Atelier régional pour l'Afrique francophone

Étude du lien entre l'intelligence artificielle et l'atténuation de la menace interne

Intrusion Typhon – Agence de réglementation atomique – (Atomic Regulatory Agency - ARA)

Usage de l'intelligence artificielle pour identifier les menaces internes

L’organisme de réglementation atomique

Vue d’ensemble : l’Agence de réglementation atomique (ARA)

L’Agence de réglementation atomique (ARA) est l’autorité compétente pour votre pays.

- Elle délivre les licences et réglemente l’utilisation des matières radioactives et nucléaires du pays.
- Elle développe et applique les réglementations en matière de sûreté, de sécurité et d’intégrité.
- Les inspecteurs recueillent des données lors des audits et archivent les activités d’inspection.

Dans le cadre de cet effort, l’autorité de réglementation a développé des outils en ligne permettant aux inspecteurs de l’ARA de transmettre, traiter et stocker les données recueillies lors des audits ou évaluations.

Le défi de sécurité :

- Les inspecteurs disposent d’un accès privilégié aux données sensibles de l’installation nucléaire.
- Les outils en ligne créent de nouvelles vulnérabilités numériques et des risques d’exposition des données.
- Les mesures de sécurité traditionnelles peuvent ne pas détecter les changements comportementaux subtils.
- Il est nécessaire d’identifier de manière proactive les menaces internes potentielles.
- **Exercice du jour :** en groupes, vous étudierez la manière dont l’intelligence artificielle peut renforcer les capacités de détection des menaces internes de l’ARA tout en maintenant l’efficience opérationnelle.

L’inspecteur à la tête d’une équipe ARA reçoit un ordinateur portable permettant d’utiliser les solutions en ligne de l’ARA, notamment l’intégration aisée des données.

Exercice en groupe : L’IA pour la détection des menaces internes

Vue d’ensemble : l’Agence de réglementation atomique (ARA)

La mission : la direction de l’ARA a chargé votre équipe d’identifier comment l’intelligence artificielle peut renforcer nos capacités de détection des menaces internes. Vous devrez réfléchir selon la perspective des professionnels de sécurité et celle des inspecteurs, pour trouver des solutions qui améliorent la protection sans perturber les opérations.

Consignes de l’exercice :

- Formez des groupes de 4 à 5 participants
- Travaillez sur chaque élément du scénario et effectuez les tâches

Vos objectifs incluent :

- *Identifier les capacités d’IA qui pourraient détecter les menaces internes*
- *Envisager divers scénarios de menaces et méthodes de détection*
- *Élaborer une liste de solutions à base d’IA hiérarchisées*
- *Préparer une présentation de groupe de 5 minutes*

4 mars 2024 : L'ARA est attaquée : ransomware Typhon

Détails de l'incident

Typhon : qu'est-ce que c'est ?

Typhon est une plateforme Ransomware-as-a-Service (RaaS) qui utilise le modèle de double extorsion (exfiltration ET chiffrement).

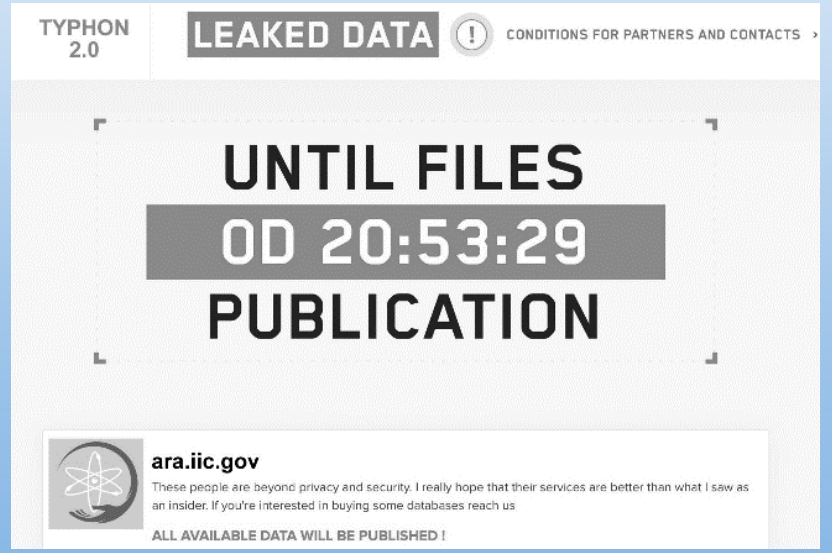
- **Exfiltration** : vole les données sensibles et menace de les publier
- **Chiffrement** : verrouille les fichiers en les rendant inaccessibles, jusqu'à ce que la rançon soit payée
- Les attaquants exigent un paiement pour : (1) déchiffrer les fichiers ET (2) ne pas divulguer les données volées
- Même si la rançon est payée, rien ne garantit que les données ne soient pas vendues ou divulguées ultérieurement

Typhon est partout dans l'ARA !

Typhon est installé sur plusieurs serveurs internes de l'ARA :

- Toutes les bases de données contenant des informations nucléaires sensibles sont chiffrées
- Toutes les sauvegardes de données sont chiffrées
- Les outils et les applications d'inspection sont chiffrés

Était-ce l'œuvre d'un agresseur interne ?



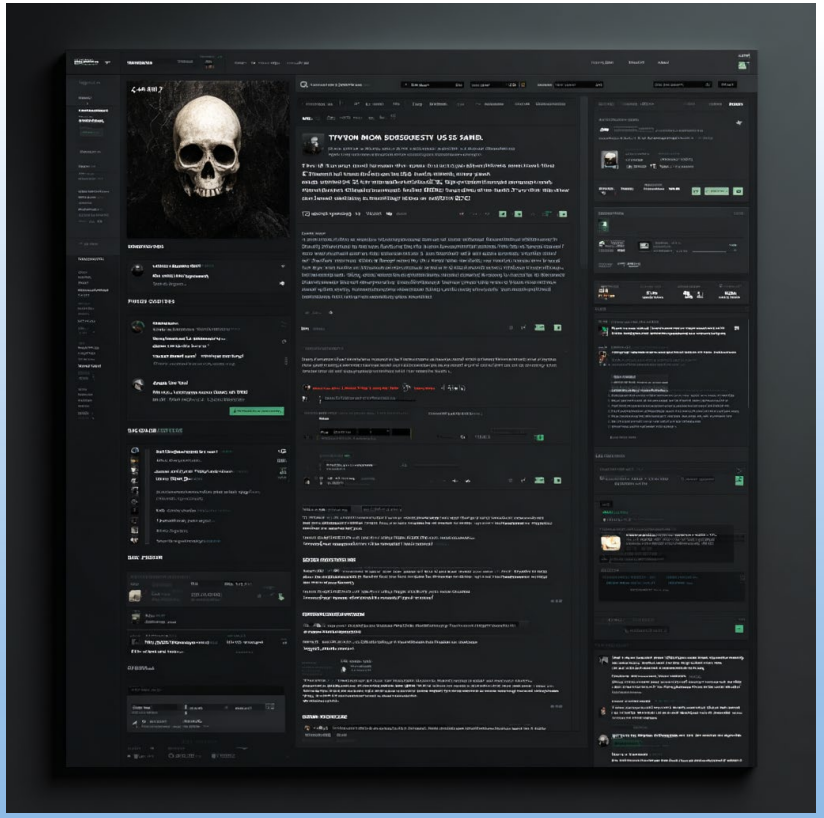
4 mars 2024 : L'ARA est attaquée : agresseur interne ?

Détails de l'incident

S'agit-il d'un agresseur interne ?

Deux mois plus tôt, Typhon a ouvertement recruté des agresseurs internes sur des forums du Dark Web :

- Proposait jusqu'à 1 million USD pour installer un logiciel malveillant sur les réseaux d'entreprise
- Ciblait les employés disposant d'un accès privilégié ou ayant des difficultés financières
- A publié des guides « d'instruction » pour contourner les contrôles de sécurité
- L'ARA était spécifiquement désignée en tant que « cible de grande valeur » en raison des données nucléaires sensibles



Pistes pour susciter la discussion :

- **Détection des tendances comportementales**
  Comment l'IA pourrait analyser des tendances comportementales numériques, afin d'identifier des personnes susceptibles de répondre à l'offre de recrutement de 1 million $ de Typhon ou qui étaient déjà compromises ?
- **Indicateurs de contraintes financières**
  Quels modèles d'IA/d'apprentissage machine pourraient établir une corrélation entre les données RH, les habitudes d'accès et les facteurs de risque externes, afin d'identifier les employés soumis à des pressions financières susceptibles d'être ciblés par les recruteurs de Typhon ?

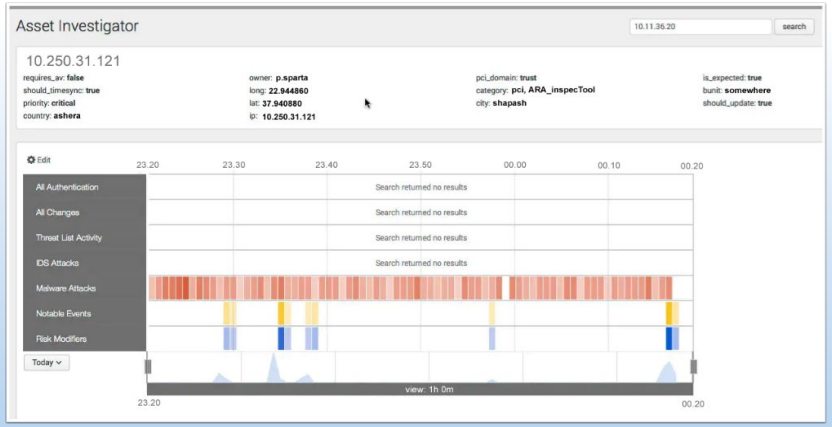
Remonter l’attaque : Investigation numérique

Artefacts collectés lors de l’analyse de l’incident

Cyber-analyse

L’équipe de cybersécurité identifie le point d’infection initial sur le réseau de l’ARA :

- À 23 h 20 le 3 mars, l’ordinateur portable (10.250.31.121) lance une analyse de réseau non autorisée
- L’ordinateur portable attribué à l’inspectrice de l’ARA commence à exécuter le logiciel malveillant Typhon
- La chronologie suggère que l’infection a eu lieu avant la connexion au réseau



Propagation de l’attaque à partir de 10.250.31.121

Une fois connecté au réseau de l’ARA, l’ordinateur portable a permis ceci :

- Typhon a immédiatement lancé une analyse pour trouver les systèmes vulnérables
- Le logiciel malveillant s’est propagé latéralement en employant les identifiants privilégiés de l’inspectrice
- Connexion aux commandes et contrôles établie vers le serveur externe de Typhon
- Début de l’exfiltration des données avant la phase de chiffrement

Le facteur humain : Enquête sur Penelope Sparta

Que savons-nous jusqu'à présent ?

Enquête préliminaire : Inspectrice Penelope Sparta

- Ordinateur portable enregistré au nom de l'inspectrice Penelope Sparta
- Est récemment revenue d'un audit de centrale nucléaire de routine (28 février au 2 mars)
- Les journaux de sécurité informatique de la centrale nucléaire confirment que l'ordinateur portable a été analysé – aucun problème détecté
- Penelope Sparta nie toute connaissance de Typhon ou d'activité suspecte

Pistes pour susciter la discussion :

- **Analyse de référence des comportements**
  Quels modèles d'IA pourraient établir des schémas de comportement normaux pour les inspecteurs comme Penelope et détecter les écarts qui pourraient indiquer une compromission ou une coercition ?
- **Corrélation multi-source**
  Comment l'IA pourrait intégrer les données provenant des badges d'accès, des journaux de réseau, des modèles d'e-mails et des systèmes financiers, afin de créer un profil de risque complet permettant la détection des menaces internes ?
- **Évaluation prédictive des risques**
  Quelles approches d'apprentissage machine pourraient attribuer des scores de risque dynamiques aux employés en fonction de leurs niveaux d'accès, des événements récents dans leur vie et de leur proximité avec des systèmes sensibles ?

Suspendue pendant l'enquête

Les cyberattaques ont l’air faciles, mais ce n’est pas le cas

Planification des attaques et arborescences d’attaques

Dilemme de l’attaquant

Pour qu’une attaque réussisse, plusieurs conditions doivent être alignées :

- Les attaquants ont besoin d’une planification et d’une reconnaissance détaillées
- Chaque étape de la chaîne d’attaque doit aboutir
- Une seule détection ou un seul échec peut compromettre l’ensemble de l’opération

La chaîne d’attaque de Typhon

1. Recruter un agresseur interne avec l’accès approprié (offre de 1 million $)
2. L’agresseur interne doit contourner les contrôles de sécurité sans être détecté
3. Le logiciel malveillant doit échapper à la protection des points de terminaison
4. Le mouvement latéral nécessite des identifiants valides
5. L’exfiltration des données nécessite des canaux non surveillés
6. Le chiffrement doit être terminé avant la détection

L’avantage du défenseur : Opportunités de l’IA

Planification des attaques et arborescences d’attaques

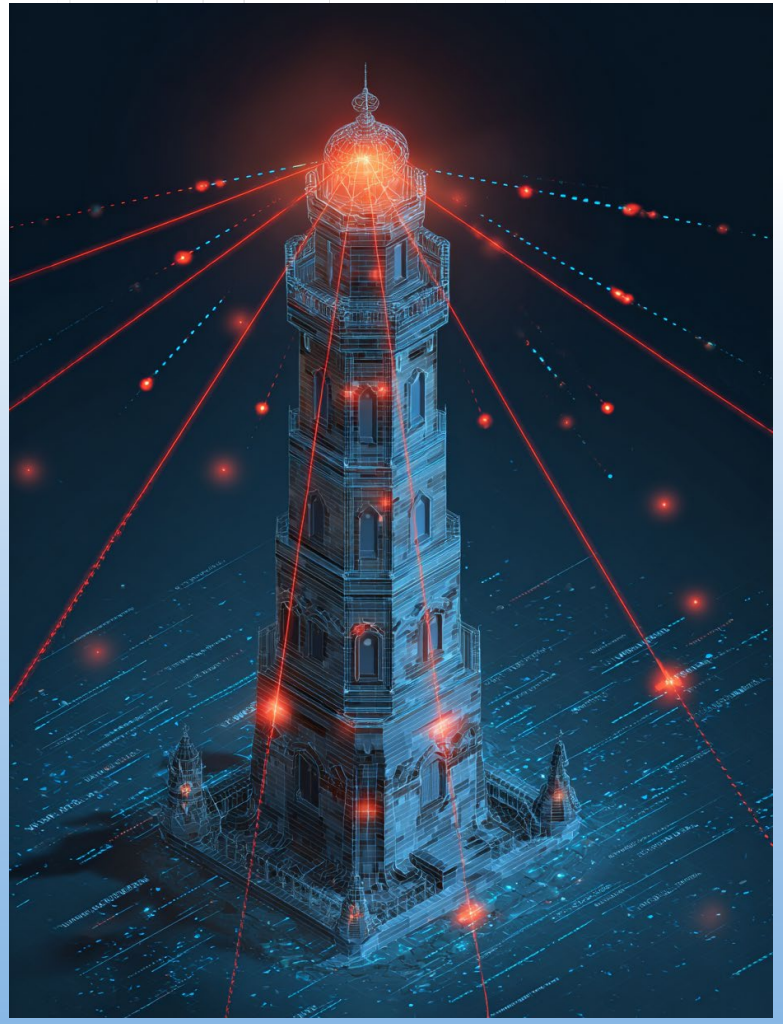
L’avantage du défenseur

- L’IA peut surveiller chaque lien dans la chaîne d’attaque simultanément
- À N’IMPORTE QUEL moment, les anomalies comportementales peuvent déclencher des alertes
- L’apprentissage machine améliore la détection à chaque tentative
- Se concentre sur ce que les attaquants DOIVENT faire, pas sur ce qu’ils pourraient faire

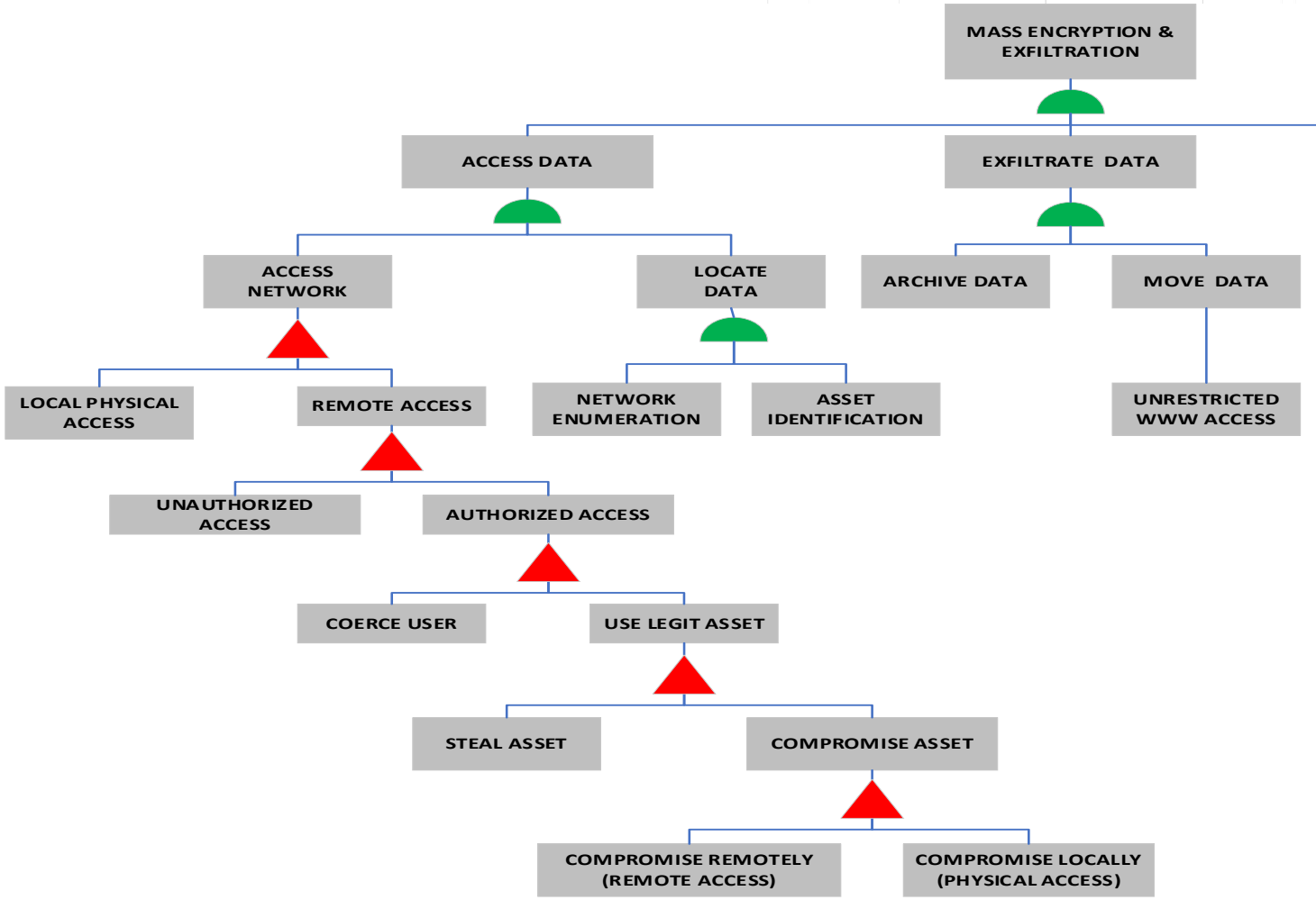
Renseignement clé : l’ordinateur portable de Penelope a dû effectuer des actions spécifiques et détectables. Quelles capacités de l’IA auraient pu identifier ces comportements d’attaque nécessaires ?

Pour susciter la discussion :

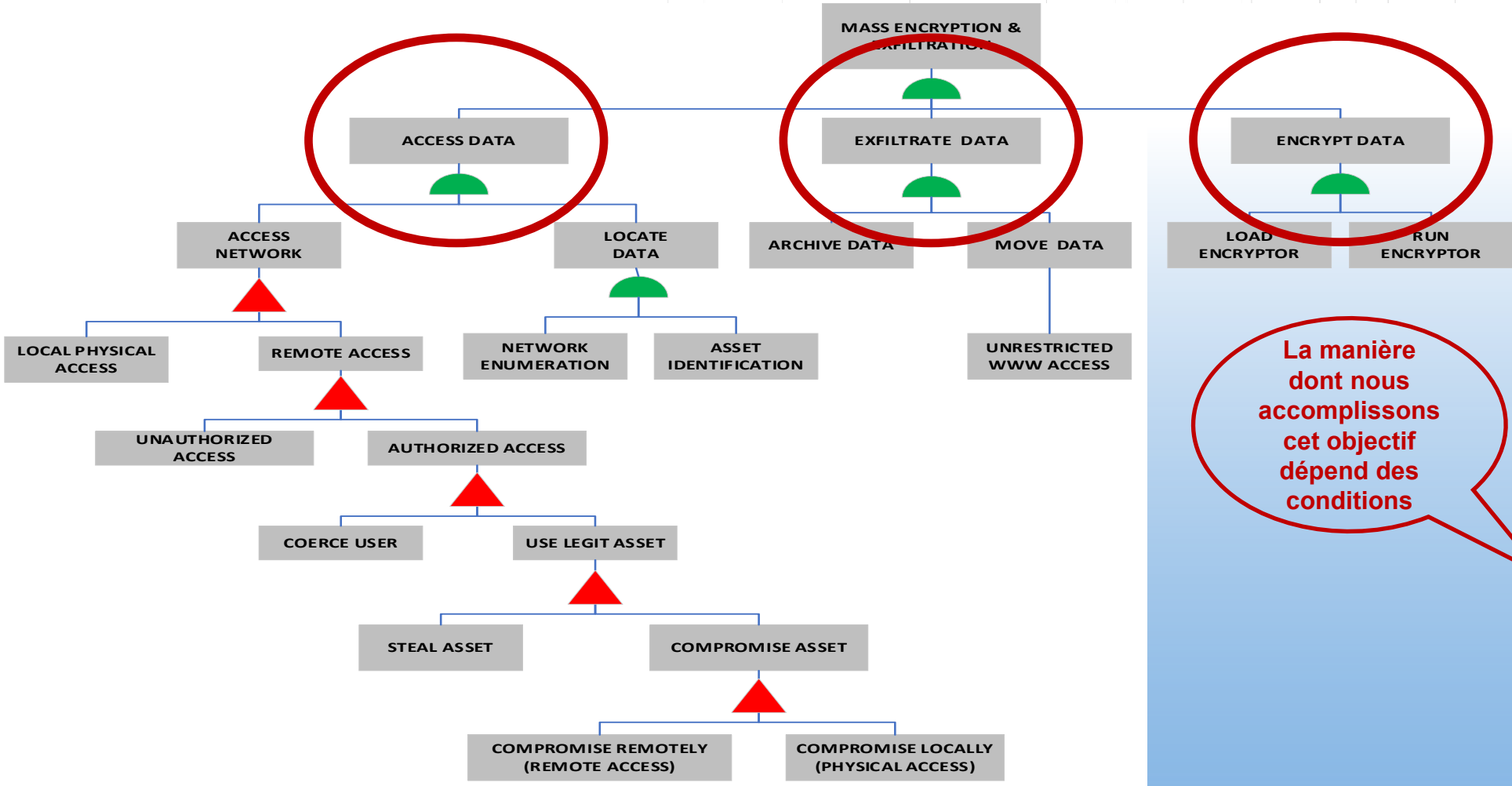
- **Analyse de la chaîne d’attaque**
  Comment les modèles d’IA pourraient-ils reconnaître le schéma séquentiel de reconnaissance → élévation des privilèges → mouvement latéral → mise en place des données que Typhon doit suivre ?
- **Détection de mouvement impossible**
  Quels algorithmes d’apprentissage machine pourraient signaler des impossibilités suspectes, comme l’utilisation simultanée des identifiants de Penelope sur le site de la centrale nucléaire et au siège de l’ARA ?



Plan d'attaque principal de Typhon

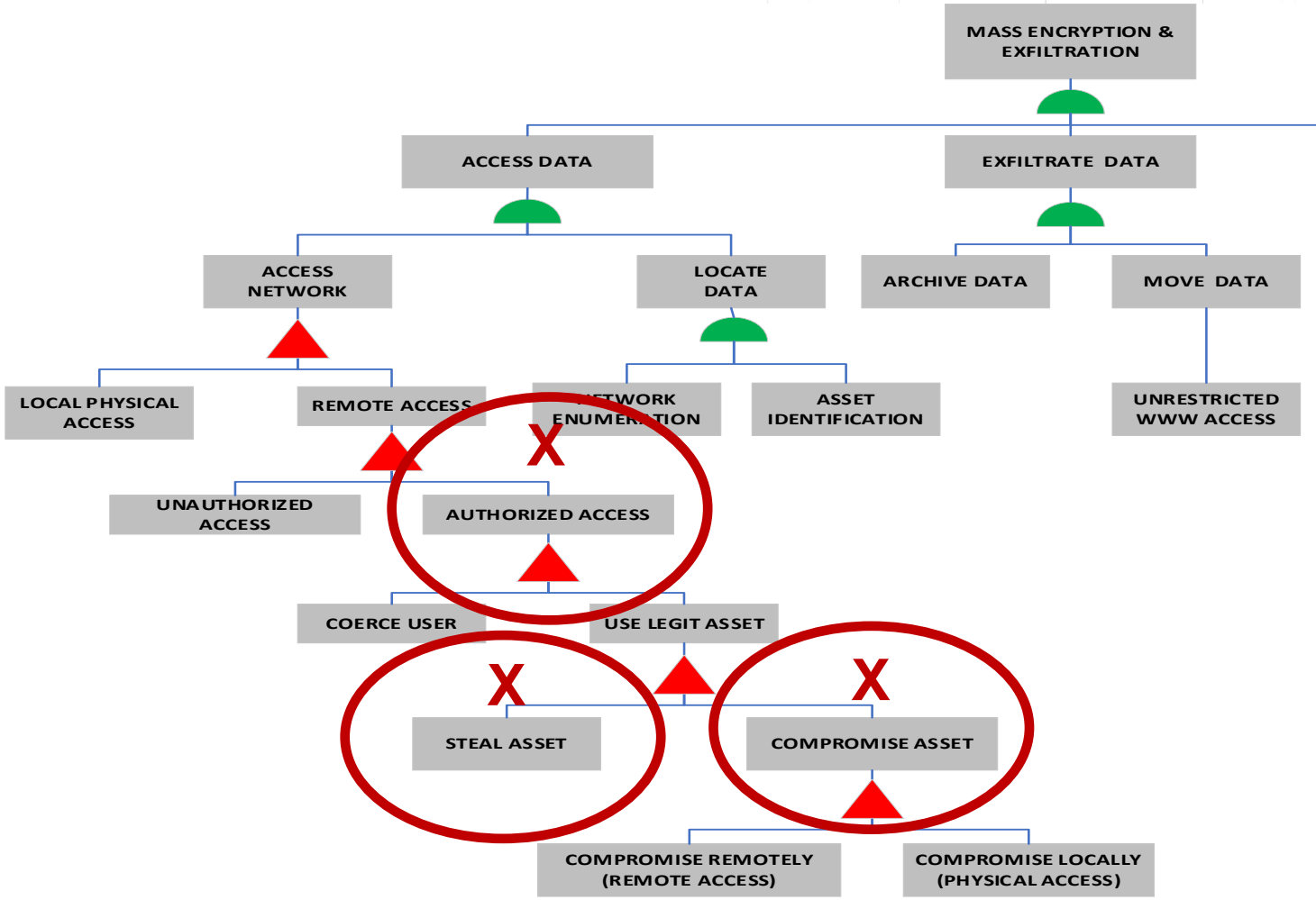


Conditions de la réussite de Typhon



La manière dont nous accomplissons cet objectif dépend des conditions

Faible probabilité pour que toutes les conditions soient remplies



Progression d'attaque contrariée sans agresseur interne

Les répercussions : L'impact de Typhon

Mesurer les conséquences de la violation

100 Go de données publiées

- Typhon publie 100 Go de données volées de l'ARA sur le Dark Web
- Les données divulguées incluent :
  - *Rapports de violation de sécurité des installations nucléaires*
  - *Documentation de conception du réacteur*
  - *Rapports d'inspection confidentiels*
  - *Informations personnelles concernant tous les inspecteurs et titulaires d'autorisation*

Analyse des impacts

- Indignation publique face à la violation de la sécurité nucléaire
- Les titulaires d'autorisation perdent confiance dans la capacité de l'ARA à protéger les données sensibles
- Des groupes illicites ciblent désormais des installations spécifiques à l'aide des données divulguées
- Les opérations de l'ARA sont paralysées – toutes les inspections sont suspendues
- Grave atteinte à la crédibilité de la réglementation internationale

Révélations de la chronologie forensique

1er mars à 14h32

Insertion de clé USB détectée lors de l'inspection de la centrale nucléaire

1er mars à 14h33

Le scan automatisé des médias de la centrale nucléaire s'est révélé « propre »

3 mars à 23h20

Typhon inactif jusqu'à la connexion réseau

Attaque déclenchée par les identifiants de connexion valides de Penelope

L’enquête se poursuit : Était-ce vraiment Penelope ?

Découvrir la vérité derrière l’attaque

Récit de Penelope

- Insiste sur le fait que tous les protocoles de sécurité ont été suivis lors de l’inspection
- La clé USB a été analysée selon la politique des médias portables de la centrale nucléaire
- Pas d’accès à des fichiers inhabituels ou de comportement suspect remarqué
- Soumission réussie au détecteur de mensonges ; les archives bancaires ne font pas état de transferts d’argent important

Conclusions clés de l’enquête

- Variante de Typhon spécialement conçue pour échapper au scanner de la centrale nucléaire
- Le logiciel malveillant est resté dormant jusqu’à la détection dans le réseau de l’ARA
- Identifiants valides d’inspecteur requis pour l’activation
- La sophistication de l’attaque suggère qu’un agresseur interne connaît les systèmes de la centrale nucléaire et de l’ARA

Stratégies d’intervention fondée sur l’IA : Analyse post-violation

Enseignements tirés de l’incident

Pistes pour susciter la discussion :

- **Biométrie comportementale**
  Comment l’IA pourrait analyser la dynamique des frappes de clavier, les mouvements de souris et les modèles d’interaction avec le système pour vérifier si c’est bien Penelope qui utilise ses identifiants ?
- **Détection d’anomalie dans les menaces dormantes**
  Quelles techniques d’apprentissage machine pourraient identifier les logiciels malveillants dormants en détectant les changements subtils du système, même lorsque le logiciel malveillant n’est pas activement en cours d’exécution ?
- **Renseignements entre organisations sur les menaces**
  Comment l’apprentissage fédéré pourrait permettre aux centrales nucléaires et à l’ARA de partager des modèles de détection des menaces par IA sans exposer de données sensibles ?
- **Modélisation de scénarios d’impact**
  Comment les grands modèles linguistiques (LLM) pourraient analyser les données divulguées, afin de générer rapidement des scénarios d’attaque potentiels, ce qui aiderait l’ARA à prévoir et à empêcher les attaques secondaires sur des installations spécifiques désormais exposées à la violation ?

Un autre agresseur interne : Technicien informatique Philippe

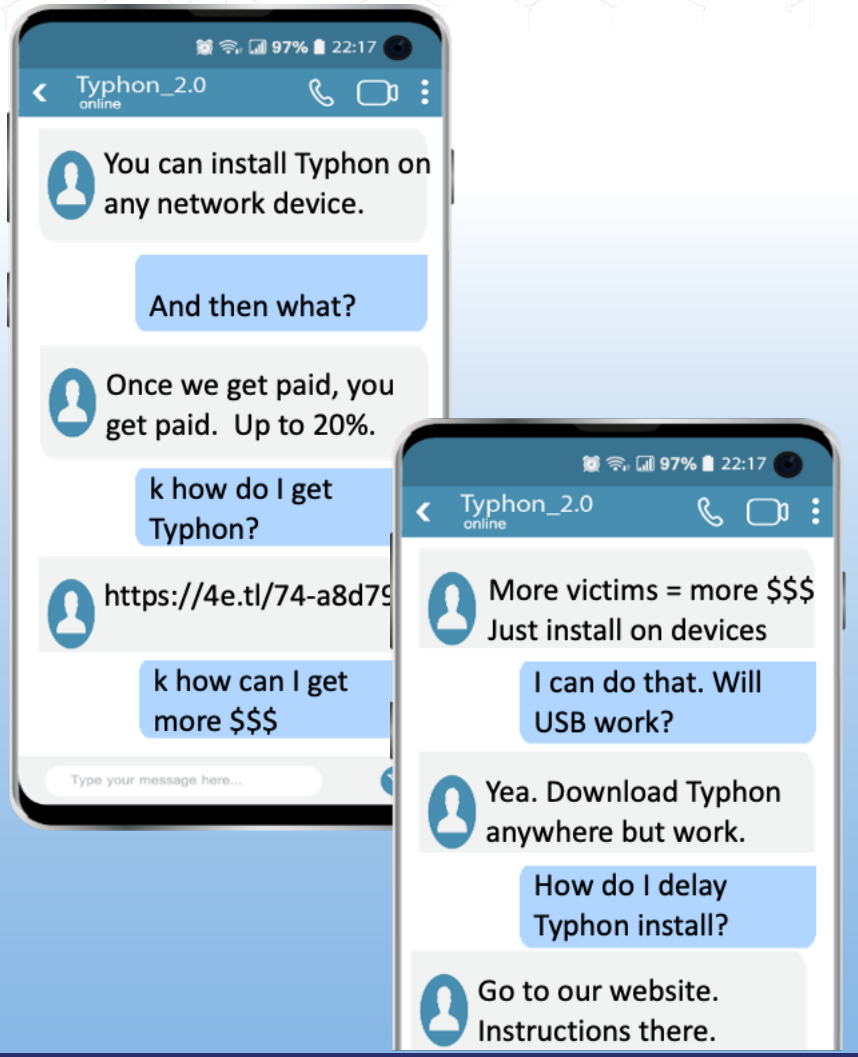
Découverte dans la centrale nucléaire

Attaque parallèle dans la centrale nucléaire

- Pendant que l'ARA s'est occupé de Typhon, la centrale nucléaire que Penelope avait auditée a détecté une attaque similaire
- L'équipe de cybersécurité de la centrale nucléaire a attrapé Typhon avant son installation sur le réseau
- Réussite en raison du partage rapide par l'ARA des renseignements sur les menaces
- Le service de sécurité de la centrale nucléaire a lancé des enquêtes sur tout le personnel ayant accès au système

À la découverte de Philippe

- Philippe : technicien informatique de la centrale nucléaire ayant un accès privilégié aux systèmes de sécurité
- Responsable de scanner l'équipement des inspecteurs en visite (y compris l'ordinateur portable de Penelope)
- A utilisé son accès au scanner pour installer Typhon sur les appareils tout en affichant des résultats d'analyse « propres »
- Le téléphone confisqué a révélé une communication directe avec le groupe Typhon
- A admis avoir installé le logiciel malveillant sur les ordinateurs portables de plusieurs inspecteurs en échange de 500 000 $

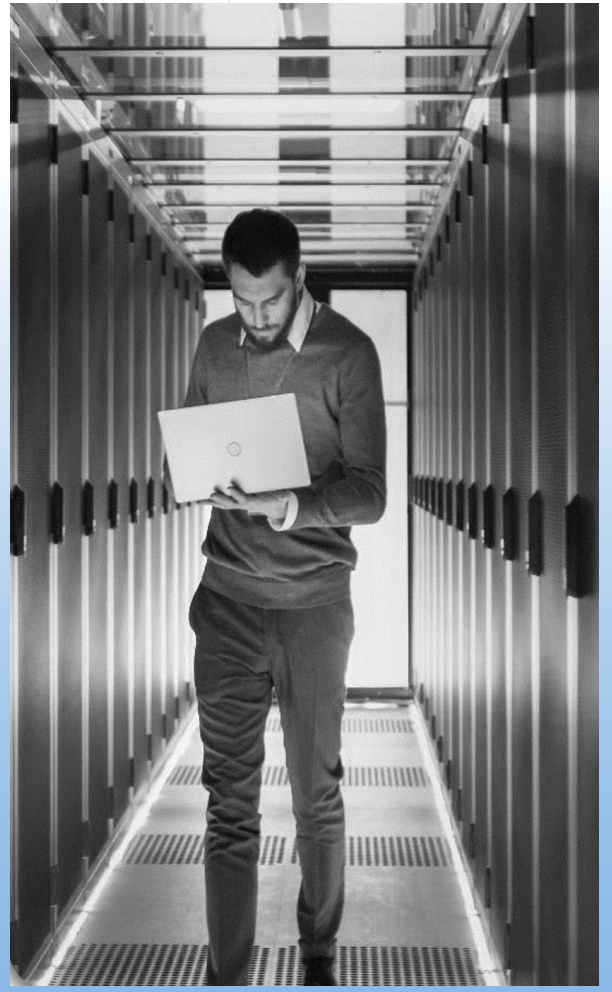


Indicateurs de menace interne : Le profil de Philippe

Reconnaître les signes d'alerte

Indicateurs comportementaux

- Le responsable a constaté des problèmes de performance persistants :
  - *Excellent sur le plan technique, mais a estimé que les tâches de routine sont « en dessous de son niveau »*
  - *Réactions hostiles aux critiques ou aux corrections d'erreurs*
  - *Récente réprimande écrite après avoir insulté des collègues*
  - *Rancœur croissante envers la direction*
- Profil d'agresseur interne classique : mécontent, compétent sur le plan technique, accès privilégié



La voie vers la compromission

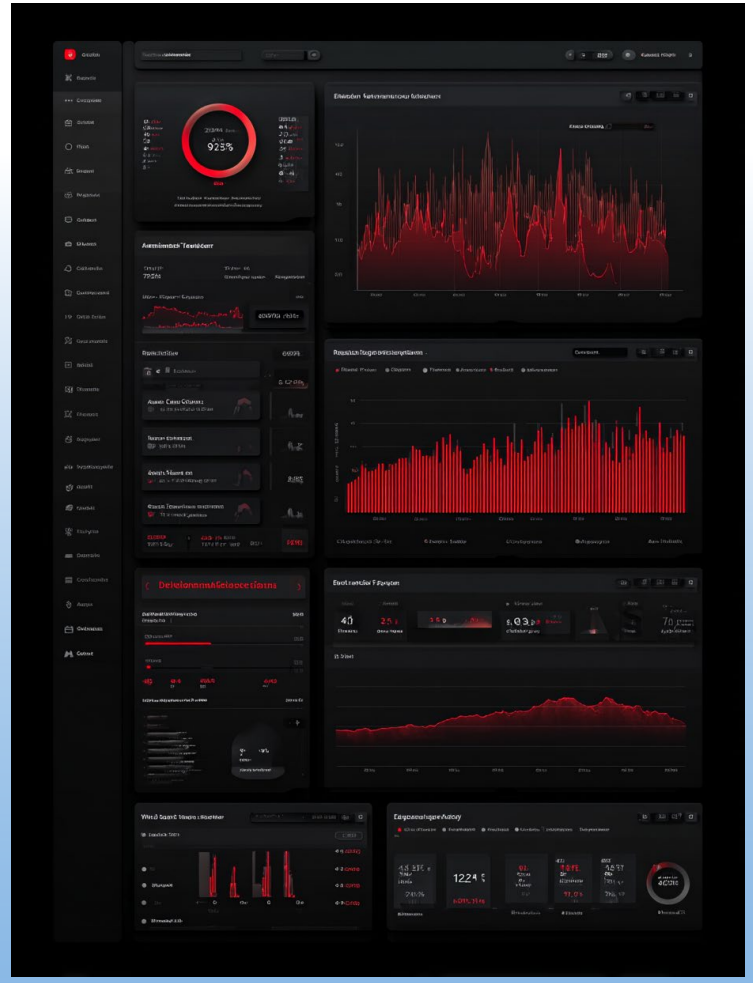
- Pression financière + rancœurs sur le lieu de travail = vulnérabilité
- Compétences techniques + accès privilégiés = capacité
- Recrutement externe + perception d'injustice = motivation
- Combinaison parfaite pour l'activation d'un agresseur interne

Détection d’agresseur interne grâce à l’IA : Analyse comportementale

Systèmes d’avertissement précoce

Pour susciter la discussion

- **Analyse des sentiments et des schémas de communication**
  Comment le traitement du langage naturel pourrait analyser les e-mails, les tickets et les journaux de chat, afin de détecter une croissance de la négativité, de l’hostilité ou du désengagement avant que la situation ne devienne un risque pour la sécurité ?
- **Cotation du risque holistique**
  Quel modèle d’IA pourrait combiner les données RH (réprimandes, évaluations), les journaux d’accès (présence en dehors des horaires de travail, systèmes inhabituels) et les indicateurs comportementaux pour créer des scores dynamiques sur les risques d’agresseurs internes ?
- **Analyse de comparaison des pairs**
  Comment l’apprentissage non supervisé pourrait identifier les employés dont les schémas comportementaux s’écartent considérablement de ceux de leurs pairs sur plusieurs plans et ce, de manière simultanée ?



Anatomie de l’attaque : Trois types d’agresseurs internes

Comprendre les profils des menaces internes

Créer des profils de menaces précis est le fondement permettant d’éviter certains comportements et de les détecter. Chaque type nécessite une stratégie de détection par l’IA différente.

Philippe : l’agresseur interne malveillant

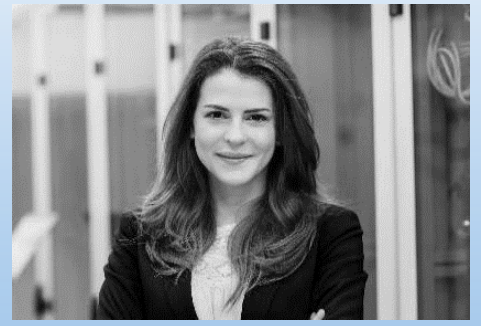
- *Technicien informatique disposant d’un accès privilégié au scanner. Motivé par l’appât du gain et des rancœurs sur le lieu de travail. A délibérément installé Typhon en échange d’un paiement de 500 000 $.*
- ***L’IA va se concentrer sur** : changements de comportement, analyse des sentiments, schémas d’accès anormaux.*

Penelope : l’agresseur interne involontaire

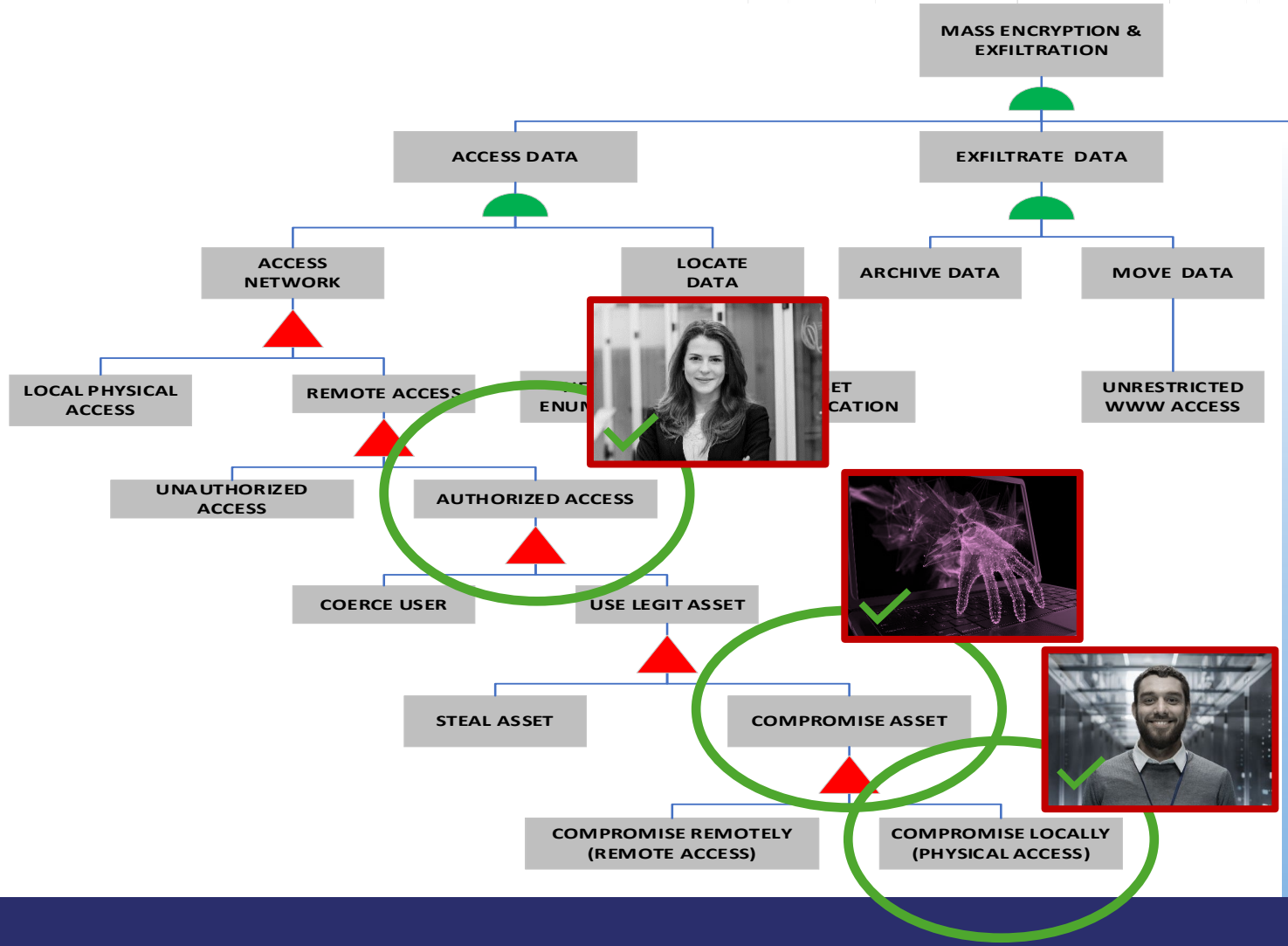
- *Inspectrice autorisée qui a respecté tous les protocoles. Ordinateur portable compromis à son insu. Est devenue un vecteur d’attaque malgré s’être conformée aux règles.*
- ***L’IA va se concentrer sur** : intégrité des appareils, comportement inhabituel du système, utilisation abusive des identifiants.*

Systèmes techniques : l’agresseur interne automatisé

- *Outils légitimes utilisés comme armes (scanners, serveurs de mise à jour). Processus de confiance exploités pour la diffusion de logiciels malveillants. Aucune intention humaine, mais permet l’attaque.*
- ***L’IA va se concentrer sur** : anomalies des processus, connexions inattendues, changements de comportement de fichiers.*



Conditions de l’attaque : Ce que Typhon doit accomplir



Philippe : compromettre localement (accès physique)
- Implanter un logiciel malveillant à l’aide de l’accès privilégié au scanner lors des inspections
- Modifier les résultats du scan pour qu’ils soient « propres » lors de l’installation de Typhon

Penelope : accès autorisé (entrée de réseau)
- Fournir des identifiants valides pour l’accès au réseau de l’ARA
- Permettre au logiciel malveillant dormant de s’activer et de se propager en utilisant ses privilèges

Systèmes techniques : compromettre les actifs (attaquer l’infrastructure)
- Permettre à Typhon d’échapper à la détection par les outils de sécurité
- Fournir des chemins non surveillés permettant l’exfiltration et le chiffrement des données

Conclusion : si N’IMPORTE LEQUEL de ces agresseurs internes est supprimé de cette chaîne = l’attaque échoue

Atténuer les agresseurs internes malveillants : Se prévenir de Philippe

Approche d'IA pluridisciplinaire

Détection comportementale

- Analyse des sentiments par l'IA à partir des communications, afin de signaler toute hostilité croissante
- Modèles d'apprentissage machine pour identifier les employés présentant des signes de mécontentement
- Alertes automatisées lorsque les scores de risques comportementaux dépassent les seuils

Intégration administrative

- Les mesures prises par les RH (réprimandes) déclenchent automatiquement une cybersurveillance renforcée
- L'IA met en corrélation les problèmes liés au personnel avec les journaux d'accès et l'utilisation du système
- Le tableau de bord des risques combine les flux de données RH, de sécurité et informatiques

Contrôles techniques

- L'IA surveille l'utilisation des comptes privilégiés à la recherche de schémas anormaux
- Les algorithmes d'apprentissage machine détectent les modifications inhabituelles de fichiers ou la falsification des résultats d'analyse
- Rapports automatisés sur les activités techniques à haut risque des employés

Défense en profondeur

Limiter la portée de l'accès privilégié en employant les modèles minimaux recommandés par l'IA

Déployer la biométrie comportementale pour vérifier l'identité lors des opérations critiques

Mise en corrélation de l'accès physique avec les activités numériques pour détecter les violations de politiques

Protéger les agresseurs internes involontaires : Les protections de Penelope

Approche d'IA pluridisciplinaire

Sécurité physique et intégrité des appareils :

- Suivi des appareils par l'IA pour alerter lorsque les ordinateurs portables ne sont plus sous le contrôle des inspecteurs
- Vérification par blockchain de l'état des appareils avant/après les scans externes
- Alertes automatisées en cas de tentatives d'accès non autorisées au matériel

Surveillance des comportements

- Base de référence d'apprentissage machine des modèles d'utilisation normaux du système d'inspection
- Détection par l'IA des commandes/processus que les inspecteurs n'exécutent généralement jamais
- Alertes en temps réel lorsque les appareils présentent des modèles de comportement non humains

Protections techniques

- Liste autorisée des applications surveillées par IA avec détection des anomalies
- Analyse par apprentissage machine du trafic réseau pour détecter les flux de données inhabituels
- Verrouillage automatisé des ports USB/ports avec surpassement biométrique uniquement

Défense en profondeur

Segmentation du réseau avec surveillance par l'IA du trafic intersegmentaire

Accès basé sur les rôles avec détection des élévations de privilèges optimisée par apprentissage machine

Surveillance continue de la santé des appareils à l'aide des analyses comportementales

Sécuriser l’accès technique : Défenses au niveau du système

Approche d’IA pluridisciplinaire

Protection des infrastructures

- Détection des anomalies par l’IA pour toutes les communications entre systèmes
- Modèles d’apprentissage machine entraînés sur le comportement normal des appareils, afin de signaler les écarts
- Surveillance en temps réel de l’intégrité des sauvegardes avec détection des manipulations

Sécurité des processus

- Base de référence par l’IA des activités et connexions normales des ordinateurs portables des inspecteurs
- Détection par apprentissage machine des apparitions de processus inhabituelles ou des élévations de privilèges
- Analyse comportementale de tous les processus automatisés d’analyse et de mise à jour

Évolution du contrôle de l’accès

- Attribution dynamique sur la base du moindre privilège basée sur la notation des risques par l’IA
- Séparation des tâches optimisée par apprentissage machine avec vérification automatisée de la conformité
- Architecture de zéro confiance avec authentification continue

Défense en profondeur

Système d’IA isolé pour la vérification des sauvegardes et l’analyse des menaces

Journaux d’audit immuables avec détection des anomalies par apprentissage machine

Capacités de restauration automatisées déclenchées par la détection des menaces par l’IA

Orientation stratégique : Perturber la chaîne d'attaque

Cibler ce que les attaquants doivent faire

L'adversaire doit :

- Recruter ou compromettre un agresseur interne (humain ou technique)
- Échapper à la détection lors de l'accès initial et du déplacement latéral
- Maintenir la persistance tout en exfiltrant les données
- Exécuter le chiffrement sans déclencher d'intervention

Pistes pour susciter la discussion

- **Modèles de détection fondés sur les conditions**
  Comment former les modèles d'IA spécifiquement sur les étapes incontournables que les agresseurs doivent réaliser, plutôt que d'essayer de prédire toutes les variations d'attaques possibles ?
- **Orchestration d'intervention automatisée**
  Quels systèmes d'apprentissage machine pourraient effectuer une détection instantanée lorsque les conditions d'une attaque critique sont réunies et déclencher automatiquement des mesures d'isolement, de restauration ou de défense ?
- **Apprentissage continu à partir des quasi-incidents**
  Comment les systèmes d'IA peuvent-ils tirer des enseignements des attaques telles que Typhon pour affiner en permanence la détection des comportements « obligatoires » des adversaires sur l'ensemble de l'infrastructure de l'ARA ?

Principe fondamental

Ne pas pourchasser des possibilités infinies.
Cibler les actions nécessaires des adversaires.

L'IA doit surveiller les quelques actions que les adversaires ne peuvent pas éviter de faire et non pas la multitude de tentatives dans lesquelles ils pourraient se lancer.

Contact

Chris Spirito, Idaho National Laboratory

E-mail : christopher.spirito@inl.gov