Die Rolle von KI in Hornetsecuritys Next-Generation E-Mail-Sicherheit

Künstliche Intelligenz (KI) ist seit einigen Jahren in aller Munde, doch als akademische Disziplin existiert sie bereits seit 1956. Jeder kennt ChatGPT, und viele nutzen es oder ähnliche Tools in ihrem persönlichen oder beruflichen Alltag, um Texte, Bilder oder sogar Videos zu erstellen. Doch wussten Sie, dass wir bei Hornetsecurity seit vielen Jahren KI in unseren Produkten einsetzen und dass sie die Grundlage für zahlreiche Funktionen bildet, die schädliche Nachrichten von Ihrem E-Mail-Posteingang fernhalten?

Zu den bemerkenswerten Technologien, die wir bei Hornetsecurity nutzen, gehören künstliche neuronale Netze, die sich ideal für die **Mustererkennung eignen, sowie Deep Learning**, das mehrere Schichten dieser Neuronen verwendet, um Computer Vision, Spracherkennung, natürliche Sprachverarbeitung und Bildklassifizierung zu verbessern.

Maschinelles Lernen (ML) ist ein Grundpfeiler der KI und bezeichnet Programme, die ihre Leistung bei der Ausführung einer Aufgabe automatisch verbessern können. Unüberwachtes maschinelles Lernen analysiert Datenströme, um Muster zu erkennen, während überwachtes maschinelles Lernen auf gekennzeichneten Daten basiert (z. B. 50 Bilder von Bananen und 50 Bilder von Äpfeln, die entsprechend gekennzeichnet sind; das Modell lernt daraus, neue Bilder korrekt zuzuordnen). Beim verstärkenden Lernen (Reinforcement Learning) wird ein Agent für korrekte Antworten belohnt und für falsche bestraft.

Hier erfahren Sie, wie wir bei Hornetsecurity diese Technologien einsetzen, um Spam und Bedrohungen von Ihren E-Mail-Posteingängen fernzuhalten.

E-Mail-Scanning und Sicherheit

Wie bei vielen technologischen Fortschritten bieten LLMs (Large Language Models) nicht nur Verteidigern zusätzliche Werkzeuge zum Schutz, sondern werden auch von Angreifern genutzt, um ihre Angriffe zu optimieren. Es ist schwierig, genaue Daten darüber zu sammeln, wie Kriminelle LLMs verwenden, aber anekdotische Hinweise zeigen eine Verbesserung der Grammatik, Übersetzungen in verschiedene Sprachen – insbesondere in Regionen, in denen die Gesellschaft weniger an Bedrohungen per E-Mails gewöhnt ist – sowie Unterstützung bei der Zielrecherche und der Generierung von Malware-Code durch KI.

Dies sind einige Bereiche, in denen wir KI und ML bei Advanced Threat Protection einsetzen:

- Analyse von Betrugsversuchen: Überprüfung der Authentizität und Integrität von Metadaten und E-Mail-Inhalten.
- Erkennung von Identitätsspoofing: Aufspüren und Blockieren gefälschter Absenderidentitäten.
- Intent-Erkennungssystem: Warnung vor Inhaltsmustern, die auf böswillige Absichten hindeuten.
- Spionageerkennung: Abwehr von Angriffen, die darauf abzielen, sensible Informationen auszuspähen.
- Erkennung gefälschter Fakten: Identitätsunabhängige Inhaltsanalyse von Nachrichten, um manipulierte oder falsche Fakten zu identifizieren.
- Erkennung gezielter Angriffe: Identifikation von Angriffen auf besonders gefährdete Personen.

Eine weitere äußerst nützliche Technik ist die Gruppierung (technisch als Clustering bezeichnet) von E-Mails mithilfe von maschinellem Lernen.

Es gibt ein breites Spektrum an Phishing-E-Mail-Kampagnen. Diese beginnen bei generischen, minimal zielgerichteten E-Mails mit geringem Wert („Klicken Sie hier, um Ihre Adresse für die FedEx-Lieferung zu bestätigen"), die in enormen Mengen versendet werden müssen, damit Angreifer eine geringe Erfolgsquote erzielen. Dann gibt es Spear-Phishing, bei dem gezielte Recherche und Vorbereitung in die E-Mails einfließen, um die Erfolgschancen zu erhöhen. Schließlich gibt es hochgradig individualisierte E-Mail-Köder, die oft auf Führungskräfte eines Unternehmens abzielen. Diese werden als Executive Phishing oder "Whaling" bezeichnet und in geringen Mengen, aber mit gut recherchierten Ansätzen versendet.

Bei den ersten beiden Kategorien ist es schwierig, eine einzelne E-Mail automatisch zu identifizieren und zu klassifizieren. Betrachtet man jedoch Millionen von E-Mails, beginnen Muster sichtbar zu werden, die die individuellen Kampagnen der Angreifer deutlich offenlegen. Hier setzen wir auf unüberwachte maschinelle Lerntechniken, um moderne Phishing-Angriffe zu bekämpfen. E-Mails werden auf Grundlage ihres Inhalts, Kontexts, der Absender-IP-Adresse, des Layouts und vieler weiterer Datenpunkte gruppiert. Das System identifiziert dann Ausreißer in diesen Clustern, um potenziell neue Phishing-Kampagnen zu erkennen.

Dies ermöglicht eine schnelle Erkennung von Phishing-Angriffen, ohne ausschließlich auf Reputationslisten (die langsam aktualisiert werden können), Heuristiken (die rechenintensiv sind, wenn jede E-Mail analysiert werden muss) oder Signaturen (die nicht mit der Geschwindigkeit moderner Phishing-Kampagnen mithalten können) angewiesen zu sein.

Dies sind einige Beispiele für Clustering:

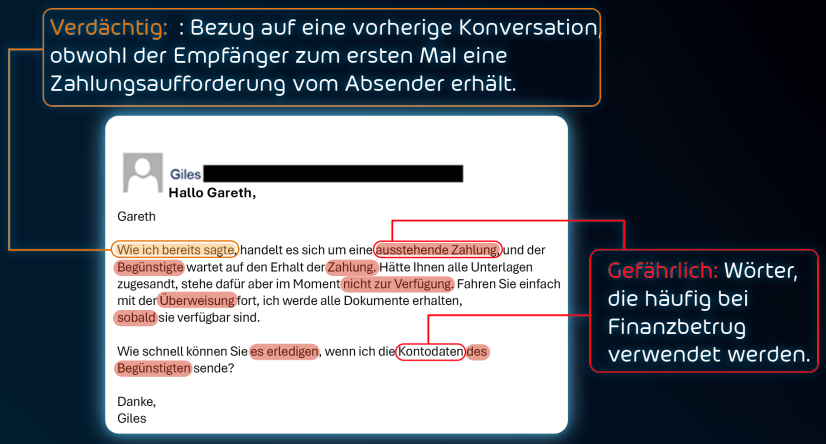
- Ein plötzlicher Anstieg ähnlicher E-Mails mit geringfügigen Variationen in den Domain-Namen kann auf eine neue Phishing-Kampagne hinweisen.
- Ein schneller Anstieg von E-Mails mit völlig unterschiedlichen Inhalten, aber einigen ähnlichen Merkmalen, wie z. B. gleichen Anhangsnamen oder ähnlichen Links in den ersten Zeilen der E-Mail, kann ebenfalls auf eine neue Phishing-Kampagne hindeuten.

Diese Methode ist auch nützlich, wenn Angreifer Nachrichten von infizierten Systemen sammeln und mit geringfügigen Änderungen erneut verwenden, um neue Opfer anzugreifen. Diese Technik wird weiterhin von verschiedenen Botnets eingesetzt (zum Beispiel QakBot im Jahr 2023).

- Extrahiere Schlüsselmerkmale aus einer E-Mail: Betreff, Absender, Anhänge, Links.
- Überprüfe E-Mail-Gruppen mit ähnlichen Merkmalen. Klassifiziere E-Mails basierend auf der Clustergruppe. Aktualisiere das Modell mit den extrahierten Merkmalen.
- Klassifiziere E-Mails basierend auf Cluster-Gruppen.
- Aktualisiere Modell mit extrahierten Merkmalen.

Dieses Diagramm veranschaulicht die Machine-Learning-Pipeline zur Identifizierung von Phishing-E-Mails anhand von Clustering

Eine weitere nützliche KI-Technik ist die natürliche Sprachverarbeitung (Natural Language Processing, NLP), die den Text von E-Mails analysiert. Dabei werden Ansätze wie Wort-Embedding und Themenmodellierung eingesetzt, um Kontext und Semantik abzuleiten. Diese Informationen können mit sequentieller bzw. Zeitreihenanalyse kombiniert werden, um abnormale Kommunikationsmuster zu erkennen. Ein typisches Beispiel wäre ein Geschäftsführer, der eine schnelle finanzielle Überweisung anfordert, was als auffällig gekennzeichnet wird, wenn er zuvor nie eine solche E-Mail gesendet hat.



Beispiel für die Analyse von Texten mittels NLP und die erkannten Signale

Wie bereits angedeutet, handelt es sich hierbei um ein fortlaufendes Anpassungsspiel, bei dem Kriminelle versuchen, unsere Abwehrmechanismen zu umgehen, während wir unsere Erkennungssysteme kontinuierlich verbessern, um neue Varianten aufzuspüren. Eine Stärke von ML-Modellen liegt in ihrer Lernfähigkeit, weshalb wir sie mittels verschiedener Datenquellen auf dem neuesten Stand halten:

- Benutzerfeedback ist wertvoll, da wir uns auf Endbenutzer verlassen, um falsch-positive Ergebnisse (bei denen eine E-Mail fälschlicherweise als bösartig markiert wurde) und falsch-negative Ergebnisse (bei denen eine verdächtige E-Mail nicht markiert wurde) zu identifizieren.
- Honeypots sind Simulationen von Zielen oder E-Mail-Postfächern, die sowohl generische als auch gezielte Angriffe anziehen und uns Trainingsdaten liefern.

Das bedeutet, dass unsere Modelle sich kontinuierlich an die sich schnell wandelnde Bedrohungslandschaft anpassen – der beste Ansatz für eine moderne E-Mail-Hygienelösung wie unsere.

Eine weitere Herangehensweise von Kriminellen besteht darin, schädliche Inhalte aus der E-Mail zu entfernen und die Nutzlast (Payload) extern auf einem Webserver zu hosten, während lediglich ein Link dazu in der E-Mail enthalten ist. Die Erkennung bösartiger Links ist ein zentraler Bestandteil unserer KI-gestützten Lösung Secure Links. Wir ersetzen jeden Link in eingehenden E-Mails durch eine Version, die über unser Secure Web Gateway (SWG) geführt wird.

Dieses Gateway nutzt maschinelles Lernen und Deep Learning sowie überwachte und unüberwachte ML-Modelle, um mehr als 47 Merkmale der URL-Links und der zugehörigen Webseitenziele zu analysieren. Es untersucht bösartiges Verhalten, URL-Weiterleitungen und Verschleierung. Darüber hinaus kommen Computer-Vision-Modelle zum Einsatz, die Bilder, einschließlich Markenlogos und QR-Codes, sowie Textinhalte, die in Bildern eingebettet sind, analysieren.

Das Ergebnis ist, dass wir schädliche Inhalte erfassen, die in E-Mails verlinkt sind, selbst bei schnellen und hochspezifischen Angriffen. Eine weitere gängige Taktik besteht darin, eine Website zu kompromittieren, deren Inhalt jedoch nicht zu verändern, um anschließend eine E-Mail-Kampagne zu starten. Sobald die E-Mails zugestellt wurden, wird die schädliche Nutzlast auf der Website bereitgestellt. Aus diesem Grund scannt Secure Links die Ziele von Links sowohl zum Zeitpunkt der Zustellung als auch beim Anklicken.

Anhangscans

Angreifer verzichten oft darauf, ihre Nutzlast direkt im Text der E-Mail zu platzieren oder Links einzufügen, die gescannt werden können. Stattdessen wird der schädliche Inhalt in einem Anhang versteckt. Anders als textbasierte E-Mails, die relativ einfach gescannt werden können, kommen Anhänge in vielen verschiedenen Dateiformaten vor und können auf unterschiedliche Weise gefährlich gemacht werden, einschließlich der Einbettung von Links zu bösartigen Inhalten.

Hier kommt unsere Sandbox Engine ins Spiel, die ebenfalls auf maschinellem Lernen basiert. Sie öffnet Anhänge, identifiziert, ob sie bösartig sind, und isoliert die E-Mail, wenn dies der Fall ist. Die Engine analysiert das Verhalten der Datei, um festzustellen, ob sie versucht, eine Sandbox-Umgebung zu erkennen (ein eindeutiges Warnsignal). Sie überprüft außerdem das Dateisystem, um zu erkennen, ob der Anhang versucht, neue Dateien zu erstellen oder bestehende zu ändern.

Der Registry-Monitor untersucht, ob ungewöhnliche Werte in der Registry erstellt werden – ein typisches Mittel, um Malware nach einem Neustart des PCs dauerhaft zu erhalten. Unser Prozess-Monitor erkennt, wenn schädliche PDF- oder Office-Dateien versuchen, Kindprozesse zu starten. Auch der Netzwerkverkehr des Anhangs wird analysiert, um Verbindungen zu Servern im Internet zu erkennen – ein weiteres verdächtiges Verhalten eines angehängten Dokuments. Schließlich wird der Speicher in der Sandbox nach dem Öffnen des Anhangs überprüft, da ungewöhnliche Speicherzugriffe ein weiteres starkes Anzeichen für Malware sind.

Insgesamt verlässt sich die ML-Engine der Sandbox auf über 500 Indikatoren, um Anhänge zuverlässig in harmlose und schädliche Dateien einzuordnen – und das in kürzester Zeit.

Ausgehender Scan

Eine weitere einzigartige, KI-gestützte Lösung von Hornetsecurity ist unsere AI Recipient Validation (AIRV).

Diese analysiert die E-Mail-Kommunikationsmuster jedes Nutzers, lernt kontinuierlich dazu und erkennt unbeabsichtigte Empfänger, E-Mails mit personenbezogenen Informationen (PII) sowie unangemessene Formulierungen. Wenn Probleme festgestellt werden, werden diese dem Nutzer gemeldet, und dessen Reaktionen auf die Warnungen fließen in die Analyse zukünftiger E-Mails ein.

AIRV warnt Nutzer in folgenden Szenarien:

- Versand einer E-Mail an einen potenziell unbeabsichtigten Empfänger.
- Ein üblicherweise häufiger Empfänger aus einer Gruppe fehlt.
- Ein Nutzer wird in einer bestehenden Gruppe hinzugefügt oder ersetzt.
- Erstmaliger Versand von E-Mails an Nutzer aus verschiedenen Organisationen oder an private E-Mail-Adressen.
- Beantwortung einer großen Verteilerliste.
- Versand einer E-Mail an einen Empfänger, mit dem der Nutzer zuvor keine Beziehung hatte.
- Versand von E-Mails mit sensiblen Informationen wie personenbezogenen Daten (PII) oder Kreditkartendaten.
- Versand einer E-Mail mit unangemessenen Formulierungen.

Schulung von Nutzern mit KI

Da keine E-Mail-Hygienelösung zu 100 % effektiv ist, bleibt in manchen Fällen der Endnutzer die letzte Verteidigungslinie, indem er wachsam ist und auf verdächtige E-Mails angemessen reagiert. Security Awareness Service von Hornetsecurity hat KI im Zentrum und bietet jedem Nutzer genau die richtige Menge an Training.

Simulierte Spear-Phishing-Kampagnen werden verschickt, und Nutzer, die auf Links klicken oder Anhänge öffnen, erhalten mehr Training, während diejenigen, die dies nicht tun, vorerst nicht weiter geschult werden. Die KI-gestützte Spear-Phishing-Engine nutzt dabei verschiedene Stufen der Raffinesse in den Simulationen (basierend auf realen Angriffen, die wir aufgezeichnet haben), um Endnutzern zu helfen, selbst die fortschrittlichsten Angriffe zu erkennen. Genau wie bei echten Angriffen führen unsere Links zu gefälschten Login-Seiten, E-Mails sind Teil eines Threads, und Dateianhänge enthalten „bösartige" Makros.

Die wahre Stärke des Security Awareness Services liegt darin, dass die KI diesen Prozess automatisch verwaltet. Dadurch werden Administratoren entlastet, die sich auf produktivere Aufgaben konzentrieren können, anstatt Phishing-Simulationen und Trainingszuweisungen manuell zu verwalten.

Fazit

Hornetsecurity nutzt Künstliche Intelligenz seit vielen Jahren und setzt verschiedene Tools zur Bewältigung unterschiedlicher Herausforderungen ein. Durch die kontinuierliche Optimierung unseres Ansatzes bieten wir effektiven Schutz vor E-Mail-basierten Bedrohungen und schulen Nutzer darin, Phishing Angriffe zu erkennen.