Comment l'IA intercepte les e-mails de phishing même lorsque nous mordons à l'hameçon

Un employé de bureau reçoit à lui seul 121 e-mails par jour en moyenne et, comme la plupart d'entre nous peuvent en témoigner, ne dispose que d'un instant pour décider si chacun d'eux mérite d'être traité.
Étant donné ce véritable déferlement, il n'est guère surprenant que 90 % des logiciels malveillants émanent de la boîte de réception, dissimulés dans des e-mails de phishing (ou hameçonnage) où les expéditeurs se font passer pour des collègues de confiance.

Naturellement, les personnes qui utilisent Internet depuis longtemps ont appris à se méfier des messages en provenance de princes étrangers qui demandent de l'aide pour transporter leur or. Cependant, près des trois-quarts des cyberattaques ciblées impliquent aujourd'hui des e-mails de « spear-phishing » (ou hameçonnage ciblé), une forme personnalisée de phishing dans laquelle les assaillants emploient la reconnaissance en ligne ou l'écoute physique pour se livrer à des falsifications convaincantes.

Tant les humains que les outils classiques de sécurité de messagerie électronique se sont avérés impuissants à repérer des menaces aussi subtiles. Une importante étude a établi que pour 150 000 e-mails de phishing expédiés dans le cadre de l'expérience, près de la moitié des destinataires ont cliqué sur le lien de l'escroquerie contenu dans les messages dans l'heure qui a suivi leur réception.

La détection des campagnes de spear-phishing exige d'adopter une approche de plate-forme pour la cyberdéfense, par opposition aux solutions cloisonnées spécifiques aux e-mails. S'appuyant sur l'apprentissage automatique non supervisé, les plates-formes de cyber IA parviennent à comprendre comment chaque utilisateur travaille et collabore au sein de l'infrastructure numérique, du service de messagerie électronique au Cloud et au réseau local.

Cette contextualisation des connaissances est impérative lorsque l'on recherche le moindre signe d'une suspicion d'hameçonnage, dans la mesure où une activité malveillante pour un utilisateur pourrait bien être anodine dans d'autres cas.

Ces plates-formes d'IA sont capables de réagir de façon autonome pour limiter les dégâts, et ce, peu importe la provenance de l'infection et l'endroit où elle se produit. Cela est fondamental dans la mesure où les assaillants motivés trouveront toujours un moyen de pénétrer l'enveloppe protectrice d'une entreprise.

Apprendre du patient zéro

Imaginez une attaque sophistiquée, mais néanmoins banale, dirigée contre une société internationale. L'attaque débute, sans surprise, par une campagne de spear-phishing ciblant des

employés dans toute l'entreprise. Les e-mails utilisent une tactique de phishing appelée « domain spoofing » (ou usurpation de domaine), qui consiste à enregistrer un nom de domaine apparemment légitime qui ressemble à l'adresse d'un expéditeur connu.
Le plus souvent, l'assaillant cherchera à se faire passer pour un cadre supérieur et enverra une demande pressante, en espérant que l'employé obtempère avant de déceler la falsification de domaine.

Dans ce cas précis, les assaillants, qui ont espionné la PDG de la société via ses tweets, ont imité son style rédactionnel afin de tromper les destinataires et de les inciter à ouvrir la pièce jointe annexée aux e-mails. Comme le nom de domaine usurpé n'apparaît pas sur les listes noires d'IP dont se servent les contrôles natifs de messagerie électronique de la société, ceux-ci se retrouvent dans les boîtes de réception de plus de 200 employés, prêts à infecter l'entreprise avec une souche de logiciel rançonneur à action rapide en un simple clic. Pour ne rien arranger, cette multinationale a des bureaux sur quatre continents. Donc, lorsque le « patient zéro » (une commerciale de Londres) découvre l'e-mail la première, son équipe de sécurité basée aux États-Unis dort encore à l'autre bout de la planète.

Pendant ce temps, la plate-forme de cyber IA de l'entreprise a analysé les e-mails et mis en corrélation leurs attributs avec le comportement en ligne typique de chaque employé, en exploitant sa connaissance de l'ensemble de l'infrastructure numérique. Cette analyse a révélé que les e-mails étaient suspects, et bien que l'IA ne soit pas encore intervenue, elle a préparé sa capacité de réaction autonome à prendre des mesures immédiates.
Revenons à Londres : le patient zéro parcourt rapidement l'e-mail et télécharge par mégarde le logiciel rançonneur, qui commence à se déplacer latéralement, à identifier les partages de fichiers et à chiffrer les documents de la société à la vitesse d'une machine. Pour la plupart des entreprises, c'est déjà trop tard.

Cependant, en quelques secondes, la plate-forme de cyber IA signale la nature inhabituelle de l'activité du logiciel rançonneur et, vu l'urgence de la menace, détermine qu'une réaction autonome s'impose. Avec une précision chirurgicale, elle neutralise uniquement le chiffrement et les mouvements latéraux irréguliers, cantonnant les appareils infectés à leur comportement normal.

La plate-forme ne s'arrête pas là. Après avoir effectué une analyse des causes profondes, l'IA retrace l'origine de l'attaque à l'e-mail de phishing, une information qui la pousse à désinfecter les autres e-mails de la campagne avant qu'ils ne dupent d'autres victimes. La commerciale poursuit son travail, sans se douter que l'IA œuvre elle aussi avec acharnement en coulisses, sauvant la société d'une grave compromission.

L'IA attaque la boîte de réception

L'intelligence artificielle n'est pas seulement à la disposition des personnes qui se défendent., l'IA promet d'amplifier le spear-phishing en rendant ces e-mails plus réalistes et bien plus extensibles, en automatisant ce qui est, pour des attaques humaines, un processus assez laborieux. Une expérience notable réalisée en 2016 a montré qu'une boîte à outils reposant sur l'IA, qui avait étudié les comportements de ses cibles sur les réseaux sociaux afin de leur envoyer des tweets de

spear-phishing personnalisés, a été capable de ridiculiser un assaillant humain en faisant tomber dans le piège 275 victimes en à peine deux heures. L'être humain n'avait fait que 129 tentatives au cours de cette même période.

Par rapport aux campagnes de phishing standard à grande échelle, qui affichent un taux de compromission de 5 % à 14 %, ce genre de spear-phishing automatisé s'est avéré efficace 30 % à 66 % du temps, alors que la technologie de l'IA ne cesse de s'améliorer de manière exponentielle.
Il n'existe aucune solution miracle pour contrer cette nouvelle vague d'attaques basées sur l'IA, quelle que soit la robustesse acquise par les protections axées sur le périmètre. Nous devons plutôt utiliser nos propres plates-formes d'IA pour sécuriser nos actifs numériques de l'intérieur vers l'extérieur.

En unissant ainsi la sécurité de la messagerie électronique à la sécurité de l'entreprise, nous pouvons combattre de façon autonome les attaques par phishing, même celles pour lesquelles nous mordons à l'hameçon, en avalant la ligne et la canne avec.