Vincent GENOT

L'Intelligence Artificielle et cybersécurité

Chapitre 1 : La découverte

L'avènement de la cybersécurité proactive : Synergie entre "deceptive technology" Analyse de données massives et Intelligence Artificielle

Université BORDEAUX MONTAIGNE
Cyber criminalités

Promotion 2024-2025

Résumé : Face à la sophistication croissante des cybermenaces, les approches de sécurité traditionnelles peinent à offrir une protection suffisante. La « Deceptive technology » ou « technologies de tromperies » en Français, émerge comme une stratégie proactive, non seulement pour détourner les attaquants des actifs réels, mais aussi pour collecter des renseignements précieux sur leurs tactiques, techniques et procédures (TTPs).

Cet article explore comment ces technologies, illustrées par une expérimentation commune avec les équipes d'apprenants Cyber et IA (TEAM 238) de l'ESILV, Ecole Supérieur d'Ingénieur Léonard de Vinci de Paris, ont permis d'explorer le concept de « Cyber Dôme » dynamique avec comme objectif d'améliorer la connaissance de l'attaquant et masquer les systèmes critiques d'un SI de SANTE : Nom de Code PI2.

Nous verrons ensuite comment l'intelligence artificielle (IA) va nous aider dans l'analyse des volumes massifs de données générés, apportant pertinence et automatisation.

Et pour finir, comment nous pourrons initier une démarche d'IA prédictive des cybermenaces sur un Système d'Information (SI) pour atteindre un niveau de défense véritablement prédictif et robuste

Mots-clés : Technologie Déceptive, Honeypots, Cyber-Renseignement, Intelligence Artificielle, Analyse de Données Massives, Cybersécurité Proactive, T-Pot, MITRE Shield, Cyber Threat Intelligence (CTI).

1. Introduction

Le paysage des cybermenaces est en constante évolution, avec des acteurs malveillants employant des méthodes de plus en plus sophistiquées pour infiltrer les systèmes d'information. Les défenses périmétriques traditionnelles, bien qu'essentielles, ne suffisent plus à contrer ces attaques avancées. Face à ce constat, la transition d'une cybersécurité réactive vers une défense proactive est une piste de réflexion nécessaire.

C'est dans ce contexte que les « deceptives technologies » offrent un changement de paradigme, passant d'une posture purement réactive à une défense proactive et axée sur le renseignement en obligeant les attaquants à révéler leurs intentions et leurs méthodes.

Ces technologies visent à créer un environnement contrôlé et surveillé pour attirer, détecter, et analyser les activités des attaquants. Le projet PI2 de notre Team 238, "Combinant une solution de déceptive technologie avec l'analyse de données pour mieux faire face aux cybermenaces", illustre concrètement cette approche par le déploiement de la plateforme T-Pot, un système multi-honeypots et fournit un exemple concret de cette démarche. Cet article vise à explorer les mécanismes par lesquels la « deceptive technology» améliore la posture de sécurité et le rôle de l'IA dans l'exploitation des données collectées.

2. La Technologie Déceptive : Un Leurre Actif pour la Cyberdéfense.

Les « deceptives technologies », telles que définies par des cadres comme MITRE Shield (une base de connaissances des techniques de défense active, complémentaire à ATT&CK), englobent un large éventail de tactiques visant à tromper, retarder et démasquer les adversaires. Elles reposent sur la création et le déploiement de leurres (decoys) et de honeypots qui imitent des systèmes, des applications, des données ou des vulnérabilités réelles pour « attirer » ou laisser entrer les cybercriminels sur ces systèmes.

2.1. Renforcement de la Connaissance de l'Attaquant : Du Bruit à l'Information Actionnable

L'un des avantages majeurs des technologies déceptives est leur capacité à collecter des informations détaillées sur les attaquants et leurs méthodes. Contrairement aux alertes de sécurité traditionnelles qui peuvent manquer de contexte, les interactions avec les honeypots fournissent des données brutes et de qualité pour comprendre le « *modus operandi* » adverse.

- **Cartographie des TTPs et des Outils :** En observant comment un attaquant explore un honeypot, quels outils il utilise, quelles commandes il exécute (exemple de commandes capturées par des honeypots comme Cowrie pour SSH/Telnet, [3, Slides p.41]), les défenseurs peuvent comprendre ses TTPs. Ces honeypots capturent les sessions interactives, incluant les commandes tapées, les scripts téléchargés et les erreurs commises par l'attaquant. Ces informations permettent de comprendre non seulement les outils (par exemple, des scanners comme Nmap, des frameworks d'exploitation comme Metasploit) mais aussi le niveau de compétence de l'attaquant.
- **Identification Précise des Vulnérabilités Ciblées :** Des honeypots spécifiques comme Log4Pot (détection d'exploitation de Log4j) ou Elasticpot (protection des serveurs Elasticsearch vulnérables) [3, Poster] permettent d'identifier les vulnérabilités activement exploitées plutôt que sur des scores CVSS théoriques.

  Dionaea, également utilisé dans le projet PI2, piège les malwares exploitant diverses vulnérabilités [3, Poster].

- **Profilage Avancé des Attaquants et des Campagnes : :** La collecte d'adresses IP, l'analyse de la géolocalisation, et l'étude des signatures d'attaques (via Suricata, par exemple [3, Slides p.30-39]) permettent de dresser un profil des sources d'attaque. Le projet PI2 a ainsi pu identifier les ports les plus ciblés (80, 443, 445, 22, 37215) et les protocoles (SMB, MSSQL, MySQL) [3, Slides p.20, p.24].
- **Extraction d'Indicateurs de Compromission (IoCs) :** Les fichiers malveillants déposés sur les honeypots, les domaines contactés pour le C2 (Command & Control), ou les adresses IP utilisées pour des scans peuvent être extraits et utilisés comme IoCs pour rechercher des compromissions sur le réseau réel.
- **Intentions et Motivations :** Bien que plus difficile à cerner, l'analyse des actions post-compromission sur un honeypot peut donner des indices sur les objectifs de l'attaquant (exfiltration de données, pivotement, etc.). C'est une piste à explorer avec de futures modèles IA enrichies par contexte géo-politique et informations économiques des établissements protégés.

Toutes ces informations sont cruciales pour adapter les stratégies de défense, prioriser les correctifs, et anticiper les menaces futures. Le projet PI2 souligne que cette analyse permet d'identifier les ports les plus vulnérables et fréquemment ciblés, et proposer des solutions pour réduire le risque. « Mener des recherches sur les IP suspectes pour recueillir des informations sur les attaquants ou le groupe APT impliqué » transformerait ainsi les données brutes des honeypots en renseignement tactique.

2.2. Dissimulation des Systèmes d'Information Critiques et Rupture du Cycle d'Attaque

Au-delà de la collecte de renseignements, les « deceptives technologies » peuvent jouer un rôle de dissimulation et de diversion ; elles protègent activement en complexifiant la tâche de l'attaquant.

- **Création d'un "Brouillard de Guerre" Numérique :** En déployant un grand nombre de leurres crédibles et diversifiés (faux serveurs de fichiers, fausses bases de données, faux postes de travail), on augmente la complexité de la phase de reconnaissance pour l'attaquant. Chaque leurre est une impasse potentielle, gaspillant ses ressources et son temps. Plus le réseau de leurres est dense et réaliste, plus il est difficile pour l'attaquant de cartographier avec certitude le véritable SI.
- **Ralentissement et Usure de l'Attaquant (Attrition) :** Un attaquant engagé avec un honeypot est un attaquant qui ne progresse pas vers les actifs critiques. Les honeypots à haute interaction, en particulier, peuvent le maintenir occupé pendant des périodes prolongées, offrant une fenêtre cruciale pour la détection et la réponse. Cette "friction" ajoutée peut décourager les attaquants moins déterminés ou moins outillés.
- **Détection Précoce et Réduction des Faux Positifs :** Puisque les leurres n'ont aucune valeur de production légitime, toute interaction est, par définition, anormale et hautement suspecte. Cela contraste fortement avec les IDS/IPS traditionnels qui génèrent souvent un volume important de faux positifs menant à une "fatigue des alertes" pour les analystes SOC. Les honeypots, ou leurres comme nos T-Pot "notifient les équipes de sécurité lorsque les attaquants interagissent avec les honeypots".

    - ⇨ **Attention toutefois,** mal configurés il peuvent devenir trop bruyant, un réseau de ce type nécessite d'être « fine-tuné » et adapté à la cible de détection et protection.

- **Risques et Limitations :** Il faut toutefois être prudent, **leur implémentation en production n'est pas sans risque**, des attaquants expérimentés peuvent parfois identifier les honeypots, en particulier les plus simples ou mal configurés. Un honeypot compromis pourrait, dans le pire des cas, être utilisé comme point de pivot pour attaquer d'autres systèmes si l'isolation n'est pas rigoureuse.

Il faut être conscient que l'utilisation de ces technologies nécessitent un certain niveau d'expertise pour être implémenter dans un Système d'information professionnel. Ces outils mal configurés peuvent servir de rebond pour atteindre la cible à protéger, ce qui aurait l'effet inverse de celui recherché. Actuellement ces technologies étant en « libre accès », elles participent au bruit ambient, mais certains acteurs plus professionnels et experts dans ces domaines savent également cartographier ce type de leurres, la « deceptive technology » comme toute technologie a ses experts.

Le déploiement par des professionnels de l'informatique, n'est pas gage de sécurité, c'est un projet ou une brique technologique à part entière dans une vision ou une solution stratégique de défense plus complète. La maintenance et la crédibilité des leurres à grande échelle représentent un défi opérationnel.

3. L'Intelligence Artificielle face à la Collecte Massive de Données Déceptives : De la Donnée Brute au Renseignement Actionnable

Le déploiement de plateformes déceptives, comme T-Pot, qui agrège les logs de multiples honeypots, ou la mise en place d'un réseau d'honeypots à grande échelle, génère des volumes considérables de données (logs d'événements, captures de trafic, échantillons de malwares). L'analyse manuelle de ces données est fastidieuse et peu scalable. C'est ici que l'Intelligence Artificielle (IA) et l'apprentissage automatique (Machine Learning, ML) interviennent.

3.1. Pertinence, Contextualisation et Automatisation grâce à l'IA et le Machine Learning (ML)

L'IA peut transformer radicalement la manière dont les données des « deceptives technology » sont exploitées :

- **Classification Intelligente et Corrélation d'Événements Complexes :** L'IA peut trier, classifier et corréler des milliers d'événements issus des honeypots pour identifier des campagnes d'attaques coordonnées ou des comportements anormaux qui échapperaient à une analyse humaine. Des algorithmes de clustering (par exemple, K-Means, DBSCAN) peuvent regrouper des alertes apparemment isolées provenant de différents honeypots pour révéler des campagnes d'attaques distribuées. Des techniques de traitement du langage naturel (NLP) peuvent être appliquées aux logs de commandes (ex: Cowrie) pour identifier des intentions ou des outils spécifiques.
- **Détection d'Anomalies Comportementales et de Signaux Faibles :** Les modèles de Machine Learning peuvent être entraînés pour reconnaître les comportements "normaux" (même s'ils sont malveillants), pour établir une "base de référence" des comportements

d'attaque courants. Toute déviation significative par rapport à cette base peut signaler des TTPs nouveaux ou émergents, ou même des erreurs commises par l'attaquant.

- **Analyse Comportementale :** Au-delà des IP et des ports, l'IA peut analyser des caractéristiques comportementales plus subtiles : séquences d'actions, temps de réponse entre les commandes, types de charges utiles utilisées.

  Et bien que l'attribution formelle reste complexe, ces profils peuvent aider à distinguer des groupes d'attaquants ou des types d'acteurs (par exemple, bots automatisés vs. opérateurs humains). Le projet PI2 a envisagé l'Adversarial Robustness Toolbox (ART) d'IBM [3, Slides p.42], qui, bien que conçu pour tester la robustesse des modèles ML, pourrait inspirer des approches pour modéliser et détecter des comportements adverses.

- **Vers l'Automatisation de la Réponse et de la Reconfiguration automatique de la Deceptive Technology:** A terme, L'IA pourrait déclencher des actions de réponse automatisées (SOAR - Security Orchestration, Automation and Response) basées sur la criticité et la confiance des alertes issues des honeypots. Par exemple, bloquer dynamiquement des IPs sur les pares-feux, isoler des segments de réseau leurrés pour une analyse forensique plus poussée, ou même reconfigurer dynamiquement le paysage déceptif pour l'adapter à une nouvelle menace détectée.

L'objectif final serait de créer une **boucle de rétroaction** où les informations tirées des leurres **par l'IA** servirait à **renforcer continuellement les défenses** en visant à "combiner sécurité déceptive et analyse basée sur l'IA pour détecter et **atténuer proactivement les cybermenaces**.

4. Vers une IA Prédictive : L'Impératif du Renseignement Augmenté et de la Fusion de Données

Bien que l'analyse des données issue des honeypots soit riche, elle reste, par nature, réactive aux menaces qui interagissent avec les leurres. Pour que l'IA devienne réellement prédictive et plus efficace dans la protection des systèmes d'information, elle doit être alimentée par une gamme plus large de renseignements.

Les données des honeypots fournissent une vue "interne" des menaces ayant atteint le périmètre (ou les leurres). Cependant, une compréhension holistique du paysage des menaces nécessite une intégration avec des flux de renseignement externes et c'est là qu'intervient les outils de CTI (Cyber Threat Intelligence) pour collecter et agréger les sources :

- **L' OSINT (Open Source Intelligence) et Renseignement Technique :** Informations publiques sur les vulnérabilités, les groupes d'attaquants, les campagnes en cours. Compléter avec une surveillance de forums publics et spécialisés, repositories GitHub des outils d'attaque publiés, listes de diffusion de sécurité pour identifier de nouvelles vulnérabilités (0-days ou n-days activement exploitées), et des discussions sur les TTPs.
- **Flux de CTI Commerciaux et Partagés :** Intégration de flux structurés (STIX/TAXII) provenant de fournisseurs spécialisés de type ISACs (Information Sharing and Analysis Centers) et d'agences gouvernementales. Ces flux fournissent des IoCs vérifiés, des analyses de campagnes d'acteurs étatiques ou criminels, et des rapports sur les tendances des menaces.

- **Renseignement sur le Dark Web :** Surveillance des « «marchés noirs » de la DATA pour la vente d'exploits, de données volées, d'accès initiaux (Initial Access Brokers), et discussions au sein de communautés cybercriminelles fermées. Cela peut donner des indications sur les cibles futures ou les vulnérabilités qui seront prochainement exploitées.
- **Analyse Géopolitique et Sectorielle :** Comprendre les motivations des acteurs menaçants (espionnage, gain financier, déstabilisation) en fonction du contexte géopolitique et du secteur d'activité de l'organisation peut aider à anticiper les types d'attaques les plus probables.

En fusionnant ces sources de données "internes" des honeypots avec ces flux "externes", l'IA peut :

- **Contextualiser les Alertes** : Une IP interagissant avec un honeypot est plus critique si elle est également listée dans un flux CTI comme appartenant à un groupe APT connu, Exemple, *si l'IA est entrainée en* ***Contextualisation Stratégique des Menaces,*** *une attaque sur un honeypot simulant un système SCADA prend une tout autre dimension si des flux CTI signalent une recrudescence des attaques contre les infrastructures critiques par un acteur étatique spécifique*.
- **Améliorer la Priorisation :** Les vulnérabilités activement exploitées "dans la nature" (selon la CTI) et également ciblées sur les honeypots, deviennent des priorités absolues pour le patching. **Avec l'IA,** nous pourrions avoir une **Priorisation Dynamique des Risques :** Une vulnérabilité présente sur le SI réel et mentionnée dans les flux CTI comme étant activement exploitée contre le secteur d'activité de l'entreprise, *même si elle n'a pas encore été vue sur les honeypots internes*, deviendrait une priorité.
- **Modélisation Prédictive des Attaques :** En apprenant des schémas d'attaques globaux et des TTPs émergents, l'IA pourrait construire des scénarios d'attaque probables contre le SI de l'organisation qu'elle défend en identifiant les points d'entrée les plus susceptibles et les chemins de propagation potentiels (jumeaux numériques).
- **Adapter Dynamiquement la « Déceptive technology » :** L'IA pourrait suggérer la création de nouveaux types de leurres, ou la modification de leurres existants, pour mimer des technologies ou des services qui, selon les renseignements externes, deviennent des cibles privilégiées. Par exemple, *si une nouvelle vulnérabilité critique sur un serveur Exchange est signalée, l'IA pourrait suggérer de déployer rapidement un honeypot Exchange crédible pour cartographier les attaquants*.

Ces travaux factuels, comme ceux que nous avons menés avec le projet PI2 de notre Team 238, ont démontré "l'efficacité de la combinaison de la sécurité déceptive et de l'analyse basée sur l'IA". Cependant, l'intégration de flux CTI diversifiés pour entraîner des modèles d'IA plus prédictifs et améliorer la protection des systèmes d'information réels nécessitera des travaux de recherche supplémentaires, L'étape suivante du projet PI2 pourrait explorer ces pistes (*projet identifié en cours et non public dans la continuité de mon travail sur le Cyber Dôme*).

5. Discussion et Travaux Futurs

La combinaison des technologies déceptives, de l'analyse de données massives, et de l'IA, ouvre des perspectives prometteuses pour une cybersécurité plus proactive. Notre Projet a fourni des

preuves concrètes de la valeur de la collecte et de l'analyse des données des honeypots pour comprendre les menaces.

Cependant, plusieurs axes de recherche méritent d'être approfondis :

- **Développement de Modèles d'IA Hybrides pour la Prédiction :** Explorer des architectures d'IA capables de fusionner efficacement des données structurées (IoCs, logs) et non structurées (rapports CTI, discussions sur le dark web) pour générer des prédictions fiables sur les menaces émergentes et les cibles probables.
- **Automatisation Intelligente de la Réponse :** Créer des systèmes où l'IA peut non seulement détecter mais aussi orchestrer des réponses défensives de manière autonome et sécurisée. L'objectif serait d'**Orchestrer une « Deceptive Technology Autonome et Adaptative »** pour concevoir des systèmes où l'IA pourrait dynamiquement reconfigurer le réseau de leurres en fonction de l'évolution du paysage des menaces et des interactions observées, créant un environnement déceptif auto-apprenant.
- **Standardisation et Interopérabilité des Données issues des Honeypots :** Développer des formats communs pour les logs et les alertes issus des honeypots afin de faciliter leur intégration dans les plateformes d'analyse et les flux CTI.
- **Mesure de l'Efficacité :** Établir des métriques claires pour évaluer l'efficacité des « deceptives technologies » déployées (par exemple, temps de détection, réduction des faux positifs, coût évité des incidents) et l'apport spécifique de l'IA.
- **Gestion de la "Contre-Déception" :** Étudier comment les attaquants tentent d'identifier et de contourner les technologies déceptives, et comment l'IA peut aider à rendre les leurres plus crédibles et plus difficiles à détecter.
- **Considérations Éthiques et Légales Approfondies :** Poursuivre la réflexion sur la collecte de données sur les activités des attaquants, notamment en ce qui concerne leur vie privée (ou pas) et les risques de dérapage (par exemple, "hacking back" même involontaire).

6. Conclusion

Les « deceptives technologies » représentent une avancée significative, transformant les réseaux en environnements hostiles pour les attaquants et en sources de renseignement pour les défenseurs. Ces technologies mises en réseaux, l'intelligence artificielle devient alors indispensable pour exploiter la masse de données générées, offrant pertinence et automatisation. Toutefois, pour transcender la simple détection et évoluer vers une défense véritablement prédictive, l'IA doit être nourrie par un spectre plus large de renseignements sur les menaces. Les travaux initiés par nos projets sont fondamentaux, mais la pleine réalisation du potentiel de cette synergie nécessitera des efforts de recherche continus et une collaboration étroite entre le monde académique et l'industrie.

Références

World Economic Forum. (2024). *Global Cybersecurity Outlook 2024*.
Spitzner, L. (2003). *Honeypots: Tracking Hackers*. Addison-Wesley.
PI2 Project Team 238. (2025). *Combining a deceptive security solution with data analysis to better face cyber threats* [Project Poster, Report, and Presentation Slides]. ESILV & Propulsar Cyber Academia. (Note : date indicative, basée sur le rapport)

Almeshekah, M., & Spafford, E. (2014). A Taxonomy of Deception for Active Cyber Defense. *Proceedings of the 2014 Workshop on New Security Paradigms Workshop*.
Fraunholz, D., Krohmer, D., & Pohl, F. (2018). The Art of Deception: A Survey and Taxonomy of Cyber Deception Techniques. *Journal of Information Security and Applications*.
Husain, M. S., Sornalakshmi, M., & Balamuralidhar, P. (2021). AI-Powered Cyber Threat Intelligence: A Survey. *Computers & Security*.
Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep Learning Approach for Intelligent Intrusion Detection System. *IEEE Access*.
Tounsi, W., & Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & Security*.
Verizon. (Annuel). *Data Breach Investigations Report (DBIR)*. Disponible sur : https://www.verizon.com/business/resources/reports/dbir/
MITRE Corporation. *MITRE Shield - Active Defense Knowledge Base*. Disponible sur : https://shield.mitre.org/
SANS Institute. (Divers articles et webcasts sur les honeypots et le threat intelligence). Par exemple, le blog SANS Internet Storm Center : https://isc.sans.edu/
CSO Online. (2023). *Alert fatigue is overwhelming cybersecurity professionals*. Disponible sur : https://www.csoonline.com/article/574977/alert-fatigue-is-overwhelming-cybersecurity-professionals.html (Note: Ceci est un exemple, la référence exacte peut varier)
The Honeynet Project. (Divers outils et recherches sur les honeypots). Disponible sur : https://www.honeynet.org/
ENISA (European Union Agency for Cybersecurity). (2022). *ENISA Threat Landscape 2022*. Disponible sur : https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022
VirusTotal. *Analyse de malwares et d'URLs*. Disponible sur : https://www.virustotal.com/