SOTI

10 YEARS
OF SECURITY INSIGHT

B10 AUSGABE 03

Web-Scraper im E-Commerce

Die Gefahr für Ihr Geschäft

Akamai

„State of the Internet"-Sicherheitsbericht

Inhaltsverzeichnis

3	Bots: Positive und negative Aspekte
4	Wichtige Erkenntnisse aus dem Bericht
5	Gute und schlechte Bots
6	Grundlagen des Scraping
6	Scraping befindet sich im Wandel – das bleibt den Kunden nicht verborgen
9	Die allgemeinen Gefahren von Web-Scraping
9	Scraping auf Auftragsbasis: Web-Scraping-Dienste von Drittanbietern
11	Scraping-Prozess für KI-Botnets
14	Fallstudie: Vorteile von Lösungen zur Erkennung von Web-Scraping
16	Schutz und Schadensbegrenzung
19	Compliance-Aspekte
20	Fazit
21	Methodik
22	Mitwirkende

Wussten Sie, dass Bots mehr als die Hälfte des gesamten Web-Traffics generieren? Der auf umsatzgenerierende Webanwendungen und -Assets angewiesene Handel ist am stärksten von risikoreichem Bot-Traffic betroffen (Abbildung 1). Wir hören oft, dass Bots sich weiterentwickeln. Derzeit erregen jedoch insbesondere Web-Scraper-Bots die Aufmerksamkeit von E-Commerce-Unternehmen, da sich ihre wirtschaftlichen Auswirkungen, die häufig nicht offensichtlich sind, von denen anderer Bots unterscheiden. Die Erkennung von Scraper-Bots ist inzwischen deutlich schwerer geworden. Das liegt unter anderem am zunehmenden Auftreten von Botnets, die auf künstlicher Intelligenz (KI) basieren, und an Headless-Browser-Technologien, wodurch Scraper äußerst gut getarnt sind. So hatte etwa einer der E-Commerce-Kunden von Akamai 99 % des risikoreichen Traffics blockiert und war sich nicht einmal bewusst, dass dieser von Scraper-Bots stammt.

Bot-Anfragen pro Monat: Die 3 am stärksten betroffenen Branchen

1. Januar 2023 bis 31. März 2024

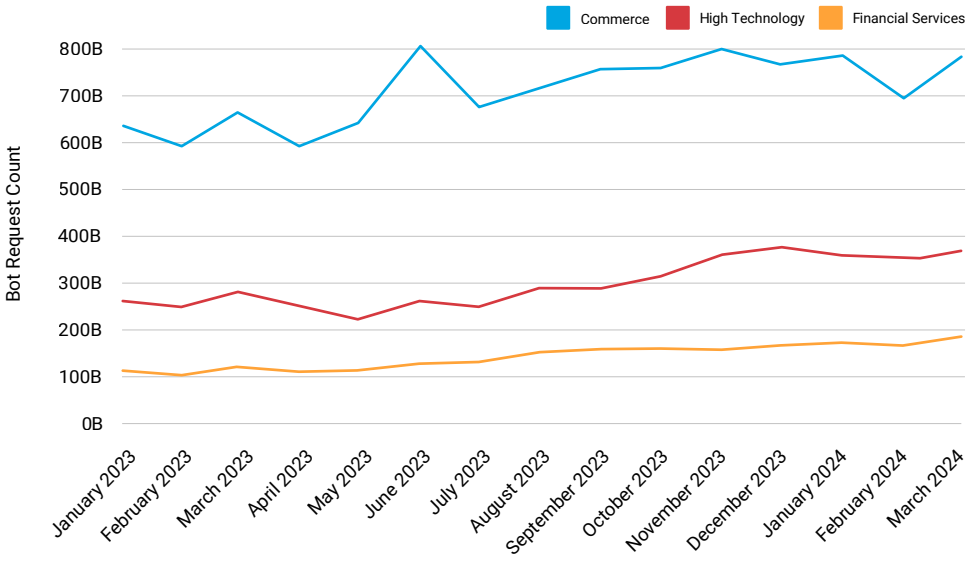

Abb. 1: Der Handel ist die am stärksten von Bot-Anfragen betroffene Branche. Zwischen Anfang 2023 und dem 1. Quartal 2024 ist ein Anstieg des globalen Bot-Traffics im Handelssektor zu beobachten

Deshalb konzentrieren wir uns in diesem „State of the Internet"-Sicherheitsbericht (SOTI) auf die Weiterentwicklung und Spezialisierung dieser Bots und ihrer Betreiber. Bots existieren schon seit geraumer Zeit und auch heute werden sie von verschiedenen Gruppierungen für kriminelle Angriffe, Betrugsversuche und die Beschaffung von wettbewerbsrelevanten Informationen genutzt. In letzter Zeit konnten wir einen Trend zur verstärkten Nutzung von Bots aller Art und einen Anstieg der negativen Auswirkungen von Scraper-Bots auf Unternehmen beobachten. In diesem Bericht sollen sowohl technische Erkenntnisse als auch Angriffsmethoden beschrieben werden, um das Bewusstsein für dieses im gesamten Handelssektor zunehmende Problem zu schärfen.

Bots: Positive und negative Aspekte

Jedes große E-Commerce-Unternehmen muss sich mit Bots auseinandersetzen, die sich ständig weiterentwickeln und entsprechend ihrer Zielsetzung spezialisieren. Innerhalb des Handelssektors wird eine Vielzahl von Bot-Typen eingesetzt, die zahlreiche verschiedene Aufgaben ausführen können. Zur einfachen Abgrenzung von Bots lassen diese sich in drei Gruppen unterteilen: gute Bots, schlechte Bots und Bots in der Grauzone. Gute Bots helfen Kunden dabei, Ihre Website zu finden. Schlechte Bots durchsuchen Ihre Website für böswillige Zwecke. Bots in der Grauzone sind legitim, neigen aber dazu, sehr auffällig zu sein. Tatsächlich sind sie eine Unterkategorie der guten Bots (z. B. Partner-Bots, die ständig Pings senden, und andere Programm-APIs mit häufigen Aufrufen).

Hilfreiche Bots wie Chatbots und Suchmaschinen-Bots, die sich positiv auswirken sollen, etwa indem sie einfache Nutzerfragen beantworten und Website-Inhalte für genauere Suchergebnisse bereitstellen, sollten optimiert werden, um IT-Kosten zu senken. In Bezug auf schädliche Bots, wie Credential-Stuffing-Bots, die unberechtigt auf Kundenkonten zugreifen und diese übernehmen, sollten vorbeugende Maßnahmen ergriffen werden, ohne dass das Kundenerlebnis insgesamt beeinträchtigt wird. Die kürzlich neu eingeführten Web-Scraper-Bots sind als besonders problematisch einzustufen, da sie dem Umsatz und der Kundentreue schaden und Kosten erhöhen.

Scraper-Bots, ein Botnet, mit dem Daten und Inhalte direkt aus Websites im Internet extrahiert werden, stellen eine gesonderte Kategorie dar. Sie verlangen unsere Aufmerksamkeit, weil sich ihre Funktionsweise, ihre geschäftlichen Auswirkungen und ihre Erkennbarkeit von anderen Bots unterscheiden. Web-Scraper haben viele Facetten, da ihre Anwendungsfälle variieren, je nachdem, wie Unternehmen und Betreiber die von diesen Bots gesammelten Informationen monetarisieren. Unabhängig von ihrem jeweiligen Ziel schaden Scraper dem Unternehmensumsatz, erhöhen die IT-Kosten und beeinträchtigen allgemein das Kundenerlebnis.

In diesem SOTI-Bericht behandeln wir die Auswirkungen von Scraping über E-Commerce hinausgehend und erörtern, warum es im allgemeinen Interesse von Geschäftsinhabern liegen sollte (etwa aus digitaler, Marketing-, Marken-, Finanz-, Risiko- und Sicherheitsperspektive), missbräuchliche Scraper zu stoppen. Um diese Auswirkungen besser verstehen zu können, sollten Sie sich ein vollständiges Bild davon verschaffen, warum Web-Scraper-Bots entwickelt wurden, wofür sie eingesetzt werden, wie sie funktionieren, welche Auswirkungen sie haben und welche Gegenmaßnahmen ergriffen werden können.

Wichtige Erkenntnisse aus dem Bericht

- Web-Scraping ist nicht nur ein Problem in Bezug auf Betrugsversuche und Sicherheit, sondern auch ein geschäftliches Problem. Scraper-Bots haben negative Auswirkungen auf mehrere Unternehmensaspekte, darunter Umsatz, Wettbewerbsvorteil, Markenidentität, Kundenerlebnis, Infrastrukturkosten und digitales Erlebnis.
- Laut einer Fallstudie von Akamai stammten 42 % der gesamten Traffic-Aktivität von Bots und 65 % dieses Bot-Traffics von bösartigen Bots. Bei insgesamt 63,1 % des von schlechten Bots verursachten Traffics wurden fortschrittliche Techniken verwendet.
- Scraper-Bots haben sich durch die Headless-Browser-Technologie, also Browser ohne Header, verändert. Diese Art von Bot-Aktivität erfordert einen komplexeren Ansatz als die gängigen auf JavaScript basierenden Gegenmaßnahmen.
- Zu den technischen Auswirkungen, denen Unternehmen durch Scraping ausgesetzt sind, unabhängig davon, ob das Scraping mit schädlichen oder guten Absichten durchgeführt wurde, gehören eine Verschlechterung der Website-Performance, einer Verwässerung der Website-Metriken, Angriffe mit kompromittierten Anmeldedaten von Phishing-Websites, erhöhte Rechenkosten und vieles mehr.
- Die verschiedenen Trafficmuster sollten beobachtet und untersucht werden, um festzustellen, ob der Traffic auf einer Website von menschlichen Nutzern oder von einfachen oder fortschrittlichen Bots stammt. Diese Muster können sich am Tagesablauf orientieren oder intermittierend oder kontinuierlich sein.

Gute und schlechte Bots

Beginnen wir mit den Grundlagen: Ein Bot, kurz für „Roboter“, ist ein Computerprogramm, das automatisierte Aufgaben schneller und genauer ausführen kann als ein Mensch. Die verschiedenen Rollen und Typen von Bots lassen sich in zwei Hauptkategorien einteilen: gute und schlechte Bots (Abbildung 2). Bots in Grauzonen sind eine Unterkategorie guter Bots, vorerst führen wir sie aber unter den guten Bots, um den Vergleich zu erleichtern.

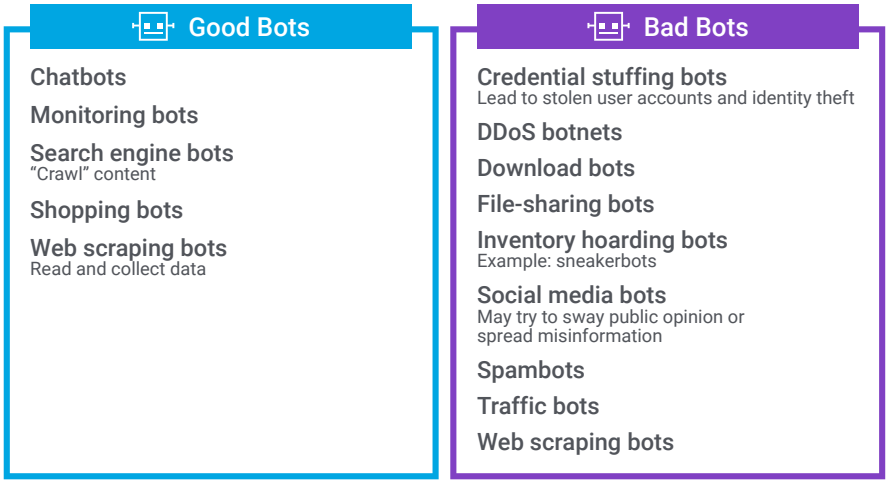

Abb. 2: Ein Vergleich von guten und schlechten Bots mit Beispielen

Gute Bots sind nützliche Bots, die Tools und Dienste bereitstellen, während schlechte Bots oft mit böswilliger Absicht von Cyberkriminellen und Betrügern eingesetzt werden. Ein Beispiel für diese Art böswilliger Bots sind Traffic-Bots, die menschliches Verhalten im Internet nachahmen, um Klicks und Traffic auf Websites zu erhöhen (d. h., um Werbebetrug zu begehen).

Web-Scraping-Bots können sowohl als gute als auch als schlechte Bots fungieren. Die Unterscheidung ergibt sich dadurch, wie Unternehmen die von diesen Bots gesammelten Informationen nutzen. Wir gehen nun genauer auf verschiedene Anwendungsfälle ein, die sowohl die guten als auch negativen Auswirkungen von Scraper-Bots beleuchten und die einige der weltweit größten Einzelhändler und E-Commerce-Marken betreffen.

Grundlagen des Scraping

Web-Scraping wird häufig von E-Commerce-Unternehmen verwendet. Im Reise- und Gastgewerbe beispielsweise nutzen Reiseveranstalter dynamische Inhalte von ihren Hotel- und Airline-Partnern, um über Verfügbarkeiten und Preise auf dem Laufenden zu bleiben. Diese Art von Scraping tritt erwartungsgemäß auf, und Unternehmen verwenden Bot-Kontrollmechanismen, um Scraper zu Tageszeiten zu drosseln, zu denen echte Nutzer Reservierungen vornehmen möchten. Zudem greifen Unternehmen auf Anbieter von Datenextraktionsdiensten zurück, um Daten zu Interessenten und andere zugehörige Informationen von Wettbewerbern zu sammeln. Darüber hinaus können Scraping-Bots zur Analyse von Daten und zum Erkennen von Trends verwendet werden. Scraping ist auch für die Überprüfung von Websites nutzbar. Dadurch lassen sich Online-Angebote und -Dienste verbessern und potenziellen Kunden wird das Auffinden von Unternehmensprodukten erleichtert, z. B. über eine Suchmaschine. All diese Maßnahmen können Unternehmen Wettbewerbsvorteile verschaffen. Jedoch werden Scraper vielfach aus weniger angemessenen Gründen eingesetzt.

Scraping befindet sich im Wandel – das bleibt den Kunden nicht verborgen

Leider hören wir oft von Verbrauchern, die Opfer von Phishing-Betrug werden. In solchen Fällen wurden möglicherweise Scraper-Bots verwendet, um Produktbilder, Beschreibungen und Preisinformationen zu erfassen und damit gefälschte Storefronts oder Phishing-Websites zu erstellen, die darauf abzielen, Anmeldedaten oder Kreditkarteninformationen zu stehlen. Bei diesen gefälschten (Phishing-)Websites handelt es sich um eine Form der Markenimitation, bei der das geistige Eigentum des betroffenen Unternehmens genutzt wird, um Vertrauen gegenüber potenziellen Kunden aufzubauen.

Einige der größten E-Commerce-Marken waren bereits von gefälschten Websites, Phishing-Kampagnen und dem Diebstahl von Webdaten der Unternehmen im Rahmen von Markenimitationskampagnen betroffen (Abbildung 3). Wenn Phishing-Websites Erfolg haben, sind Folgen für die rechtmäßigen Markeninhaber der Verlust von Kundenvertrauen und Kundenbindung.

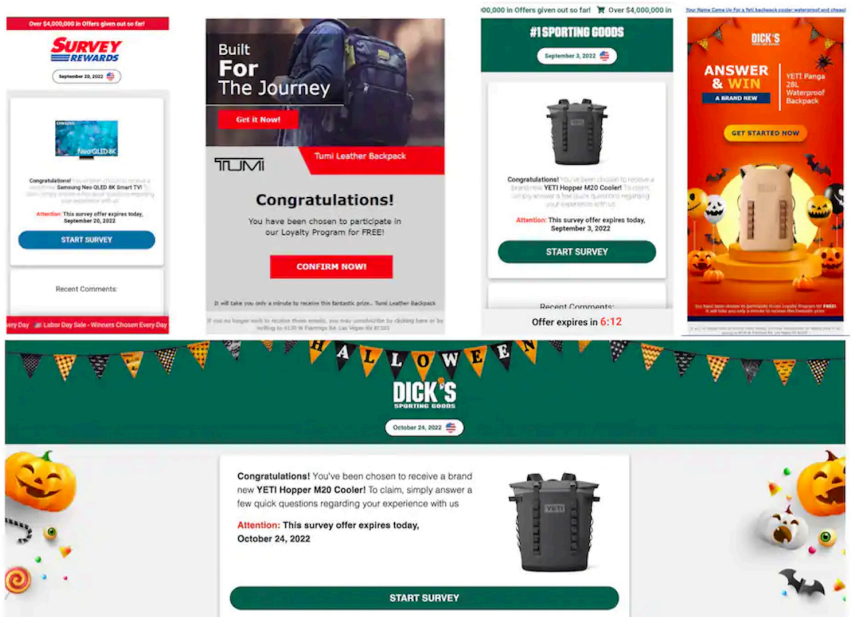

Abb. 3: Beispiel für große E-Commerce-Unternehmen, die Markenimitationen zum Opfer gefallen sind

Auch Scalping ist eine Form des Web-Scraping, da Scalper eine Website nach verfügbaren Produkten mittels Scraper-Bots durchsuchen und diese kaufen bevor legitime Kunden die Möglichkeit dazu haben (Abbildung 4).

Abb. 4: Anwendungsfälle für Scraping

Akteure, die derartige schadhafte Scraping-Aktivitäten durchführen, wissen genau, wie sich ihre boshaften Absichten auf ihre Opfer auswirken. Dazu gehören die negativen Auswirkungen von Wettbewerbsinformationen/Spionage, Inventory Hoarding/Scraping, Fälschungen und Nachahmen von Websites/Produkten sowie Scraping und Weiterverbreitung der Inhalte von Medienseiten (Tabelle 1). Der Einsatz von Scraper-Bots ist nicht ausdrücklich durch bestehende Gesetze verboten.

Auswirkungen	Beschreibung
Informationsgewinnung/ Spionage durch Mitbewerber	Wettbewerber nutzen Informationen von der Website eines Unternehmens, um Preise zu unterbieten, ihr eigenes Angebot zu ändern und Einblicke für neue Chancen und Bedrohungen zu gewinnen.
Inventory Hoarding/Scraping	Scalper pingen stetig gezielt Websites an, um auf die Verfügbarkeit von Produkten zu prüfen und diese zu Einkaufswagen hinzuzufügen, sodass sie echten Kunden nicht zur Verfügung stehen.
Fälschungen und betrügerische Websites/Waren	Inhalte können von Betrügern mittels Scraping erfasst und genutzt werden, um gefälschte Websites und Produktkataloge zu erstellen und Nutzer auf diese Weise im Glauben zu wiegen, dass sie legitime Produkte und keine Fälschungen kaufen.
Scraping und Weiterverbreitung der Inhalte von Medienseiten	Angreifer können Nachrichtenartikel, Blogs und andere Inhalte durch Scraping erfassen und auf ihren eigenen Websites platzieren. Dadurch verliert das Unternehmen, das die Inhalte ursprünglich erstellt hat, Besucher und potenzielle Werbeeinnahmen. Werbeeinnahmen basieren häufig auf der Anzahl der Besucher/dem Publikumsverkehr auf Websites. Geringere Besucherzahlen bedeuten deshalb, dass Medienseiten Umsatz verlieren, den sie durch höhere Werbeeinahmen erzielt hätten.

Tabelle 1: Verwendung von Web-Scrapern mit beabsichtigt negativer Auswirkung

Die allgemeinen Gefahren von Web-Scraping

Unabhängig von der Absicht des Web-Scraping müssen Unternehmen die Ausgaben handhaben, die durch unerwünschte Auswirkungen von Scraping entstehen. Einige Unternehmen zahlen für erwünschte Scraping-Dienste, jedoch erzeugen diese bei den Ziel-Unternehmen des Scraping wiederum eigene Kosten. Dazu gehören Kosten für Lösungen zur Bot-Abwehr sowie negative wirtschaftliche Auswirkungen durch die Verschlechterung der Website-Performance und die Verwässerung wichtiger Metriken (Tabelle 2).

Auswirkungen	Beschreibung
Bot-Traffic erhöht die Kosten für Server, CDN und Cloudservices	Wenn Inhalte von Wettbewerbern, Angreifern und Fälschern missbraucht werden, wirkt sich dies auf den Umsatz aus und schadet der Reputation.
Verschlechterung der Websiteperformance	Scraper-Bots laufen kontinuierlich, bis sie gestoppt werden. Für Unternehmen, die unerwünschten Bot-Traffic bedienen, erhöhen sich somit die Server- und Bereitstellungskosten und das Nutzererlebnis wird aufgrund der langsameren Website- und App-Performance beeinträchtigt.
Ungenaue Kennzahlen	Unerkannte Bot-Aktivitäten verzerren wichtige Kennzahlen wie die Website-Konversion, auf die sich Geschäftsteams verlassen, um Investitionsentscheidungen hinsichtlich Produktstrategie und Marketingkampagnen zu treffen.

Tabelle 2: Verwendung von Web-Scrapern mit unbeabsichtigt negativer Auswirkung

Scraping auf Auftragsbasis: Web-Scraping-Dienste von Drittanbietern

Wie bereits erwähnt, können Web-Scraper-Bots mit guter oder schlechter Absicht eingesetzt werden. Neben Bots, die für Credential-Stuffing-Angriffe verwendet werden, bei denen es sich bekanntermaßen um schädliche Bots handelt, die zu Recht blockiert werden, werden auch legitime Web-Scraping-Bots durch Unternehmen angeboten. Viele Unternehmen nutzen diese Web-Scraping-Dienste von Drittanbietern, um Daten zu extrahieren und intern bereitzustellen, was insbesondere in der Welt des wettbewerbsorientierten Marketings von Vorteil sein kann.

Die verschiedenen Arten von Web-Scraping-/Datenextraktionsdiensten werden durch zahlreiche Unternehmen angeboten und sogar im Rahmen von Konferenzen beworben. Bright Data veranstaltet zum Beispiel eine Konferenz namens ScrapeCon, auf der Experten für die Umgehung von Bot-Erkennungen ihr Know-how zum Scraping von Daten anpreisen. Tabelle 3 enthält Beispiele für die Serviceniveaus, die von Web-Scraping-Drittanbietern erbracht werden.

Serviceniveau 1	Proxy-Dienste können Teil des Scraping sein und eine Infrastruktur bieten, die mobile und private IP-Adressen sowie IP-Adressen von Rechenzentren umfasst.
Serviceniveau 2	Dieses zweite Niveau kann eine automatisierte Datenextraktion für das Bereinigen und Strukturieren der Daten umfassen, damit diese durch die Data Science-Teams des Kunden leichter verwendet werden können, um wertvolle Informationen für die Entscheidungsfindung des Unternehmens zu gewinnen.
Serviceniveau 3	Auf dem höchsten Niveau kann zusätzlich die direkte Extraktion von Geschäftsinformationen bereitgestellt werden, wodurch sich Entscheidungsprozesse in Unternehmen weiter verbessern lassen. Diese werden als „KI-Botnets“ bezeichnet.

Tabelle 3: Serviceniveaus, die von Drittanbietern für Web-Scraping angeboten werden

Kunden können eines dieser Serviceniveaus wählen, von grundlegenden bis hin zu erweiterten Diensten, und die Häufigkeit der Datenerfassung und ihre Ziele angeben. Oft hängt das Niveau des angebotenen Service oder des gewählten Botnets davon ab, welches Schutzniveau überwunden werden muss. Einfache Botnets können Daten über ein fortschrittliches Skript erfassen, wobei mehrere Tausend Proxyserver in Rechenzentren genutzt werden, um die Trafficlast auszugleichen. Wenn das Schutzniveau zu gering ist, könnte das Botnet diese Technik verwenden, um die Mechanismen zur Bot-Abwehr und die Web Application Firewall der Sicherheitsinfrastruktur zu überwinden.

Bei fortgeschritteneren Schutzmechanismen, kann jedoch ein ausgeklügelter Ansatz für das Scraping erforderlich werden, wie z. B. ein Angriff durch Browser ohne Header. Dies gilt unabhängig davon, ob das Scraping von einem Akteur mit guter oder schlechter Absicht durchgeführt wird. Dies ist mit erheblichen Kosten verbunden, da Unternehmen für die erweiterte Infrastruktur im Allgemeinen deutlich höhere Kosten entstehen als für das grundlegende Serviceniveau. Eine fortschrittliche Verteidigung kann Challenge-Technologien umfassen (z. B. CAPTCHA oder Proof of Work), mehrere Erkennungsebenen zur clientseitigen Fingerprint-Beurteilung sowie eine Analyse der Merkmale des Hypertext Transfer Protocol (HTTP) und der Transport Layer Security (TLS).

Scraping-Prozess für KI-Botnets

Einfache Web-Scraper weisen im Allgemeinen konsistente Scraping-Techniken auf. Dagegen können KI-Botnets unstrukturierte Daten und Inhalte mit weniger einheitlichen Formaten und Speicherorten erkennen und scrapen. Darüber hinaus können KI-Botnets mithilfe von Geschäftsinformationen die Entscheidungsfindung optimieren. Die erweiterten KI-Botnets, die in Tabelle 3 unter Serviceniveau 3 aufgeführt sind, nutzen für das Scraping von Daten einen dreistufigen Prozess, der das Erfassen, Extrahieren und Verarbeiten der Daten umfasst (Abbildung 5).

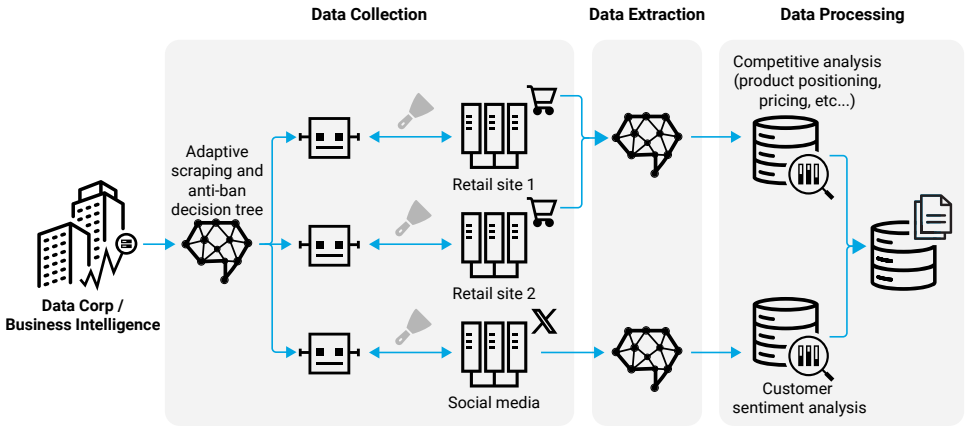

Abb. 5: KI-Botnet mit dreistufigem Prozess

Im Folgenden möchten wir diese drei Schritte und ihre Bedeutung eingehender untersuchen.

Datenerfassung

Web-Scraping umfasst das Organisieren von Daten, die von einer oder mehreren Websites extrahiert wurden, damit Unternehmen neue Datensätze erstellen und diese nach eigenem Ermessen verwenden und analysieren können. Am Anfang steht das Erfassen der Daten.

Damit sie schnell und reibungslos funktioniert muss bei der Datenerfassung adaptives Scraping verwendet werden, in Kombination mit Technologien, die das Sperren oder Erkennen von Bots verhindern. Diese Technologien werden mit einer Entscheidungshierarchie eingerichtet und erkennen dadurch verschiedene Aspekte etwaiger Schutzmaßnahmen. Hierbei nimmt die Resilienz eine zentrale Rolle ein. Der Schutz vor Bots kann verschiedene Mechanismen umfassen, etwa JavaScript-, HTTP- und TLS-Fingerprinting (Bewertung der HTTP-Header und TLS-Handshakes) und Internet Protocol (IP) Reputation Detection (Abbildung 6). Diese Workflows verwenden zum Teil maschinelles Lernen (ML), insbesondere beim Erfassen von Statistiken zur Erfolgsrate, beim Anpassen an die Cookie-Strategie, den HTTP-Header und die TLS-Parameter und beim Auswerten des JavaScript-Fingerabdruckcodes. Hierbei können auch Headless-Browser-Technologien eine Rolle spielen.

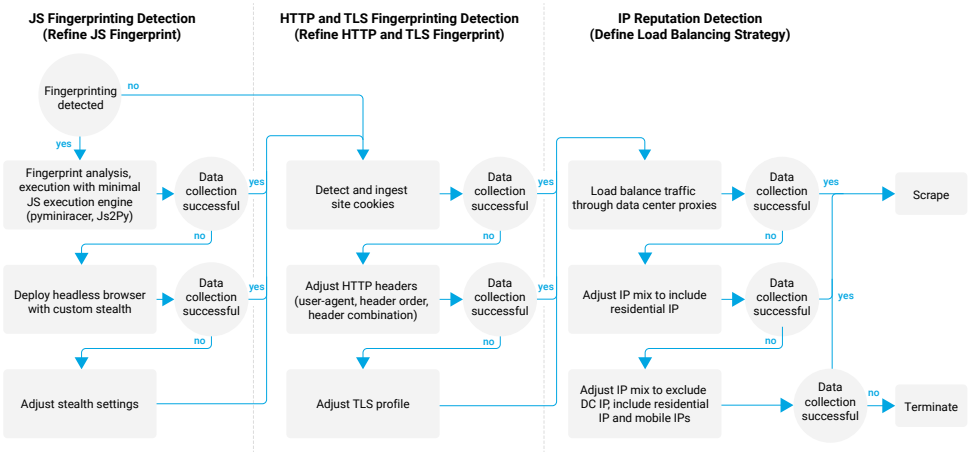

Abb. 6: Im Rahmen der versuchten Datenerfassung zielt diese Entscheidungshierarchie zur Anti-Bot-Erkennung auf das Umgehen von JavaScript-Fingerprinting, HTTP- und TLS-Fingerprinting und IP Reputation Detection ab.

Browser ohne Header

Ein Browser ohne Header ist ein Webbrowser, der ohne grafische Nutzeroberfläche (GUI) auskommt. Menschliche Nutzer können also nicht direkt mit der Webseite interagieren, auf der der Browser ohne Header angezeigt wird. Der Browser wird stattdessen über eine Befehlszeilenschnittstelle (CLI) oder eine Netzwerkkommunikation ausgeführt. Der Open-Source-Browser Selenium, ein häufig für Web-Scraping verwendeter Browser ohne Header, wird automatisiert eingesetzt. Dies kann für Datensuchende sehr hilfreich sein, die dynamische Inhalte durch Scraping erfassen möchten.

Browser ohne Header können auch für Screenshots und das effiziente Kopieren von Website-Code genutzt werden, wobei sich die ausgewählten Daten ohne Rendering der gesamten Seite extrahieren lassen. Angriffe durch Browser ohne Header sind jedoch kostspielig in der Durchführung und können teilweise durch hinterlassene Fingerprints erkannt werden. Ausgaben für andere ausgereifte Infrastrukturen fallen jedoch ähnlich wie bei Technologien mit Browsern ohne Header aus, sind also vergleichbar hoch.

Datenextraktion und Datenverarbeitung

Die extrahierten Informationen bestehen im Allgemeinen aus HTML- und JSON-Inhalten. Von der Gesamtheit der extrahierten Daten ist möglicherweise nur ein Bruchteil hilfreich für die Analyse. Beispielsweise umfassen Wettbewerbsanalysen in der Regel Preise, Rabatte, Lagerbestände sowie die SKU-Nummern, Kategorien und Beschreibungen von Produkten. Wichtige Informationen lassen sich automatisch durch ML-Modelle extrahieren, die auf Basis mehrerer Strukturen und Datenformate trainiert werden, die Informationen zu erkennen. Dadurch wird der zusätzliche Aufwand für die manuelle Extraktion der Daten vermieden und die Untersuchung der Code-Struktur von HTML- und JSON-Inhalten wird überflüssig. Darüber hinaus kann sich die Code-Struktur der Inhalte mit der Weiterentwicklung des Designs der Website verändern. ML-Logik ist auch dann für die Verarbeitung erforderlich, wenn der Analyseumfang mehrere Websites umfasst.

Fallstudie: Vorteile von Lösungen zur Erkennung von Web-Scraping

Forscher von Akamai untersuchten die Infrastruktur mehrerer E-Commerce-Kunden, die durch eine Web-Scraping-Lösung zur Erkennung von Scraping-Aktivitäten geschützt waren, und analysierten für den Zeitraum von einer Woche die Aktivität des Traffics. Die Stichprobe umfasste etwa 6,9 Milliarden Anfragen. Bei der Analyse wurden nur HTML- und AJAX-Anfragen berücksichtigt. Statische Inhalte (Bilder, JavaScript, Stylesheets) wurden nicht in die Analyse einbezogen, da die meisten Bots keine statischen Inhalte anfordern und dadurch eine unnötige Zunahme des Datenumfangs vermieden wurde.

Die Gesamtaktivität, die mithilfe von Akamai Content Protector klassifiziert wurde, bestand zu 49 % aus risikoarmem menschlichem Traffic, zu 42 % aus Bot-Traffic (27,5 % risikoreiche schlechte Bots und 14,6 % gute Bots) und zu 8,7 % aus nicht klassifiziertem Traffic mit mittlerem Risiko (Abbildung 7).

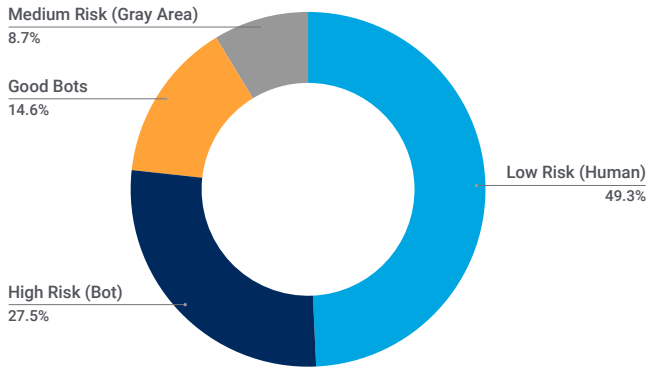

Abb. 7: Aufschlüsselung zur Klassifizierung der Trafficaktitivät

Abbildung 8 zeigt, dass von den 42,1 % des Bot-Traffics 65,3 % von Scrapern stammten, die als schlechte Bots gelten, und die restlichen 34,7 % von Scrapern, die als gute Bots eingestuft wurden (z. B. Websuchmaschinen, SEO, soziale Medien und Online-Werbung).

Abb. 8: Traffic von guten bzw. schlechten Bots

Das Maß der technischen Ausreifung wurde auch anhand der risikoreichen schlechten Bots, die mit 65 % zum gesamten Bot-Traffic beitrugen, ermittelt. 37 % dieses Traffics stammten von grundlegenden skriptbasierten Botnets, die durch einfache zustandslose Methoden leicht zu erkennen sind. 47 % stammten von fortschrittlicheren skriptbasierten Botnets, die fortschrittlichere zustandsbehaftete Erkennungsmethoden mit ML erfordern und 15,5 % stammten von Browsern ohne Header, die erweitertes JavaScript-Fingerprinting und zustandsbehaftete Erkennungsmethoden erfordern (Abbildung 9).

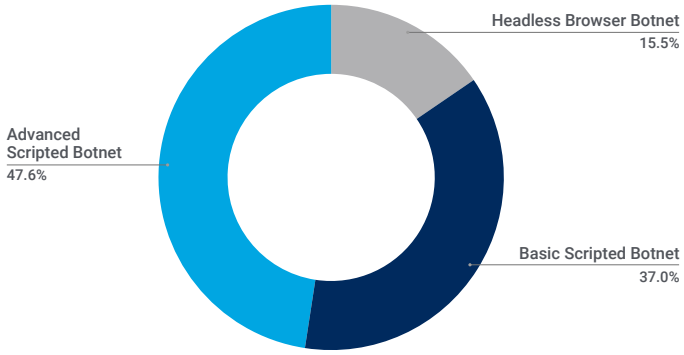

Abb. 9: Verteilung des Traffics von schlechten Bots basierend auf deren Ausreifung (die Summe ergibt aufgrund der Rundung keine 100 %)

Anhand dieser Daten ist ersichtlich, dass schlechte Bots deutlich zahlreicher vorkommen als gute Bots, und dass knapp die Hälfte des gesamten Traffics auf Bots zurückzuführen war, wobei erweiterte Scripting-Botnets das höchste Maß an schlechtem Bot-Traffic generierten (47 %).

Website-Aktivitäten werden deutlich schneller und effizienter ausgeführt und die Website-Metriken sind besser zu interpretieren, sobald Abwehrmechanismen gegen diese Bots vorhanden sind und Scraper entfernt werden. Diese Ergebnisse führen wiederum zu besseren Nutzer-/Kundenerlebnissen. Wie in Abbildung 10 dargestellt, nahm die Anzahl der risikoreichen Bot-Anfragen nach Aktivieren der Abwehrmaßnahmen erheblich ab.

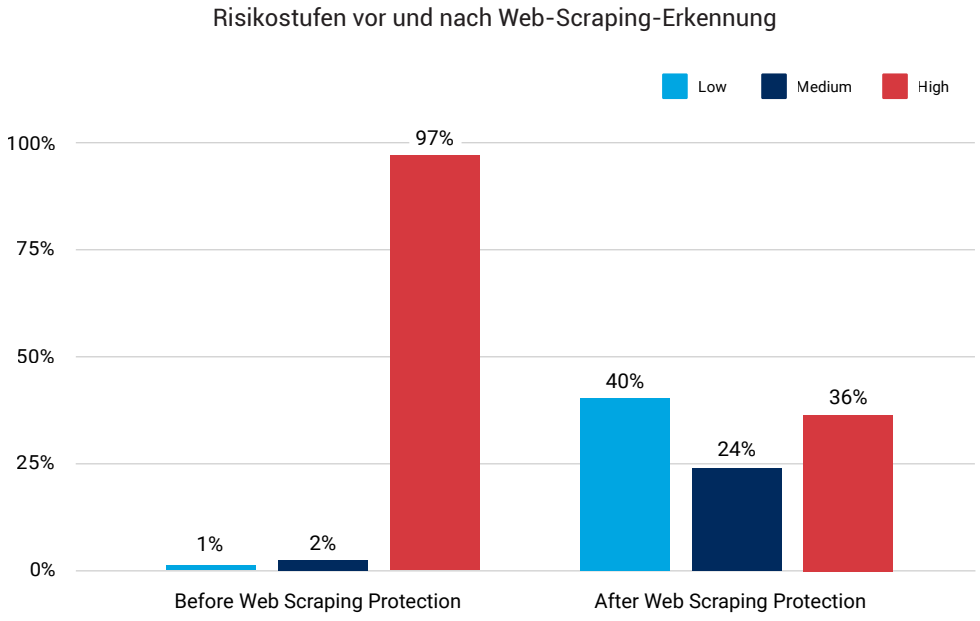

Abb. 10: Risikostufen vor und nach der Abwehr durch Content Protector

Schutz und Schadensbegrenzung

Dieser Abschnitt behandelt wichtige Indikatoren für die Erkennung von Web-Scrapern und Informationen zu Tools, die Abwehrmaßnahmen bieten.

Erkennung einfacher Scraper

Obwohl fortschrittliche Scraper schwer zu erkennen sind, können Bot-Management-Lösungen vor der Erfassung von Daten durch angreifende Scraper aller Art schützen. Diese Lösungen untersuchen insbesondere die folgenden Eigenschaften, um einfachere Web-Scraper-Bots zu erkennen:

- Anforderungen mit älteren Browser- und Betriebssystemversionen
- Anomalien in der HTTP-Header-Signatur
- Die Verwendung alter HTTP-Versionen (z. B. v1.1) anstelle des gebräuchlicheren HTTP v2 oder des zunehmend auftretenden HTTP v3
- Anfragen, die von Tausenden von Cloudservices/Rechenzentren stammen

Erkennung fortschrittlicherer Scraper

Keines der oben aufgeführten Merkmale ist bei den fortschrittlicheren Scrapern zu erkennen. Hier also einige Merkmale fortschrittlicher Scraper:

- Anfragen stammen von den neuesten Browser- und Betriebssystemversionen
- Der HTTP-Header gleicht dem legitimer Browser
- Verwendung von HTTP v2
- Anfragen stammen von Hunderttausenden privaten und mobilen IP-Adressen

Identifizieren von Trafficmustern

Anhand einiger Schlüsselindikatoren kann ermittelt werden, ob der Traffic einer Website menschlich ist (Abbildung 11) oder von einfachen (Abbildung 12) oder fortschrittlichen Bots (Abbildung 13) stammt.

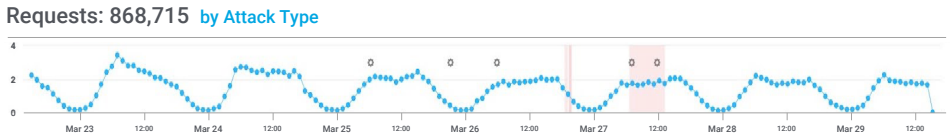

Abb. 11: Traffic legitimer Nutzer weist im Allgemeinen einen am Tagesablauf orientierten Aktivitätszyklus auf

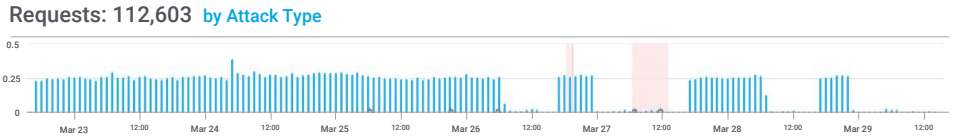

Abb. 12: Typischer Bot-Traffic weist regelmäßige Aktivitäten mit gelegentlichen Pausen auf

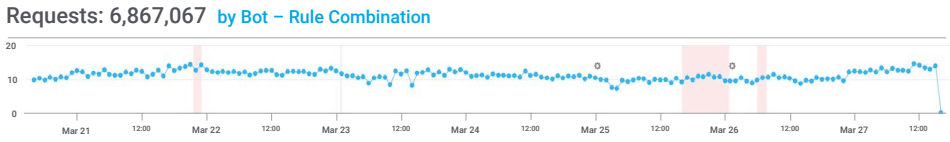

Abb. 13: Fortschrittliche Bots weisen kontinuierlichen Traffic am Tag und in der Nacht auf

Häufig beobachten wir auch Botnets, die in der Mitte des Spektrums einzugliedern sind, da sie eine schwache Lastausgleichstrategie, aber eine ausgeklügelte Fingerprinting-Strategie aufweisen (oder umgekehrt). Fortschrittlichere Botnets können jedoch so ausgereift sein, dass sie perfekte Fingerprints aufweisen oder sogar das Trafficmuster legitimer Nutzer reproduzieren.

Akamai

Tools zum Schutz vor Web-Scraping wie Content Protector erkennen derartige Scraper-Bots und bieten zusätzliche Vorteile in einer Online-Welt, das zunehmend durch Scraping belastet ist. Vorteile:

- Höhere Konversionsraten und geringere IT-Kosten
- Präzise Kennzahlen, die bessere Investitionsentscheidungen fördern und den Umsatz steigern können
- Verringerter Preisdruck und dadurch weniger Umsatzverluste durch Mitbewerber, die Preise unterbieten
- Steigern der Zufriedenheit von Kunden, da diese auf die gewünschten Artikel zugreifen können, sowie höherer Upselling-Umsatz, wenn Kunden zusätzliche Produkte zum Einkaufswagen hinzufügen, nachdem sie sich begehrte Artikel gesichert haben
- Der Ruf der Marke bleibt erhalten, da Kunden vor Fälschungen schlechter Qualität geschützt sind, die sie für legitime Waren des ursprünglichen Verkäufers halten
- Gewinnung von Produktumsätzen und Aufrechterhaltung der Kundenbindung
- Schutz/Steigerung der Werbeeinnahmen
- Bindung der Zielgruppen und Websitebesucher

Compliance-Aspekte

Der Payment Card Industry Data Security Standard (PCI DSS) v4.0 ist jetzt in Kraft. Viele der Änderungen wurden durch Bedrohungstrends ausgelöst, von denen Unternehmen nach wie vor betroffen sind. Sichtbarkeit ist der Schlüssel zur Bekämpfung dieser Angriffe. Diese Angriffe müssen unbedingt schnell erkannt und abgewehrt werden, unabhängig davon, ob sie sich in Ihrer historischen JavaScript-Umgebung oder in APIs befinden, die zur Erleichterung der Transformation verwendet werden.

Neue Compliance-Trends beobachten wir auch im neuen NIST Cybersecurity Framework Version 2.0, mit dem eine Governance-Funktion eingeführt wurde. Das NIST dient als Grundlage diverser behördlicher Vorschriften und prägt dadurch zahlreiche Frameworks für Cybersicherheit. Aktuell bietet es sich also an, die neue Vorschrift zu lesen und auf deren Grundlage entweder die eigenen Richtlinien zu überarbeiten oder die aktuelle Dokumentation zu überprüfen, um etwaige Schwachstellen aufzudecken.

Für börsennotierte Unternehmen und solche, die die allgemein anerkannten Rechnungslegungsgrundsätze (GAAP) verwenden, bilden materielle Aspekte der Cybersicherheit einen weiteren Bereich, den es zu berücksichtigen gilt. Die Notwendigkeit, materielle Risiken und Bedrohungen zu definieren, erfordert eine Zusammenarbeit im gesamten Führungsteam. Sobald Sie materielle Bedrohungen (wie Ransomware) identifiziert haben, müssen Sie Abwehrmaßnahmen (wie Mikrosegmentierung) einleiten. Stellen Sie sicher, dass Ihre Krisenmanagementpläne Fristen zur Offenlegung berücksichtigen und dass Sie eine Strategie für das Worst-Case-Szenario haben, in dem Sie das Formular „Cyber Incident Form 8-K“ der Security and Exchange Commission vorlegen müssten.

Fazit

Wir hoffen, dass Ihnen dieser Bericht Einblicke in einen Bereich bieten konnte, der sich wirtschaftlich negativ auf Ihr Unternehmen auswirken könnte. Bots beeinflussen Websites in immer größerem Umfang. Es ist wichtig, nützliche Bots zu optimieren, schädliche Bots zu bekämpfen und ein insgesamt reibungsloses Kundenerlebnis zu gewährleisten. Hierbei handelt es sich um ein Sicherheitsproblem mit geschäftsrelevanten Auswirkungen. Wie bei allen Sicherheitsproblemen besteht der erste Schritt darin, Transparenz zu gewinnen, der zweite in der Analyse der Auswirkungen und der letzte Schritt in der Ermittlung des ROI für Risiko und Umsatz, damit sich geeignete Sicherheitskontrollen implementieren lassen.

Sie können nicht schützen, was im Verborgenen liegt. Daher ist es jetzt an der Zeit, Lücken in der Sichtbarkeit aufzudecken. Dazu müssen Sie den Umfang der Web-Scraping-Aktivitäten auf Ihren Websites ermitteln und herausfinden, welcher Absicht dahintersteckt. Die Bot-Landschaft besteht sowohl aus guten als auch aus schlechten Bots. Scraper-Bots lassen sich je nach ihrem Zweck einer der beiden Kategorien zuordnen. Obwohl die Grenze zwischen nützlichen und schädlichen Scraper-Bots unscharf verlaufen kann, steht fest: Bots werden immer raffinierter (z. B. Web-Scraper, die Angriffe durch Browser ohne Header durchführen). Gerade für E-Commerce-Unternehmen verursachen Web-Scraper-Bots erhebliche Auswirkungen auf die IT-Kosten und das Kundenerlebnis. Tools, um Bot-Aktivitäten und deren Auswirkungen auf Ihre Website zu analysieren, sind da unerlässlich.

Zu vermeiden sind Angreifer, die über Ihre Websites kriminelle Geschäftsmodelle und böswillige Aktivitäten durchführen, wie das Einlösen von Treuepunkten, das betrügerische Aufgeben von Bestellungen oder sogar Retourenbetrug. Außerdem sollten Sie verhindern, dass Ticketbots begrenzte Events bis zum Ausverkauf der Tickets buchen oder Shopping-Bots begehrte Produkte erwerben. Bots können zudem verwendet werden, um betrügerisch neue Konten zu erstellen, indem sie Aktionsangebote ausnutzen, was sich auf Kampagnenanalysen und Kosten auswirken kann. Große DDoS-Botnets (Distributed Denial-of-Service) können Web-Anwendungen überlasten und dadurch zu einem schlechten Nutzererlebnis führen oder verhindern, das Bestellungen aufgegeben oder Reservierungen vorgenommen werden können, was wiederum zu Umsatzverlusten und Kundenproblemen führt. Bots können sogar menschliches Verhalten im Internet nachahmen, um Klicks und Traffic auf Websites zu erhöhen, wodurch sowohl das Marketing als auch die Performance Analytics sorgfältig gestalteter digitaler Erlebnisse verzerrt werden. All dies sind keine wünschenswerten Szenarien.

Wie bereits erwähnt, besteht schon heute mehr als die Hälfte des globalen Webtraffics im Handel aus Bots, und Anteil von Bots steigt stetig an. Die Erkenntnisse und Ratschläge in diesem Bericht stützen sich auf die Akamai-Sicherheitsplattform, die Content Protection und Schutz vor Web-Scraping umfasst. Aufbauend auf unserer Zusammenarbeit mit zahlreichen führenden Unternehmen im E-Commerce haben wir hier Sicherheitsvorkehrungen und Abwehrmechanismen besprochen, die Unternehmen zum bestmöglichen Schutz ihrer Kunden einsetzen können. Wir erwarten eine Zunahme des Einsatzes, der unterschiedlichen Serviceniveaus und der Arten verfügbarer Web-Scraper-Bots. Daher sollten Sie kontinuierlich die Risikolage Ihres Unternehmens bewerten und überprüfen, ob Ihre aktuellen Sicherheitskontrollen der Risikobereitschaft Ihrer Führungskräfte entsprechen.

Bleiben Sie auf dem Laufenden über unsere neuesten Forschungsergebnisse – besuchen Sie unseren Security Research Hub.

Methodik

Daten von Content Protector

Dieses Datenbeispiel beschreibt die Klassifizierungen der Risikostufen, die dem überwachten Traffic durch unseren Content Protector zugewiesen werden. Diese Klassifizierungen werden verwendet, um Bot-Scraping-Aktivitäten zu erkennen und zu bestimmen, ob es sich um gute oder schlechte Bots handelt. Da die meisten Bots keinen statischen Inhalt anfordern, wurden bei dieser Analyse nur HTML- und AJAX-Anforderungen berücksichtigt, um eine unnötige Zunahme der Datenmenge zu vermeiden.

Diese Datenstichprobe bezieht sich auf den einwöchigen Zeitraum vom 12. April bis zum 19. April 2024. Die Stichprobe besteht aus mehr als 6,5 Milliarden Anfragen.

Bot-Angriffe

Diese Daten beschreiben Warnungen auf Anwendungsebene über Traffic, der unsere Web Application Firewall (WAF) und unser Bot-Management-Tool durchläuft. Die Warnungen werden ausgelöst, wenn wir innerhalb einer Anfrage an eine geschützte Website, Anwendung oder API eine schädliche Bot-Payload erkennen. Diese Bot-Warnungen können sowohl von schädlichen als auch von gutartigen Bots ausgelöst werden. Die Warnungen zeigen nicht an, ob ein Angriff erfolgreich war. Obwohl diese Produkte ein hohes Maß an Anpassung ermöglichen, haben wir die hier dargestellten Daten auf eine Weise erfasst, bei der keine nutzerdefinierten Konfigurationen der geschützten Ressourcen berücksichtigt werden. Die Daten stammen aus einem internen Tool zur Analyse von Sicherheitsereignissen, die in der Akamai Connected Cloud erkannt wurden, einem Netzwerk aus ca. 340.000 Servern an mehr als 4.000 Standorten in fast 1.300 Netzwerken in über 130 Ländern. Diese Daten werden in Petabyte pro Monat gemessen und von unserem Sicherheitsteam verwendet, um Angriffe zu untersuchen, schädliches Verhalten aufzudecken und zusätzliche Informationen in die Lösungen von Akamai einzuspeisen.

Diese Daten decken den Zeitraum von 15 Monaten vom 1. Januar 2023 bis zum 31. März 2024 ab.

Mitwirkende

Chefredakteur

Lance Rhodes

Redaktion und Text

David Senecal
Maria Vlasak

Prüfung und Fachleute

Mitch Mayne
Susan McReynolds
Christine Ross
Badette Tribbey
Steve Winterfeld

Datenanalyse

Chelsea Tuttle

Werbematerialien

Annie Brunholzl

Marketing und Veröffentlichung

Georgina Morales
Emily Spinks

Weitere „State of the Internet"-Sicherheitsberichte

Lesen Sie vorherige Ausgaben und informieren Sie sich über bevorstehende Veröffentlichungen der renommierten „State of the Internet"-Sicherheitsberichte von Akamai. **akamai.com/soti**

Weitere Informationen zur Bedrohungsforschung von Akamai

Halten Sie sich unter diesem Link zu neuesten Threat-Intelligence-Analysen, Sicherheitsberichten und Cybersicherheitsforschung auf dem Laufenden: **akamai.com/security-research**

Greifen Sie auf Daten aus diesem Bericht zu

Sehen Sie sich die hochauflösenden Versionen der Diagramme und Grafiken an, auf die in diesem Bericht verwiesen wird. Diese Bilder können kostenlos verwendet und referenziert werden, vorausgesetzt, Akamai wird ordnungsgemäß als Quelle genannt und das Akamai-Logo wird beibehalten.
akamai.com/sotidata

Weitere Informationen zu Akamai-Lösungen

Weitere Informationen zu Lösungen von Akamai für das Erkennen und zum Schutz vor Web-Scrapern finden Sie auf unserer **Content Protector Seite**.