Cyber Angriffe im Zeitalter der KI: Eine Analyse der Bedrohungslandschaft und Schutzstrategien - Patrick Schläpfer und Oliver Nass

*[Ein "Hacker" sitzt mit dem Rücken zum Publikum und generiert live eine Demo Phishing-Attacke. Er dreht sich um und startet mit seiner Präsentation.]*So einfach und schnell lassen sich grossflächige und doch einigermassen gezielte Phishing-Angriffe mit Hilfe von AI generieren. Wie Sie gesehen haben, war das ein sehr simples Beispiel. Wir haben in diesem Fall lediglich eine E-Mail mit ein paar wenigen Anweisungen generiert. Wir haben die E-Mail dann schön in Form einer HTML-E-Mail formatieren lassen. Sieht schön aus und wirkt vertrauenswürdig. Und wir haben es dann direkt in verschiedene Sprachen übersetzt. Wir könnten nun noch einen Schritt weitergehen und sogar ein Skript generieren lassen, das die Phishing-E-Mail dann an alle Empfänger automatisch versendet. Unser Beispiel war jedoch sehr einfach. Der Grund dafür ist, dass ChatGPT wie auch andere Online-AI-Services, sogenannte Security Policies einsetzen. Diese Security Policies schützen die Systeme vor Missbrauch. Das heisst, es ist nicht einfach möglich, Phishing zu generieren oder Malware damit zu entwickeln. Aber dies hindert Angreifer nicht daran, dies dennoch zu tun. Denn es gibt sehr viele open-source Large Language Models (LLM). Man kann diese offline auf eigener Infrastruktur hosten und so Malware entwickeln oder auch Phishing generieren. Oft ist das jedoch gar nicht nötig, denn es gibt relativ viele Services und Tools in jeglichen Hackerforen und Cybercrime-Marktplätzen, die man einfach nutzen kann. Und diese sind alle relativ einfach zugänglich und ziemlich günstig.

Wir machen solche Analysen seit ein paar Jahren und stellen fest, es ist wirklich alles sehr günstig. Wie Sie sehen, kann man in solchen Hackerforen oder Cybercrime-Marktplätzen Exploits kaufen, zu verschiedensten Schwachstellen in diversen Software-Packages. Angreifer können gesamte Malware-Pakete kaufen, mit jeglichen Funktionen. Von Information Stealern bis zu Cryptern. Alles Mögliche und alles ist sehr einfach zugänglich. Aber Angreifer können auch direkt Zugriff zu Unternehmen kaufen. Netzwerkzugang zu verschiedenen Firmennetzwerken überall auf der Welt. Na, wie das aber jetzt wirklich in der Realität aussieht, will ich Ihnen hier kurz zeigen. Wir nutzen einen sogenannten Tor-Browser um ins Darknet zu gelangen. Der Verbindungsaufbau dauert normalerweise etwas länger, denn der Tor-Browser verbindet sich über verschiedene

Knotenpunkte zum Server, damit die Anonymität gewährleistet ist. Kennt man dann die spezifischen Adressen, so kommt man auf einen Hacker-Marktplatz wie hier. Sie haben gesehen, man kann anonym mit verschiedenen Kryptowährungen bezahlen. Hier gibt es Zugänge zu verschiedenen Servern, die exponierte Remote Desktop Protokoll Zugänge haben. Es werden gehackte E-Mail-Accounts sehr günstig angeboten. Oder Angreifer können Phishing Kits erwerben. Das heisst, sie kaufen die gesamten Skrips und müssen diese gar nicht mehr selbst entwickeln und kaufen so auch das komplette Design. Das Design in diesem Fall ist relativ einfach und doch sehr überzeugend.

Das heisst, auch mit dem Zugang und der Verfügbarkeit von allen diesen Tools und Services werden Cyber Angriffe immer wie gefährlicher. Damit zeigt sich, dass auch wenn Security Policies implementiert sind, können Angreifer solche Services erwerben und nutzen. Aktuell sehen wir zudem eine starke Zunahme von Angeboten solcher AI-Services in verschiedenen Cybercrime-Marktplätzen.

Wir haben bisher primär über eher grossflächige und einfache, schnelle Angriffe gesprochen, die sich alle im Bereich Social Engineering bewegen. Angriffe können aber auch noch viel gezielter durchgeführt werden. Ich gehe davon aus, dass die meisten von Ihnen einen Social-Media-Account haben. Ich gehe auch davon aus, dass möglicherweise die meisten Firmen einen solchen haben. Und wenn nicht, dann haben Sie sicher eine Firmenwebseite. Da findet man Mitarbeiter, möglicherweise Standorte und Referenzprojekte, alles interessante Informationen. Diese Informationen sammeln die Angreifer. Sie füttern die Informationen dem LLM und dieses erstellt eine sehr gezielte Kampagne, die ein sehr gezieltes Unternehmen angreift. So ist das viel glaubwürdiger als der Angriff, den wir hier demonstriert haben. Wir können aber noch einen Schritt weiter gehen. Angreifer können ein Deep Fake einer Person generieren. In diesem Fall wird nicht nur die Sprache imitiert, sondern es wird auch das Bild imitiert, quasi die gesamte Persönlichkeit. Und dass es sich dabei nicht nur Theorie handelt, sondern Praxis, zeigte ein Fall von Anfang dieses Jahres.

Ein Angreifer gab sich als CFO aus und hat einen Finanzmitarbeiter eines internationalen Konzerns angerufen. Er bat diesen höflich, eine Überweisung auf ein entsprechendes Konto zu tätigen. Da der Finanzmitarbeiter dachte, er spreche effektiv mit dem CFO, hat er die Überweisung getätigt. Dieser hat ja die Stimme gehört und auch das Bild gesehen und der vermeintliche CFO hat sich sogar entsprechend verhalten. So wurde dem Unternehmen 25 Millionen gestohlen. Und das ist effektiv Realität. LLMs und AI können aber auch für andere Angriffe verwendet werden.

Bisher haben wir primär Phishing und Social Engineering angeschaut. Angreifer können damit aber auch Malware entwickeln oder Schwachstellen suchen. Das alles ist möglich. Dieses Video ist ein einfaches Beispiel. Wir haben eine Datei. In dieser Datei ist ein Passwort versteckt. Die Idee dieser Capture the Flag Challenge ist es, das Passwort zu finden. Wir gehen in diesem Szenario davon aus, Angreifer haben nicht das Fachwissen, wie man so eine Datei oder gar Maschinen Code analysiert. Dies spielt aber keine Rolle. Wir öffnen die Datei in einem Disassembler, nehmen den Code und beauftragen das LLM: «Gib mir eine Zusammenfassung, was das Programm macht und das richtige Passwort.» Teilweise erhalten wir nicht alle Informationen, also fragen wir einfach erneut. Nun rekonstruiert die AI das Programm, welches wir in Assembly hineingegeben haben.

Die Idee ist aber, das Passwort zu finden. Wir wollen nicht wissen, was das Programm macht. Das ist dem Angreifer eigentlich auch meistens egal. Wir wollen aber quasi den Exploit der Schwachstelle finden. In diesem Fall hat mich das LLM nicht so ganz verstanden. Das bin ich mir aber auch mit Menschen gewohnt und daher frage ich es einfach nochmals. Und in diesem Fall schreibt es effektiv quasi den Exploit in C Code. Wir können das Programm einfach kompilieren, ausführen und erhalten so ein Passwort. Wir prüfen, ob das Passwort stimmt, und es ist effektiv korrekt. Das kann alles mit Hilfe von LLMs ohne viel Fachwissen gemacht werden. Das heisst, der Einstieg in diesen Bereich wird immer wie einfacher.
Wir müssen aber auch anfügen, dass nicht alle Angriffe nur noch AI verwenden. Das klingt interessant und cool, aber die Bedrohungslage hat sich über die letzten Jahre immer wieder verändert. Ich bin Teil des Thread Research Teams von HP und wir analysieren Threats und deren Trends über die Zeit. Jedes Quartal publizieren wir einen Threat Insights Report und über die Jahre schauen wir, wie sich die Lage verändert. Würden wir zusammenfassen, wie sich die Lage über die letzten Jahre verändert hat, wären das unsere Kernpunkte.

Der Verteilmechanismus von Malware hat sich stark diversifiziert. Sie halten Malware in verschiedenen Dateitypen und auf verschiedene Weise. Sei das per E-Mail, via Webdownload oder gar per USB-Stick. Wir sehen auch, dass immer wie mehr Schwachstellen ausgenutzt werden. Sei dies zum Beispiel in Office-Programmen, PDF-Readern, Web-Browsern oder auch VPN-Gateways. Wie wir zu Beginn der Präsentation gesehen haben, sind die Tools sehr günstig und sehr einfach zu erhalten. Das heisst auch «Initial-Access» ist sehr günstig und immer wie häufiger zu erwerben. Und zu guter Letzt stellen wir fest, dass Angreifer immer wie stärker zusammenarbeiten. Das heisst, verschiedene Angreifer Gruppierungen fokussieren sich auf einen Teilbereich, in dem sie sehr stark sind und kümmern sich gar nicht mehr um den Rest. Klingt schon fast wie in der Industrie, wo man professionell zusammenarbeitet. Und AI ist in allen diesen Bereichen ein starker Enabler. Das heisst, überall kann man AI nutzen und AI wird die Bedrohungslage auch noch entsprechend verändern und diese Veränderungen beschleunigen.

Wie Sie sich und Sie Ihr Unternehmen entsprechend vor dieser Bedrohungslage schützen können, wird Ihnen Oliver nun etwas näherbringen.

Vielen Dank, Patrick. Mein Name ist Oliver Nass. In meiner Rolle als Security Advisor helfe ich Unternehmen, Cyber Risiken zu erkennen und mit wirksamen Massnahmen zu begegnen. Was Sie hier nun sehen, ist eine Darstellung, wie man Verteidigungsmassnahmen strukturieren kann. Angefangen von einer guten Security Governance, das bedeutet eine Transparenz darüber: Was habe ich überhaupt in meinem Unternehmen für Unternehmensprozesse, für IT Assets? Über die Identifizierung von Cyberangriffen, wie sie Patrick verdeutlicht hat, über natürlich die Wiederherstellung von den Systemen im Falle eines erfolgreichen Angriffs können wir hier alle Produkte und Dienstleistungen auch einordnen. Diese scheinbar simple Darstellung wurde vom National Institute of Standards and Technology in den USA entwickelt und geht als Gold-Standard unter Cyber Security Experten. Wir von HP sind der Überzeugung, dass alle unsere Produkte und alle unsere Dienstleistungen diese Bedürfnisse auch abdecken und jegliche Aktivitäten im Bereich Cyber Security auch dabei berücksichtigen. Vielleicht erkennen Sie bereits einige der Produkte wieder, die Sie im Alltag nutzen oder zumindest vom Namen her gehört haben. Egal ob die Bedrohung platformbasiert ist, also vom sogenannten Root of Trust, Firmware oder auf der Software-Ebene, wir haben für alle Bedrohungsszenarien eine Antwort. Ich möchte Ihnen gleich gerne anhand von Sure Click ein Beispiel aufzeigen, wie das denn konkret aussieht.

Stellen Sie sich den Alltag vor, Sie klicken auf eine Einzeldatei, Sie öffnen eine E-Mail. In dieser E-Mail befindet sich vielleicht ein Anhang. Überall Angriffsvektoren für potenzielle Chart-Software, für potenzielle Chart-Code in Expertenterminology. Wir möchten durch SureClick verhindern, dass diese Fehler ein grösseres Ausmass haben als für die Einzelperson. Wir sehen hier ein ganz gewöhnliches Alltagsbeispiel. Links eine Excel-Datei und rechts SureClick Enter im Liveview. Jede einzelne Task, den Sie auf dem Computer gerade ausführen, wird dort dargestellt anhand einer sogenannten mikro virtuellen Umgebung. So können wir zum Beispiel eine PDF öffnen, einen Link anklicken und diesen Link auch anschauen. Sie werden in Ihren Aktivitäten, in Ihrem ihr Alltag nicht beeinträchtigt. Wir können auch Anhänge öffnen, die werden auch natürlich dann dort dargestellt. USB-dateien werden natürlich dort auch berücksichtigt. Wie Sie sehen, keine Beeinträchtigung Ihres Alltags. Was hier wichtig ist: Sie haben jeden einzelnen Task dort verfügbar und für Ihre Security-Teams im Hintergrund eine Transparenz darüber, was diese Dateien, die Sie gerade ausführen, auch auf der Prozessebene durchführen. Können weiterarbeiten, kein Problem. Und dieser Liveview, der ist nicht immer da. Den kann man einfach hochholen und dann wissen, was denn gerade passiert.

Doch was passiert, wenn wir uns in einem Szenario befinden, wo wir unbeabsichtigt auf eine Phishing-E-Mail klicken. Sie haben gesehen, wie leicht es ist, eine Phishing-E-Mail zu generieren und wie leicht es ist, auch Schadsoftware zu integrieren. Hier ein Beispiel aus dem Bereich HR. Wir haben hier einen Lebenslauf angeklickt, ganz unscheinbar. Was wir aber nicht wissen, ist, die Schadsoftware wird bereits ausgeführt. Es passiert noch nichts, es passiert noch nichts und gleich sollten wir natürlich das Problem haben. Das sieht natürlich erst mal sehr gefährlich aus, ist es auch. Das ist eine sogenannte Ransom Note. Das ist eine Aufforderung, Geld zu bezahlen für die Entschlüsselung Ihrer Dateien. Wenn Ihre Prozesse nicht richtig segmentiert sind, Ihre Netzwerke nicht segmentiert sind, haben Sie aufgrund dieser Ransomware Ihre Dateien verschlüsselt. In einem Normalfall würde dies auch eintreten, doch zum Glück haben wir SureClick. Durch SureClick und die Segmentierung in einer Mikroumgebung haben wir kein Problem, dass die Datei angeklickt wird, da sich die Schad-Software in der Datei sich nicht mehr ausbreiten kann. Und hier eine Übersicht darüber, was ihre Security-Teams dann auch im Nachhinein in der sogenannten Kill Chain auch nachvollziehen können. Dadurch wird es einfacher Ihre Unternehmensrisiken zu verstehen und dem auch entgegenzuwirken.

Nichtsdestotrotz ist ohne sichere und resistente Plattform-Security eine Software-Security anfällig für potenzielle Angriffe. Security muss vom sogenannten Root of Trust herkommen. Nur so kann sichergestellt werden, dass auch die Software-Integrität gewährleistet ist. In den letzten Jahren haben wir einen Trend festgestellt hinsichtlich Firmware-Malware. Das bedeutet Malware, die auf der Plattform-Security-Ebene vorhanden ist. Diese ist viel schwieriger zu erkennen und viel schwieriger zu entfernen als gewöhnliche Malware. Was Sie hier sehen auf der Timeline, ist eine Darstellung, wie sich die Frequenz der Firmware-Malware in den letzten Jahren vorgefunden haben. An den weissen Punkten erkennen Sie einige einige Beispiele darüber, was fortgeschrittene Bedrohungsakteure an Firmware-Malware kreiert haben. Wie Sie sehen, es gibt viele Beispiele, wie KI die Bedrohungslage verändert, zum Beispiel, wie Phishing-E-Mails kreiert werden können, bis hin zur Erstellung von dedizierten Payloads. Wir können dabei helfen, Ihre Risiken dabei zu verstehen. Wir können dabei Ihre Abwehrmechanismen stärken und dabei natürlich auch Lösungen finden, wie wir diese denn konkret begegnen. Das alles im Einklang mit Ihren Unternehmensprozessen, mit Ihren Ressourcen und natürlich Ihren Prioritäten. Was Sie am Ende erhalten, ist ein detaillierter Security Report, der zugeschnitten von Ihren Security-Teams bis hin zum Executive Level verstanden werden kann, damit Sie die notwendigen Ressourcen haben, Cyberrisiken zu begegnen.

Doch zu guter Letzt, was ich Ihnen noch mitgeben möchte, ein Take away, warum Sie uns denn damit vertrauen sollten. Zum einen sind wir Spezialisten über unsere Produkte und natürlich auch, wie man Sie am besten schützt. Und zum anderen haben wir Security Compliance auch messbar

gemacht. So sind wir stolz darauf, dass wir bei über 200 Kunden im Schnitt die Security Compliance 24% verbessern konnten. Sie werden bestimmt einige Fragen jetzt haben und ich würde Sie im Namen von Patrick auch ganz herzlich einladen, gleich beim Apero auf uns zuzukommen und uns diese Fragen zu stellen. Vielen Dank.