Étape 1 : La Genèse (L'Infrastructure)

- Avant de créer le compte, il faut préparer l'environnement.
- **Consigne** : Lancez votre VM (CSI Linux) et assurez-vous que votre VPN est actif pour ne pas corréler votre IP réelle avec la création du compte.

Étape 2 : L’Identité (Cohérence)

- **Outil** : DataFakeGenerator
- **Action** : Générer une identité complète (Nom, Adresse, Date de naissance).
- **Consigne OSINT** : Ne pas se contenter de copier-coller. Il faut "incarner" le profil.
    - Si le générateur donne une adresse à Lyon, le participant doit vérifier sur Google Maps que l'adresse existe et ressemble à une zone résidentielle cohérente avec le métier fictif choisi.

Travaux Pratiques (2 / 2)

Étape 3 : Le Visage (L'IA Générative)

- **Outil** : Perchance AI Character Generator
- **Consigne Avancée** (Prompting) : Évitez la "perfection".
  - Mauvais prompt : "Portrait of a man, professional photo".
  - Bon prompt (selon le cours) : "Casual selfie, bad lighting, grainy photo, mirror selfie".
- **Objectif** : Créer une imperfection humaine pour tromper les algorithmes anti-bot et la méfiance humaine.

Étape 4 : L'Identité Numérique (Email)

- **Outil** : Proton.me
- **Action** : Créer l'adresse email correspondant au nom de la légende (ex: jean.dupont.1985@proton.me).
- **Attention OPSEC** : Ne jamais utiliser cet email pour vos comptes personnels. Ne jamais utiliser un email de récupération personnel.

Le Défi de la Vérification Téléphonique

C'est le **point de friction majeur**. La plupart des plateformes (Twitter/X, Google, Facebook, et parfois Proton lors de l'inscription sous VPN) demandent une validation SMS.

À la question : "**Faut-il acheter un téléphone avec un abonnement acheté dans un bureau tabac ?**", la réponse est nuancée.

- **En théorie**, OUI car cela garantit une séparation physique totale (Air Gap) entre votre vie réelle et votre Sock Puppet. Aucune application ne peut fuiter de données de l'un vers l'autre.
- **En pratique**, la loi oblige l'activation de la ligne via une pièce d'identité. MAIS même si vous devez donner une identité à l'opérateur téléphonique, cette ligne reste séparée de votre numéro personnel principal; Pour une cible (OSINT), ce numéro n'est pas lié à votre identité publique immédiate.

Le Défi de la Vérification Téléphonique

Acheter un téléphone physique est lourd. Vous pouvez utiliser des services de numéros temporaires (SMSPVA, 5sim, OnOff), **MAIS**:

- C'est une solution risquée (numéros souvent blacklistés par les plateformes).
- C'est une solution temporaire (si la plateforme redemande une validation dans 6 mois, le numéro sera perdu).

Recommandation pour les Travaux Pratiques : Utilisez une solution en ligne (type 5sim ou SMS Activate qui acceptent les cryptos ou petites sommes).

Pour aller plus loin : Vérification de Profils & Anti-Fraude

FaceCheck.id : Reconnaissance Faciale IA "À qui appartient vraiment ce visage ?"

- Utilise l'IA pour matcher un visage avec tout le web public.
- Retrouve la personne réelle (victime de vol de photos) derrière le faux profil (ex: un influenceur, un mannequin).
- Très efficace même avec des photos recadrées ou floues.

Scamdigger : Base de Données Criminelle "Ce profil est-il un escroc connu ?"

- Ne cherche pas le "vrai" propriétaire, mais les signalements d'arnaques.
- Vérifie si la photo, l'email ou le pseudo sont fichés dans les bases de "Romance Scams" (Brouteurs).
- Donne souvent le scénario exact utilisé par l'escroc.

Conclusion

Votre sock puppet est un outil.

Traitez-le avec la même rigueur OPSEC qu'un agent de terrain.

La moindre corrélation peut invalider une enquête et vous exposer .

Deep Web & Analyse de Surface Approfondie

Mythe vs. Réalité : Le "Deep Web"

Mythe : Le Deep Web, c'est le Dark Web (Tor, .onion).

Réalité : Le Deep Web représente 90% du web. C'est tout ce qu'un moteur de recherche ne peut pas ou ne veut pas indexer :

- Vos emails et comptes bancaires (derrière login).
- Les bases de données internes.
- Les pages non-liées (orphelines).
- Les API (Application Programming Interfaces).
- Les résultats de recherche de bases de données.

Notre objectif : Trouver les portes d'entrée de ce Deep Web.

Les "Plans" (robots.txt)

Qu'est-ce que c'est ? Un fichier à la racine d'un site (cible.com/robots.txt) qui demande poliment aux robots (comme Google) de ne pas indexer certaines parties du site.

Pourquoi c'est une mine d'or ? C'est une liste, fournie par la cible, des répertoires qu'elle souhaite cacher.

Exemple d'analyse robots.txt (ci-contre)

- Analyse : La cible a des répertoires /admin/, /api/v1/, /backup/…
- Action : Tenter de visiter ces pages. Sont-elles protégées ? (Ne jamais forcer l'accès !)

```
User-agent: *

Disallow: /admin/

Disallow: /private_uploads/

Disallow: /api/v1/

Disallow: /backup/

Disallow: /wp-admin/
```

La Carte Routière (sitemap.xml)

Qu'est-ce que c'est ?

L'opposé de robots.txt. C'est un fichier (cible.com/sitemap.xml) qui donne à Google la liste de toutes les pages que le site veut voir indexées.

Pourquoi c'est une mine d'or ?

- Fournit un index complet de toutes les pages, y compris celles difficiles à trouver.
- Permet de découvrir des "patterns" d'URL (ex: /profils/utilisateur/123).
- Idéal pour le scraping : donne une liste propre de toutes les URL à aspirer.

Fouille de Code Source (Le "Bâti")

Action : Sur une page, faites "Clic-droit > Afficher le code source" (ou Ctrl+U).

Ce que l’on cherche :

- Commentaires des développeurs :
- Liens vers des fichiers "cachés" : <script src="main.bundle.js"></script> <script src="app.1a2b3c.js"></script>
- Ces fichiers JavaScript sont notre prochaine cible.

Le JavaScript (Les "Tuyaux" de l'API)

Les sites web modernes ne sont bien souvent que des "coquilles". Les données (posts, profils) sont chargées via des API (Application Programming Interfaces).

Les adresses de ces API sont écrites en clair dans les fichiers .js.

Action :

1. Ouvrez les Outils de Développement (F12) -> Onglet "Sources".
2. Ouvrez les fichiers .js (ex: app.js, main.js).
3. Cherchez (Ctrl+F) des mots-clés :
    - api/
    - v1/
    - user/
    - token
    - apiKey

Exemple : Vous venez de trouver un "endpoint" d'API. Vous pouvez maintenant l'étudier (passif) pour voir comment le site récupère les données.

Travaux Pratiques

Objectif : Cartographier l'infrastructure cachée d'une cible sans lancer un seul scan actif (Nmap/Dirbuster), uniquement en lisant les fichiers de configuration publics.

Cible : https://www.lemonde.fr

TP 1 - Zones Interdites

Le fichier robots.txt est une liste fournie par la cible des endroits où elle ne veut pas que vous alliez. **C'est donc exactement là où nous voulons regarder**.

Concept à valider :

- Comprendre que ce fichier est une mine d'or pour découvrir des répertoires administratifs ou de test.

Action :

1. Allez sur la page d'accueil de la cible.
2. Ajoutez /robots.txt à la fin de l'URL (ex: cible.com/robots.txt).
3. Analysez les lignes Disallow.

Questions :

- Trouvez-vous des répertoires sensibles (ex: /admin, /wp-admin, /backup, /api) ?.
- Voyez-vous des règles spécifiques pour certains "User-Agents" (Googlebot vs Twitterbot) ?

Rappel Juridique : Respecter ce fichier est une question de "bonne conduite", mais il n'a pas de valeur légale contraignante.

TP 2 : La carte routière

Si le robots.txt est ce qui est caché, le sitemap est l'index complet de ce qui existe. **C'est l'outil idéal pour préparer un scraping massif.**

Concept à valider :

- Utiliser le sitemap pour trouver des pages orphelines ou comprendre la structure des URLs pour un futur scraping.

Action :

1. Tentez d'accéder à cible.com/sitemap.xml.
2. Astuce : Si cela ne marche pas, retournez voir le fichier robots.txt du TP précédent. Souvent, le lien vers le sitemap y est indiqué tout en bas !
3. Naviguez dans le XML.

Questions :

- Quelle est la date de dernière modification (<lastmod>) des pages ? Cela indique-t-il un site actif ou abandonné ?
- Identifiez-vous des "patterns" (motifs) intéressants dans les URLs (ex: /users/1001, /product/v2/) ?.

TP 3 : Les coulisses

C'est la partie la plus avancée ("Deep Web"). Les sites modernes chargent leurs données via des API. Ces connexions sont écrites en clair dans le code JavaScript.

Concept à valider : Trouver des endpoints d'API, des clés oubliées ou des commentaires de développeurs.

Action A (Le Bâti) :

1. Faites Ctrl+U (Afficher le code source).
2. Cherchez des commentaires (souvent en vert). Les développeurs laissent-ils des notes "TODO" ou "FIXME" ?.

Action B (Les Tuyaux - JavaScript) :

1. Ouvrez les Outils de Développement (F12).
2. Allez dans l'onglet "Sources" (ou "Debugger" sur Firefox).
3. Cherchez les fichiers .js (souvent dans un dossier /assets/ ou /js/).
4. Utilisez la fonction de recherche (souvent Ctrl+F ou une icône de loupe) pour trouver les mots-clés du cours : api/, v1/, token, key.

Le "Graal" à trouver : Une URL ressemblant à https://api.cible.com/v1/users cachée dans un script.

Monitoring d'Infrastructure (crt.sh)

Problème : Comment trouver des sous-domaines (comme dev. ou vpn.) que DNS Dumpster ne connaît pas encore ?

Solution : La Transparence des Certificats (CT)

Concept : Pour avoir un site en https://, un administrateur doit demander un Certificat SSL.

Fait : Toutes les demandes de certificats sont enregistrées dans un journal public mondial.

Outil : crt.sh (un moteur de recherche pour ces journaux).

Action :

- Aller sur crt.sh
- Entrer %.cible.com (le % est un joker)

Résultat : Vous voyez tous les sous-domaines (passés et présents) pour lesquels un certificat a été émis (jira.cible.com, test.cible.com, vpn-dev.cible.com...).

Corrélation (Shodan / Censys)

Maintenant, nous lions toutes nos découvertes.

Scénario :

- Sur crt.sh, vous avez trouvé dev.cible.com.
- Vous récupérez son Adresse IP (Ping ou nslookup).
- Vous mettez cette IP dans Shodan / Censys.

Analyse :

- Shodan vous dit que cette IP n'a pas que le port 443 (web) ouvert.
- Elle a aussi le port 3389 (RDP - Bureau à distance) ou 27017 (MongoDB - Base de données).
- **Conclusion** : Vous avez trouvé un serveur de développement, probablement non sécurisé, grâce à une fuite de métadonnées (le certificat SSL).

Synthèse : La Stratégie d'Exploration

1. Domaine Cible (cible.com)
2. Lire les Plans (robots.txt, sitemap.xml)
    - → Découverte de répertoires (/admin, /api)
3. Analyser le Code (.js, Code Source)
    - → Découverte d'endpoints API, clés, commentaires
4. Monitorer l'Infra (crt.sh)
    - → Découverte de sous-domaines (dev.cible.com, test.cible.com)
5. Corréler (Shodan, Censys)
    - → Profiler les IPs des sous-domaines découverts (ports, services).

Travaux Pratiques

Objectif : Découvrir une infrastructure de développement non sécurisée oubliée par une entreprise, en utilisant la transparence des certificats et les moteurs de recherche d'objets connectés.

Pré-requis :

- Navigateur (Tor Browser ou Firefox dans la VM).
- Terminal (dans CSI Linux) pour la résolution DNS.
- Optionnel : Un compte gratuit sur Shodan/Censys (recommandé pour voir plus de détails).

Scénario : Vous auditez l'entreprise "Le Monde". Votre client affirme avoir sécurisé son site principal, mais vous suspectez l'existence de serveurs de test oubliés exposés sur le net.

TP 1 : Les sous-domaines

Concept : Utiliser les journaux de Transparence des Certificats (CT Logs) pour trouver des sous-domaines que DNSDumpster ne voit pas.

Actions :

1. Allez sur crt.sh.
2. Entrez la requête magique avec le joker % : %.lemonde.fr.
3. Le joker est crucial : il demande "tout ce qui finit par https://www.google.com/url?sa=E&source=gmail&q=.lemonde.fr".

Analyse :

- Cherchez des mots-clés "suspects" : dev, staging, test, vpn, jira, jenkins.
- Note : Un sous-domaine dev-api.lemonde.com indique souvent un environnement moins sécurisé que la production.

TP 2 : Pivot DNS

Concept : Transformer la donnée "Administrative" (Nom de domaine) en donnée "Technique" (IP).

Action (Dans le terminal CSI Linux) :

- Choisissez un sous-domaine intéressant trouvé à l'étape précédente (ex: dev.cities.lemonde.fr).
- Utilisez la commande dig vue dans le cours pour obtenir l'IP :
  - dig +short dev.cities.lemonde.fr

Copiez l'adresse IP affichée (ex: 205.234.x.x).

TP 3 : Scanning Passif

Concept : Au lieu de scanner vous-même l'IP avec Nmap (ce qui serait illégal), vous demandez à Shodan ce qu'il sait déjà sur cette IP.

Action :

1. Allez sur Shodan.io.
2. Collez l'IP trouvée à l'étape précédente dans la barre de recherche.

Analyse :

- Quels ports sont ouverts ? (80/443 sont normaux).
- Voyez-vous des ports d'administration dangereux ?
  - 3389 (RDP - Bureau à distance Windows).
  - 22 (SSH).
  - 27017 (MongoDB) ou 9200 (Elasticsearch).
- Regardez les "Vulnerabilities" (CVE) listées par Shodan. Le serveur est-il patché ?

TP 4 - Corraboration

Concept : Shodan scanne l'internet, mais Censys le fait différemment. Il peut voir des choses que Shodan a manquées ou avoir une date de scan plus récente.

Action :

1. Allez sur search.censys.io.
2. Recherchez la même IP.

Comparaison :

- Censys voit-il les mêmes ports que Shodan ?
- Regardez les détails du certificat SSL sur Censys : donne-t-il une adresse email ou une localisation différente ?

TheHarvester : L'Automatisation de la Collecte Passive

Qu'est-ce que c'est ?

- Un outil en ligne de commande (Python) pré-installé sur Kali Linux et CSI Linux.
- Son rôle : C'est un agrégateur. Il interroge simultanément des dizaines de sources publiques pour cartographier la "Surface d'Exposition" d'une cible.

Positionnement dans le Cycle OSINT :

- Phase : Collecte (Reconnaissance).
- Type : Principalement OSINT Passif.
  - Il ne scanne pas directement la cible (sauf option activée).
  - Il interroge des intermédiaires (Moteurs de recherche, Bases de données).
- L'Alternative Moderne : Il remplace les outils vieillissants comme Dmitry ou Metagoofil en centralisant la recherche d'Emails, de Sous-domaines et d'Hôtes virtuels.

Pourquoi l'utiliser ?

- Vitesse : Récupère en 30 secondes ce qui prendrait 2 heures manuellement via Google Dorks.
- Vision 360° : Combine la recherche Corporate (LinkedIn), Technique (DNS/Shodan) et Web (Bing/Google).

TheHarvester : L'Automatisation de la Collecte Passive

Comment fonctionne-t-il ? Il agit comme un méta-moteur. Quand vous lui donnez un domaine (ex: cible.com), il lance des requêtes vers trois types de sources :

1. Les Moteurs de Recherche

- Google, Bing, DuckDuckGo, Yahoo, Baidu.
- Objectif : Trouver des pages indexées mentionnant des emails (@cible.com) ou des sous-domaines.
- Note : Souvent limité par les CAPTCHA sans l'usage de proxies.

2. Les Sources Techniques & Certificats

- crt.sh, DNSDumpster, Netcraft.
- Objectif : Trouver l'infrastructure oubliée (sous-domaines de dev, VPN, etc.) via les logs de transparence.

3. Les API Spécialisées

- Shodan, Hunter.io, IntelX, SecurityTrails.
- Condition : Nécessite d'ajouter vos clés API (gratuites ou payantes) dans le fichier de configuration (api-keys.yaml).
- Résultat : C'est ici que l'outil devient surpuissant, accédant à des bases de données que Google ne voit pas.

TheHarvester : L'Automatisation de la Collecte Passive

La Syntaxe de base: theHarvester -d [domaine] -b [source] -l [limite]

Exemples Concrets :

- Recherche Rapide (Emails & Domaines) : theHarvester -d entreprise.com -b google,bing,linkedin -l 500
  - Récupère les 500 premiers résultats sur les moteurs classiques et LinkedIn.
- Recherche Technique (Infrastructure) : theHarvester -d entreprise.com -b crtsh,anubis,dnsdumpster
  - Focalisé sur la découverte de sous-domaines (TECHINT).
- Le "Full Scan" avec Export (Pour Maltego/Rapport) : theHarvester -d entreprise.com -b all -f rapport_cible
  - Interroge TOUT et sauvegarde les résultats en XML/JSON.

Conseils OPSEC :

- Attention : Bien que "passif", interroger Google 500 fois en 3 secondes est suspect. Utilisez un VPN et ne le faites pas depuis votre IP personnelle.
- Intégration : Le fichier de sortie XML peut être importé directement dans Maltego pour visualiser les liens.

La Vérification Manuelle avec dig

Pourquoi l'utiliser ? Quand les outils automatiques échouent ou donnent des résultats contradictoires, l'analyste doit revenir aux "fondamentaux" pour interroger la source brute.

Les 3 Commandes de Survie :

1. L'Information Brute (Tous les enregistrements) : *dig cible.com ANY +noall +answer*
   - Récupère tout ce que le serveur veut bien donner (MX, TXT, A...).
2. La Chasse aux Vulnérabilités (Transfert de Zone) : *dig axfr @ns1.cible.com cible.com*
   - Tente de télécharger la carte complète du réseau. Rarement ouvert, mais critique si ça marche.
3. L'Analyse de Sécurité Email (SPF/DMARC) : *dig cible.com TXT +short*
   - Affiche instantanément les règles de sécurité email (utile pour voir si le domaine est usurpable).

Avantages de dig

Contrairement à TheHarvester qui est un "chalutier", dig est un "scalpel".

La "Vérité" Technique : Les outils en ligne (DNSDumpster) ou automatisés peuvent avoir du cache. dig interroge directement les serveurs en temps réel.

Le Debugging : Si un outil automatique vous donne un résultat étrange, dig permet de comprendre pourquoi (ex: voir la chaîne de redirection complète).

Features Avancées : Il permet de faire des choses spécifiques comme le transfert de zone (AXFR) ou le traçage de résolution (+trace) que les outils simplifiés masquent.

Comparatif: nslookup vs TheHarvester vs dig

Comparé à nslookup ou à TheHarvester, dig est considéré comme comme un outil de "vérification chirurgicale", non de "collecte massive".

Outil	Rôle OSINT	Statut
nslookup	Interrogation simple	*Obsolète / Trop basique pour un analyste avancé*.
TheHarvester	Collecte massive & automatisée	*Indispensable (votre module 3)*.
dig	Vérification manuelle & Précise	*Utile pour confirmer une hypothèse*.

Travaux Pratiques

Objectif : Automatiser la collecte d'emails et de sous-domaines, puis vérifier manuellement la sécurité DNS d'une cible critique.

Environnement : Terminal de la VM (CSI Linux).

Cible recommandée :

- Pour la partie TheHarvester, utilisez une grande cible (ex: tesla.com).
- Pour la partie dig (Transfert de Zone), nous utiliserons zonetransfer.me (un site conçu pour être vulnérable) afin de garantir un résultat positif.

TP 1 : Collecte Massive (1 / 2)

Concept : Interroger simultanément Google, Bing, et crt.sh pour extraire des emails et des hôtes virtuels .

Commande de base : theHarvester -d [domaine] -b [source] -l [limite]

Action (Dans le terminal) : Lancez une recherche ciblant les moteurs de recherche

- theHarvester -d lemonde.fr -b duckduckgo,crtsh,hackertarget -l 500 -f rapport_lemonde
- Explication des flags :
  - -d : Domaine cible.
  - -b : Sources (Google, Bing, crt.sh). Utiliser all est possible mais plus long.
  - -l : Limite de résultats (500).
  - -f : Sauvegarde le résultat dans un fichier XML/HTML pour le rapport.

TP 1 : Collecte Massive (2 / 2)

Analyse :

- Regardez la section "Emails found". (Souvent vide sans API payantes, mais parfois fructueuse sur Bing).
- Regardez la section "Hosts found". Comparez avec ce que vous aviez trouvé manuellement sur crt.sh.
- Note : Pour une puissance maximale, il faudrait ajouter des clés API (Hunter.io, IntelX) dans le fichier api-keys.yaml .

TP 2 : Vérification Chirurgicale

Concept : Vérifier si le domaine est protégé contre l'usurpation d'identité (Email Spoofing) en lisant les enregistrements TXT .

Action :

- dig lemonde.fr TXT +short

Question : Cherchez la ligne commençant par v=spf1.

- Si elle finit par -all : C'est sécurisé (Rejet strict).
- Si elle finit par ~all : C'est "mou" (Soft fail - souvent marqué comme spam mais accepté).
- Si pas de record : Le domaine est usurpable.

TP 3 : Vérification Chirurgicale - Transfert de Zone (AXFR)

Concept : Demander au serveur DNS de nous donner la carte complète du réseau (tous les sous-domaines d'un coup). C'est une faille critique de configuration.

Cible : zonetransfer.me (Car tout bon site est sécurisé contre ça).

Action 1 (Trouver le Serveur de Noms - NS) :

- dig zonetransfer.me NS +short (Notez l'un des serveurs, ex: nsztm1.digi.ninja.)

Action 2 (Tenter le transfert) :

- dig axfr @nsztm1.digi.ninja. zonetransfer.me

Analyse :

- Si vous voyez une longue liste défiler (robinwood, office, vpn...) : Succès. Vous avez la cartographie complète du réseau interne.
- Sur une cible réelle sécurisée, vous recevriez : *Transfer failed* ou *Refused*.

TP 4 : IP Pivot

Concept : Transformer les résultats de TheHarvester en cibles pour Shodan.

Action :

1. Prenez un sous-domaine intéressant trouvé dans le rapport TheHarvester.
2. Utilisez dig pour obtenir son IP précise : *dig +short dev.cities.lemonde.fr*
3. C'est cette IP que vous entrerez ensuite dans Shodan (comme vu dans le TP précédent).

Intelligence X : Le "Google" du Deep Web

Contrairement à Google qui n'indexe que le Web de Surface actuel, IntelX archive tout, y compris l'éphémère.

Omniscient : Indexe le Dark Web (Tor/I2P), les Pastes (Pastebin), les Leaks de données et les fichiers publics.

Historique : La "Killer Feature". IntelX conserve une copie archivée. Même si le post est supprimé ou le site .onion fermé, la donnée reste accessible chez eux.

Fouille Profonde : Recherche dans le contenu des fichiers (PDF, Excel) et pas juste les titres.

Pourquoi IntelX est indispensable à l'analyste avancé

Le "Buffer" Juridique

Télécharger un leak (.sql) est illégal (recel). IntelX permet de consulter la donnée sur leur serveur sans jamais la posséder sur votre machine. Vous restez dans la zone grise légale.

Le Satellite : Phonebook.cz

Un outil de l'écosystème IntelX. Il liste tous les emails, domaines et URLs liés à un nom de domaine cible. Idéal pour reconstruire un organigramme.

Pivot Crypto & Dark Web

Permet de lier une adresse Bitcoin à un post sur un forum Dark Web archivé il y a 3 ans. C'est souvent le seul moyen de retrouver l'origine d'un wallet.

Travaux Pratiques

Objectif : Utiliser Intelligence X pour retrouver des données historiques effacées et cartographier une organisation, tout en respectant le "Buffer Juridique" (ne pas télécharger les leaks).

Outils :

- Navigateur (Firefox/Tor).
- Phonebook.cz (Satellite d'IntelX).
- Intelx.io (Moteur principal).
- Optionnel : Maltego.

Cible : Nous reprenons lemonde.fr pour comparer avec les résultats de TheHarvester obtenus précédemment.

TP 1 : L’Annuaire

Concept : Phonebook.cz est un outil d'IntelX qui liste tous les emails, domaines et URLs liés à un nom de domaine.

Action :

1. Allez sur phonebook.cz.
2. Entrez le domaine : lemonde.fr.
3. Cochez uniquement "Email" pour filtrer.
4. Lancez la recherche.

Comparaison:

- Combien d'emails voyez-vous ? (Probablement des centaines/milliers).
- Comparez cela avec votre fichier rapport_lemonde.xml de l'atelier précédent. La différence est le "Gap" entre le Web de Surface (Google/Bing) et le Deep Web (Bases de données).

TP 2 : Deep Web et Historique

Concept : IntelX archive tout (Tor, I2P, Pastebin). Même si un post est supprimé, la copie reste accessible chez eux.

Action :

1. Allez sur intelx.io.
2. Entrez le domaine : lemonde.fr.
3. Observez les catégories dans la barre latérale (ou les onglets) :
    - Pastes : (Code fuité sur Pastebin).
    - Darknet : (Mentions sur le réseau Tor/.onion).
    - Leaks : (Bases de données compromises).

Le test du "Buffer Juridique" :

- Cliquez sur un résultat "Paste" ou "Darknet" (vieux de préférence).
- Observation : Vous lisez le contenu sur le site d'IntelX. Vous ne téléchargez pas le fichier sur votre disque dur.
- Rappel Légal : C'est ce qui vous maintient dans la zone grise légale (consultation) plutôt que dans l'illégalité (recel de données volées).

TP 3 : Intégration - Pivot

Option A : Via TheHarvester

- Dans votre terminal CSI Linux, si vous avez une clé API IntelX, vous l'ajoutez au fichier api-keys.yaml.
- La commande serait : theHarvester -d lemonde.fr -b intelx ...
- Cela permettrait à TheHarvester d'aller chercher ces milliers d'emails automatiquement.

Option B : Via Maltego

1. Créez une entité Domain (lemonde.fr).
2. Clic-droit -> Transform -> Search [IntelX].
3. Observez les entités qui apparaissent : vous verrez non seulement des emails, mais potentiellement des Alias (pseudos) ou des Mots de passe (hashs) liés.

   Note : C'est ici que l'on voit la puissance du pivot : Email -> IntelX -> Mot de passe Hashé -> Pseudo -> Sherlock.

DDoSecrets : Le Nouveau WikiLeaks

La Source des Hacktivistes

Distributed Denial of Secrets est le point de chute majeur des données exfiltrées lors de conflits géopolitiques (Russie/Ukraine) ou d'opérations hacktivistes (Anonymous).

Qualité vs Quantité

Contrairement aux bases de "Credential Stuffing" (DeHashed) qui ne contiennent que des mots de passe, DDoSecrets héberge de l'Intelligence Brute :

- Emails internes & Mémos confidentiels.
- Chats (Slack/Teams) d'entreprises compromises.
- Bases de données RH et Financières.

DDoSecrets : Risques & Précautions

L'accès à DDoSecrets n'est plus de la "zone grise". **C'est une zone de danger immédiat.**

Risque Juridique

- Le site fonctionne souvent par Torrents. Vous ne "consultez" pas la donnée (comme sur IntelX), vous la téléchargez.
- Conséquence : Vous êtes techniquement en possession du produit d'un délit (Recel de données - Art 323-3).

Risque Technique

- Les "Dumps" bruts ne sont pas nettoyés. Ils contiennent souvent les malwares et ransomwares qui ont servi à l'attaque initiale, ou des documents piégés.
- Danger : Infection immédiate de votre machine d'analyse.

Protocole Obligatoire : Machine "Burner" (Jetable) totalement isolée (Air Gap) + VPN + Autorisation Juridique/Mandat explicite avant tout téléchargement.

Imageboards Non-Modérés :8kun & 4chan

Une Mine pour la "Threat Intel"

Si votre cible est "Corporate", passez votre chemin. Si votre cible est idéologique ou extrémiste, c'est une source primaire.

- **Radicalisation** : Berceau de mouvements comme QAnon et lieu de publication de nombreux "manifestos" avant des tueries de masse.
- **Fuites "Sales"** : Les documents bannis des plateformes mainstream (Twitter, Reddit) ou des forums de hacking "propres" atterrissent souvent ici.
- **Culture "Chan"** : Comprendre les codes, les mêmes et le langage pour anticiper des actions (raids, désinformation).

8kun & 4chan : Protocole de Sécurité

Attention : Vous entrez dans une zone de non-droit numérique.

Risque Légal

- Ces forums ne sont pas modérés. Vous pouvez tomber accidentellement sur du contenu illégal (CSAM / Pédopornographie). La simple mise en cache sur votre PC peut constituer un délit.

Risque Psychologique

- L'exposition à la haine pure, au racisme et à la violence graphique (Gore) peut causer des traumas à l'analyste. Préparez-vous.

Sécurité Technique

- Ces sites sont hostiles. Ils scannent souvent leurs visiteurs. **Ne jamais y accéder sans une couche d'anonymisation robuste.**

Règle d'Or : VM + VPN + Navigation en mode "TEXTE SEUL" (Désactiver les images dans le navigateur)

Conclusion

Le Deep Web n'est pas le Dark Web. C'est l'ensemble des informations non-indexées (API, pages non-liées, bases de données).

Les fichiers robots.txt et sitemap.xml sont des outils de reconnaissance, pas de simples fichiers techniques.

Le vrai "secret" d'un site web moderne est dans son code JavaScript (API, endpoints).

On ne cherche pas seulement ce qui existe (Google), on monitore ce qui va exister (crt.sh).

Prochaine Étape : Nous avons exploré le Deep Web. Maintenant, nous allons explorer le Dark Web.

Dark Web Opérationnel

Rappel: Le Dark Web ne pardonne pas

- Tor protège votre IP, pas vos actions.
- NE JAMAIS se connecter à un compte personnel (Gmail, X...) via Tor.
- NE JAMAIS télécharger de fichiers (PDF, DOCX) sur votre machine hôte. (Risque de leak de métadonnées ou de malware).
- NE JAMAIS faire confiance à un lien .onion sans l'avoir vérifié (risque de phishing).
- TOUT doit être cloisonné dans une VM dédiée.

Tor Browser vs. Whonix vs. TAILS

Outil	Niveau OPSEC	Comment ça marche ?	Idéal Pour...
Tor Browser (sur Hôte)	**Faible (Niveau 1)**	Simple navigateur. Votre OS est exposé. Risque de fuites (DNS, IP via faille).	**NON RECOMMANDÉ** pour une enquête sérieuse.
Tor Browser (dans une VM)	**Moyen (Niveau 2)**	Isolé de votre Hôte. Si la VM est compromise, votre machine réelle est protégée.	Consultation ponctuelle, analyse de *leaks* simple (non-connecté).
TAILS	**Élevé (Niveau 3)**	OS "amnésique" sur clé USB. Ne laisse **aucune trace** sur le PC. Force *tout* le trafic via Tor.	**Consultation Anonyme :** Visiter un site sensible sans laisser de traces.
Whonix	**Élevé (Niveau 3)**	Solution à **deux VM** : 1 "Gateway" (Tor) + 1 "Workstation" (votre travail).	**Enquête Approfondie :** Télécharger (en sécurité), analyser, écrire des rapports.

Cartographie : L'Économie du Dark Web (1/2)

Rappel : Il y a des forums et des marketplaces. Ils forment une économie structurée.

Les Acteurs Clés (Les noms changent, les rôles restent) :

- **Forums** (Information & Vente) :
  - Exemples : BreachForums (nouvelle version), XSS, Exploit.in
  - Rôle : Vente de leaks de BDD, d'exploits, d'accès initiaux (IABs). C'est là que la Threat Intelligence commence.
- **Marketplaces** (Produits & Services) :
  - Exemples : (Les noms changent constamment : AlphaBay, ASAP, etc.)
  - Rôle : Vente de produits (drogues, malwares "as-a-service", faux papiers).

Cartographie : L'Économie du Dark Web (2/2)

Les "Hubs" de Discussion :

- Exemple : Dread (le "Reddit" du Dark Web).
- Rôle : C'est là que l'on prend le "pouls". Les utilisateurs font la revue des marketplaces, signalent les arnaques (scams), discutent des méthodes.
- Valeur OSINT : Source incroyable pour évaluer la réputation d'un vendeur ou d'un site.

Les "Règles" de l'Économie :

- Réputation : Tout est basé sur la réputation. Un vendeur inconnu ne vend rien. On cherche les "vendeurs" établis.
- Escrow (Tiers de confiance) : Le marketplace (ou un admin du forum) garde l'argent (crypto) jusqu'à ce que l'acheteur confirme la réception.
- PGP (Chiffrement) : Utilisé pour chiffrer les communications et, surtout, pour vérifier l'identité d'un vendeur (via sa clé PGP publique).

Recherche Avancée : Au-delà d'Ahmia

Phase 1 : Le Pivot Journalistique

- **Scénario** : Un groupe de ransomware "RansomX" revendique une attaque.
- **Action** :Un Googler "RansomX" "leak".
- **Résultat** : Des blogs de sécurité (BleepingComputer, KrebsOnSecurity, ZDNet) vont couvrir l'événement. Ils ne donneront pas le lien .onion, mais ils citeront le nom du forum (ex: "sur BreachForums...").

Phase 2 : Le Pivot "Wiki"

- **Action** : Une fois le nom du forum connu ("BreachForums"), chercher un annuaire de liens fiable sur le Surface Web (ex: dark.fail) pour trouver l'URL .onion actuelle et légitime.
- **Attention** : Googler "Hidden Wiki" vous mènera à 99% de sites de phishing.

Phase 3 : Le Pivot Crypto

- Une adresse BTC/XMR est trouvée sur un post de blog.
- **Action** : Rechercher cette adresse (Google, Ahmia). Elle peut être liée à un profil sur un forum.

Travaux Pratiques

Contexte : Vous enquêtez sur un groupe de hackers. Votre seule piste est une adresse email trouvée dans un leak (via DeHashed) : foxtrotter@sc.rr.com

Mission : Trouver le profil de cet utilisateur sur un forum du Dark Web.

Hypothèse : L'utilisateur réutilise des parties de son email (foxtrotter) comme pseudo.

Travaux Pratiques

Action 1 (Surface Web) :

- Googler "foxtrotter" breachforums
- Googler "foxtrotter" xss
- **Analyse** : On trouve peut-être des discussions sur lui, ou un lien vers son profil (rare).

Action 2 (Moteurs DW) :

- Lancer TOR dans une VM
- Aller sur Ahmia.fi.
- Rechercher foxtrotter.
- **Analyse** : Ahmia indexe les profils publics. Il peut trouver onion-forum-xyz.onion/user/foxtrotter.

Action 3 (Vérification) :

- Naviguer (en sécurité) vers le lien .onion trouvé.
- **Corroboration** : Le profil utilise-t-il PGP ? La date d'inscription est-elle ancienne (crédibilité) ?
- **Pivot** : Lire ses posts. Avec qui parle-t-il ? Quelles adresses crypto poste-t-il ?
- **Résultat** : Nous avons lié un email (Surface) à un profil (Dark).

Conclusion

L'OPSEC est reine : Utilisez **TAILS** (consultation) ou **Whonix** (analyse). Le Tor Browser simple n'est pas suffisant pour une enquête.

Le Dark Web n'est pas "introuvable", il est juste "non-indexé". C'est un écosystème avec ses propres règles (réputation, PGP, escrow).

Oubliez la recherche "Google", **la méthode la plus efficace est le pivot** :

- Surface (Blogs, X) → Nom du Forum
- Wiki Fiable (dark.fail) → URL .onion légitime
- Recherche interne (Forum) → Cible

Prochaine Étape : Nous avons vu la collecte manuelle avancée (Deep & Dark). C'est lent. Nous allons maintenant voir comment automatiser cette collecte.

Web Scraping Avancé

Le Problème : La "Noyade de Données"

Rappel : L'analyste est face à 50+ onglets ouverts et des notes partout.

Le problème s'aggrave :

- **Volume** : Vous devez analyser 500 profils d'employés, pas 2.
- **Vitesse** : Vous devez suivre 20 forums, pas 1.
- **Complexité** : L'information est dans un sitemap.xml de 10 000 lignes.

La collecte manuelle est une perte de temps.

La solution : L'extraction automatisée (Web Scraping).

Statique (HTML) vs. Dynamique (JS)

C'est la distinction la plus importante. L'outil dépend de la nature du site.

Statique (Le Papier Imprimé)

- **Quoi** : Le contenu (texte, liens) est directement dans le code HTML de la page.
- **Exemple** : Wikipédia, un blog simple, la plupart des sites d'actualité.
- **Outils** : Simple. requests (Python), extensions No-Code.

Dynamique (L'Ardoise Magique")

- **Quoi** : La page (HTML) arrive vide. C'est du JavaScript (JS) qui s'exécute ensuite dans votre navigateur pour aller chercher les données (posts, commentaires) à afficher sur une API.
- **Exemple** : X (Twitter), Facebook, LinkedIn.
- **Outils** : Complexe. Il faut un vrai navigateur (Selenium) ou interroger l'API.

Statique ou Dynamique : Le Test en 10 secondes

Scénario : Vous voulez scraper un post sur un site.

1. Allez sur la page.
2. Faites Clic-droit > Afficher le code source (Ctrl+U).
3. Recherchez (Ctrl+F) un morceau du texte du post.
    - Résultat A : Vous trouvez le texte ?
        - → STATIQUE. C'est facile.
    - Résultat B : Vous ne trouvez pas le texte ? (Mais vous le voyez à l'écran).
        - → DYNAMIQUE. Le JavaScript l'a chargé. C'est difficile.

Partie 1 : Les Outils No-Code

Objectif : Extraire des données structurées (listes, tableaux) sans écrire une ligne de code.

Idéal pour :

- Scraper des résultats de recherche Google.
- Scraper une liste d'employés sur un site simple.
- Scraper les prix d'un produit.

Outils :

- Extensions de navigateur (ex: Web Scraper, Data Miner).
- Applications visuelles (ex: ParseHub, Octoparse).

Outil 1 : Extensions Navigateur (Web Scraper)

1. Vous ouvrez l'extension (dans les Outils Développeur F12).
2. Vous créez un "Sitemap" (un plan de scraping).
3. Vous "pointez et cliquez" pour apprendre à l'outil :
4. "Ceci est le titre que je veux" (Sélecteur 1).
5. "Ceci est le lien que je veux" (Sélecteur 2).
6. Vous lancez le scraping.
    - Il parcourt la page (ou les pages) et génère un CSV/Excel.
    - **Pros** : Visuel, rapide, intégré au navigateur.
    - **Cons** : Se "casse" facilement si le site change de design.

Outil 2 : Applications Visuelles (ParseHub)

C'est une application dédiée (pas une simple extension).

Même principe de "pointer-cliquer".

Fonctionnalités Avancées (Pros) :

- Gère mieux les sites dynamiques (JS) complexes.
- Gère la pagination ("Cliquer sur le bouton 'Suivant'").
- Gère la navigation ("Cliquer sur chaque titre, aller sur la page, scraper le contenu").
- Peut exécuter le scraping "dans le cloud" (sur leurs serveurs).

Cons : Plus lourd, le "free tier" (version gratuite) est souvent limité.

Partie 2 : Low-Code/Code

Objectif : Comprendre ce qui se passe "sous le capot" pour gagner en flexibilité, en puissance et en automatisation.

Le "Pourquoi" du Code :

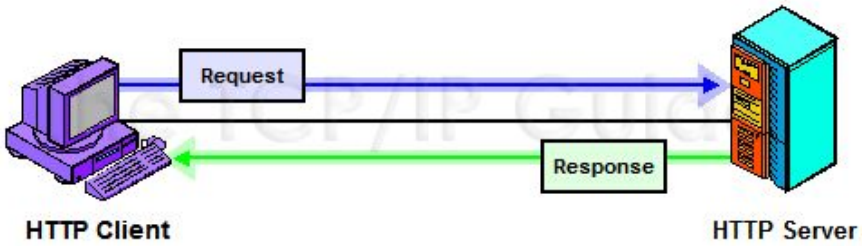
- Les outils No-Code échouent sur des sites complexes.
- Vous devez intégrer le scraping dans un script plus large (ex: scraper, puis analyser, puis envoyer une alerte).
- Vous devez gérer les blocages (Cf. Partie 3).

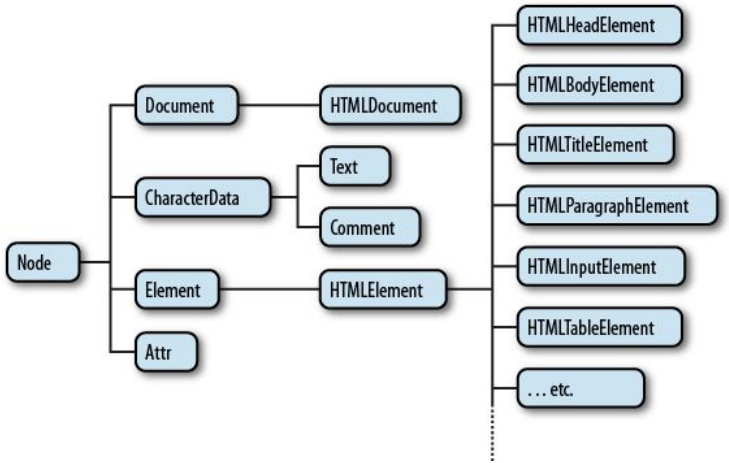
La "Boîte à Outils" : Python

- C'est le langage "standard" de l'OSINT et de la Data Science.

HTTP - HTML

Avant d'utiliser les libraires de scraping, il est important de bien comprendre comment fonctionne une requête HTTP, et quel type de structure nous allons traiter.





L'Écosystème Python

Vous n'avez pas besoin d'être un "développeur", mais vous devez savoir quels outils existent.

- **requests** (Le "Facteur")
  - Rôle : Va chercher la page (le HTML brut). Incroyablement rapide.
  - Limite : Ne comprend pas le JavaScript. Inutile sur les sites Dynamiques.
- **BeautifulSoup4** (bs4) (Le "Traducteur")
  - Rôle : Prend le HTML brut (difficile à lire) de requests et le transforme en un objet propre que l'on peut interroger ("Donne-moi tous les liens <a>").
- **Selenium** (Le "Robot Navigateur")
  - Rôle : Ouvre et contrôle un vrai navigateur (Firefox, Chrome).
  - Force : Exécute le JavaScript (JS), donc il voit les sites Dynamiques.
  - Faiblesse : Beaucoup plus lent (car il charge tout : images, pubs...).
- **Pandas** (L'"Analyste")
  - Rôle : Prend vos résultats (listes) et les sauvegarde proprement en CSV/Excel.

Exemple : Script Statique (Requests + BS4)

Objectif : Scraper les titres d'un blog (Statique).

```
import requests
from bs4 import BeautifulSoup

# 1. Le "Facteur" (Requests) va chercher la page
url = "http://exemple-blog-statique.com"
page = requests.get(url)
# 2. Le "Traducteur" (BS4) lit le HTML
soup = BeautifulSoup(page.content, 'html.parser')
# 3. On demande ce qu'on veut
# "Trouve toutes les balises <h2> qui ont la classe
'titre-article'"
titres = soup.find_all('h2', class_='titre-article')
for titre in titres:
    print(titre.text.strip())
```

Exemple : Script Dynamique (Selenium)

Objectif : Scraper les posts d'un site (Dynamique).

```
from selenium import webdriver
from selenium.webdriver.common.by  import By
import time

# 1. Le "Robot Navigateur" (Selenium) ouvre Firefox
driver = webdriver.Firefox()
# 2. Il visite l'URL (comme un humain)
driver.get("http://exemple-site-dynamique.com" )
# 3. IMPORTANT : On attend que le JS se charge !
time.sleep(5) # Attendre 5 secondes
# 4. On demande ce qu'on veut
# "Trouve tous les éléments qui ont la classe 'post-contenu'"
posts = driver.find_elements(By.CLASS_NAME,  'post-contenu' )
for post in posts:
   print(post.text)
driver.quit()  # Ferme le navigateur
```

Travaux Pratiques

Objectif : Créer un script Python capable de scraper automatiquement des données, de les nettoyer et de les exporter proprement pour une analyse ultérieure (Excel/Maltego).

Cible d'entraînement : http://quotes.toscrape.com (Site bac-à-sable légal conçu pour apprendre le scraping).

TP 1 - Préparation de l’environnement

1. Cloner le projet
   - git clone https://github.com/akoudri/osint-training.git
2. Dans le dossier créé, instancié un environnement virtuel
   - cd osint-training
   - virtualenv -p python3 .venv
3. Activer l’environnement virtuel
   - source .venv/bin/activate
4. Installer les dépendances
   - pip install -r requirements.txt
5. Ouvrir le projet avec un IDE (pycharm ou studio code)

TP 2 - Requests

Point clé : Montrer l'importance du dictionnaire HEADERS.

Exercice : Dans le fichier “static_scraping”, retirez le HEADERS et de lancer le script sur un site protégé (ex: Amazon ou Google). Observez le résultat.

TP 3 - BeautifulSoup

Action : Ouvrez quotes.toscrape.com dans leur navigateur.

Inspecter l'élément : Faites Clic-droit > Inspecter sur une citation.

Lien Code/UI : Observez que la balise dans le navigateur correspond à soup.find_all('div', class_='quote') dans le Python.

TP 4 - Pandas

Pourquoi Pandas ?

La gestion des fichiers texte (.txt ou .csv) est un cauchemar dès qu'il y a des virgules ou des retours à la ligne.

L'avantage OSINT :

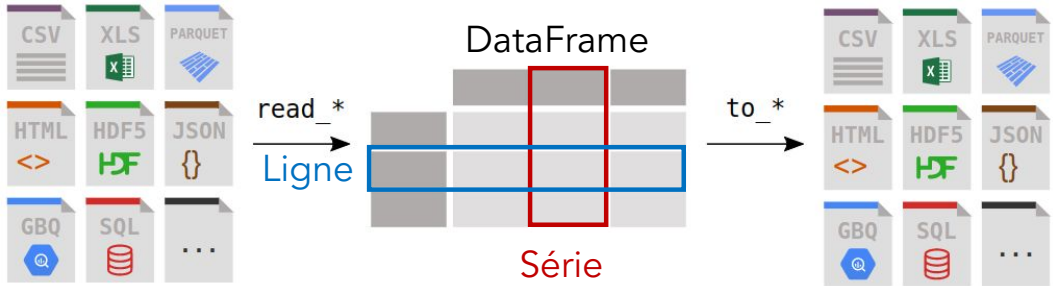
Le fichier .csv généré peut être :

- Ouvert dans Excel pour filtrer.
- Importé dans Maltego pour transformer les "Auteurs" en entités.

Exercice bonus: Le script ne scrape que la page 1. Modifiez le code pour qu'il trouve le lien 'Next' (bouton Suivant) et qu'il scrape automatiquement les 5 premières pages.

Précisions sur Pandas

- Pandas permet de travailler sur des données tabulaires
  - Données stockées dans des feuilles de calcul
  - Données stockées des bases de données
- Pandas vous permet d'explorer, nettoyer et traiter vos données de manière simple et efficace
- Le **DataFrame** est le concept central de Pandas



https://pandas.pydata.org/pandas-docs/stable/reference/index.html

Travaux Pratiques

Objectif : Créer un robot capable de s'authentifier sur un espace membre protégé par mot de passe, de vérifier que la connexion a réussi, et d'extraire des données réservées aux membres.

Cible d'entraînement : http://quotes.toscrape.com/login

TP 1 - Reconnaissance

Avant de coder, vous devez comprendre comment le site fonctionne. Un robot ne peut pas deviner où cliquer.

1. Ouvrez http://quotes.toscrape.com/login dans votre navigateur (Firefox de préférence).
2. Faites Clic-droit > Inspecter (ou F12) sur la case "Username".
    - Question : Quel est l'attribut unique qui permet d'identifier ce champ ? (Cherchez id="..." ou name="...").
3. Faites de même pour la case "Password".
4. Connectez-vous manuellement.
    - Observation : Qu'est-ce qui change sur la page une fois connecté ?
    - Indice : Regardez en haut à droite. Que devient le lien "Login" ?

TP 2 - Développement du Robot

1. Inspectez et complétez éventuellement le code “dynamic_scraping.py”
2. Exécutez le code et observez le résultat
3. Modifiez le code si le résultat ne correspond pas à vos attentes
4. Itérez si besoin

TP 3 - Analyse

Pourquoi wait.until est-il important ici ?

Pourquoi utiliser Keys.RETURN plutôt que de chercher le bouton "Login" et cliquer dessus ?

Comment ce script réagirait-il si le mot de passe était faux ?

Question bonus: Certains sites protègent leurs formulaires avec un champ caché appelé CSRF Token. Comment feriez-vous pour contourner ce problème ?

Partie 3 : Le Jeu du Chat et de la Souris

Le Scénario : Votre script (ou outil No-Code) fonctionnait... et soudain, il ne fonctionne plus.

- Vous recevez une "Erreur 403 Forbidden" (Accès refusé).
- Vous recevez une "Erreur 429 Too Many Requests" (Trop de requêtes).
- Vous recevez un CAPTCHA ("Je ne suis pas un robot").

Le site a détecté que vous êtes un script. C'est là que l'analyse avancée commence.

Problème 1 : Blocage "User-Agent"

Le Problème : Par défaut, votre script requests se présente au serveur en disant : "Bonjour, je suis python-requests/2.28"

- Réponse du Serveur : "Un script ? Dehors. (Erreur 403)"
- Le User-Agent (UA) : C'est la signature de votre navigateur. Un vrai navigateur dit : "Bonjour, je suis Mozilla/5.0 (Windows NT 10.0; Win64; x64) ..."

La Solution : Le "Spoofing" (Usurpation)

- **Action** : Vous mentez. Vous copiez un vrai UA (cherchez "My User Agent" sur Google) et vous l'ajoutez aux "headers" (en-têtes) de votre script.
- **Résultat** : Le serveur pense que vous êtes un vrai navigateur Chrome ou Firefox.

Problème 2 : Blocage d'IP

Le Problème : Vous faites 1000 requêtes en 10 secondes depuis la même adresse IP.

- Comportement Humain : Impossible.
- Réponse du Serveur : "Cette IP est un bot. Bloquée pour 24h. (Erreur 429)"

Solution 1 : Le Ralentissement (Être poli)

- **Action** : Ajoutez time.sleep(2) entre chaque requête.
- **Résultat** : Vous avez l'air plus humain. Simple, mais lent.

Solution 2 : Les Proxies Rotatifs (Changer d'IP)

- **Action** : Vous utilisez un service de "Proxy" payant qui vous donne accès à un pool de milliers d'IPs (ex: https://www.scraperapi.com)
- **Résultat** : Chaque nouvelle requête (requests.get) est faite depuis une nouvelle adresse IP. Le serveur ne peut pas corréler les requêtes.

Problème 3 : Le CAPTCHA

Le Problème : Le serveur a encore des doutes et vous présente le test "Cliquez sur les feux de circulation".

- Résultat : Votre script requests ou Selenium est bloqué.

Solutions (Complexes) :

- Services de Résolution (API) :
  - Exemples : 2Captcha, Anti-Captcha.
  - Action : Votre script envoie le CAPTCHA à l'API, un humain (ou une IA) le résout en quelques secondes et renvoie la solution.
  - Coût : Payant. Efficace, mais éthiquement "gris".
- Selenium "Warmed-Up" (Chauffé) :
  - Action : Utiliser Selenium avec un profil de navigateur qui est déjà connecté à un compte Google (un "sock puppet").
  - Résultat : Google vous fait confiance et vous donne un CAPTCHA plus simple (juste une case à cocher) que Selenium peut cliquer.

Conclusion

Le scraping automatisé est essentiel pour passer de "l'artisanat" à "l'industrie" de l'OSINT.

La distinction Statique (HTML) vs. Dynamique (JS) est la clé qui détermine votre outil.

Commencez toujours No-Code (Extensions). C'est 80% des besoins.

Gardez le Code (Python/Selenium) pour les 20% (sites complexes, automatisation).

Soyez prêt pour votre investigation : préparez vos User-Agents et vos Proxies.

Rappel Éthique : Respectez les CGU (ou comprenez les risques), ne submergez pas les serveurs et soyez TRÈS prudent avec le scraping de Données Personnelles (RGPD !).

Pour aller plus loin: https://rapidapi.com/hub https://github.com/public-apis/public-apis

Travaux Pratiques
Profilage d'Entreprise Avancé

Posture OPSEC et Légalité

Pendant cet atelier, vous devez respecter :

Votre OPSEC :

- Toutes les actions doivent être faites depuis votre VM d'enquête.
- Utilisez un VPN ou Tor (si l'outil le permet) pour masquer votre IP.
- N'utilisez aucun compte personnel (LinkedIn, Google).

La Légalité :

- **Scraping** : Vous scrappez des données publiques. C'est une zone grise contractuelle (CGU), mais vous ne forcez aucun accès.
- **Leaks** : Vous utilisez des services tiers (DeHashed, HIBP). Vous ne téléchargez pas les bases de données volées.
- **Scans** : Vous utilisez Shodan (Passif). N'utilisez PAS Nmap (Actif/Illégal).

Objectifs et Scénario

Scénario : Votre cabinet d'intelligence économique doit auditer l'empreinte numérique de Tesla Inc. avant une fusion stratégique. Le client veut connaître l'état de leur infrastructure oubliée et l'exposition de leurs employés clés.

Votre Mission : Vous devez, en groupe, collecter des informations clés en utilisant les techniques avancées vues aujourd'hui.

Cible: tesla.com

Objectifs de l'atelier :

1. Appliquer la recherche de sous-domaines.
2. Utiliser le scraping No-Code pour extraire une liste.
3. Analyser les offres d'emploi pour la stack technique.
4. Pivoter sur les leaks de données.

Pourquoi cette cible ?

Richesse : Il y a de la matière sur tous les plans (Infra, Humain, Tech).

Sécurité : Tesla possède un programme de Bug Bounty très actif. Si vous trouvez par hasard une vraie faille critique (peu probable en passif, mais possible), vous êtes protégé tant que vous la signalez de manière responsable.

Plan de Mission

Tâche 1 : Infrastructure

- Outil : crt.sh
- Mission : Trouvez 3 sous-domaines "non-évidents" (ex: dev, test, vpn, git, jira...).
- Question Bonus : Que nous dit Shodan sur l'IP de l'un de ces sous-domaines ?

Tâche 2 : Stack Technique

- Outils : Google, LinkedIn Jobs, APEC...
- Mission : Trouvez des offres d'emploi (actuelles ou passées) pour l'entreprise.
- Analyse : Listez 3 technologies de leur stack technique (ex: "AWS", "Python", "React", "MongoDB"...).

Plan de Mission

Tâche 3 : Personnel

- Outil : Extension Web Scraper (ou ParseHub).
- Cible : Chercher une cible sur LinkedIn - société AlphaDynamics
- Mission : Scrapez la liste des 10 premiers "Employés" (Nom + Titre du poste).
- Résultat : Produire un CSV.

Tâche 4 : Vulnérabilités

- Outils : HIBP, DeHashed, IntelX.
- Mission : En vous basant sur le format d'email de l'entreprise (ex: prenom.nom@alphadynamics.com), vérifiez si des employés (trouvés en Tâche 3) sont présents dans des leaks.

Restitution

Chaque groupe devra présenter :

- Trouvailles Infrastructures : "Nous avons trouvé dev.tesla.com sur crt.sh, et Shodan montre un port RDP ouvert."
- Trouvailles Techniques : "Leurs offres d'emploi mentionnent 'Elasticsearch v6', qui est obsolète et vulnérable."
- Succès du Scraping : "Nous avons réussi à extraire une liste de 10 employés au format CSV en 5 minutes."
- Trouvaille de Leaks : "L'email du CTO, trouvé sur la liste, est présent dans le leak 'Adobe' de 2013, révélant un mot de passe récurrent."

Conclusion Module 1

Ce que nous avons accompli :

- Franchi la "Ligne Grise" : Nous avons défini les limites légales du scraping et de l'exploitation des leaks.
- Exploré le "Caché" : Nous avons utilisé les robots.txt, crt.sh et le code source pour trouver des informations non-indexées.
- Opéré dans le "Sombre" : Nous avons établi une méthodologie sécurisée pour investiguer sur le Dark Web.
- Automatisé la Collecte : Nous avons appris à utiliser des outils No-Code et compris les concepts (Statique/Dynamique) du scraping avancé.

Nous avons collecté énormément de données. Dans la suite, nous allons apprendre à les analyser (SOCMINT avancé, Graphes) et à les visualiser.

Module 2
De la Donnée à l'Intelligence

Maîtrise des Outils d’Analyse

Rappel

Maltego est une plateforme d'analyse de liens.

Tout est basé sur 3 concepts :

- Entités : Les "nœuds" (un Email, une IP, une Personne) .
- Transforms : Les "scripts" qui trouvent des liens (ex: Domaine → IP) .
- Graphes : Le "tableau blanc" visuel .

La puissance vient des clés API (Shodan, VT, etc.) .

Problème : Le "clic-droit" est limité. Un analyste avancé ne se contente pas des transforms par défaut. Il importe et analyse.

Importer vos Données (CSV/Excel)

Le Scénario : Vous avez scrapé un CSV de 100 employés et leurs titres. Comment l'analyser ? Un tableur est mauvais pour voir les connexions.

Solution : L'importation Maltego

- Maltego permet d'importer des fichiers plats (CSV, XLSX).
- Vous "mappez" (associez) vos colonnes à des Entités Maltego.
  - Colonne "Nom" → Entité maltego.Person
  - Colonne "Poste" → Entité maltego.Phrase
  - Colonne "Entreprise" → Entité maltego.Company

Résultat : Votre CSV plat devient un graphe interactif.

Exemple d'Importation

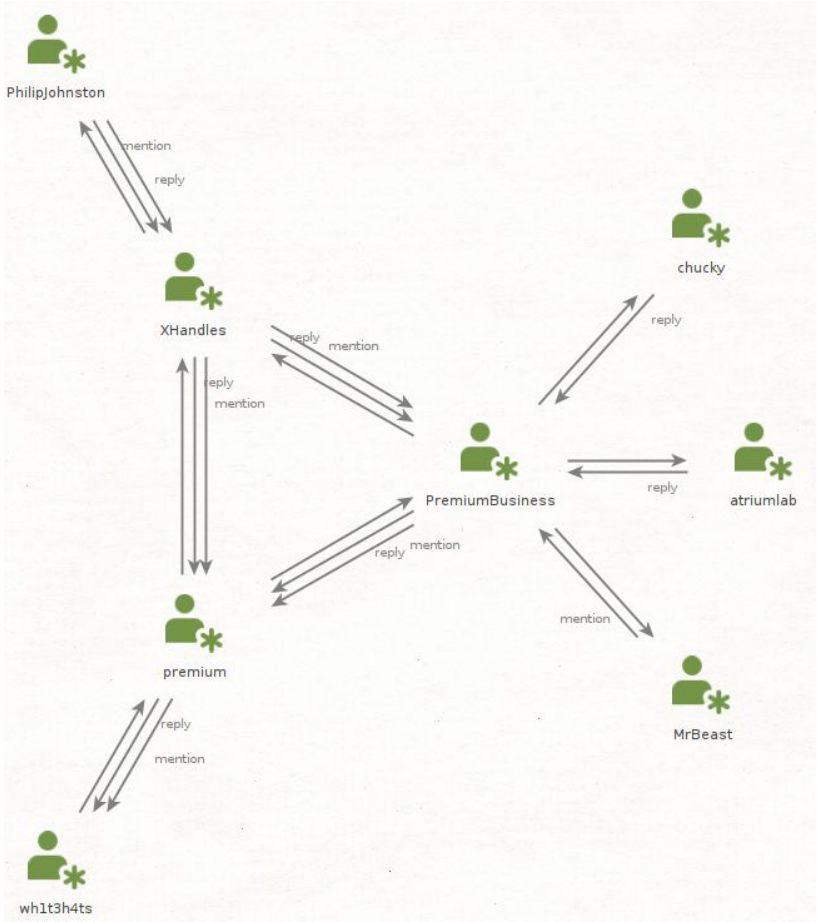
Étape 1 : Vos Données (CSV)

| Prénom.Nom | Poste | Entreprise | | :--- | :--- | :--- | | Jean Dupont | CEO | AlphaTech | | Marie Durand | CTO | AlphaTech | | Paul Martin | Dev | AlphaTech | | Alice Leroy | CEO | BetaCorp | | Jean Dupont | Conseil | BetaCorp |

Étape 2 : Le Graphe Maltego (Après Import)

(Le graphe montrera "Jean Dupont" connecté à "AlphaTech" ET "BetaCorp", le révélant instantanément comme un pont (SNA), ce qu'un tableur cache.)

Analyse : Le graphe révèle immédiatement que "Jean Dupont" est le lien (le pont) entre les deux entreprises.



Stratégie des Transforms Payantes

Rappel : Il faut des clés API . **Niveau Avancé** : Ce ne sont pas des "plus", ce sont des multiplicateurs de pivot.

Focus : Les "Leaks" (IntelX / DeHashed)

- Ces transforms sont la version "légale" et intégrée de l'analyse de leaks (Bloc 1).
- Elles permettent le pivot ultime.

Focus : Shodan - Workflow Avancé

- Trouver une IP (via Domaine).
- Lancer la transform Shodan "Query Host".
- L'IP est taguée comme "MongoDB" ou "RDP".
- Pivot : Créer une Entité "Tag:MongoDB" et chercher (via Shodan) d'autres IPs avec le même tag appartenant à la même organisation.

Workflow : Pivot via "Leak" (IntelX)

1. **Entrée** : Vous avez une Entité Email (cible@email.com).
2. **Transform** : Clic-droit → Lancer "Search [IntelX]".
3. **Sortie** (Nœuds Verts) : La transform retourne automatiquement :
    - maltego.Alias (Pseudo) : SuperCible99
    - maltego.Alias (Pseudo) : Cible_Fan_Jeu
    - maltego.Password (Hash) : P@ssword123!
    - maltego.IPAddress : 123.45.67.89 (IP du leak)
4. **Analyse** : Vous avez 3 nouveaux pivots instantanés. Vous pouvez maintenant lancer des transforms Sherlock/WhatsMyName sur ces nouveaux pseudos.

Analyse de Graphe (Layouts)

Le Problème : Votre graphe a 500 nœuds. C'est illisible (un "Hairball" ou "Plat de spaghettis").

La Solution : Les "Layouts" (Mises en page).

- Ce sont des algorithmes qui réorganisent votre graphe pour révéler des patterns (liens directs avec SNA).
- L'outil ne fait que visualiser. C'est l'analyste qui analyse.

Les "Layouts" Clés

Organic / Force-Directed (Par défaut)

- **Quoi** : Simule la physique (les nœuds liés s'attirent, les autres se repoussent).
- **Révèle** : Les Clusters (groupes) et les Ponts (connexions entre groupes).
- **Usage** : Parfait pour l'analyse SNA.

Block / Hierarchical (En Blocs)

- **Quoi** : Regroupe les nœuds par "Type d'Entité".
- **Révèle** : "Toutes mes IPs", "Tous mes Emails".
- **Usage** : Idéal pour "nettoyer" le graphe.

Les "Layouts" Clés

Circular (Circulaire)

- **Quoi** : Place les nœuds en cercle.
- **Révèle** : Les Hubs. Un nœud au centre avec des "rayons" vers l'extérieur est un connecteur central.
- **Usage** : Trouver les influenceurs ou les pivots (SNA).

Timeline (Chronologique)

- **Quoi** : Aligne les nœuds par date (si l'entité a une date).
- **Révèle** : La chronologie d'une attaque ou d'une enquête.
- **Usage** : Analyse temporelle.

Transforms Locales (TDS)

Le Problème :

- Vous avez une API (interne, ou non-supportée) que Maltego ne connaît pas.
- Vous avez un script (Python, Bash) que vous exécutez toujours manuellement (ex: un scraper custom).
- Vous voulez automatiser une recherche complexe.

La Solution : Les Transforms Locales (TDS - Transform Distribution Server)

Concept : Vous écrivez un simple script (Python, etc.) qui prend une Entité en entrée (ex: un Domaine) et retourne du texte en sortie (ex: une IP).

Maltego "habille" ce script pour qu'il apparaisse dans votre menu "clic-droit".

Transforms Locales - Pratique

Avant (Manuel)

- Clic-droit sur cible.com → Copier l'Entité.
- Ouvrir un terminal.
- Taper python mon_script.py cible.com.
- Lire le résultat (1.2.3.4).
- Retourner dans Maltego.
- Créer une Entité IPAddress manuellement.
- Lier les deux.

Après (Transform Locale)

- Clic-droit sur cible.com → "Lancer [Ma Super Transform]".
- Maltego exécute votre script et ajoute l'IP 1.2.3.4 au graphe.

Les transforms locales seront l’objet du prochain atelier pratique.

L'Alternative : Lampyre

Modèle Économique : Pay-per-Request. Vous payez pour ce que vous utilisez (les "Lampyres").

Focalisation : Considéré comme plus puissant sur certains axes très spécifiques.

Gestion du Big Data : Conçu pour gérer des graphes de très grande taille (plus que Maltego CE).

Maltego vs. Lampyre

Utilisez Maltego (Standard de l'Industrie) pour :

- L'Infrastructure / Cyber (TECHINT) : Son écosystème de transforms (Shodan, VT, SecurityTrails) est imbattable.
- La Flexibilité : Les transforms locales (TDS) le rendent adaptable à l'infini.
- L'Intégration : La version "Community" (CE) est gratuite et puissante.

Utilisez Lampyre pour :

- Le SOCMINT Avancé : Il dispose de modules intégrés très puissants pour Telegram, VK, etc.
- Le BLOCKCHAIN-INT (Financier) : Modules natifs pour l'analyse Crypto (BTC, ETH).
- Le PIVOT de Leaks : Très fort sur l'analyse de leaks et PII (Emails, Téléphones).

SOCMINT Avancé
Analyse Comportementale et RS

Rappel : SOCMINT

Formation Initiale : L'Art de Trouver

- Quoi : Énumération de pseudos (Sherlock).
- Quoi : Recherche simple sur plateformes (X, LinkedIn).
- Quoi : Utilisation d'opérateurs (from:, site:).
- Résultat : Une liste de points de données (des profils).

Formation Avancée : L'Art de Comprendre

- Quoi : Analyser les relations entre ces profils.
- Quoi : Analyser les patterns de comportement.
- Résultat : Un réseau d'intelligence (des connexions et des habitudes).

Pilier 1 : L'Analyse de Réseau Social (SNA)

Définition : Le SNA n'est pas "regarder les réseaux sociaux". C'est une méthode d'analyse issue de la sociologie qui étudie les **structures sociales** en utilisant la **théorie des graphes**.

En OSINT : Nous l'utilisons pour visualiser et comprendre les relations entre des entités (personnes, groupes).

L'objectif : Répondre à "Qui est le vrai leader ?", "Qui connecte deux groupes différents ?", "Quels sont les 'cliques' ?".

Concepts Clés de la SNA

Le Cluster (La "Clique") :

- Un groupe de nœuds (personnes) fortement connectés entre eux. Ils interagissent beaucoup, mais peu avec l'extérieur.
- Ex: Une équipe projet, un groupe d'amis proches.

Le Hub (Le "Connecteur") :

- Un nœud avec un nombre très élevé de connexions. Il est central et populaire.
- Ex: Un influenceur, un manager, un compte d'actualité.

Le Pont (Le "Pivot") :

- Le plus important pour l'OSINT !
- Un nœud qui est la seule connexion entre deux Clusters différents.
- Ex: L'ingénieur qui est dans un cluster "Collègues de travail" ET dans un cluster "Forum de hacking".
- C'est votre meilleur pivot.

https://www.youtube.com/watch?v=UX7YQ6m2r_o

Outils de SNA

Manuel / Cérébral (Petits Graphes) :

- Pour 1 à 10 cibles, vous le faites "à la main" : "Je vois que @CibleA parle toujours à @AmiB et @AmiC... ils forment un cluster."

Maltego / Lampyre (Graphes Moyens) :

- Parfait pour l'OSINT.
- Vous importez vos données (ex: CSV de followers) ou utilisez des transforms pour extraire les "amis", "followers", etc.
- Les layouts (mises en page) visuelles révèlent les clusters et hubs.

Gephi (Big Data / Académique) :

- L'outil "standard" open-source pour l'analyse de milliers de nœuds.
- Permet des calculs statistiques complexes (centralité, modularité).
- Courbe d'apprentissage élevée, mais extrêmement puissant.

Visualisation de la SNA

Si vous avez scrapé 10 000 followers Twitter, Maltego va ramer. **Gephi** est un logiciel open-source conçu pour les mathématiques des graphes.

- **Pourquoi ?** Il calcule mathématiquement la "Centralité" (qui est important ?) et la "Modularité" (quelles sont les communautés ?).
- **Usage** : Réservé aux analystes confirmés qui veulent faire des statistiques sur des foules.

Travaux Pratiques

Objectif : Importer des données brutes d'interactions (Qui parle à Qui ?) dans Maltego pour révéler visuellement la structure d'un groupe criminel et identifier son chef.

Outil : Maltego CE + Un fichier CSV (que nous allons créer).

Scénario : Vous avez récupéré le graph d’un réseau. Vous voulez caractériser ce graph: clusters, hubs, ponts.

TP 1 - DataSet

1. Inspectez, complétez éventuellement le code “twitter_extractor.py”
2. Exécutez le code et observez le résultat
3. Modifiez le code si le résultat ne correspond pas à vos attentes
4. Itérez si besoin

TP 2 - Chargement des données

1. Dans Maltego, allez dans l'onglet "Import" > "Import Graph from Table".
2. Sélectionnez le fichier reseau_x.csv.
3. L'Assistant de Mapping :
    - Maltego va vous demander : "Qu'est-ce que la colonne Source ?".
    - Mappez la colonne Source vers l'Entité Person (ou Alias).
    - Mappez la colonne Cible vers l'Entité Person.
    - Mappez la colonne Type vers le "Link Label" (l'étiquette de la flèche).
4. Cliquez sur Finish.

TP 3 - Analyse SNA

1. Allez dans l'onglet "Layout".
2. Testez le "Circular Layout".
    - Observation : Qui est au centre ou a le plus de rayons ?
3. Testez le "Organic Layout".
    - Observation : Voyez-vous des groupes se détacher ?
    - Y a-t-il des clusters ? Des Hubs ? Qui sont les ponts ?

TP 4 - Automatisation avec Maltego

Inspectez, complétez éventuellement le code “twitter_transform.py”

Intégration dans Maltego

1. Ouvrez Maltego.
2. Allez dans l'onglet du ruban supérieur "Transforms".
3. Cliquez sur le bouton "New Local Transform".
4. Un assistant s'ouvre. Remplissez-le comme indiqué dans la page suivante

TP 4 - Automatisation avec Maltego

Page 1 : Informations Générales

- Display Name : To Twitter Profile [Local] (C'est le texte du menu clic-droit).
- Description : Génère l'URL X à partir d'un pseudo.
- Input Entity Type : Cherchez et sélectionnez maltego.Alias (C'est crucial : la transform n'apparaîtra que si vous cliquez sur un Alias).
- Author : Vous.

Page 2 : Paramètres d'Exécution (Le Lien Technique)

- Command : Le chemin vers votre exécutable Python.
  - Linux/Mac : /usr/bin/python3 (ou le chemin de votre venv).
  - Windows : python.exe (ou le chemin complet ex: C:\Python39\python.exe).
- Parameters : C'est ici qu'on construit la commande. Il faut le chemin du script et la variable d'entrée.
  - Écrivez : Chemin/Vers/Votre/local_transform.sh
  - Exemple Linux : /home/osint/scripts/local_transform.sh
- Working Directory : Le dossier où se trouve le script.
- Cliquez sur Finish.

TP 4 - Automatisation avec Maltego

Test de la Transform

1. Créez un nouveau graphe.
2. Glissez une entité Alias sur le graphe.
3. Renommez-la wh1t3h4ts.
4. Faites Clic-Droit sur l'entité.
5. Dans le menu contextuel, vous devriez voir une section "Local Transforms" (ou le nom que vous avez donné au "Set").
6. Cliquez sur To Twitter Profile [Local].

Résultat attendu : Une nouvelle entité URL (https://x.com/wh1t3h4ts) doit apparaître, reliée à votre Alias.

TP 4 - Bonus

Pour transformer ce script simple en outil puissant (comme vu dans le cours), il suffit de modifier l'étape #3 du code Python. Au lieu de faire une simple concaténation de texte, vous pouvez :

- Importer requests.
- Interroger l'API de HaveIBeenPwned ou Sherlock.
- Créer des entités si le compte existe.

C'est ainsi que vous créez vos propres outils d'enquête sur mesure.

Pilier 2 : L'Analyse Comportementale

On ne regarde plus avec qui ils parlent, mais comment et quand ils parlent. L'objectif est de construire une "signature comportementale" de la cible.

Deux axes principaux :

- Analyse Temporelle (Le "Quand")
- Analyse Linguistique (Le "Comment")

Analyse Temporelle (Quand ?)

Objectif : Déduire un fuseau horaire, un cycle de sommeil et des habitudes.

La Méthode :

1. Extraire les timestamps (dates et heures) des 50-100 derniers posts/actions (via scraping ou manuellement).
2. Placer les heures sur un graphique de 24h.
3. Analyser les "trous".

Exemple d'analyse :

- Scénario : Une cible prétend être à Paris (UTC+2).
- Données : Elle poste massivement et régulièrement entre 01h00 et 05h00 (heure de Paris).
- Hypothèse 1 (Faible) : C'est un insomniaque.
- Hypothèse 2 (Forte) : Elle n'est pas à Paris. Ces heures (01h-05h UTC+2) correspondent à 19h-23h à New York (UTC-4). Le "trou" (sommeil) est probablement entre 06h et 14h (heure de Paris).
- Corroboration : Lier à l'analyse des ombres (IMINT) ou à des IP de leaks

Analyse Linguistique (Comment ?)

Objectif : Déduire un état émotionnel, un niveau d'éducation, des intérêts ou confirmer une identité.

Analyse de Sentiment :

- La cible est-elle de plus en plus agressive ? Paranoïaque ?
- Utiliser des outils (N-grams, Word Clouds) pour voir les mots les plus fréquents ("arnaque", "vol", "peur" vs. "opportunité", "gagnant").

Analyse de Jargon (Pivot d'Identité) :

- C'est la version SOCMINT du pivot "chat Garfield" .
- Scénario : Vous suspectez que @DevMaster (Twitter) = alpha_user (Forum Hacking).
- Analyse : @DevMaster utilise des termes très spécifiques (ex: crypto-slang "LFGB", "WAGMI") ET des fautes d'orthographe particulières (ex: "il son partis").
- Corroboration : Vous retrouvez le même jargon et les mêmes fautes sur les posts de alpha_user.
- Résultat : La confiance dans le lien entre les deux identités augmente drastiquement.

Pilier 3 : Les Nouveaux Terrains - Telegram

Problématique : La recherche est difficile, non-centralisée.

Canaux :

- Diffusion (1-à-Plusieurs). Ex: Canaux de leaks, de désinformation.
- OSINT : Contenu public. La difficulté est de trouver le canal.
- Technique : Google Dorking (site:t.me "mot-clé") ou utilisation de moteurs spécialisés (ex: Telemetr.io, TGStat).

Groupes :

- Discussion (Plusieurs-à-Plusieurs). Ex: Groupes de trading, communautés de hackers.
- OSINT : L'or se trouve ici.
- Difficulté : Les groupes publics sont rares. Les groupes privés (majorité) nécessitent un lien d'invitation.

OPSEC :

- NE JAMAIS utiliser votre numéro de téléphone personnel.
- Utiliser des numéros virtuels ou prépayés (achetés en espèces) pour le sock puppet .

Pilier 3 : Les Nouveaux Terrains - Discord

Problématique : Encore plus fermé que Telegram. C'est un "Deep Web" conversationnel.

Structure : Organisé en "Serveurs" (privés ou publics).

La **Difficulté** : Google n'indexe rien du contenu d'un serveur.

Technique 1 (Découverte) :

- Trouver des serveurs "publics" via des annuaires (ex: Disboard, Top.gg).
- Google Dorking (site:discord.gg "mot-clé") pour trouver des liens d'invitation expirés ou publics.

Technique 2 (Infiltration) :

- C'est la "zone grise" éthique.
- Rejoindre un serveur avec un sock puppet crédible pour observer passivement (monitoring).
- Risque : L'interaction est du social engineering, PAS de l'OSINT passif. C'est une limite à ne pas franchir sans mandat clair.

Pour aller plus loin : WhoPostedWhat

Le Problème : La Forteresse Facebook

- Depuis 2019, Facebook a verrouillé le "Graph Search". La plupart des outils "magiques" (Stalkface, Scan) ne fonctionnent plus.
- L'interface de recherche native est chaotique : impossible de filtrer facilement par date précise ou par lieu spécifique.

La Solution : WhoPostedWhat.com

- Ce n'est pas un outil de hacking. C'est une interface qui génère des URL de recherche avancées que Facebook comprend encore, mais ne montre pas aux utilisateurs.

WhoPostedWhat : Cas d’usage

Scénario : La Recherche Temporelle

Trouver des témoins ou des preuves d'un événement à une date précise, sans le bruit du flux d'actualité.

- Mot-clé : "Explosion"
- Lieu : "Beyrouth"
- Date : "04/08/2020"

Résultat : L'outil force Facebook à n'afficher que les posts publics publiés ce jour-là.

WhoPostedWhat : Cas d’usage

Scénario : Ciblage Profil (ID)

Pour rechercher dans les posts d'une cible spécifique, WPW a souvent besoin de l'ID numérique unique.

- Les pseudos (@jean.dupont) ne fonctionnent pas toujours.
- L'Astuce : Utilisez d'abord un outil comme Lookup-ID.com (ou le code source de la page `Ctrl+U` => `userID`) pour extraire l'ID (ex: `10000234...`).
- Injectez cet ID dans WPW pour filtrer l'historique de la cible.

L'Investigation par l'Email : Holehe vs Epieos

L'objectif : Vérifier sur quels sites (120+) un email est inscrit, sans alerter la cible.

Epieos : Version Web (SaaS)

- No-Code : Accessible à tous, interface visuelle.
- Enrichi : Ajoute Google Maps, Avis, Calendrier.
- Limité : Version gratuite restreinte.
- OPSEC Faible : Le serveur d'Epieos sait qui vous cherchez.

Holehe : Version CLI (Python) - https://github.com/megadose/holehe

- OPSEC Maximale : Tourne localement sur votre VM.
- Illimité : Gratuit et scriptable (automatisation).
- Moteur : C'est la technologie qui propulse Epieos.
- Technique : Nécessite le terminal.

Travaux Pratiques

Prenez le compte X (Twitter) de la cible choisie.

Mission 1 : Analyse Temporelle

- Regardez les timestamps (heures) de ses 20 derniers posts (réponses incluses).
- Quel est son fuseau horaire probable ?
- Quelles sont ses heures d'activité ?
- Quel est son "trou" (cycle de sommeil) apparent ?

Mission 2 : Analyse de Réseau (SNA)

- Regardez ses 10 dernières réponses (pas ses posts, ses réponses).
- Répond-il toujours aux mêmes 2-3 personnes ? (Formation d'un cluster).
- Répond-il à des personnes très différentes ? (Agit-il comme un pont).
- Notez 3 comptes avec qui il interagit le plus.

Partagez vos découvertes.

Bonus: Vous pouvez automatiser avec Pandas et NLTK.

Due Diligence et Vérification

Due Diligence : Le Volet Judiciaire

Le Mythe du "Casier Judiciaire"

- En OSINT civil, vous n'avez JAMAIS accès au casier judiciaire officiel (B2/B3). C'est une donnée privée réservée aux autorités.

La Réalité de l'Enquêteur

- Nous cherchons les traces publiques de conflits avec la loi ou de sanctions administratives.
  - Listes de Sanctions (Terrorisme, Blanchiment)
  - Jugements Commerciaux & Civils
  - Articles de Presse ("Adverse Media")

Niveau 1 : Sanctions Internationales

C'est la base de la conformité (AML/CTF). Avant de chercher un petit délit, vérifiez si la cible est un acteur menaçant global.

Quoi : Terrorisme, Oligarques, Blanchiment, Prolifération.

L'Outil Clé : OpenSanctions.org

- C'est le "Google" des personnes sanctionnées. Il agrège les listes de l'ONU, de l'UE, de l'OFAC (USA), etc.

Action : Une simple recherche sur le nom suffit. Si positif, c'est un "Red Flag" critique immédiat.

Niveau 2 : Archives Judiciaires

USA (Open Records) : Le système est très ouvert. On accède souvent aux détails.

- JudyRecords : Moteur de recherche gratuit pour des millions de dossiers de tribunaux.
- Pacer : (Payant) Accès officiel aux documents fédéraux.
- BlackBookOnline : Recherche de "Mugshots" (photos d'arrestation).

France / Europe (RGPD) : L'accès est restreint (Droit à l'oubli). On vise le commercial.

- Doctrine.fr / Pappers Justice : Pour voir si une personne a été partie prenante dans un procès civil ou commercial.
- Ce qu'on trouve : Faillites, litiges prud'hommaux, contentieux commerciaux.
- Ce qu'on ne trouve pas : Les petites condamnations pénales.

Niveau 3 : Adverse Media (Réputation)

La Technique

- Cherchez l'association du nom de la cible avec des mots-clés négatifs précis.
- "Jean Dupont" AND (arnaque OR tribunal OR police)

Mots-Clés Clés

- Ne cherchez pas juste "Crime".
- **Essayez** : Escroquerie, Blanchiment, Plainte, Mis en examen, Faillite, Condamnation, Suspect.

Mise en Garde

- **Attention** : Une arrestation ou un article de presse n'est pas une preuve de culpabilité.
- Respectez toujours la présomption d'innocence dans vos rapports.

Résumé: Workflow de Validation

1. Sanctions
   - Check OpenSanctions.org (Terrorisme / AML)
2. Judiciaire
   - Check Bases Nationales (Doctrine / JudyRecords)
3. Réputation
   - Check Adverse Media (Google Dorks)
4. Rapport
   - Contextualiser les faits (Pas de jugement hâtif)

TECHINT Avancé

Le Pivot par Identifiants (Tracking Codes)

Le Concept: Les infrastructures ne sont pas liées que par des IPs, elles sont **liées par l'argent et le marketing**. Les administrateurs réutilisent souvent les mêmes scripts de suivi sur tous leurs sites.

Ce qu'on cherche (dans le Code Source - Ctrl+U) :

- Google Analytics / GTM : UA-123456-1, G-XXXXXX.
- Google AdSense (La Monétisation) : pub-1234567890.
- Amazon Affiliate : Identifiants partenaires.

L'Outil : **DNSlytics** (ou **BuiltWith**) Contrairement à dig qui interroge le serveur, DNSlytics archive le code source de millions de sites.

DNSLytics : Méthodologie

1. **Extraction** : Vous trouvez un ID AdSense sur site-arnaque.com.
2. **Reverse Search** : Vous entrez cet ID dans DNSlytics.
3. **Résultat** : L'outil liste les 15 autres sites utilisant ce MÊME ID.
4. **Impact** : Vous passez d'un site isolé à un réseau complet, exposant potentiellement un site "vitrine" légitime qui trahit l'identité de l'opérateur.

Alternative:

PublicWWW (le "Google du code source") est une excellente alternative pour chercher un bout de code ("UA-123456") dans le code source de tout le web.

Élargir le Champ de Vision : Domain Codex & ViewDNS

Le Problème : Parfois, vous n'avez pas de domaine cible, juste un nom d'entreprise ou un email suspect. Les outils techniques (dig) ne servent à rien ici.

L'Outil : Domain Codex (Moteur de Recherche de Domaines) C'est une base de données qui indexe les "Whois" et les noms de domaines.

Domain Codex : Cas d’usage

1. Détection de Phishing (Typosquatting) :
    - Vous cherchez le nom de votre client (ex: "TotalEnergies").
    - L'outil liste tous les domaines contenant ce mot-clé.
    - Analyste : Vous repérez totalenergies-support-login.com enregistré hier. C'est une menace immédiate.
2. Reverse Whois (Le Pivot Administratif) :
    - Vous avez un email : admin@bad-hacker.com.
    - Vous le cherchez dans Domain Codex.
    - Résultat : Cet email a été utilisé pour enregistrer 50 autres domaines, dont certains datent de 2015 (avant le RGPD strict).
3. Alternative à DNSlytics :
    - Domain Codex permet aussi de chercher par IP ou par ID Analytics, servant de "backup" gratuit si DNSlytics vous bloque.

Domain Codex vs DNSLitycs vs crt.sh

Domain Codex complète parfaitement la "trinité" du TECHINT :

- crt.sh (Technique pure)
- DNSlytics (Marketing/Trackers)
- Domain Codex (Administratif/Noms)

Outil	Point Fort Principal	Usage Cible
DNSlytics	**Reverse Analytics/AdSense**	Trouver des réseaux de sites liés par l'argent/pub.
Domain Codex	**Recherche par Mots-clés & Email**	Trouver des domaines frauduleux (phishing) ou liés à une identité (email).
crt.sh	**Sous-domaines (Certificats)**	Trouver l'infrastructure technique *d'un* domaine connu.

Remonter le temps (Bypass CDN/WAF)

Le Problème : Une cible utilise Cloudflare (ou un autre CDN). Son IP actuelle est masquée. Si vous scannez cible.com, vous tapez sur les serveurs de Cloudflare, pas sur le serveur réel.

La Solution SecurityTrails :

- L'outil enregistre les changements DNS depuis des années.
- Vous regardez l'historique des enregistrements "A" (IP).
- Le Pivot : Il y a 6 mois, avant d'installer Cloudflare, le site pointait vers 1.2.3.4. Il est très probable que le serveur réel soit toujours à cette adresse.
- **Résultat** : Vous avez trouvé l'IP d'origine (Origin IP), contournant la protection WAF.

SecurityTrails : La "Mémoire" du DNS

Pourquoi l'utiliser ? Contrairement aux outils qui vous donnent l'état actuel du réseau, SecurityTrails vous donne son historique. C'est la "Wayback Machine" des DNS.

Le Cas d'Usage Avancé : Contourner Cloudflare/WAF

- Situation : Votre cible est protégée par un CDN (Cloudflare/Akamai). L'IP publique est celle du protecteur, pas de la cible.
- Action : Consultez l'onglet "Historical Data" (Records A).
- Analyse : Cherchez l'IP utilisée juste avant le passage sur Cloudflare.
- Exploitation : Tentez d'accéder directement à cette vieille IP (via fichier hosts ou curl). Si le serveur répond, vous avez contourné la protection.

L'Intégration (Automatisation)

- API : SecurityTrails possède une API très rapide.
- Usage : Indispensable d'ajouter votre clé API (gratuite) dans TheHarvester (api-keys.yaml) ou Maltego pour débloquer la puissance maximale de ces outils.

SecurityTrails : Un Complément à crt.sh

crt.sh vous montre les sous-domaines qui ont eu un certificat SSL (ou l'intention).

SecurityTrails vous montre les sous-domaines qui ont eu un enregistrement DNS (la réalité technique).

Les deux ensembles ne se chevauchent pas toujours à 100%. Utiliser les deux garantit une couverture maximale.

L'Analyse par Procuration : urlscan.io

Le Concept : "Ne pas toucher". Plutôt que de visiter le site cible avec votre navigateur (même via Tor/VPN), vous demandez à un robot tiers de le faire pour vous.

Pourquoi l'utiliser ?

1. Sécurité (Sandboxing) : Si le site contient un malware ou un script de tracking agressif, c'est urlscan.io qui est infecté, pas vous.
2. Rayons X Technique (DOM Analysis) : L'outil capture ce que l'utilisateur ne voit pas :
    - Les IPs vers lesquelles le site envoie des données.
    - Les scripts JavaScript chargés (Google Analytics, trackeurs, mineurs de crypto).
    - La technologie serveur utilisée.
3. Machine à Remonter le Temps : L'onglet "Search" permet de voir si quelqu'un d'autre a déjà scanné ce site il y a 6 mois. Le site a-t-il changé d'apparence ? D'IP ?

Urlscan.io : Avertissement OPSEC

Par défaut, tout scan est PUBLIC.

N'utilisez JAMAIS urlscan.io pour scanner :

- Un lien contenant un token ou un mot de passe.
- Un document interne confidentiel.
- Une URL que vous voulez garder secrète (l'admin du site verra l'IP de urlscan dans ses logs et saura qu'il est analysé).

Bonne pratique : Utilisez l'option "Private Scan" (nécessite un compte/API) si disponible, ou cherchez simplement si l'URL a déjà été scannée sans lancer de nouveau scan.

Recherche de Code Avancée : grep.app

Le Problème de la Recherche GitHub La recherche native de GitHub a des limites majeures pour l'analyste :

- **Bruit** : Elle mélange code, issues, wikis et descriptions.
- **Syntaxe** : Elle gère mal les caractères spéciaux (_, ., =) essentiels pour trouver des variables techniques.
- **Vitesse** : Elle peut être lente et limiter le nombre de résultats.

La Solution : grep.app

- **Qu'est-ce que c'est ?** Un moteur de recherche ultra-rapide qui indexe plus d'un demi-million de dépôts publics.
- **Philosophie** : "Code Only". Il ne cherche que dans le contenu des fichiers.
- **La "Killer Feature"** : Le support des **Regex** (Expressions Régulières).
    - Au lieu de chercher un mot précis, vous cherchez la forme d'une donnée (ex: une clé API, un numéro de téléphone, un IP interne).

grep.app : Cas d’usage OSINT

Cas 1 : Infrastructure Cachée (TECHINT)

- **Action** : Chercher le domaine racine cible.com (avec l'option "Case Sensitive" désactivée).
- **Résultat** : Découverte de sous-domaines "hardcodés" (dev-api.cible.com, staging.cible.com) dans des scripts de configuration (Terraform, Ansible) appartenant souvent à des prestataires tiers, invisibles via DNS.

Cas 2 : Chasse aux Fuites (Leaks / Regex)

- **Action** : Utiliser le mode Regex pour trouver des clés spécifiques.
- **Query** : AIza[0-9A-Za-z-_]{35} (Pattern d'une clé Google API).
- **Analyse** : Vous trouvez des clés valides laissées par erreur dans des fichiers de test ou des commentaires.

Cas 3 : Attribution (SOCMINT)

- **Action** : Chercher @cible.com.
- **Résultat** : Identification de développeurs utilisant leur email pro dans des projets personnels, permettant de pivoter vers leurs profils sociaux ou d'autres pseudos.

grep.app : Avantage et Limitation

Avantage

Souvent, on ne trouve pas le code dans l'organisation de la cible (qui est sécurisée), mais dans les dépôts personnels de ses freelances ou stagiaires qui ont "forké" du code. Grep.app excelle pour trouver ces connexions indirectes.

Limitation

grep.app ne cherche "que" dans ~500k repos (les plus populaires). Pour une recherche exhaustive sur TOUT GitHub, l'API GitHub (via script Python) reste nécessaire, mais grep.app couvre 90% des besoins "quick win".

VirusTotal : Bien plus qu'un Antivirus

La Perception Grand Public : "Un site web où j'envoie un fichier pour voir s'il contient un virus."

La Réalité pour l'Analyste OSINT : VirusTotal (VT) est la plus grande base de renseignement (CTI) publique au monde sur les infrastructures malveillantes.

- Ce n'est pas un simple scanner, c'est un moteur de corrélation.
- Il enregistre l'historique de chaque fichier, domaine et IP analysés depuis plus de 15 ans.

Pourquoi l'utiliser en OSINT ?

- **Attribution** : Lier un fichier malveillant à une infrastructure spécifique (IP/Domaine).
- **Historique** DNS passif : Voir sur quelles IPs un domaine pointait il y a 2 ans.
- **Chasse** : Trouver d'autres variantes d'une attaque grâce aux métadonnées similaires.

VirusTotal : Avertissement OSINT

La Règle d'Or : Tout fichier uploadé sur VirusTotal est partagé avec la communauté sécurité (et les abonnés payants).

Le Danger : Si vous uploadez un document interne confidentiel ("Plan_Fusion_2025.pdf") pour vérifier s'il est sain, vous venez de le leaker au monde entier.

La Bonne Pratique : N'uploadez jamais de documents sensibles. Calculez le Hash (empreinte numérique) du fichier sur votre machine (localement) et cherchez ce Hash sur VT. Si le fichier est inconnu, analysez-le dans votre VM (sandbox locale), pas dans le Cloud.

VT : Pour aller plus loin

Recherche Avancée (Intelligence) Utilisez les modificateurs de recherche ("VT Dorks") pour chasser :

- content:"pass" : Trouver des documents contenant des mots de passe.
- similar-to:[hash] : Trouver des variantes d'un malware.
- engine:"phishing" p:10+ : Trouver les nouveaux sites de phishing détectés par au moins 10 antivirus.

VirusTotal est à l'analyse de malware ce que LinkedIn est à l'analyse RH : une base de données relationnelle géante.

L'API gratuite de VT s'intègre dans Maltego et TheHarvester, rendant ces pivots automatiques.

VT: Méthodologie de Pivot

L'Art du Pivot : Hash → Infra → Acteur La puissance de VT réside dans ses onglets "Relations" et "Behavior". L'analyste ne regarde pas le résultat du scan (rouge/vert), il regarde les liens.

Le Workflow de l'Enquêteur :

1. Point de départ : Vous avez un fichier suspect (ou son Hash MD5/SHA256).
2. Analyse (Onglet Relations) : VT vous dit : "Ce fichier communique avec update-server-fake.com".
3. Le Pivot : Vous cliquez sur update-server-fake.com.
4. Expansion : VT vous montre que 50 autres fichiers communiquent aussi avec ce domaine.
5. Résultat : Vous êtes passé d'un fichier isolé à la découverte de toute la campagne d'attaque.

Outil Visuel : VT propose une interface graphique (VT Graph, similaire à Maltego) pour visualiser ces connexions complexes en un clic. Indispensable pour cartographier une menace.

Qualification d'IP : Tor Exonerator & GreyNoise

Le Problème : "Qui est vraiment derrière cette IP ?" Une adresse IP n'est pas toujours une identité. Elle peut être un relais. Si vous enquêtez sur une IP (issue de logs ou de DNS) sans vérifier sa nature, vous risquez de profiler un serveur innocent (relais Tor) au lieu de la cible.

Tor Exonerator (metrics.torproject.org)

- La Question : "Cette IP était-elle un relais Tor ce jour-là ?"
- Pourquoi l'historique compte : Une IP peut être un nœud Tor le lundi, et un serveur web normal le mardi (DHCP, Cloud). Shodan donne l'état actuel, Exonerator donne l'état passé.
- Verdict : Si Exonerator dit "Oui", l'IP est une impasse pour la géolocalisation . L'attribution géographique est impossible.

GreyNoise (Alternative Moderne, vu en Partie 1)

- La Question : "Cette IP attaque-t-elle tout le monde ou juste moi ?"
- Concept : GreyNoise écoute le "bruit de fond" d'Internet.
- Usage : Si votre IP suspecte est connue de GreyNoise comme un "Scanner de masse", c'est probablement un bot automatisé, pas une attaque ciblée contre votre client.

Précisions sur Tor Exonerator

L'Analyse Forensique : Exonerator est aussi très utilisé par les forces de l'ordre ou les avocats pour prouver qu'un téléchargement illégal n'a pas été fait par le propriétaire de la ligne, mais par un utilisateur Tor anonyme passant par là (d'où le nom "Exonerator").

Complémentarité : Il complète Shodan (qui tague souvent les IPs "Tor Exit Node", mais pas toujours avec l'historique précis).

Importance de Tor Exonerator

Vous avez appris à pivoter sur des adresses IP.

- **Le Piège** : L'analyste trouve une IP suspecte (ex: dans des logs ou via une résolution DNS). Vous la géolocalisez à "Berlin, Allemagne". Vous commencez à enquêter sur un hébergeur allemand.
- **La Réalité** : Cette IP était un Nœud de Sortie Tor (Exit Node) ce jour-là. L'attaquant n'est pas à Berlin. Vous avez perdu votre temps.
- **La Solution Exonerator** : Il permet de vérifier si une IP spécifique était un nœud Tor actif à une date précise.

Travaux Pratiques

Contexte

Un employé de "l'Entreprise Cible" a reçu un email suspect concernant une commande d'iPhone qu'il n'a jamais passée. Il l'a transféré au SOC (vous).

Mission

Analyser l'email, extraire les indicateurs techniques (IOCs) et pivoter pour cartographier l'infrastructure de l'attaquant.

TP 1 - Analyse "Header" (Le Point de Départ)

Objectif : Identifier la véritable source technique derrière l'usurpation.

Actions :

1. **Identifier l'Usurpation** : Comparer le champ From ("Apple Store" noreply@apple-secure-notifications.com) avec le Return-Path (admin@upskillforge.com).
2. **Analyser la Route** : Repérer le premier saut dans les Received headers. L'email vient de upskillforge.com (IP: 34.204.186.209) avant de passer par un relai mail-relay.unknown-isp.net.
3. **Repérer le Script** : Noter le header X-Mailer: PHP/8.1 (MassMailer Script v2.0), qui trahit l'utilisation d'un script automatisé plutôt qu'un serveur mail d'entreprise légitime.
4. **Extraire les Liens** : Identifier l'URL de phishing principale https://upskillforge.com/?ref=email_campaign_nov25 et le lien de tracking https://upskillforge.com/track/8821

TP 2 - Infrastructure (Le Pivot Technique)

Objectif : Cartographier l'infrastructure liée à ce domaine.

Actions :

1. Historique DNS (SecurityTrails / ViewDNS) :
    - Entrer upskillforge.com dans l'outil.
2. Sous-domaines (crt.sh) :
    - Rechercher %.upskillforge.com sur crt.sh.

TP 3 - Code & Marketing (Le Pivot "Tracking")

Objectif : Trouver d'autres sites liés à cet attaquant via des identifiants partagés.

Actions :

1. Tracking Codes (PublicWWW / Inspection) :
    - Allez sur upskillforge.com et faire Ctrl+U.
    - Cherchez un ID Google Analytics (UA-XXXX), un Pixel Facebook, ou un ID AdSense.
    - Action : Cherchez cet ID sur PublicWWW (ou simulent la recherche) pour trouver si cet ID est utilisé ailleurs (simulez un réseau de sites de phishing).
2. Fouille de Code (grep.app / GitHub) :
    - L'email mentionne un outil spécifique : MassMailer Script v2.0 et un paramètre URL ref=email_campaign_nov25.
    - Action : Chercher upskillforge.com ou ces chaînes spécifiques sur GitHub/grep.app.

TP 4 - Threat Intel (La Réputation)

Objectif : Voir ce que le monde sait déjà.

Actions :

1. Urlscan.io :
    - Scannez upskillforge.com.
    - Regardez l'onglet "HTTP Transactions" pour voir si le site redirige ou charge des ressources externes suspectes.
2. VirusTotal :
    - Cherchez l'URL ou le domaine pour voir s'il est déjà flaggé comme "Phishing".

IMINT & GEOINT Avancé

Rappel : Quand les Outils Échouent

Ce que nous savons faire :

- Recherche Inversée (Yandex/TinEye) : Trouver l'origine d'une photo connue .
- Analyse EXIF : Lire les métadonnées (GPS, Date, Appareil).

Le Problème Avancé :

1. Les réseaux sociaux suppriment les EXIF (99% des cas).
2. La recherche inversée échoue si l'image est originale (ex: une photo de "vacances" postée à l'instant).

Conclusion : Quand les métadonnées sont absentes et que la recherche inversée échoue, l'analyste doit utiliser la géolocalisation manuelle.

La Méthodologie "Bellingcat"

C'est un processus de déduction, pas une simple recherche. Nous allons répondre à 3 questions (de manière itérative, détaillé dans les diapos suivantes) :

1. **QUOI** ? (Analyse)
    - Quels indices puis-je voir et entendre ?
    - Ex: Langue, plaques d'immatriculation, style de bâtiment, soleil.
2. **OÙ** ? (Hypothèse & Vérification)
    - Où dans le monde ces indices co-existent-ils ?
    - Ex: "Conduite à gauche + langue portugaise + végétation tropicale" → Hypothèse : Brésil ?
    - Vérification : Google Maps / Earth.
3. **QUAND** ? (Datation)
    - Quand cette image a-t-elle été prise ? (Saison, heure).
    - Vérification : SunCalc, Google Earth (Images historiques).

Piste 1 : “QUOI ?”

Langue :

- Panneaux de signalisation, noms de magasins, affiches, graffitis.

Infrastructure :

- Plaques d'immatriculation (Couleur, format).
- Sens de circulation (Conduite à gauche ou à droite ?).
- Poteaux électriques (forme), bornes d'incendie, marquage au sol.

Architecture :

- Style (Haussmannien, Victorien), matériaux (brique rouge).
- Lieux de culte (Églises, Mosquées, Temples).

Géographie & Végétation :

- Montagnes, mer, terrain plat ?
- Type d'arbres (Palmiers, Sapins, Oliviers).

Piste 2 : "OÙ ?"

Google Maps (Street View) :

- L'outil n°1. Permet de "marcher" virtuellement pour trouver l'angle de vue exact.

Google Earth Pro (Logiciel) :

- Indispensable. Ne pas utiliser la version web.
- Vue 3D : Permet de confirmer la hauteur des bâtiments et les lignes de vue (ce que Maps ne fait pas bien).
- FONCTION CLÉ : Images Historiques. Permet de "remonter le temps" pour voir si un bâtiment existait en 2010.

Alternatives Spécifiques :

- Yandex Maps : Meilleure couverture Street View pour la Russie, l'Europe de l'Est.
- Baidu Maps : Indispensable pour la Chine.
- Mapillary / Wikimapia : Sources alternatives (collaboratives).

Piste 3 : "QUAND ?"

La Méthode :

1. Trouver un objet vertical (poteau, personne) et son ombre.
2. Ombres courtes = Proche du midi solaire.
3. Ombres longues = Matin ou soir.
4. La direction de l'ombre indique l'opposé du soleil.

L'Outil de Vérification : SunCalc.net

- C'est un "Google Maps" qui simule la position du soleil (et la direction des ombres) pour n'importe quel lieu, à n'importe quelle date/heure.
- Usage :
    a. Placez le marqueur sur votre lieu (hypothèse).
    b. Faites glisser le curseur de l'heure.
    c. Question : "À quelle heure l'ombre correspond-elle à ce que je vois sur ma photo ?"
    d. Vous venez de dater l'image et de confirmer son orientation.

VIDINT : Analyse Vidéo

Technique 1 : Lecture Frame-by-Frame

- Outil : VLC (Media Player).
- Action : Touche 'E' (par défaut) pour avancer image par image.
- Objectif : Chercher l'indice fugace (une plaque d'immatriculation, un panneau flou).

Technique 2 : Le "Couteau Suisse" (InVID / WeVerify)

- Outil : Une extension navigateur (Firefox/Chrome) indispensable.
- Fonctions Clés :
  - "Keyframes" : Découpe la vidéo en images clés (miniatures).
  - Recherche Inversée : Lance une recherche inversée (Google, Yandex...) sur ces images clés automatiquement.
  - Analyse : Permet une analyse fine des métadonnées de la vidéo.

L'Analyse Audio

Langue & Accents :

- Évident, mais crucial. Les gens parlent-ils français ? Avec quel accent ?

Sons Ambiants (Signatures Uniques) :

- Sirènes (Police, Ambulance, Pompiers) :
  - Elles sont différentes dans chaque pays (et souvent chaque ville).
  - Action : Il existe des bases de données en ligne de sons de sirènes pour comparer.
- Annonces Publiques :
  - Le son d'une annonce de gare (SNCF vs. Deutsche Bahn).
- Cloches d'église :
  - Leur "mélodie" peut parfois être unique.

Caméras en Ligne

Objectif : Vérifier si un événement a lieu maintenant.

- **Problème** : Une photo/vidéo prétend montrer une manifestation aujourd'hui sur la Place de la République. Est-ce vrai ?
- **Solution** : Trouver une caméra en direct pointant sur la place.

Où les trouver ?

- YouTube : Rechercher live cam "Nom de la ville".
- Windy.com : Affiche une surcouche de webcams météo/trafic publiques du monde entier.
- Insecam.org : Référence les caméras de sécurité IP non sécurisées.
    - **ATTENTION** : Zone Grise Légale/Éthique ! C'est de l'OSINT, mais ne pas interagir avec.

Cartographie de la Surveillance

Surveillance Watch : Corporate Intel - surveillancewatch.io

- Ne montre pas les caméras.
- Cartographie les liens financiers et les contrats entre les entreprises de surveillance et les États.
- **Usage** : "Qui fournit la reconnaissance faciale à la police de Nice ?"

Surveillance under Surveillance : Physical Intel - kamera.rs / sunders.uber.space

- Montre les caméras.
- Basé sur OpenStreetMap. Affiche la position exacte des caméras CCTV, Dômes et ALPR signalés par la communauté.
- **Usage** : "Y a-t-il une caméra à l'angle de cette rue pour corroborer ma photo ?"

Travaux Pratiques

Où la photo suivante a pu être prise ? Fichier *resources/challenge_image.jpeg*

Travaux Pratiques - Étapes

Étape 1 : Analyse (QUOI ?)

- Identifiez et listez 3 à 5 indices clés (langue, plaques, architecture, végétation...).

Étape 2 : Hypothèse (OÙ ?)

- Émettez une hypothèse de pays ou de ville.

Étape 3 : Vérification

- Utilisez Google Maps, Google Earth Pro (3D), Yandex Maps pour "chasser" le lieu .
- Cherchez des points de repère uniques (une église, un pont).

Étape 4 : Confirmation

- Trouvez l'URL Street View exacte ou les coordonnées GPS.
- Bonus : Si c'est une photo, utilisez SunCalc pour confirmer l'heure/orientation.

Restitution : Chaque groupe présente sa trouvaille et sa méthodologie.

Travaux Pratiques

Objectif:

Voici une vidéo interceptée. Nous ne savons rien d'elle. Votre client veut savoir OÙ et QUAND (Date et Heure) elle a été tournée. Vous avez 20 minutes.

Support :

Fichier *resources/challenge_video.mp4*

TP 1 - Où sommes-nous ?

Q1 : Quels sont les marqueurs géographiques "macro" ?

- Indice à chercher : Regardez l'architecture et les infrastructures.

Q2 : Pouvez-vous lire quelque chose ? (SOCMINT / Google Dorking)

- Indice à chercher : Cherchez du texte, même flou. Enseignes, panneaux.
- Rechercher ces termes sur Google Images pour confirmer le lieu exact.

TP 2 - Quelle période ?

Q3 : C'est un marché de Noël, mais sommes-nous vraiment à Noël ?

- Indice à chercher : Regardez les arbres (les platanes le long de la rive).
- Question clé : Les arbres sont-ils nus (Hiver) ou ont-ils encore des feuilles (Automne) ?
- Observation : Les arbres ont encore beaucoup de feuilles, éclairées par les spots.

Q4 : Regardez les gens. Ont-ils vraiment froid ?

- Indice à chercher : La buée (condensation) quand ils parlent ou respirent. L'épaisseur des vêtements.
- Question clé : Voyez-vous de la fumée sortir des bouches ? Les manteaux sont-ils fermés jusqu'en haut ?
- Observation : Pas de buée visible. Manteaux souvent ouverts.

TP 3 - Quelle Heure ?

Q5 : Regardez la population. Qui est là ?

- Indice à chercher : La taille des personnes.
- Question clé : Voyez-vous des enfants ? Des poussettes ?
- Observation : Oui, beaucoup de familles et de jeunes enfants.
- Deduction : Il est peu probable qu'il soit 23h00 ou minuit.

Q6 : Analysez la densité de la foule.

- Indice à chercher : Est-ce qu'on circule bien ?
- Observation : C'est bondé ("Peak time").
- Deduction : C'est une heure de pointe pour un marché. Probablement entre 17h00 et 20h00.

Q7 : Quand tombe la nuit à cette période ?

- Outil : SunCalc.
- Action : Vérifier l'heure du coucher de soleil à l’endroit pressenti à la mi-novembre.
- Donnée : Le soleil se couche vers ?

Conclusion

- GEOINT est un puzzle. L'échec de la recherche inversée n'est pas une fin, c'est le début de l'enquête manuelle.
- Votre "checklist" doit être systématique : Langue → Infra → Archi → Végétation.
- Google Earth Pro (Historique) et SunCalc sont les outils avancés qui font la différence pour dater et confirmer.
- VIDINT = IMINT + AUDIO-INT. N'oubliez pas par exemple d'écouter les sirènes.
- La pratique est la clé : entraînez-vous 10min par jour.

BLOCKCHAIN-INT Avancé

Rappel : Le Mythe de l'Anonymat

La Réalité : Le Bitcoin (et la plupart des cryptos) est PSEUDO-ANONYME.

Le Registre : Toutes les transactions sont PUBLIQUES, immuables et traçables à vie.

Le Pseudo : Votre "Adresse" (ex: 0xAb58... ou 1A1zP...) est votre nouveau "pseudo".

Notre Objectif (Initial) : Lier ce "pseudo" (l'adresse) à une identité réelle (ex: un email, un compte Twitter).

Nos Outils (Initiaux) : Les Block Explorers (Etherscan, Blockchain.com) .

Le Problème de l'Analyste Avancé

Un Block Explorer (Etherscan) vous montre :

- Adresse A (Suspect) a envoyé 10 ETH à Adresse B.

Et alors ?

- L'Adresse B est-elle un autre portefeuille du suspect ?
- Est-ce une victime ?
- Est-ce une plateforme d'échange (un "exit") ?
- Est-ce un mixeur (un "trou noir") ?

L'analyse de flux (Flux Analysis) consiste à répondre à ces questions.

Technique 1 : Le "Clustering" (Heuristiques)

Question : Comment savoir si Adresse A et Adresse B appartiennent à la même personne ?

Réponse : L'Heuristique "Common-Input-Ownership" (Dépense Commune)

- **C'est le concept le plus important de l'analyse Bitcoin.**
- **Principe** : Si une seule transaction (Transaction 3) utilise des fonds (inputs) provenant à la fois de l'Adresse A et de l'Adresse B...
- **Conclusion** : ... alors Adresse A et Adresse B sont contrôlées par le même portefeuille (la même clé privée).
- **Action** : Vous pouvez les "clusteriser". Elles sont la même entité.

Technique 2 : "Following the Hops" (Suivi de Flux)

L'argent (surtout illicite) ne va jamais directement du Point A au Point B. Il transite par des portefeuilles intermédiaires ("hops" ou "sauts").

Le Schéma Classique :

1. Adresse A (Ex: Victime de Phishing)
2. Adresse B (Portefeuille du Hacker - 1er Hop)
    - Action : Le hacker regroupe les fonds de 10 victimes.
3. Adresse C (Portefeuille Intermédiaire 1 - 2e Hop)
4. Adresse D (Portefeuille Intermédiaire 2 - 3e Hop)
    - Action : Le hacker "peel" (épluche) les fonds en plus petites quantités.
5. Adresse E (L'Exit)

Notre travail : Suivre méthodiquement ces "hops" à l'aide d'un Block Explorer.

Technique 3 : "Spotting the Exit" (Identification KYC)

Qu'est-ce qu'un "Exit" ? C'est une plateforme d'échange centralisée (CEX) comme Coinbase, Binance, Kraken.

Pourquoi c'est l'Exit ? Pour échanger des cryptos contre des Euros/Dollars, l'utilisateur doit prouver son identité via KYC (Know Your Customer) (passeport, facture) .

L'Indice : Les "Labels" (Étiquettes)

- Les Block Explorers (Etherscan, Whale Alert) connaissent les adresses publiques de ces plateformes.
- Si vous voyez votre flux arriver sur une adresse étiquetée "Binance: Hot Wallet 7"...
    - ... BINGO. Vous avez trouvé le point de de-anonymisation.
- **Légalité** : L'information sur qui possède ce compte n'est accessible que sur réquisition judiciaire, mais vous avez prouvé le lien.

Outils Avancés : La Visualisation

Suivre les "hops" manuellement sur Etherscan est long et complexe. Les outils avancés automatisent le "clustering" et la "visualisation de flux".

Outils (Gratuits / Freemium) :

- Breadcrumbs.app : Excellent pour Ethereum/EVM. Visualise les flux et les labels.
- Arkham Intelligence : Très puissant (IA), visualise les flux, identifie les entités.
- OXT.me (Samourai) : Puissant pour l'analyse heuristique de Bitcoin.
- Lampyre : Possède des modules d'analyse crypto.

Ce qu'ils font : Ils transforment la "liste" Etherscan en "graphe" Maltego.

Angle Mort : Les "Mixers"

Problème : Vous suivez le flux... et il arrive sur une adresse étiquetée "Tornado Cash" (ETH) ou "Wasabi Wallet" (BTC).

Qu'est-ce qu'un "Mixer" ?

- Un service qui "brouille" la piste.
- Comment : Vous déposez 10 ETH. 100 autres personnes déposent aussi 10 ETH. Le "contrat" mélange tout ("pool").
- Plus tard, vous retirez 10 ETH vers une nouvelle adresse.
- Résultat : Le lien transactionnel entre votre ancienne et votre nouvelle adresse est (en théorie) brisé.

Valeur OSINT : La piste est "froide". MAIS, le fait même d'utiliser un mixer est un indicateur de comportement suspect.

Angle Mort : Les "Privacy Coins"

Problème : Un groupe de ransomware ne demande pas du Bitcoin, mais du Monero (XMR) ou du Zcash (ZEC).

Qu'est-ce qu'un "Privacy Coin" ?

- Une crypto-monnaie où l'anonymat est intégré par défaut.
- Monero : Les adresses de l'expéditeur, du destinataire ET le montant sont cachés sur le registre public.

Valeur OSINT :

- Aucune analyse de flux possible. C'est un "trou noir" complet.
- L'information de renseignement est le choix de l'outil : La Cible X utilise Monero, ce qui indique un haut niveau de maturité OPSEC et une volonté active de masquer ses flux financiers.

Pivot Avancé : Les NFTs

N'analysez pas que l'argent (Fongible). Analysez "l'art" (Non-Fongible).

Un NFT est un jeton UNIQUE. Il agit comme un "marqueur".

- Le Pivot : Identité Sociale.
- Les gens n'affichent pas leur solde bancaire, mais ils affichent (revendiquent) leurs NFTs (ex: Bored Ape, CryptoPunk).

Pivot Avancé : Les NFTs

1. Vous avez une Adresse A (Suspect).
2. Vous regardez son "Token Portfolio" (ERC-721) sur Etherscan.
3. Vous voyez qu'il possède le "CryptoPunk #1234".
4. Vous allez sur OpenSea (marketplace) et cherchez "CryptoPunk #1234".
5. Le profil OpenSea a un nom d'utilisateur : SuperHacker99.
6. Ce profil est lié à un compte Twitter (@SuperHacker99) qui utilise le NFT en photo de profil (PFP).
7. **Résultat** : Vous avez lié Adresse A à @SuperHacker99.

Travaux Pratiques

Scénario :

Votre client, un fonds d'investissement crypto, a repéré des mouvements suspects autour d'un nouveau protocole prometteur sur le réseau Hyperliquid. Un portefeuille inconnu semble accumuler des actifs rares bien avant le grand public. Le client veut savoir : Qui est cette "Baleine" ? Est-ce un développeur du projet (délit d'initié) ou un investisseur anonyme ?

Votre Point de Départ (IOC) :

- Réseau : HyperEVM (Hyperliquid)
- Contrat de la collection : 0x9125e2d6827a00b0f8330d6ef7bef07730bac685
- Actif Cible : Token ID #5

TP 1 - Le Piège du "Single Chain"

L'objectif est de comprendre que l'analyse ne se limite pas à Ethereum Mainnet.

1. La fausse piste : Allez sur Etherscan.io et cherchez l'adresse du contrat.
    - Question : Que voyez-vous ?
    - Constat attendu : Rien, ou une adresse vide sans transactions.
    - Leçon : L'adresse existe mathématiquement sur toutes les chaînes EVM, mais l'actif est ailleurs.
2. La bonne porte : Comment visualiser cet actif ?
    - Action : Utilisez un explorateur multi-chain ou une marketplace agnostique comme OpenSea.
    - Recherche : Construisez l'URL ou cherchez la collection via le contrat sur OpenSea.
    - Résultat : Vous tombez sur la collection "Hypurr". Trouvez l'item #5.

TP 2 - Le Profilage par l'Actif

L'objectif est de déduire le profil psychologique et technique de la cible sans connaître son nom.

1. **Analyse de la Rareté :**
    - Regardez le Token ID : #5.
    - Question : Sur une collection de plusieurs milliers, que signifie avoir le numéro 5 ?
    - Déduction : Ce n'est pas un achat aléatoire. C'est souvent réservé à l'équipe (Team), aux fondateurs ou aux partenaires stratégiques.
2. **Analyse de l'Acquisition (Provenance) :**
    - Regardez l'historique de l'item ("Item Activity" en bas de page).
    - Question : A-t-il été acheté (Sale) ou généré (Mint/Transfer) ?
    - Observation : Il s'agit probablement d'un Airdrop ou d'un Mint lors du "Genesis Event".
    - Conclusion Profilage : La cible est un "OG" (Original Gangster). Elle était là au jour 0. Ce niveau d'accès indique une forte sophistication technique.

TP 3 - Le Pivot Identité

L'objectif est de passer de la Blockchain au Web Social (Twitter/X).

1. **Inspection du Propriétaire** (Owner) :
    - Cliquez sur le profil du propriétaire actuel du Hypurr #5.
    - Cas A (Facile) : Le profil a un nom (ex: CryptoWhale.eth ou un pseudo). -> BINGO.
    - Cas B (Difficile - Fréquent) : Le profil est "Unnamed" (0x123...).
2. **La "Chasse aux Miettes"** (Crumbs) :
    - Regardez les autres NFT détenus par ce même wallet.
    - Action : Cherchez un ENS (Ethereum Name Service, un NFT qui finit en .eth) ou un Lens Protocol.
    - Action Alternative : Regardez les "Offres" reçues. Souvent, les amis font des offres "blagues" ou des transferts de test entre leurs propres comptes identifiés.
3. **Le Saut vers Twitter** :
    - Prenez le pseudo trouvé (ou le nom ENS).
    - Cherchez-le sur Twitter (X) ou Telegram.
    - Confirmation : Le compte Twitter utilise-t-il ce chat (Hypurr) en photo de profil (PFP) ? Si oui, l'attribution est confirmée avec un haut degré de certitude.

TP 4 - Visualisation de l'Empire (Graph Analysis)

L'objectif est de voir l'invisible.

Outil : Arkham Intelligence (ou outil de visualisation graph).

Action : Entrez l'adresse du propriétaire du Hypurr #5.

Analyse :

- Regardez la section "Top Counterparties". Avec qui échange-t-il le plus d'argent ?
- Voyez-vous des liens avec des CEX (Binance, Coinbase) ? -> Points de sortie potentiels pour une réquisition judiciaire.
- Voyez-vous des liens avec Tornado Cash ? -> Indicateur de volonté d'anonymisation.

Conclusion

Le BLOCKCHAIN-INT Avancé n'est pas de la "magie", c'est une méthodologie de comptabilité forensique.

Vos 3 Techniques Clés :

- Clustering (Heuristique de Dépense Commune).
- Flux Analysis ("Following the Hops").
- Exit Identification (Trouver les plateformes KYC).

Vos Limites : Les Mixers (Tornado) et les Privacy Coins (Monero).

Votre Pivot Caché : Les NFTs sont un pont direct entre une adresse anonyme et une identité sociale.

Travaux Pratiques

Objectifs et Scénario

Le Scénario

L'acteur menaçant 'VoidWalker' se pense intouchable. Il ne poste jamais son visage et utilise un VPN. Cependant, son ego l'a poussé à deux erreurs :

- Il a posté une photo de la vue depuis sa chambre d'hôtel pour narguer les autorités ("Catch me if you can").
- Il a brièvement partagé une adresse Ethereum pour recevoir des 'dons', avant de supprimer le tweet.

Votre Mission

Retrouver l'adresse, identifier la ville où il se trouve grâce à la photo, et prouver techniquement sa présence physique dans cette ville grâce à la Blockchain.

Les indices

Adresse ETH
0x954f66945240a532e8978aac6ac198f8da13bd1d

Vue de l’hôtel

TP 1 - SOCMINT & Identité

Outil : Etherscan ou OpenSea.

Action : Entrez l'adresse 0x954f66945240a532e8978aac6ac198f8da13bd1d

Question : Cette adresse a-t-elle un nom "humain" ?

Analyse : La cible utilise un pseudonyme persistant. C'est un pivot vers d'autres réseaux sociaux potentiels.

TP 2 - GEO-INT

Utilisez vos compétences acquises dans l’analyse des images afin de déterminer:

- Le lieu
- La date
- L’heure

TP 3 - BlockChain-INT (Preuve de Présence)

L'objectif est d'utiliser un NFT non pas pour sa valeur financière, mais pour sa valeur de "Preuve de Présence" (Proof of Attendance).

1. **Exploration des POAP** (Les Badges) :
    - Concept : Un POAP est un badge numérique remis uniquement aux participants d'un événement.
    - Outil : POAP Scan (https://collectors.poap.xyz/scan).
    - Action : Entrez le nom ENS trouvé (suripuri23.eth) ou l'adresse du wallet.
2. **Analyse Temporelle et Géographique** :
    - Consigne : Scrollez dans sa collection. Cherchez un événement qui correspond à la ville identifiée en Mission 2 (Singapour).
    - Découverte Clé : Trouvez le Token #7512417 (ou l'événement récent correspondant).
    - Déduire la date et l’heure approximative

Conclusion Module 2

Ce que nous avons accompli :

- Nous ne faisons plus que "chercher", nous analysons des réseaux (SNA) et des comportements (temporels, linguistiques).
- Nous savons visualiser des données complexes (Maltego) et les importer (CSV).
- Nous pouvons géolocaliser une image originale sans métadonnées, en utilisant les indices visuels (GEOINT) et le soleil (SunCalc).
- Nous pouvons suivre des flux financiers sur la blockchain, identifier des exits (KYC) et utiliser les NFTs comme pivots SOCMINT.

À suivre : Nous passerons de l'enquête "ponctuelle" à l'OSINT Opérationnel : la surveillance en temps réel, l'automatisation et l'étude de cas finale.

Module 3
OSINT Opérationnel et Étude de Cas

Rappel

Le Premier Ennemi de l'Analyste... c'est lui-même

Rappel Formation M1 : Le Cycle du Renseignement

- La Planification est la première étape, la plus importante.
- L'erreur du débutant est de commencer par la Collecte.

Rappel Formation M3 : Les Biais Cognitifs

- Votre pire ennemi est le Biais de Confirmation : La tendance à chercher ce qui confirme votre hypothèse et à ignorer ce qui l'infirme .
- Le Problème de l'Analyste Avancé : Plus vous avez d'outils, plus vous collectez de "bruit" et plus vous risquez de vous noyer (Data Drowning) ou de confirmer un faux scénario.

Partie 1 : Définir la Cible (KIQ)

Le Problème : Le Mandat Flou

Votre client/manager vous demande :

- "Fais-moi une enquête sur l'Entreprise X."
- "Regarde ce que tu trouves sur cet acteur 'Shadow'."
- "On a une IP suspecte, dis-moi tout."

Risque : Vous partez dans 10 directions, collectez 10Go de données et ne répondez à aucune vraie question.

La Solution : Les KIQ (Key Intelligence Questions)

- Ce sont les "Questions Clés" que vous définissez avant de commencer la collecte.
- Elles transforment le "flou" en "plan de collecte ciblé".

Exemple : De "Flou" à "KIQ"

Le Mandat Flou : "Enquête sur le groupe de ransomware 'LockShadow'."

Votre Planification (KIQ) :

- KIQ 1 (Identité) : Qui sont les opérateurs (pseudos, identités) ?
- KIQ 2 (Infrastructure) : Quels sont leurs domaines, IPs, adresses crypto ?
- KIQ 3 (TTPs) : Comment opèrent-ils (quels outils, quels vecteurs d'attaque) ?
- KIQ 4 (Victimologie) : Qui ciblent-ils (secteurs, pays) ?
- KIQ 5 (Intention) : Quelle est leur motivation (financière, politique, autre) ?

Résultat : Votre collecte a maintenant un but. Vous ne cherchez plus "tout", vous cherchez des réponses à ces 5 questions.

Partie 2 : Combattre les Biais

Le Biais de Confirmation en action :

1. Un indice pointe vers un pays X.
2. Votre cerveau décide : "C'est le Pays X".
3. Vous passez le reste de l'enquête à chercher des preuves qui confirment "Pays X".
4. Vous ignorez inconsciemment les preuves qui pointent vers le Pays Y.
5. Échec de l'analyse.

La Solution (ACH) :

- Une méthode structurée (créée par la CIA) qui vous force à chercher à infirmer vos hypothèses.

Méthodologie ACH : Analyse des Hypothèses Concurrentes

Processus structuré pour évaluer objectivement de multiples explications possibles.

1. Identifier les Hypothèses

Formuler toutes les hypothèses possibles (H1, H2, H3...)

2. Collecter les Preuves

Rassembler toutes les informations et indices disponibles (E1, E2, E3...)

3. Créer la Matrice

Construire un tableau croisant Hypothèses et Preuves

4. Évaluer la Cohérence

Noter chaque croisement : Cohérent (+), Incohérent (-), Neutre (N)

5. Analyser et Conclure

Identifier l'hypothèse la plus robuste (moins d'incohérences). Rejeter les autres.

Objectif : Minimiser les biais cognitifs et éviter la conclusion prématurée.

La Matrice ACH (Outil)

- L'objectif n'est pas de prouver une hypothèse, mais de réfuter les autres.
- On note la cohérence de chaque preuve (E) avec chaque hypothèse (H).
- C = Cohérent / N = Neutre / I = Incohérent

Preuve	**H1 (Initié)**	**H2 (Criminel)**	**H3 (Concurrent)**
E1 : Leak à 3h du matin	**C** (peut le faire)	**C** (fuseau horaire)	**C** (fuseau horaire)
E2 : Que données clients	**I** (Incohérent, un initié prendrait aussi la R&D)	**C** (Cohérent, c'est ce qui a de la valeur)	**I** (Incohérent, un concurrent prendrait la R&D)
E3 : Revendication forum .onion	**I** (Incohérent, un initié ne fait pas ça)	**C** (Cohérent, TTP classique)	**I** (Incohérent, l'espionnage est discret)

La Matrice ACH - Analyse

- H1 et H3 sont infirmées (plusieurs "I").
- H2 (Criminel) est la seule hypothèse cohérente avec toutes les preuves. C'est l'hypothèse la plus probable.

Preuve	**H1 (Initié)**	**H2 (Criminel)**	**H3 (Concurrent)**
E1 : Leak à 3h du matin	**C** (peut le faire)	**C** (fuseau horaire)	**C** (fuseau horaire)
E2 : Que données clients	**I** (Incohérent, un initié prendrait aussi la R&D)	**C** (Cohérent, c'est ce qui a de la valeur)	**I** (Incohérent, un concurrent prendrait la R&D)
E3 : Revendication forum .onion	**I** (Incohérent, un initié ne fait pas ça)	**C** (Cohérent, TTP classique)	**I** (Incohérent, l'espionnage est discret)

Gestion de la Connaissance

Problème : Une enquête complexe dure des mois. Vos KIQ et matrices ACH s'accumulent.

Outils : Notion, OneNote, Mind Maps .

- Limite : Souvent sur le Cloud (mauvais OPSEC) ou pas assez connectés.

La Solution Avancée : Le "Second Cerveau" (Zettelkasten)

- Outil : Obsidian.md.

Pourquoi Obsidian ?

1. Local-First (OPSEC) :
    - Le "Vault" (coffre) de votre enquête reste sur votre VM chiffrée. Aucune donnée ne part sur un cloud.
    - C'est la seule option viable pour des enquêtes sensibles.
2. Liens Bidirectionnels (Le "Cerveau") :
    - C'est mieux qu'un tableur ou un doc Word.
    - Vous créez une Note par Entité : [[Pseudo Alpha]], [[IP 1.2.3.4]], [[Entreprise X]].
    - Quand vous écrivez dans votre journal : "Aujourd'hui, j'ai lié [[Pseudo Alpha]] à [[IP 1.2.3.4]]", l'outil crée le lien automatiquement.
3. Le Graphe Visuel (Le "Maltego des Idées") :
    - Obsidian génère un graphe visuel non pas de données (comme Maltego), mais de vos notes (votre analyse).
    - Vous voyez comment vos idées et vos entités sont connectées.

Workflow : Obsidian pour une Enquête

1. Créer un "Vault" (Coffre) par enquête (ex: Enquete_LockShadow).
2. Créer une Note "Hub" (ex: 00_KIQ.md) qui liste vos questions.
3. Créer un Journal (ex: Journal_de_Bord.md) pour noter vos actions (OPSEC).
4. Créer une Note par Entité :
    - Pseudo_VoidWalker.md
    - IP_8.8.8.8.md
    - Entreprise_Victime.md
    - Leak_X.md
5. Lier les Notes :
    - Dans Pseudo_VoidWalker.md, vous écrivez : "Est lié à [[IP_8.8.8.8.md]] (Source: Leak_X). A attaqué [[Entreprise_Victime.md]]."
6. Analyser le Graphe : Le graphe visuel montre que Pseudo_VoidWalker est un nœud central (un Hub) connectant plusieurs entités.

Travaux Pratiques

Scénario :

Lundi matin, des plans confidentiels du nouveau drone de l'entreprise 'AeroTech' sont publiés sur un forum du Dark Web. La direction est en panique. Votre équipe OSINT/CTI doit déterminer QUI est le plus probablement responsable avant la conférence de presse de 18h.

Outil : Un simple tableur ou même papier/crayon.

Travaux Pratiques - Les Faits

E1 : Le fichier a été posté à 03h00 du matin (heure de Moscou).

E2 : Le post utilise un argot typique des hackers russophones.

E3 : Les métadonnées PDF montrent que le fichier a été exporté par un utilisateur nommé "J.Smith".

E4 : "J.Smith" est un ingénieur d'AeroTech récemment licencié pour faute grave, très actif et en colère sur LinkedIn.

E5 : Un groupe d'hacktivistes écologistes ("GreenStrike") revendique l'action sur Twitter pour protester contre l'usage militaire du drone.

E6 : L'analyse technique du fichier montre qu'il a été stocké sur une clé USB avant d'être uploadé.

E7 : Le groupe "GreenStrike" n'a jamais réalisé d'intrusion technique complexe par le passé, uniquement des défigurations de sites web.

E8 : Les services IT confirment que le compte de J.Smith a été désactivé avant la date de création du PDF.

Travaux Pratiques - Le Déroulé

Étape 1 : Brainstorming des Hypothèses (H)

Il vous est demandé de définir les scénarios possibles.

- H1 : C'est une cyberattaque d'État (Russie, vu l'heure et l'argot).
- H2 : C'est une vengeance interne ("Insider Threat" - J.Smith).
- H3 : C'est le groupe hacktiviste "GreenStrike".

Étape 2 : Construction de la Matrice

Vous construisez un tableau : Hypothèses en colonnes, Preuves (E1 à E8) en lignes.

Étape 3 : L'Évaluation

C'est là que vous combattez les biais. La consigne est : "Pour chaque case, demandez-vous : si cette hypothèse est vraie, cette preuve est-elle cohérente ?"

Note : Utilisez une notation simple (+ = cohérent, - = incohérent, N = neutre).

Conclusion

La collecte sans planification est inutile.

La stratégie avancée repose sur 3 piliers :

- KIQ (Focus) : Transformer le "flou" en plan de collecte.
- ACH (Rigueur) : Utiliser la matrice pour infirmer les hypothèses et vaincre vos biais.
- Obsidian (Mémoire) : Utiliser un "Second Cerveau" local-first pour gérer la complexité à long terme.

Prochaine Étape : Maintenant que nous avons un plan stratégique, nous allons voir comment l'exécuter en temps réel (OSINT Opérationnel).

OSINT Opérationnel : Automatisation de la Veille

Enquête vs. Veille

L'Enquête

- Posture : "Historien".
- Question : "Que s'est-il passé ?"
- Action : Collecte ponctuelle de données (scraping, analyse de leaks).
- Résultat : Un rapport (statique).

La Veille Opérationnelle

- Posture : "Sentinelle" / "Gardien".
- Question : "Que se passe-t-il maintenant ?" et "Alerte-moi quand X se passera."
- Action : Collecte continue et automatisée.
- Résultat : Des Alertes (dynamiques).

L'analyste avancé ne cherche pas l'information, il met en place des systèmes pour que l'information vienne à lui.

De la KIQ à la Veille

Votre travail de planification définit quoi surveiller.

KIQ : "Quels sont les nouveaux sous-domaines de cible.com ?"

- Veille : "Alerte-moi quand crt.sh trouve un nouveau sous-domaine pour cible.com."

KIQ : "Qui sont les nouveaux employés de l'Entreprise X ?"

- Veille : "Alerte-moi quand la page 'Carrières' de l'Entreprise X change."

KIQ : "Quelles clés API l'Entreprise X fuite-t-elle ?"

- Veille : "Alerte-moi quand le mot-clé 'apiKey_cibleX' apparaît sur GitHub."

Partie 1 (No-Code) : Les Fondamentaux

1. **Google Alerts**

- Outil : google.com/alerts
- Usage : Surveiller des mots-clés (votre pseudo, votre entreprise, un nom de cible, un Dork site:pastebin.com "mot-clé").
- Limite : Ne surveille que ce que Google indexe. Lent.

2. **Le RSS (Le Standard Oublié)**

- Quoi : La plupart des blogs, sites d'actu, et même des forums/subreddits ont un flux RSS (/feed, /rss.xml).
- Outil (Agrégateur) : Feedly, Inoreader.
- Usage Opérationnel :
    a. Abonnez-vous à 50 blogs de cybersécurité (ex: BleepingComputer, Krebs).
    b. Créez un "filtre" dans Feedly/Inoreader pour le mot-clé de votre cible (ex: "LockShadow").
    c. Résultat : Vous recevez une alerte en temps réel quand votre cible est mentionnée.

Partie 1 (No-Code) : La Surveillance Visuelle

Le Problème : Que faire si le site n'a pas de flux RSS ? (ex: une page de profil LinkedIn, la page "Notre Équipe" d'un concurrent).

La Solution : La Surveillance de Changement (Visuel)

Outils : Distill.io, Visualping.io (extensions navigateur).

Comment ça marche :

1. Vous sélectionnez une zone de la page à surveiller (ex: le bloc "Employés").
2. L'outil prend une "photo" de cette zone.
3. Il vérifie toutes les heures (ou minutes).
4. Si un seul pixel change, vous recevez une alerte (email, Slack, Discord).

Usage : "Alerte-moi si un nouvel employé apparaît sur la page 'Équipe' de la Cible X."

Partie 2 (Code/Low-Code) : Les "Agents"

Le Problème : Les outils No-Code sont limités (pas d'API, pas de logique complexe).

La Solution : Le Modèle "Agent"

Concept : Vous créez des "ouvriers" (agents) qui s'enchaînent.

Outil "Low-Code" : n8n (Open Source).

- C'est le "Maltego" de l'automatisation. Un IFTTT open-source que vous hébergez.

Exemple de flux (Scénario) :

- Agent 1 (RSS) : "Lire le flux RSS de crt.sh pour %.cible.com."
- Agent 2 (Filtre) : "Ne garder que les nouveaux résultats."
- Agent 3 (Webhook) : "Envoyer une alerte dans notre canal Discord/Slack."

Résultat : Une chaîne de renseignement 100% automatisée et personnalisée.

Partie 2 (Code) : La Surveillance GitHub

Le Problème : Les développeurs fuitent des secrets (clés API, mots de passe, IP internes) dans leur code public sur GitHub.

La Collecte : Utiliser les "GitHub Dorks".

- "cible.com" "apiKey"
- "cible.com" password
- filename:.env "prod_password"
- org:NomDeLaCible "internal_ip"

La Veille :

- Outils : truffleHog, GitGuardian (commercial), ou des scripts Python custom utilisant l'API GitHub.
- Action : Ces outils scannent GitHub en temps réel (via l'API) et vous alertent (par email/Slack) à la seconde où un commit contient un de vos mots-clés.
- C'est de l'OSINT offensif/défensif critique.

Partie 2 (Code) : Le Script Python

Parfois, l'outil le plus simple est un script que vous avez déjà écrit.

Le Scénario :

- Vous avez écrit un script requests/BeautifulSoup pour scraper la liste des employés.
- Comment le transformer en "Veille" ?

La Solution :

1. Modifier le script :
    - Le script doit sauvegarder son résultat dans un resultats_hier.txt.
    - À chaque exécution, il compare son nouveau résultat avec resultats_hier.txt.
    - S'il y a une différence, il envoie un email ou un message (via une API simple).
2. Planifier le script - Voir Slide Suivante.

Partie 3 (OPSEC Infra) : Où tourne la Veille ?

Le Problème :

- Vous ne pouvez pas faire tourner 50 scripts 24/7 sur votre PC portable.
- Vous ne devez JAMAIS faire une veille depuis votre IP personnelle/domicile. C'est une violation OPSEC fondamentale (votre IP sera loguée, bannie).

La Solution Opérationnelle :

1. Infrastructure : Un VPS (Virtual Private Server).
    - Un petit serveur Linux loué 5€/mois (ex: Vultr, DigitalOcean, Scaleway).
    - Il est anonyme (payé avec une carte virtuelle) et jetable.
    - C'est votre "base d'opérations" dans le cloud.
2. Planification : Cron / Cronjob.
    - C'est l'outil "standard" de Linux pour dire :
    - "0 3 * * * python3 /home/osint/check_cible.py"
    - Traduction : "Exécute mon script Python tous les jours à 3h00 du matin."

Travaux Pratiques

Scénario :

Vous devez surveiller le nouveau PDG d'un concurrent, 'M. Durant'. Votre directeur veut être alerté uniquement s'il parle de 'fusion', 'acquisition' ou de 'notre entreprise', et il veut recevoir l'alerte sur un canal Telegram dédié, pas par email pour ne pas être noyé.

Outils : N8N

TP 1 - Source de Données

Objectif : Définir où écouter.

Action : Au lieu d'un simple Google Alert (trop basique), nous allons utiliser un flux RSS.

Exemple : Créer une alerte Google News sur "M. Durant" AND (PDG OR CEO), récupérer le flux RSS de cette alerte.

Dans l'outil No-Code : Utiliser le module "RSS Monitoring" comme déclencheur.

TP 2 - Filtrage

Objectif : Éliminer le bruit. C'est l'étape clé du niveau "Avancé".

Action : Intercaler un module de "Filtre" ou de "Router" après le RSS.

La Règle : L'automatisation ne doit continuer QUE SI le titre ou la description de l'article contient des mots-clés critiques : fusion, rachat, [Nom de de l’entreprise].

Bonus IA : Insérer un module OpenAI pour demander à GPT : "Lis ce titre, et dis-moi sur une échelle de 1 à 5 si c'est une nouvelle stratégique urgente". Filtrer si score > 4.

TP 3 - Notification Ciblée

Objectif : Alerter au bon endroit.

Action : Si le filtre est passé, utiliser un module final :

- Telegram Bot (Très apprécié en CTI pour les canaux d'alerte).
- Slack / Discord Webhook.
- Google Sheets (Pour l'archivage, pas pour l'alerte).

Consigne : Le message d'alerte doit être propre : "⚠️ ALERTE STRATÉGIQUE : [Titre de l'article] - Lien : [URL]".

Conclusion

L'OSINT avancé n'est pas une "enquête", c'est une "veille". L'information doit venir à vous.

Commencez avec le No-Code (80/20) :

- Google Alerts et RSS (Feedly) couvrent le "bruit" du web.
- Distill.io / Visualping couvrent les sites statiques sans RSS.

Passez au Code (le 20%) :

- n8n ou des scripts Python pour les API, la logique complexe et les sources comme GitHub.

L'OPSEC est reine :

- Votre veille tourne sur un VPS (IP neutre), planifiée avec Cron.
- Ne jamais faire de veille depuis votre machine personnelle.

Prochaine Étape : Nous avons tous les blocs. Il est temps de compléter le tableau.

Problème Final : La “Noyade d’Alertes”

Votre "veille" fonctionne. Elle génère 100 alertes par jour.

Votre "Second Cerveau" (Obsidian) est parfait pour votre analyse (texte, liens).

Mais où stockez-vous les données brutes (Indicateurs) pour les corréler sur 5 ans ?

Solution: Plateformes de CTI (Threat Intelligence)

Plateformes de CTI

Ce sont des bases de données structurées conçues pour le renseignement, la corrélation et le partage. L'OSINT est leur principal carburant.

MISP (Malware Information Sharing Platform)

- Rôle : Le standard (surtout Blue Team) pour partager des Indicateurs de Compromission (IOCs).
- Votre OSINT : L'IP de Shodan, le domaine du blog ou le hash d'un malware deviennent des "Attributs" dans un "Événement" MISP.

Plateformes de CTI

OpenCTI (Open Cyber Threat Intelligence)

- Rôle : Un "Maltego stratégique". Il ne stocke pas seulement les IOCs, il cartographie les relations entre tout.
- Votre OSINT :
  - @RedShadow_Truth n'est pas une "note", c'est une entité "Acteur" (Intrusion Set).
  - red-shadow-analytics.com est une entité "Domaine".
  - L'adresse ETH est une entité "Wallet".
  - OpenCTI vous montre visuellement comment l'Acteur utilise le Domaine et le Wallet.

Conclusion : L'OSINT Opérationnel ne s'arrête pas à l'alerte. Il alimente ces plateformes pour transformer vos découvertes en intelligence structurée, partageable et corrélable dans le temps.

Contre-OSINT (Red & Blue Team)

Définition : Qu'est-ce que le Contre-OSINT ?

C'est la discipline qui consiste à réduire et à protéger sa propre "Surface d'Attaque OSINT".

Pourquoi est-ce vital pour VOUS (l'analyste) ?

- Doxing : Votre cible (ou ses alliés) peut tenter de vous identifier en retour.
- Contamination : Une erreur OPSEC peut lier votre enquête à votre identité réelle.

OSINT (Red Team - Offensif)	**CONTRE-OSINT (Blue Team - Défensif)**
Question : "Que puis-je trouver sur la Cible X ?"	**Question :** "Que peut trouver la Cible X sur *moi* ?"
Objectif : Collecter des informations.	**Objectif :** Empêcher la collecte d'informations.
Action : Scraping, Analyse de Leaks, GEOINT.	**Action :** Cloisonnement, Minimisation, Audit.

Partie 1 : Penser "Red Team" (S'auditer soi-même)

La meilleure défense est un bon audit. Utilisez les KIQ et les techniques contre vous-même.

KIQ 1 (Identité) : Que trouve-t-on avec mon email personnel ?

- Outil : DeHashed, HIBP.
- Révèle : Mots de passe de leaks, pseudos liés.

KIQ 2 (Pseudos) : Que trouve-t-on avec mes vieux pseudos (forums, jeux) ?

- Outil : Sherlock, WhatsMyName.
- Révèle : Comptes oubliés, photos, opinions.

Partie 1 : Penser "Red Team" (S'auditer soi-même)

La meilleure défense est un bon audit. Utilisez les KIQ et les techniques contre vous-même.

KIQ 3 : Que trouve-t-on sur mon entreprise ?

- Outil : crt.sh, GitHub Dorks, Shodan.
- Révèle : Sous-domaines de dev, clés API fuitées, serveurs ouverts.

KIQ 4 : Mes photos de vacances sont-elles géolocalisables ?

- Outil : Google Earth, SunCalc.
- Révèle : Ma maison, mon hôtel, mes habitudes.

Partie 2 : Penser "Blue Team" (Le Playbook Défensif)

Comment réparer ce que le "Red Team" (vous) a trouvé ?

Les 3 Piliers de la Défense :

1. **Minimisation** : Réduire ce qui est public. "Moins il y en a, moins on en trouve."
2. **Cloisonnement** : Séparer les identités. "Ne jamais croiser les flux."
3. **Obfuscation / Inondation** : Rendre la collecte difficile. "Cacher une aiguille dans une botte de foin."

Minimisation (Défense Personnelle)

Paramètres de Confidentialité (SOCMINT) :

- LinkedIn : Masquez votre liste de relations (empêche le scraping).
- Facebook/Instagram : TOUT en "Amis seulement". Ne rendez pas votre liste d'amis publique.
- Twitter/X : Passez en compte privé ou, si public, activez "Protéger vos posts" (empêche la suppression).

Métadonnées (IMINT) :

- Désactivez le "Tag GPS" sur l'application Appareil Photo de votre téléphone.
- Utilisez des "EXIF Scrubber" (ex: exiftool) avant de poster une image.

Historique (GEOINT) :

- Allez sur Google Maps Timeline (google.com/timeline) et supprimez votre historique de localisation.

Droit à l'Oubli (RGPD) :

- Demandez activement aux "Data Brokers" (Pipl, etc.) de supprimer vos informations.

Minimisation (Défense Corporate)

Contre-GitHub :

- NE JAMAIS hardcoder de clés API, mots de passe, ou IP dans le code.
- Utilisez des variables d'environnement (.env) et ajoutez .env au fichier .gitignore.
- Utilisez des outils (TruffleHog, GitGuardian) en défense (Blue Team) pour scanner vos propres dépôts.

Contre-Scraping :

- Demandez aux employés de ne pas lister les versions logicielles spécifiques sur LinkedIn (ex: "Administrateur de base de données" au lieu de "Admin MongoDB v3.4 non-patchée").

Contre-Infrastructure :

- N'utilisez pas de sous-domaines évidents (dev., test., vpn.).
- Utilisez des noms obscurs (portal-ext-v7., gw-8b.) pour rendre crt.sh moins parlant.

Le Cloisonnement (OPSEC Analyste)

C'est la règle la plus importante pour vous. Vos identités ne doivent jamais se croiser.

Identité 1 : Personnelle (Protégée)

- Votre vrai nom (Jean.Dupont@gmail.com).
- Vos vrais amis/famille.
- Défense : Minimisation maximale.

Identité 2 : Professionnelle (Publique)

- Votre nom de travail (J.Dupont@entreprise.com).
- Votre LinkedIn.
- Défense : Image professionnelle, minimisation des détails techniques.

Identité 3 : "Sock Puppet" (Anonyme/Jetable)

- Faux nom, faux email (WarriorGeek92@proton.me).
- Utilisé dans votre VM d'enquête.
- Défense : N'existe pas.

La Faute OPSEC Critique : Utiliser le même email de récupération, le même mot de passe, ou la même photo de profil entre ces identités.

L'Inondation (Défense "Gray Team")

Concept : Si vous ne pouvez pas cacher l'information, vous pouvez la noyer dans le bruit (désinformation).

Objectif : Augmenter le coût et le temps de l'analyste adverse (lui faire perdre du temps).

Exemple (Personnel) :

- Un acteur menaçant (ex: un PDG) crée 10 faux profils LinkedIn avec le même nom, mais des carrières et des villes différentes.
- L'analyste qui le scrape obtient 11 résultats et doit passer des heures à déterminer lequel est le vrai.

Exemple (Corporate) :

- Publier de fausses offres d'emploi mentionnant des technologies que vous n'utilisez pas (ex: "Expert Java" alors que vous êtes 100% Python) pour tromper les concurrents.

Avertissement : C'est une tactique "Red Team" / "Gray Team". Éthiquement très discutable et potentiellement illégale (usurpation d'identité).

Travaux Pratiques

Consigner : Vous n'êtes plus vous-même. Vous êtes un attaquant motivé qui veut détruire votre réputation ou voler votre identité. Vous avez 45 minutes pour remplir ce tableau de chasse.

IMPORTANT : Cet atelier est privé. Ne partagez pas vos résultats avec le groupe. C'est pour votre propre bénéfice.

TP1 - Kill List 1/2

Le Vecteur "Identifiant Unique" : Prenez le pseudo que vous utilisiez il y a 10 ans (forums de jeux, vieux Twitter). Utilisez des outils comme WhatsMyName ou Sherlock. Où est-il encore actif ?

Le Vecteur "Fuite de Données" : Ne vous contentez pas de HaveIBeenPwned. Si vous avez accès à des bases de données (DeHashed, Intelligence X), cherchez vos vieux emails.

L'objectif : Trouver un vieux mot de passe en clair.

TP1 - Kill List 2/2

Le Vecteur "Image" : Faites du Reverse Image Search (PimEyes si disponible, ou Yandex/Google Lens) sur votre photo de profil actuelle, mais aussi sur une photo de vous datant d'il y a 5-10 ans.

Le Vecteur "Familial/Proche" (Le plus dangereux) : Regardez les profils Facebook/Instagram publics de vos parents, frères/sœurs ou conjoint(e).

L'objectif : Trouver votre date de naissance exacte, votre ville d'origine, ou le nom de votre animal de compagnie (réponses aux questions de sécurité secrètes).

TP 2 - Weaponization

Consigne : "Choisissez vos 3 meilleures trouvailles de la Phase 1. Pour chacune, écrivez un scénario d'attaque crédible en 3 lignes."

Exemples de livrables attendus :

- Trouvaille : J'ai trouvé le nom de mon chien sur le Facebook public de ma mère + ma date de naissance sur un vieux forum.
  - Scénario : "Je peux tenter une réinitialisation de mot de passe ("Mot de passe oublié") sur le compte Gmail principal en répondant aux questions de sécurité, ou appeler le service client de l'opérateur mobile pour tenter un SIM-Swapping en utilisant la date de naissance comme validation."
- Trouvaille : J'ai trouvé un mot de passe en clair de 2014 ("Rover1234").
  - Scénario : "Je vais tenter du Credential Stuffing sur les comptes Netflix, Spotify et Amazon actuels avec ce mot de passe et ses variantes (Rover1234!, Rover2025)."

TP 3 - Remédiation

Consigne : "Pour le scénario le plus critique, proposez une mesure corrective immédiate."

Exemples :

- Utiliser un gestionnaire de mot de passe, activer le 2FA partout,
- "Empoisonner" les données (mettre une fausse date de naissance sur les réseaux sociaux non critiques).

Conclusion

Le Contre-OSINT n'est pas une option, c'est une nécessité pour la survie de l'analyste.

Vous ne pouvez pas défendre ce que vous ne connaissez pas.

La "Blue Team" (Défense) commence par la "Red Team" (Audit).

Vos 3 Actions Clés:

- Minimiser : Verrouillez la confidentialité de vos réseaux sociaux (listes d'amis).
- Cloisonner : Assurez-vous qu'aucun lien (email de récup, mdp) n'existe entre votre identité personnelle et vos sock puppets.
- Nettoyer : Supprimez les vieux comptes (identifiés à l'instant) que vous n'utilisez plus.

L'analyste OSINT est sa propre première cible.

Module 4 : L'Ère de l'Analyste Augmenté

Introduction

De la Recherche à la Synthèse

L'OSINT consiste à trouver l'aiguille dans la botte de foin. L'OSINT augmenté par l'IA ne cherche plus l'aiguille : il utilise un aimant géant pour trier la botte de foin instantanément.

- **Avant** : Vous passiez 4 heures à lire des forums pour trouver une mention pertinente.
- **Maintenant** : Vous collectez tout le forum, et l'IA vous dit en 30 secondes : "Voici les 3 posts pertinents et pourquoi ils le sont."

Les rôles de l’IA en OSINT

Nous ne parlons pas ici de remplacer l'analyste, mais de l'augmenter via **trois rôles distincts** :

L'Assistant (LLM - Large Language Model) : C'est votre binôme junior infatigable. Il résume des rapports de 50 pages, traduit des textes complexes et explique des lignes de code obscur.

L'Ouvrier (Agents & Automation) : C'est votre main-d'œuvre. Il exécute les tâches répétitives sans se plaindre (via n8n).

L'Expert (Vision & Audio) : C'est vos yeux et vos oreilles bioniques. Il voit des détails dans les images satellites ou transcrit des heures d'audio que l'humain raterait par fatigue.

La Ligne Rouge (OPSEC & Éthique)

L'utilisation de l'IA introduit deux risques mortels pour l'enquêteur.

Hallucinations : L'IA est un menteur confiant. Elle peut inventer des faits, des citations ou des liens juridiques.

- Règle d'Or : Human in the Loop (HITL). Ne jamais copier-coller une conclusion d'IA sans vérifier la source primaire.

Fuite de Données (OPSEC) :

- Tout ce que vous tapez dans ChatGPT (OpenAI) ou Claude (Anthropic) part sur leurs serveurs.
- Interdit : Ne JAMAIS envoyer de noms, emails, documents confidentiels ou secrets industriels à une IA en ligne. Pour cela, nous utiliserons les LLM Locaux.

LLM au service du Renseignement

TECHINT Augmenté (Code Assistant) Vous n'avez pas besoin d'être un développeur expert pour créer des outils sur mesure.

- **Génération de Regex** : Demandez à l'IA : "Génère une Regex pour grep.app qui capture les clés API Google Maps, mais ignore les exemples de documentation."
- **Scripting de Scraping** : Demandez : "Écris un script Python utilisant Selenium pour extraire tous les liens 'href' de la classe '.user-profile' sur cette page HTML ci-dessous."
- **Analyse de Code** : Collez un script suspect trouvé sur GitHub et demandez : "Que fait ce code ? Y a-t-il une fonction malveillante ou une exfiltration de données ?"

SOCMINT Augmenté

Au-delà de la lecture, l'IA permet de "sentir" la masse de données.

- **Analyse de Sentiment** : Sur un dataset de 1000 tweets extraits (via scraping), l'IA peut classer chaque message : Positif, Négatif, Agressif, Urgent. Cela permet de détecter une radicalisation ou un changement de ton soudain chez une cible.
- **Stylométrie & Profilage** : En analysant le style d'écriture (vocabulaire, fautes récurrentes, structure de phrase), l'IA peut suggérer si deux pseudonymes différents sont probablement tenus par la même personne.

Traduction & Contexte Culturel

Les traducteurs classiques (Google Translate) échouent sur l'argot, le "Leetspeak" ou les dialectes régionaux.

Usage : Un LLM bien prompté ("Agis comme un expert en argot russe des années 2020") peut décrypter le sens caché d'une conversation sur un forum Dark Web, là où un traducteur littéral ne verrait que du non-sens.

OPSEC : L'IA en Local (LocalLLM)

Pourquoi le Local ? Pour traiter des documents confidentiels (leaks, rapports internes) sans briser l'OPSEC. En local, les données ne quittent jamais votre machine (fonctionne même en "Air Gap" sans internet).

LM Studio / Ollama (+ WebUI) :

- **Quoi** : Des logiciels simples pour télécharger et faire tourner des modèles Open Source (Llama 3, Mistral, Gemma) sur un PC portable standard.
- **Configuration** : Nécessite idéalement 16Go de RAM (et un GPU Nvidia pour la vitesse, mais fonctionne sur CPU).

GPT4All (L'IA Documentaire) :

- **Concept** : RAG (Retrieval Augmented Generation).
- **Usage** : Vous lui donnez un dossier contenant 50 PDF (leaks). Vous lui posez des questions ("Qui est mentionné en lien avec le Projet X ?"). Il répond en citant les sources précises dans vos documents, sans internet.

Automatisation Intelligente

Le Concept d'Agent

- Chatbot : Il ne fait que parler (ChatGPT).
- Agent : Il a des "bras". Il peut effectuer des recherches Google, utiliser une calculatrice, interroger une base de données ou envoyer un email.

n8n (L'Orchestrateur)

- Nous remplaçons les outils vieillissants par n8n, l'outil d'automatisation visuel (Low-Code) de référence.
- Pourquoi n8n ? Il s'installe en local (Docker) ou sur votre VPS, possède une interface "drag & drop" claire, et intègre nativement des Nœuds IA (AI Nodes).

Automatisation Intelligente : Scénario Type

1. **Trigger** (Déclencheur) : Un flux RSS (ex: BleepingComputer) détecte un nouvel article.
2. **Agent IA** (Traitement) :
   - Le nœud IA lit l'article.
   - Prompt système : "Analyse cet article. Est-ce qu'il parle d'une faille critique ? Si oui, résume-le en 3 points et extrais les noms des logiciels touchés."
3. **Routeur** (Logique) : Si la réponse est "Non critique" -> Stop. Si "Critique" -> Continuer.
4. **Action** (Sortie) : Envoie une alerte formatée sur votre canal Discord privé ou Telegram.

GEOINT Augmenté

L'analyse manuelle de cartes est fastidieuse. La Computer Vision (CV) permet de scanner de vastes zones.

Détection d'Objets : Utilisation de modèles pour repérer automatiquement des formes spécifiques sur des images satellites (ex: identifier des silos à missiles, des types d'avions sur un tarmac, ou des changements de terrain après une catastrophe).

Teachable Machine (Google)

Le Concept : Entraîner son propre modèle de reconnaissance d'image sans écrire une ligne de code.

Cas d'Usage OSINT :

- Vous avez 10 heures de vidéos de manifestations.
- Vous voulez savoir quand apparaît le logo d'un groupe extrémiste spécifique.
- Vous donnez 50 photos du logo à Teachable Machine + 50 photos "bruit".
- Vous exportez le modèle pour scanner vos vidéos automatiquement.

Précaution: Ne pas envoyer des données sensibles !

Transcription - OpenAI Whisper (Version Locale)

Outil révolutionnaire qui transcrit l'audio (et la vidéo) en texte avec une précision quasi-humaine, dans de multiples langues.

Usage : Transformer des podcasts, des écoutes ou des vidéos YouTube en texte pur pour pouvoir faire du Ctrl+F (Recherche) sur des mots-clés (noms, lieux).

https://github.com/openai/whisper

Travaux Pratiques

Contexte

Votre veille automatisée sur le PDG (le TP précédent) fonctionne. Mais elle génère trop de faux positifs. Votre directeur reçoit des alertes inutiles à chaque fois que le mot 'achat' apparaît dans un contexte banal.

Votre Mission

Reprenez votre scénario Make/n8n existant. Remplacez le filtre basique à 'mots-clés' par un cerveau IA (OpenAI/Mistral API) qui va lire l'article et décider lui-même de son importance.

TP - Déroulé

1. Ouvrez votre précédent scénario.
2. Supprimez le module "Filtre" (le losange avec les conditions IF contains).
3. Insérez un module HTTP / OpenAI à la place.
4. Rédigez le prompt pour l'IA.
    - Exemple de prompt à écrire : "Tu es un analyste CTI senior. Voici le titre d'un article : [Variable Titre RSS]. Analyse-le. Si cet article parle d'une cyberattaque critique ciblant le secteur bancaire, réponds uniquement 'OUI'. Sinon, réponds 'NON'."
5. Ajoutez un nouveau filtre simple après l'IA : Si la réponse de l'IA est "OUI", alors envoyer l'alerte.

Défi Final

Philosophie de l'Exercice

Ce n'est pas un TP guidé. Jusqu'à présent, vous avez appris à utiliser des outils spécifiques pour des tâches précises. Dans cette étude de cas finale, vous êtes seuls face à la complexité du réel.

Vos objectifs ne sont plus techniques, ils sont stratégiques :

1. **Briser les silos** : Vous devrez combiner SOCMINT, GEOINT, FININT et TECHINT. Aucune discipline ne suffira seule à résoudre le cas.
2. **Gérer l'impasse** : Vous allez rencontrer des murs (dead ends). Votre capacité à pivoter créativement sera évaluée.
3. **Maintenir l'OPSEC** : Vous enquêterez sur une cible potentiellement hostile. Une seule erreur d'OPSEC (like accidentel, mauvaise IP) et la mission est considérée comme échouée.
4. **Produire du Renseignement** : Trouver le nom du coupable ne suffit pas. Vous devez évaluer sa dangerosité, ses capacités et son réseau.

Scénario

Contexte Client : Une ETI (Entreprise de Taille Intermédiaire) européenne dans le secteur de la logistique de défense subit une campagne de harcèlement et de tentatives d'intrusion très ciblées. Le RSSI est persuadé qu'il ne s'agit pas d'un groupe étatique, mais d'un "mercenaire" ou d'un petit groupe privé très compétent qui cherche à monnayer des accès.

Le Point d'Entrée : La seule trace tangible est un compte Telegram qui a contacté un employé de l'entreprise pour lui proposer de vendre ses identifiants.

- Handle Telegram : @Kael_Sec_Ops (Le compte semble actif par intermittence).
- Message intercepté : "On sait que votre VPN Cisco n'est pas à jour. Payez 5 ETH sur [adresse supprimée par le client] ou on leak l'architecture réseau d'ici 48h. On n'est pas des amateurs, checkez notre réputation sur Dread."

Vos Objectifs :

- Attribution : Qui se cache derrière @Kael_Sec_Ops ? Est-ce un individu ou un groupe ?
- Évaluation de la menace : Sont-ils crédibles ("checkez notre réputation") ou est-ce du bluff ?
- Localisation : D'où opèrent-ils principalement ?

Règles d'Engagement

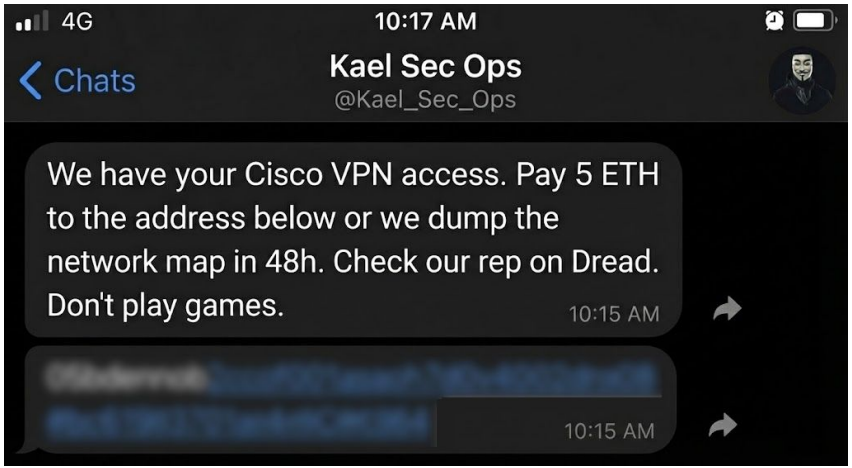
Temps imparti : 2h30 (simulant une réponse à incident sous pression).

OPSEC : Niveau Rouge. Utilisation obligatoire de machine dédiée (VM), VPN, et Sock Puppets "chauffés". Interdiction formelle d'interagir directement avec la cible (pas de message).

Livrable : Vous ne rendrez pas une liste de liens. Vous rendrez une Note de Renseignement d'une page maximum, destinée au CODIR du client, incluant :

- Résumé exécutif.
- Identité/Localisation suspectée avec niveau de confiance (High/Medium/Low Confidence).
- Chronologie des faits techniques.
- Recommandations immédiates.

Matériel



0x732c8beaa96b9416b205de654e93eac63c3e8d9e

Conclusion

Synthèse

Vous avez terminé la formation OSINT Avancée.

- **Jour 1** (Tactique) : Nous avons appris à franchir la "zone grise" (Légalité, Scraping, Deep/Dark Web) pour collecter l'information.
- **Jour 2** (Analyse) : Nous avons appris à comprendre l'information (SNA, GEOINT, FININT) et à la visualiser (Maltego).
- **Jour 3** (Stratégie) : Nous avons appris à penser comme des analystes (Planification, ACH), à opérer (Veille) et à nous défendre (Contre-OSINT).

Rappel Final : Les outils changent chaque jour. La méthodologie (Planifier, Collecter, Analyser, Rapporter) et l'Éthique sont les seules choses qui restent.

Ne cessez jamais d'être curieux, mais ne cessez jamais d'être rigoureux et prudent.

Pour aller plus loin - Lecture

"Open Source Intelligence Techniques" (par Michael Bazzell)

- La référence absolue, mise à jour régulièrement. C'est une encyclopédie dense de techniques, d'outils et de méthodologies d'enquête. Incontournable pour approfondir.

"Extreme Privacy" (par Michael Bazzell)

- Pour maîtriser l'autre côté de la pièce : l'OPSEC et la protection de la vie privée, un sujet que nous avons couvert dans le Module 1.

Pour aller plus loin - Sites et Forums

Bellingcat - https://fr.bellingcat.com

- Le site de référence pour le journalisme d'investigation basé sur l'OSINT. Essentiel pour voir l'application concrète des techniques sur des sujets géopolitiques, de conflit ou criminels.

Sector035 (Week in OSINT) - https://sector035.nl/articles/category:week-in-osint

- L'incontournable "revue de presse" hebdomadaire de l'OSINT. Si vous ne devez suivre qu'une seule chose pour rester à jour sur les nouveaux outils et techniques, c'est celle-ci.

Pour aller plus loin - Sites et Forums

Nixintel - https://nixintel.info

- Un blog personnel rempli de guides techniques approfondis (IMINT, SOCMINT) et de réflexions méthodologiques précieuses.

OSINT Curious - https://www.osintcurio.us

- Un collectif d'experts qui publie des articles, des "10-Minute Tips" et des analyses. Très pédagogique et axé sur la communauté.

Pour aller plus loin - Réseaux sociaux

X (Twitter) & Mastodon : L'écosystème OSINT est extrêmement actif sur ces plateformes. C'est là que les nouveaux outils et les nouvelles techniques sont partagés en premier.

- @bellingcat (Collectif)
- @Sector035 (Pour le "Week in OSINT")
- @Nixintel (Techniques avancées)
- @osintcurious (Collectif)
- @jakecreps (Expert SOCMINT)

Discord / Reddit :

- r/OSINT : Un subreddit très actif pour les questions, partages d'outils et études de cas.
- Serveur Discord "Project Owl" (OSINT Curious) : Une communauté très active pour échanger et poser des questions.

Pour aller plus loin - Pratique

OSINT games - https://www.myosint.training

- Des challenges variés (IMINT, SOCMINT, etc.) avec différents niveaux de difficulté pour s'entraîner sur des cas concrets et ludiques.

Challenges de @Sector035 (Quiztime)

- Suivez le hashtag #quiztime sur X. Des challenges de géolocalisation (IMINT) sont postés quasi-quotidiennement par la communauté. C'est le meilleur moyen de s'entraîner 10 minutes par jour.

CTF (Capture The Flag)

- Participez à des CTF qui ont une catégorie OSINT (ex: HackTheBox, TryHackMe, et de nombreux CTF événementiels).