Prévenir et détecter

Par **Jacques DE LA RIVIÈRE**
Gatewatcher

L'hétérogénéité des systèmes d'information, la migration des données dans le *Cloud* ou encore le nomadisme sont autant de paramètres qui rendent difficile la définition du périmètre de protection des entreprises. À cela s'ajoute la professionnalisation de la cybercriminalité qui rend les menaces plus nombreuses et polymorphes.

Dans cet environnement sinueux, et alors que les analystes des Centres d'Opérations de Sécurité (SOC) et experts de la sécurité du *Cloud* sont des denrées rares, il est logique de vouloir utiliser la technologie pour automatiser des tâches de détection, d'évaluation et de réponse.

Plusieurs solutions se dessinent aujourd'hui, à la fois innovantes et complémentaires, mais chacune avec des limites : le SOAR (*Security Orchestration, Automation and Response*), la CTI (*Cyber Threat Intelligence*) et enfin l'intelligence artificielle et plus spécifiquement le *machine learning*.

Les enjeux de l'automatisation et de l'orchestration dans la cybersécurité

À l'heure où les volumes d'alertes et d'informations sont parfois trop importants à traiter pour l'humain, l'orchestration et l'automatisation de la gestion des incidents sont devenues incontournables.

L'acronyme SOAR rassemble toutes les technologies qui permettent de collecter les données et alertes de sécurité à partir de multiples sources et, surtout, aident à industrialiser l'analyse et le triage des incidents. L'idée est ici de combiner les apports de l'humain et de l'informatique pour définir, hiérarchiser et intégrer des activités de réponse à incident dans un processus standardisé.

Si les entreprises s'intéressent à ces outils, c'est notamment pour essayer d'améliorer la productivité de leurs équipes en charge de la sécurité, alors que le budget manque pour en gonfler les effectifs, ou que les tentatives de recrutement se heurtent à la pénurie de compétences. Et si le recours à des prestataires de services extérieurs peut être une solution, faire appel à des outils d'orchestration et d'automatisation peut en constituer une autre.

Outre les gains de productivité, les outils de SOAR peuvent également aider à réduire les délais de réponse aux incidents de confinement et de remédiation, mais aussi à libérer les analystes de certaines tâches routinières, parfois redondantes, et souvent chronophages. Et celles-ci ne manquent pas de peser sur la motivation des équipes des SOC. Cela concerne notamment tout le travail de triage initial des alertes, parmi lesquelles, souvent, de nombreux faux positifs. Les outils de SOAR enrichissent automatiquement les alertes, ajoutent des informations de contexte clé, pour permettre un triage automatisé ou, *a minima,* un triage manuel plus rapide et plus simple.

L'optimisation du triage peut dès lors permettre d'améliorer les capacités de détection en limitant le risque de voir des alertes significatives passer inaperçues. Et cela vaut aussi pour tout le travail de documentation des processus et d'audit, le suivi des performances du SOC, ou encore pour le temps de formation initiale des analystes.

L'automatisation et l'orchestration en sécurité sont rendues possibles par la multiplication des API nécessaires à l'intégration. Fut un temps où quelques outils proposaient des API, mais peu d'entre elles étaient aussi standardisées et simples que les API Rest qu'on utilise aujourd'hui. L'automatisation totale des processus de sécurité nécessitera encore du temps, si tant est qu'elle soit possible. En attendant, les outils de SOAR peuvent apporter beaucoup dans de nombreux domaines.

Cela commence notamment par la gestion des alertes remontées par le système de gestion des informations et des événements de sécurité (SIEM). Là, les outils de SOAR permettent d'automatiser des tâches d'enrichissement, comme la recherche d'informations supplémentaires sur des indicateurs de compromission, ou la soumission d'échantillons suspects à des systèmes d'analyse externes. Les tickets correspondant à des alertes simples, ne relevant pas d'une menace réelle, peuvent être fermés automatiquement.

Des outils d'EDR [(1)] peuvent également être mis à contribution de manière transparente pour la collecte de données supplémentaires sur les hôtes concernés et la recherche de corrélation avec le trafic réseau peut ainsi être lancée. Concrètement, il peut être question de traitement de courriels suspectés de relever d'une tentative de *phishing*, et d'enrichissement d'une alerte générée par le SIEM à l'aide d'une plateforme de gestion du renseignement sur les menaces. On peut aussi se pencher sur les outils de gestion de tickets et plus généralement de services IT (ITSM), mais aussi de travail collaboratif, comme Slack, ou encore d'administration de systèmes, comme SCCM et SSH.

Le principe de la *Security Automatisation,* et non de l'IA, est bel et bien de remplacer le travail de l'analyste par celui d'une machine. La machine collecte les informations, les agrège en les confrontant à une base de suspicions connues pour enclencher *in fine* les protocoles de remédiation et la communication aux utilisateurs, dès lors évidemment que la réponse à ce type d'alertes est d'ores et déjà identifiée.

Zoom sur la *Cyber Threat Intelligence*

La *Cyber Threat Intelligence* a pour but de collecter et d'organiser toutes les informations liées aux cybermenaces afin de dresser un portrait des attaquants et d'en dégager des tendances (méthodes et techniques d'exploitation, secteurs touchés…). La CTI permet de connaître, de mieux se défendre et d'anticiper pour détecter les prémices d'une attaque.

Le terme *Threat Intelligence* est apparu début 2011, lorsque les premières APT (*Advanced Persistent Threats*) ont été largement médiatisées. Certaines organisations et entités étatiques utilisent ce concept depuis plus longtemps.

Les informations collectées peuvent être de différentes natures : marqueurs, IOC (indicateurs de compromissions tels que des *hash,* des noms de domaines ou des adresses IP), historiques d'attaques, réutilisation d'architecture ou de platement ayant servi antérieurement, utilisation de services, techniques et méthodes spécifiques (signatures communes, registrant identique).

Les moyens de collecte existants sont nombreux. On retrouve notamment le renseignement *Open Source* (OSINT), la *Social Media Intelligence* (SMI ou SOCMINT), les flux de données commerciaux et communautaires, les informations provenant du *Deep* ou du *Dark Web*, ou encore le renseignement d'origine humaine et la capacité d'analyse et de corrélation (HUMINT).

(1) *Endpoint Detection and Response.*

Dans une volonté globale de développer des standards réutilisables par le plus grand nombre, les moyens d'échanges ont tendance à s'harmoniser. De nombreuses initiatives, notamment *Open Source*, ont vu le jour pour favoriser la communication des informations CTI : STIX (*Structured Threat Information Expression*), MISP (*Malware Information Sharing Platform*), TAXII (*Trusted Automated eXchange of Indicator Information*). Cette transmission d'information permet également de générer des règles de détection pour des outils de supervision tels que les IPS.

L'intérêt d'implémenter un programme de *Cyber Threat Intelligence* au sein de son entreprise ou organisation est donc réel. Mais il faut tout d'abord se poser les bonnes questions. En effet, toutes les données ne sont pas significatives, il faut donc cadrer le programme de renseignements sur les cybermenaces en amont du projet. Il faut dans un premier temps définir les enjeux et les conséquences d'une atteinte à la sécurité de l'entreprise : qu'est-ce qui doit être protégé ? Quelles informations un attaquant voudrait-il obtenir ou détruire ? Lorsque les réponses à ces questions ont été formalisées, il faut déterminer les objectifs liés à la mise en place d'un programme de CTI au sein de son organisation. Ils doivent être clairs et réalisables, mais doivent également être facilement mesurables grâce à des indicateurs et des critères fixés au préalable.

Une fois les besoins exprimés et les objectifs fixés, vient le temps de la collecte des données avec la mise en place de différents « capteurs », en exploitant par exemple des sources ouvertes accessibles sur Internet (OSINT).

Il faut ensuite passer sur la phase de traitement des données pour simplifier leur exploitation par les analystes. Toutes les informations collectées (IOC, *malwares*, contexte géopolitique, méthodes de groupes d'attaquants) doivent être contextualisées et enrichies pour se transformer en véritable renseignement. C'est à ce moment-là que nous entrons dans la phase d'analyse, qui repose encore aujourd'hui en grande partie sur l'expertise de l'analyste. À l'issue de son travail, il produira un rapport, qu'il conviendra de diffuser aux bonnes personnes, au bon moment et dans le bon format.

Les équipes qui peuvent être intéressées par la CTI sont multiples :

- la direction générale, les RSSI ou le DSI, doivent être alimentés par du renseignement stratégique sur les cybermenaces de leur secteur d'activité ;
- le *Risk Manager* et ses équipes doivent avoir une vision globale des véritables menaces cyber qui pèsent sur les différents métiers de leur organisation ;
- le *Security Operation Center* (SOC) a besoin de renseignement très opérationnel et contextualisé sur les dernières cybermenaces, sous forme d'IOC, de règles de détection / corrélation voire de stratégies de réponse / remédiation adaptées (bloquer telle IP, lancer une investigation sur tel périmètre, bloquer ce protocole vulnérable, etc.) ;
- le CERT/CSIRT (interne ou externe) aura tant besoin de renseignements stratégiques qu'opérationnels très contextualisés sur les groupes d'attaquants ciblant son organisation ou ses clients, afin d'accélérer ses investigations et ses réponses aux incidents de sécurité.

L'intelligence artificielle : le futur de la détection ?

Commençons cette dernière partie avec un peu de théorie, en définissant ce qui se cache derrière le terme « intelligence artificielle ». On peut la définir comme un ensemble de concepts, de théories et de techniques permettant de résoudre des problèmes à forte complexité logique ou algorithmique. On trouve plusieurs disciplines associées, comme la neurobiologie computationnelle, la logique mathématique ou l'informatique. Tout cela peut sembler complexe mais comme avec toutes les technologies, il faut se reconcentrer sur les besoins et exploiter l'IA à bon escient.

Zoomons sur les possibilités offertes par le *machine learning* pour la sécurité et la détection des menaces. Il y a eu de nombreux débats sur les différences entre *deep learning* et *machine learning*. Nous avons fait le choix du *machine learning*. Cela nous a semblé être une bonne solution pour la résolution de problèmes considérés comme insolubles par des algorithmes classiques.

Le *machine learning* nous aide à modéliser la normalité : c'est ce qu'on appelle une intelligence artificielle supervisée. L'algorithme va permettre de qualifier une situation donnée et d'en déterminer la distance par rapport à la normalité. Si l'écart est trop important, il faut qualifier cette « anormalité ». Cela peut donner lieu à une alerte (incident de sécurité ou fraude par exemple) ou à un faux-positif. L'avantage du *machine learning* réside dans le traitement des faux-positifs : on va apprendre à la machine que c'est un cas « normal » et que le modèle de normalité doit être enrichi pour éviter de faire deux fois l'erreur.

Les apports de l'intelligence artificielle et plus spécifiquement du *machine learning* sont multiples.

L'IA peut être utilisée pour mettre à jour les bases de données de sécurité. L'analyse des journaux provenant de plusieurs sources permet à l'intelligence artificielle de détecter de nouvelles menaces imminentes. Autrement dit, l'intelligence artificielle est capable de collecter des données exhaustives à partir de plusieurs journaux et enregistrements, et de faire les rapprochements qui permettent d'identifier de nouvelles menaces diffusées par les pirates.

Côté logiciels malveillants et logiciels espions, l'intelligence artificielle peut également identifier les tendances en analysant les données de plusieurs canaux. Grâce à la détection plus rapide des nouveaux systèmes malveillants, l'intelligence artificielle empêche les dommages de prendre une ampleur démesurée. On dispose ainsi de plus de temps pour rechercher les méthodes de prévention qui permettent de corriger les bogues ou les failles de sécurité susceptibles d'être exploités par le *malware* ou le virus.

Hormis la détection des transferts de *malwares* à grande échelle, l'intelligence artificielle peut également être utilisée pour analyser un système afin d'y détecter toute activité anormale. Le fait d'analyser un système en permanence permet de recueillir suffisamment de données permettant de conclure au caractère anormal de certaines activités et de créer un modèle de normalité.

Les utilisateurs peuvent être surveillés en permanence afin de détecter tout accès non autorisé. Si le système repère une activité anormale, l'IA peut exploiter certains paramètres pour déterminer en amont si la menace est réelle ou pas afin de la signaler.

Le *machine learning* peut être utilisé pour aider l'intelligence artificielle à distinguer ce qui doit être considéré comme une activité « normale » d'une activité « anormale ». Plus le *machine learning* se perfectionne, plus l'IA gagne en efficacité pour déceler de légères anomalies potentiellement révélatrices d'un problème. Comme évoqué plus haut, l'important est de pouvoir faire les rapprochements nécessaires. Certaines anomalies mineures paraîtront insignifiantes en soi, mais prises ensemble, elles permettent de se faire une idée plus complète des causes sous-jacentes. L'IA est capable d'effectuer une analyse permanente d'un système, d'analyser plusieurs activités différentes et de les comparer, et de déclencher des alertes. L'intelligence artificielle met l'accent sur l'identification des points faibles, des bogues et des failles de sécurité potentiels. L'apprentissage machine peut, par exemple, être utilisé pour détecter à quel moment des données non fiables sont envoyées par une application. Les vulnérabilités de type « Injection SQL » sont l'une des failles les plus couramment exploitées par les logiciels malveillants et les virus pour dérober des données et accéder aux systèmes. Autre faiblesse que l'intelligence artificielle peut aider à détecter : le débordement de mémoire tampon ou transfert d'un volume de données inhabituellement élevé par une application dans une mémoire tampon. L'intelligence artificielle peut aussi avoir son utilité pour limiter l'erreur humaine. En effet, l'une des principales causes de violation de

la protection des données reste l'erreur commise par certains collaborateurs : l'IA permet d'en prévenir les dommages.

L'intelligence artificielle peut plus globalement déterminer les éventuelles vulnérabilités du système en effectuant un suivi des menaces actuelles, et notamment des *malwares*. En évoluant, l'intelligence artificielle ne détecte pas seulement les défauts d'un système ou d'une mise à jour en particulier, mais elle empêche aussi automatiquement l'exploitation de ces failles. Qu'il s'agisse d'ajouter des pare-feux supplémentaires ou de corriger des erreurs de codage à l'origine de vulnérabilités, l'IA offre un excellent moyen de prévention des problèmes.

Si la réponse peut ressembler à la prévention, cette phase a lieu ultérieurement : au moment où les logiciels malveillants se sont déjà introduits dans le système. Comme évoqué plus haut, l'intelligence artificielle peut être utilisée pour détecter les comportements anormaux et établir des corrélations afin de déterminer le profil d'un *malware* ou d'un virus dans le système.

Cette étape consiste à mettre en place la réponse appropriée face au *malware* ou au virus. Il s'agit de maîtriser les dommages, d'éliminer le virus du système, de corriger les éventuelles failles de sécurité et de mettre en place des protections supplémentaires pour éviter que le virus n'infecte à nouveau le système.

Si l'IA présente de nombreux avantages pour la cybersécurité, il y a encore une certaine marge de progression. La détection des anomalies empêche certes les accès non autorisés à un compte ou détecte les logiciels malveillants dès les prémices d'une attaque, mais elle peut également produire des faux positifs. L'intelligence artificielle peut réellement s'améliorer pour mieux repérer les activités vraiment anormales (car une connexion à partir d'un nouvel emplacement peut tout simplement signifier que l'utilisateur est en déplacement).

Les sociétés de sécurité et les éditeurs de logiciels continueront néanmoins à s'appuyer sur l'apprentissage machine pour réduire les délais de détection, augmenter les taux de détection, empêcher la propagation des logiciels malveillants, protéger les systèmes et accroître la sécurité des clients. Et si l'IA a encore beaucoup de chemin à parcourir, son impact commence à être tangible dans le domaine de la cybersécurité.