Messagerie d'entreprise : l'internationalisation des attaques est un casse-tête pour les algorithmes

Contrairement aux autres types de menaces véhiculées par les emails, les attaques BEC se composent généralement de quelques lignes de texte, sans URL, fichier joint, ni autre élément analysable.

Pour les combattre, les éditeurs de solutions de sécurité de l'email ont donc opté pour des algorithmes basés sur l'intelligence artificielle. L'intérêt est que l'IA, par l'intermédiaire du traitement du langage naturel, va pouvoir détecter efficacement les attaques BEC.

De fait, les algorithmes vont analyser le texte et parvenir à détecter un sentiment d'urgence exprimé dans un email, ainsi que les expressions et mots clés couramment utilisés dans ces attaques, comme les demandes de virement bancaire, les paiements de factures et les cartes cadeaux.

Pendant de nombreuses années, la plupart des emails BEC étaient rédigés en anglais, mais depuis peu, on observe que de nouvelles langues telles que le français, l'italien, l'espagnol, l'allemand, etc. commencent à émerger dans ce type d'attaques. Cette évolution constitue en réalité un changement de stratégie face au développement de l'IA dans la sécurité de l'email et pose une difficulté majeure aux algorithmes pensés avant tout pour l'anglais.

Multiplication des attaques BEC dans différentes langues

L'apparition d'attaques BEC dans de nouvelles langues s'inscrit dans la droite lignée de la montée en sophistication globale des attaques. Même si de nombreux emails malveillants restent écrits avec une orthographe plus qu'approximative, les plus sophistiqués sont quant à eux parfaitement bien écrits et ne comportent pas les signes évocateurs des attaques BEC.

De plus, les hackers prennent désormais plus de temps pour initier un contact et un échange avec leurs victimes plutôt que de formuler leur demande dès le premier contact.
Cette tactique s'explique par deux raisons : tout d'abord, le pretexting incite la victime à baisser sa garde. Ensuite, en échangeant des emails avec le hacker, la victime apprend sans le savoir à certains algorithmes que l'expéditeur est légitime : l'adresse email du hacker peut ainsi être inscrite en liste blanche.

Les algorithmes d'IA et notamment de traitement du langage naturel détectent de mieux en mieux ces stratégies. Le problème, c'est que les algorithmes majoritairement pensés pour l'anglais sont naturellement plus efficaces dans cette langue et moins dans les autres.

Il a été récemment mis en évidence dans un article du Times aux États-Unis, que les algorithmes de

détection de discours haineux de Facebook offraient une efficacité limitée. Alors que Facebook affirme pouvoir analyser des contenus dans 40 langues, ses algorithmes ne détectent que 80 % des publications malveillantes. Un taux de détection non seulement très mauvais mais surtout dangereux en matière de sécurité de l'email.

Pour être efficaces dans l'analyse d'autres langues, les algorithmes d'IA ont besoin de jeux de données volumineux. L'anglais étant la langue la plus parlée dans le monde, les éditeurs de solutions de sécurité de l'email disposent de nombreuses données à l'aide desquelles entraîner leurs algorithmes. La taille des jeux de données dans les autres langues, en particulier de celles qui ne sont pas parlées partout dans le monde, est probablement bien inférieure.
Cette différence est essentielle, car plus les jeux de données sont réduits, moins les données sont fiables.

Pour améliorer les capacités linguistiques de leurs algorithmes d'IA, les éditeurs doivent non seulement renforcer leurs jeux de données, mais aussi réaliser d'importants investissements pour mettre à jour leurs moteurs de détection, deux tâches à la fois longues et coûteuses.
Par ailleurs, les données doivent être actualisées en permanence par de nouveaux échantillons de la langue cible.

Le nombre de boîtes aux lettres protégées par un éditeur et l'importance de sa présence internationale constituent au final les meilleurs indicateurs de la qualité de ses algorithmes d'IA, car ces outils sont entraînés à l'aide d'échantillons réels de texte dans de nombreuses langues différentes.