SOTI

10 YEARS
OF SECURITY INSIGHT

V10, NUMERO 03

Nuovi modi per erodere la redditività:

l'impatto dei web scraper sull'e-commerce

Akamai

Stato di Internet - Security

Sommario

Sapevate che i bot generano più della metà di tutto il traffico web? Il commercio, in particolare, che si basa su risorse e applicazioni web remunerative, è il settore maggiormente colpito dal traffico dei bot ad alto rischio (Figura 1). Inoltre, anche se spesso sentiamo che i bot si evolvono, sono i i bot dei web scraper ad attirare oggi l'attenzione delle organizzazioni basate sull'e-commerce perché il loro impatto economico, spesso nascosto sotto la superficie, differisce da quello di altri tipi di bot. Il rilevamento dei bot scraper è diventato molto più complesso a causa dell'aumento delle botnet basate sull'intelligenza artificiale (AI) e delle tecnologie basate sui browser headless, che li rendono estremamente elusivi. Uno dei clienti di Akamai che operano nel settore dell'e-commerce, ad esempio, si è visto bloccare il 99% del traffico ad alto rischio senza nemmeno sapere che proveniva dai bot scraper.

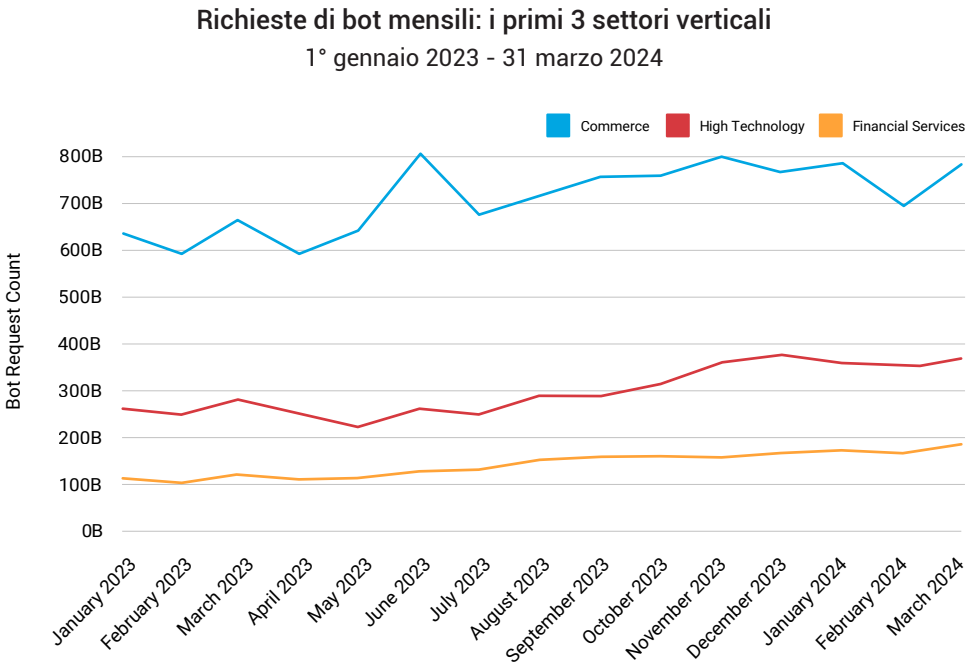


Figura 1. Il commercio è il primo settore per numero di richieste dei bot, in cui abbiamo assistito a un aumento del traffico dei bot globale dall'inizio del 2023 al 1° trimestre del 2024

Pertanto, in questo rapporto sullo stato di Internet (SOTI), ci siamo focalizzati sull'evoluzione e sulla specializzazione di questi bot e dei loro autori. Anche se i bot sono in circolazione da qualche tempo, osserviamo che vengono ancora utilizzati in vari gruppi allo scopo di sferrare attacchi criminali o applicare sistemi fraudolenti, nonché a fini di competitive intelligence. Di recente, abbiamo osservato una tendenza verso un incremento nell'uso di tutti i bot e un aumento dell'impatto negativo esercitato dai bot scraper sulle aziende. Questo rapporto è stato concepito allo scopo di condividere le informazioni tecniche e le metodologie di attacco per aumentare la consapevolezza nei confronti di un problema che colpisce sempre di più il settore del commercio.

I bot: il buono, il brutto e il cattivo

Ogni organizzazione basata sull'e-commerce subisce la presenza dei bot che si evolvono in continuazione, diventando sempre più specializzati a seconda dell'obiettivo preso di mira. Nel settore del commercio, esiste un'ampia varietà di tipi di bot che eseguono molte attività diverse. Possiamo suddividere semplicemente i bot in tre gruppi: bot legittimi, bot dannosi e bot "neutrali". I bot legittimi aiutano i clienti a trovare il vostro sito. I bot dannosi esfiltrano dati dal vostro sito per scopi illeciti. I bot neutrali tendono a dare fastidio anche se sono comunque legittimi, anzi fanno parte di una sottocategoria di bot legittimi (ad es., i bot dei partner che monitorano costantemente il vostro sito e le API di altri programmi che effettuano frequenti chiamate).

Quindi, se pensiamo a chatbot utili e a bot dei motori di ricerca che possono avere un impatto benefico, ad es. rispondere alle domande di base dei clienti e fornire contenuti dei siti web in grado di restituire risultati di ricerca più accurati, il nostro intento è ottimizzare questi tipi di bot, contenendo, al contempo, i costi dell'IT. Per i bot dannosi, come i bot del credential stuffing che cercano di ottenere un accesso non autorizzato agli account dei clienti per conquistarne il controllo, ci proponiamo di attuare misure preventive senza influire sulle customer experience complessive. Esiste un tipo di bot osservato di recente che sta creando particolari problemi perché riduce il fatturato, diminuisce la fidelizzazione e aumenta i costi: si tratta dei bot dei web scraper.

I bot scraper, una botnet utilizzata per estrarre direttamente dati e contenuti dai siti web presi di mira su Internet, sono esclusivi, ossia richiedono una specifica attenzione perché operano in modo diverso e i loro sistemi di rilevamento, insieme al loro impatto sulle aziende, variano rispetto a quelli di altri bot. I web scraper presentano, inoltre, molte sfaccettature nel senso che i loro casi di utilizzo variano a seconda del modo con cui le organizzazioni e gli operatori capitalizzano sulle informazioni raccolte da questi bot. Indipendentemente dall'obiettivo specifico, gli scraper causano perdite di ricavi, aumentano i costi dell'IT e peggiorano le customer experience complessive.

In questo rapporto SOTI, esaminiamo l'impatto dello scraping sull'e-commerce e il motivo per cui i proprietari delle aziende (che operano, ad es., nei settori del digitale, marketing, brand, finanza, gestione dei rischi e sicurezza) devono tutti interessarsi a fermare gli abusi degli scraper. Per comprendere meglio questo impatto, è fondamentale avere un quadro completo del motivo per cui i bot dei web scraper si sono evoluti, per cosa vengono utilizzati, come operano, qual è il loro impatto e cosa possono fare le organizzazioni dell'e-commerce per contrastarli.

I principali risultati emersi dal rapporto

Il web scraping non è solo una frode o un problema di sicurezza, ma anche una questione aziendale. I bot scraper hanno un effetto negativo su molti aspetti dell'organizzazione, tra cui, ad esempio, ricavi, vantaggio competitivo, identità del brand, customer experience, costi dell'infrastruttura ed experience digitali.

In un recente case study riportato in uno studio di Akamai, il 42,1% delle attività del traffico complessivo proveniva dai bot, di cui il 65,3% era stato generato da bot dannosi, mentre il 63,1% del traffico dei bot dannosi utilizzava tecniche avanzate.

La tecnologia dei browser headless ha cambiato lo scenario degli scraper, richiedendo un approccio di gestione più sofisticato rispetto ad altri tipi di mitigazione basati su JavaScript.

Le problematiche tecniche che le organizzazioni devono affrontare in seguito a uno scraping, fatto sia con buone che con cattive intenzioni, includono il deterioramento delle performance dei siti web, la manipolazione delle metriche del sito, gli attacchi con credenziali compromesse dai siti di phishing, l'aumento dei costi di elaborazione e molto altro ancora.

È importante osservare e comprendere i diversi modelli di traffico per identificare se un sito web è interessato da un traffico di bot provenienti da utenti oppure da bot semplici o sofisticati. Questi modelli possono essere di vario tipo (circadiani, intermittenti o continui).

Bot legittimi e bot dannosi

Iniziamo dalle nozioni fondamentali. Un bot (abbreviazione di "robot") è un programma informatico in grado di eseguire attività automatizzate più velocemente e con maggiore precisione di un essere umano. I vari ruoli e tipi di bot rientrano in due categorie principali: i bot legittimi e i bot dannosi (Figura 2). I bot neutrali sono una sottocategoria dei bot legittimi, tuttavia, per ora, li considereremo come parte dei bot legittimi per semplificare il nostro confronto.

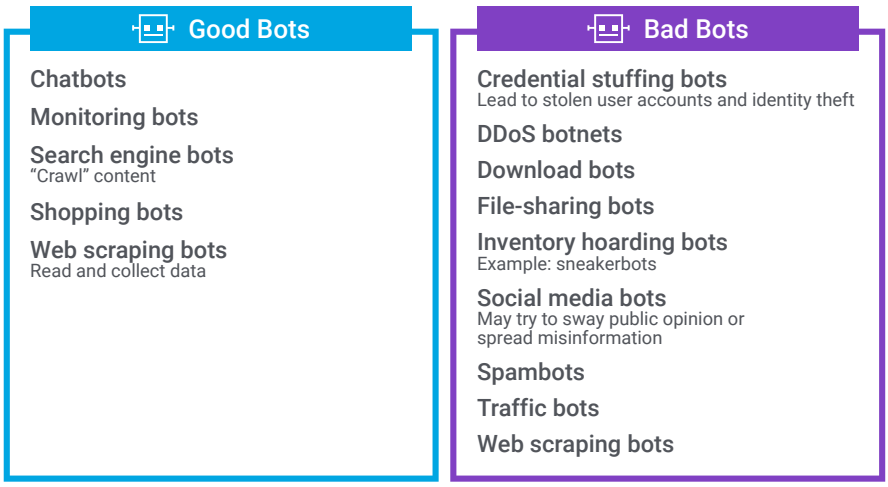


Figura 2. Un confronto esemplificativo tra bot legittimi e bot dannosi

I bot legittimi sono bot utili perché forniscono strumenti e servizi, mentre i bot dannosi sono spesso usati per scopi illeciti da parte di criminali informatici e truffatori. Un esempio di questo tipo di comportamento dannoso è un bot del traffico che imita il comportamento umano online per aumentare il numero di clic e il traffico su un sito web (ossia, commettere frodi pubblicitarie).

I bot di web scraping sono inseriti sia nelle categorie di bot legittimi che in quelle di bot dannosi, ma si distinguono per il modo con cui le organizzazioni utilizzano le informazioni da essi raccolte. Ci focalizzeremo ora più da vicino sui vari casi di utilizzo associati agli effetti positivi e negativi dei bot scraper che alcuni tra i principali retailer e brand di e-commerce al mondo si trovano ad affrontare.

Scraping 101

Il web scraping è comunemente usato dalle società di e-commerce. Ad esempio, nei settori dei viaggi e del turismo, gli aggregatori di viaggio esfiltrano contenuti dinamici dai propri partner di hotel e compagnie aeree per tenersi aggiornati sulle disponibilità e sui prezzi. Questo tipo di scraping è previsto e le aziende usano comuni controlli dei bot per limitare gli scraper in alcune ore della giornata in cui gli utenti reali cercano di effettuare le prenotazioni. Le organizzazioni usano anche provider di servizi di estrazione dei dati per raccogliere informazioni sui potenziali clienti e altri dati legati alla concorrenza. Inoltre, è possibile usare i bot di scraping per analizzare i dati e identificare le tendenze. Lo scraping può anche risultare utile per esaminare un sito al fine di migliorare le soluzioni e i servizi offerti online e per consentire ai potenziali consumatori di trovare più facilmente i prodotti dell'azienda tramite, ad esempio, un motore di ricerca. Tutte queste operazioni possono aiutare le aziende a raggiungere un vantaggio competitivo. Tuttavia, non si può negare che molte organizzazioni stiano usando gli scraper per motivi meno lodevoli.

Lo scraping cambia radicalmente e i clienti se ne accorgono

Spesso purtroppo sentiamo parlare di clienti che sono caduti vittime delle truffe di phishing. In tal caso, è probabile che i bot scraper siano stati usati per catturare immagini, descrizioni e informazioni sui prezzi dei prodotti per creare store o siti di phishing contraffatti volti a rubare credenziali o dati sulle carte di credito. Questi siti di phishing/contraffatti sono una forma di impersonificazione dei brand, una truffa in cui la proprietà intellettuale delle organizzazioni prese di mira viene utilizzata per fidelizzare potenziali clienti.

Alcuni dei principali brand di e-commerce al mondo sono stati colpiti da siti contraffatti, campagne di phishing e dal furto di dati aziendali come parte delle campagne di impersonificazione dei brand (Figura 3). Sfortunatamente, i siti di phishing riescono nel loro intento e i brand legittimi subiscono la perdita della fiducia e della fedeltà dei clienti.

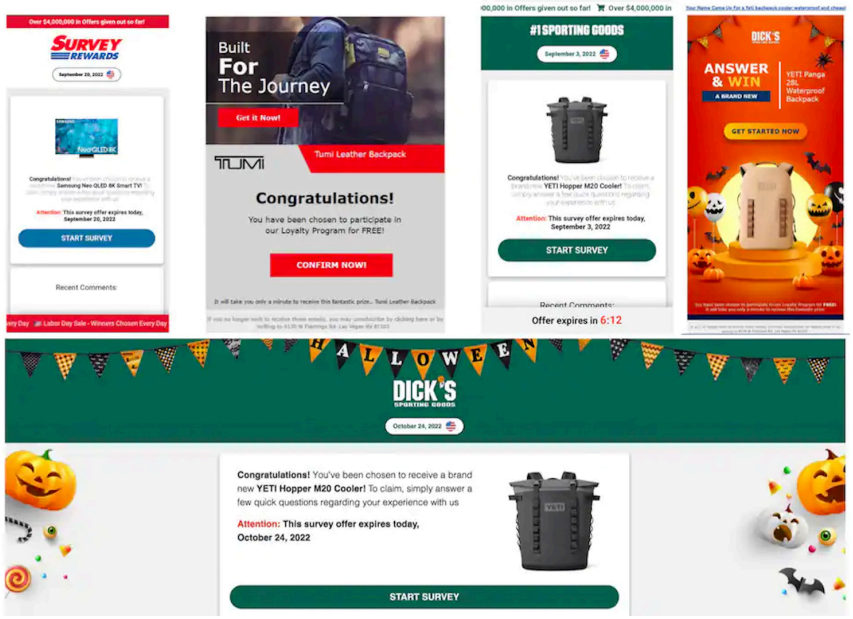


Figura 3. Un esempio di alcune delle principali società di e-commerce che hanno subito campagne di impersonificazione dei brand

Anche il bagarinaggio può essere attribuito al web scraping poiché gli scalper possono estrarre informazioni da un sito riguardo ai prodotti disponibili ed acquistarli prima che i clienti legittimi abbiano la possibilità di farlo (Figura 4).

Figura 4. Casi di utilizzo degli scraper

I criminali che eseguono questi tipi di attività di scraping dannose sono consapevoli degli effetti esercitati dai loro obiettivi malevoli sulle vittime prese di mira, inclusi gli effetti negativi causati da competitive intelligence/spionaggio, furto/scraping di inventari, contraffazione e creazione di siti/prodotti falsi e scraping e ripubblicazione di contenuti multimediali (Tabella 1). Infine, non esistono leggi che vietano esplicitamente l'uso dei bot scraper.

Impatto	Descrizione
Intelligence competitiva e spionaggio	Le aziende concorrenti utilizzano le informazioni ricavate dal sito di un'organizzazione per influenzarne i prezzi, apportare modifiche alle loro offerte e farsi un'idea su nuove opportunità e minacce.
Furto/Scraping di inventari	Gli scalper monitorano costantemente i siti presi di mira per trovare i prodotti diventati disponibili, quindi li aggiungono al carrello, rendendo questi prodotti non disponibili per i clienti reali.
Contraffazione e creazione di siti/prodotti falsi	Gli autori di atti di contraffazione utilizzano i contenuti esfiltrati per creare siti e cataloghi di prodotti falsi al fine di convincere gli utenti che si tratta di beni legittimi e non di contraffazioni.
Scraping e ripubblicazione di contenuti multimediali	I criminali possono esfiltrare dai siti articoli, blog e altri contenuti e aggiungerli ai loro siti, facendo in modo che l'organizzazione legittima perda visitatori e potenziali ricavi pubblicitari. Poiché le percentuali pubblicitarie spesso si basano sul numero di visitatori/utenti del sito, un numero inferiore di visitatori significa meno ricavi derivanti dalla pubblicità.

Tabella 1. Gli effetti negativi causati intenzionalmente dai web scraper

Gli effetti collaterali generali del web scraping

Indipendentemente dallo scopo del web scraping, le organizzazioni devono affrontare le spese derivanti dai suoi effetti collaterali. Alcune aziende utilizzano servizi di scraping legittimi a pagamento, tuttavia, le società che subiscono lo scraping devono sostenere propri costi, tra cui le spese richieste per le soluzioni anti-bot e gli effetti economici negativi dovuti al deterioramento delle performance del sito e alla manipolazione delle metriche chiave (Tabella 2).

Impatto	Descrizione
Aumento dei costi legati ai server, alle CDN e al cloud per la gestione del traffico dei bot	Influisce sui ricavi e causa la perdita della reputazione dell'azienda per l'utilizzo di contenuti da parte di aziende concorrenti, criminali e autori di atti di contraffazione.
Deterioramento delle performance del sito	Poiché i bot scraper non interrompono il loro lavoro finché non vengono bloccati, fanno aumentare i costi legati ai server e alla delivery poiché le organizzazioni si trovano a gestire un traffico dei bot indesiderato e a subire user experience compromesse, come il rallentamento delle performance dei siti e delle app.
Manipolazione delle metriche chiave	Le attività dei bot non rilevate possono inquinare significativamente le metriche chiave, come i tassi di conversione, a cui si affidano i team aziendali per prendere decisioni di investimento, come strategie di posizionamento dei prodotti e campagne di marketing.

Tabella 2. Gli effetti negativi causati non intenzionalmente dai web scraper

Scraping-for-hire: i servizi di web scraping di terze parti

Come abbiamo detto, i bot dei web scraper possono essere usati a scopi legittimi o dannosi. A differenza dei bot utilizzati per gli attacchi di credential stuffing, che sono bot dannosi noti e, quindi, legittimamente bloccati, i bot di web scraping legittimi vengono offerti da alcune aziende. Molte organizzazioni utilizzano questi servizi di web scraping di terze parti per estrarre e fornire dati utili e vantaggiosi, specialmente nel settore del marketing competitivo.

Alcune di queste aziende forniscono vari tipi di servizi di web scraping/estrazione dei dati, anzi organizzano apposite conferenze per promuoverli. Ad esempio, Bright Data organizza la conferenza ScrapeCon, in cui alcuni esperti del settore spiegano come eludere i sistemi di rilevamento dei bot per consentire alle aziende di apprendere come esfiltrare i dati. Nella Tabella 3, vengono riportati alcuni esempi di livelli di servizi forniti da aziende di web scraping di terze parti.

Livello di servizio 1	I servizi proxy possono far parte delle attività di scraping e offrire un'infrastruttura che potrebbe includere gli indirizzi residenziali e degli IP mobili dei data center.
Livello di servizio 2	Anche questo secondo livello può includere l'estrazione automatizzata dei dati che pulisce e struttura i dati per un utilizzo più semplice da parte dei membri del team addetti alla data science dei clienti, che si occupano di estrarre le informazioni più importanti per prendere decisioni aziendali oculate.
Livello di servizio 3	Il livello più elevato può aggiungere l'estrazione della business intelligence effettiva, migliorando ulteriormente il processo decisionale delle aziende. In questo contesto, si parla di "botnet AI".

Tabella 3. Vari livelli di servizi forniti da aziende di web scraping di terze parti

I clienti possono scegliere il livello di servizio desiderato, da quello standard a quello più avanzato, nonché la frequenza con cui deve avvenire la raccolta dei dati e i loro obiettivi specifici. Spesso, il livello di servizio fornito o le botnet scelte dipendono dal livello di protezione che i clienti devono conseguire. Una botnet standard può raccogliere i dati tramite uno script avanzato con poche migliaia di server proxy dislocati in data center che bilanciano il carico del traffico. Se la protezione è basilare, la botnet potrebbe utilizzare questa tecnica per superare i sistemi di difesa della gestione dei bot e il WAF (Web Application Firewall) dell'infrastruttura di sicurezza.

Se, tuttavia, la protezione è più avanzata, potrebbe essere richiesto un approccio allo scraping più sofisticato, come un attacco basato sui browser headless, sia se lo scraping è effettuato a scopi legittimi che dannosi. Si tratta di una scelta non economica per cui le aziende devono sostenere i costi, generalmente molto più elevati, implicati da un'infrastruttura più sofisticata rispetto ad un livello di servizio standard. Un sistema avanzato di difesa può includere tecnologie complesse, come CAPTCHA o PoW (Proof of Work), alcuni livelli di rilevamento progettati per una valutazione del fingerprinting lato client e un'analisi delle caratteristiche dei protocolli HTTP (Hypertext Transfer Protocol) e TLS (Transport Layer Security).

Lo scraping delle botnet basate sull'AI

Se i web scraper standard possono risultare più coerenti nell'utilizzo delle tecniche di scraping, le botnet AI sono in grado di rilevare ed esfiltrare dati e contenuti non strutturati in molteplici formati o posizioni. Inoltre, le botnet AI possono utilizzare la business intelligence effettiva per migliorare il processo decisionale. Le botnet AI più sofisticate, menzionate nella Tabella 3, livello di servizio 3, prevedono un processo per lo scraping dei dati costituito da 3 fasi: raccolta, estrazione ed elaborazione dei dati (Figura 5).

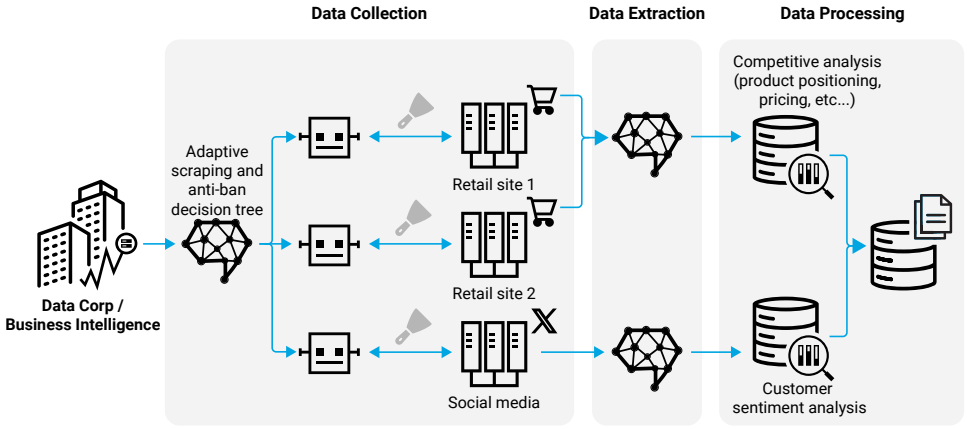
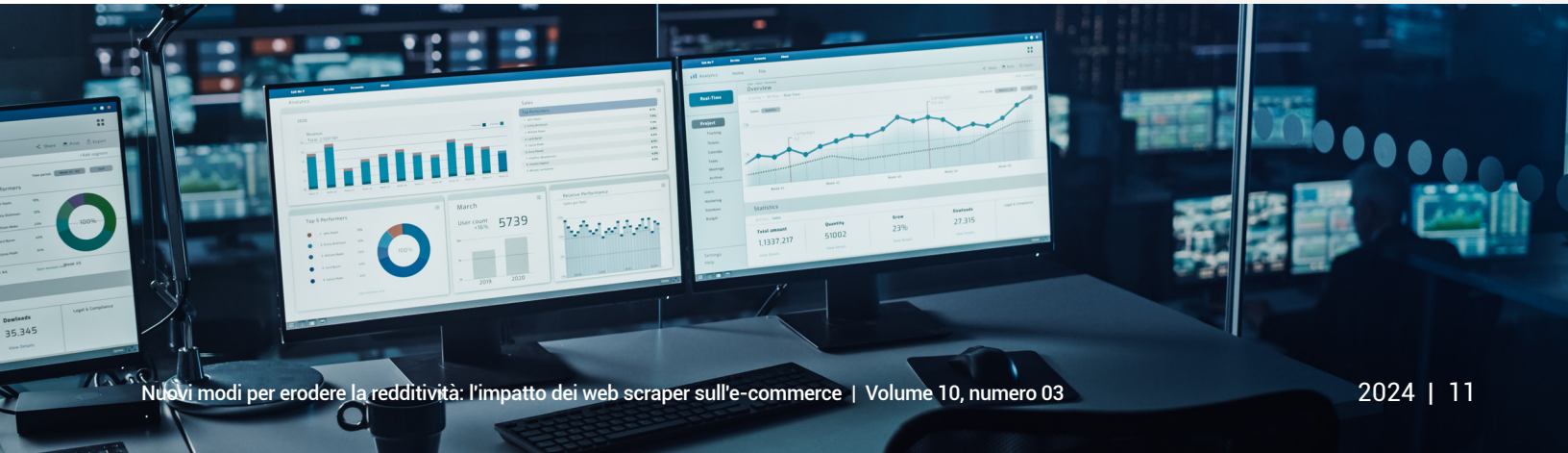


Figura 5. Rappresentazione di una botnet AI e del suo processo costituito da 3 fasi

Esaminiamo in modo più approfondito queste 3 fasi per comprendere meglio cosa implicano.

Raccolta dei dati

Il web scraping riguarda l'organizzazione dei dati estratti da un sito o più siti web che consentono alle organizzazioni di generare nuovi dataset da poter applicare e analizzare nel modo desiderato. La prima fase è rappresentata dalla raccolta dei dati.



Per la raccolta dei dati, è necessario combinare lo scraping adattivo con tecnologie di rilevamento "anti-blocco" o "anti-bot" per garantire operazioni rapide e semplici. Queste tecnologie vengono configurate come schemi decisionali per rilevare vari aspetti dei sistemi di protezione messi in atto. In questo contesto, parliamo di resilienza. La protezione dei bot può includere tecnologie di fingerprinting JavaScript, fingerprinting HTTP e TLS (valutazione delle intestazioni HTTP e dell'handshake TLS), nonché il rilevamento della reputazione dell'IP (Internet Protocol) (Figura 6). Alcuni di questi workflow possono includere l'apprendimento automatico (ML), specialmente durante la raccolta di statistiche sul tasso di successo; l'adattamento alla strategia dei cookie, l'intestazione HTTP e i parametri TLS, nonché la valutazione del codice di fingerprinting JavaScript. In questi casi, inoltre, può entrare in gioco un browser headless.

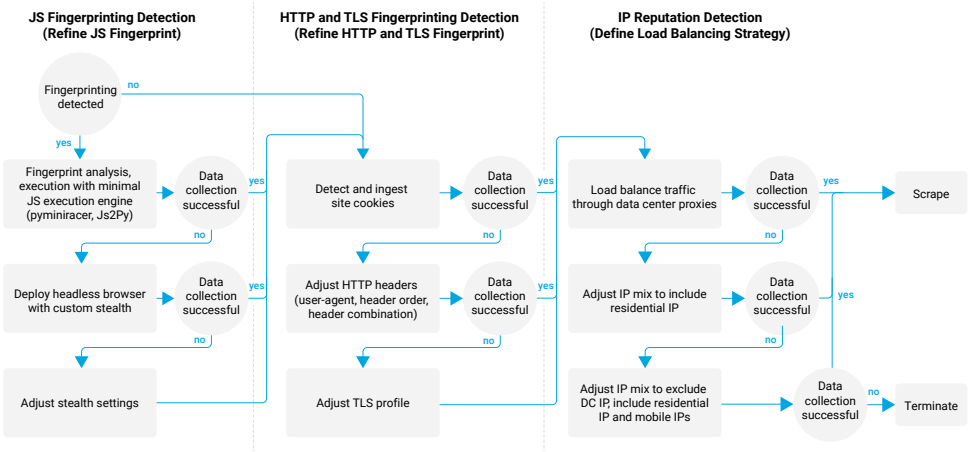


Figura 6. Durante il tentativo di raccogliere i dati, questo schema decisionale sul rilevamento anti-bot cerca di evitare le tecnologie di fingerprinting JavaScript, HTTP e TLS, nonché il rilevamento della reputazione dell'IP

Il browser headless

Un browser headless è un browser web che non presenta un'interfaccia utente grafica (GUI), quindi gli utenti non possono interagire direttamente con la pagina web su cui viene visualizzato il browser headless, che, invece, viene eseguito tramite un'interfaccia della riga di comando (CLI) o una comunicazione di rete. Nel caso di Selenium, un browser headless open-source molto diffuso, si tratta di un browser automatizzato e ampiamente utilizzato per il web scraping, che può risultare molto utile per chi cerca dati con lo scopo di esfiltrare contenuti dinamici.

I browser headless possono anche consentire la copia di screenshot e codici di siti web in modo efficiente, nonché l'estrazione dei dati desiderati senza effettuare il rendering dell'intera pagina. Tuttavia, gli attacchi basati sui browser sono costosi da effettuare e possono, a volte, essere comunque rilevati dalle "impronte digitali" che lasciano. Le spese da sostenere per altre infrastrutture sofisticate, tuttavia, sono simili a quelle relative ai browser headless, ossia di solito elevate.

L'estrazione e l'elaborazione dei dati

Le informazioni estratte sono. generalmente, costituite da contenuti HTML e JSON. Tra tutti i dati estratti, solo una parte può risultare utile ai fini dell'analisi. Ad esempio, l'analisi della concorrenza, di solito, include prezzi, sconti, inventari e numeri SKU dei prodotti, categorie e descrizioni. È possibile estrarre automaticamente informazioni importanti tramite i modelli di ML che possono essere addestrati con più strutture e formati di dati per riconoscerle. In tal modo, si può evitare tutto l'ulteriore lavoro di elaborazione richiesto per estrarre manualmente i dati e i requisiti necessari per esaminare la struttura del codice dei contenuti HTML e JSON. Questa struttura, inoltre, può cambiare man mano che si evolve la progettazione del sito. Per l'elaborazione è anche necessaria una logica di ML aggiuntiva se sono coinvolti più siti web come parte dell'ambito dell'analisi.

Case study: i vantaggi delle soluzioni di rilevamento del web scraping

I ricercatori di Akamai hanno osservato un sottoinsieme di clienti di e-commerce che erano protetti da una soluzione di web scraping in grado di rilevare le attività di scraping e di analizzare la suddivisione delle attività del traffico relative ad una settimana. Il campione di dati era costituito da circa 6,9 miliardi di richieste. L'analisi ha considerato solo le richieste HTML e AJAX. I contenuti statici (immagini, JavaScript, fogli di stile) non sono stati inclusi nell'analisi poiché la maggior parte dei bot non richiede questo tipo di contenuti, pertanto è stato più facile evitare di gonfiare inutilmente i dati.

L'attività complessiva è stata classificata da Akamai Content Protector ed è stata costituita per il 49,3% da traffico di utenti a basso rischio, per il 42,1% da traffico di bot (27,5% di bot dannosi ad alto rischio e 14,6% di bot legittimi) e per l'8,7% da traffico non classificato a rischio medio (Figura 7).

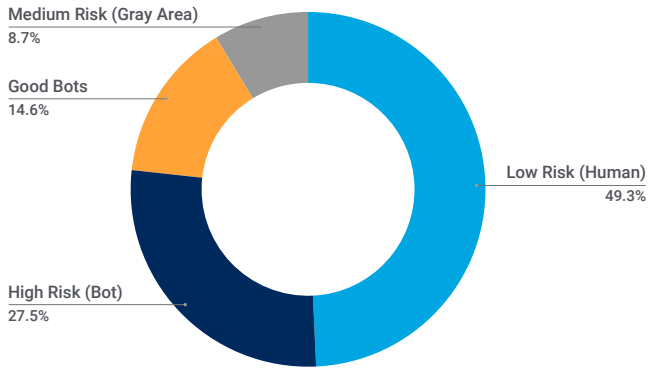


Figura 7. Suddivisione della classificazione dell'attività del traffico

Nella Figura 8, viene mostrato come, del 42,1% del traffico che proveniva dai bot, il 65,3% era originato da scraper considerati bot dannosi e il rimanente 34,7% proveniva da scraper classificati come bot legittimi (ossia, motori di ricerca web, SEO, social media e pubblicità online).

Figura 8. Confronto tra il traffico di bot legittimi e di bot dannosi

Sono stati misurati anche i livelli di sofisticazione per i bot dannosi ad alto rischio che hanno contribuito a raggiungere il 65,3% del traffico dei bot complessivo. Il 37% di questo traffico proveniva da botnet con script standard facilmente rilevabili tramite semplici metodi stateless, il 47,6% proveniva da botnet con script più avanzati rilevabili tramite metodi stateful più avanzati basati sul ML e il 15,5% proveniva da browser headless rilevabili tramite metodi stateful e avanzate tecnologie di fingerprinting JavaScript (Figura 9).

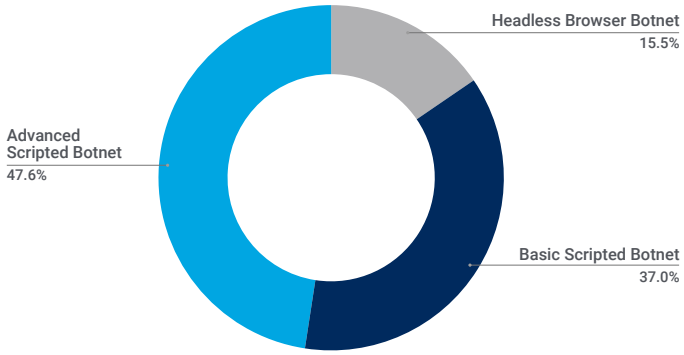


Figura 9. La distribuzione del traffico dei bot dannosi in base al loro livello di sofisticazione (i totali non arrivano al 100% a causa dell'arrotondamento)

Da questi dati, è emerso come gli scraper di bot dannosi sono molto più numerosi di quelli legittimi (quasi la metà di tutto il traffico dei bot), di cui le botnet con script avanzati hanno generato la maggior parte del traffico dei bot dannosi (47,6%).

Una volta messe in atto le difese contro questi bot e rimossi gli scraper, le attività del sito verranno eseguite molto più velocemente e in modo più efficiente, mentre le metriche del sito saranno più semplici da leggere. Questi risultati porteranno ad un miglioramento delle user/customer experience. Come mostrato nella Figura 10, il numero delle richieste di bot ad alto rischio è diminuito notevolmente una volta attivata la procedura di mitigazione.

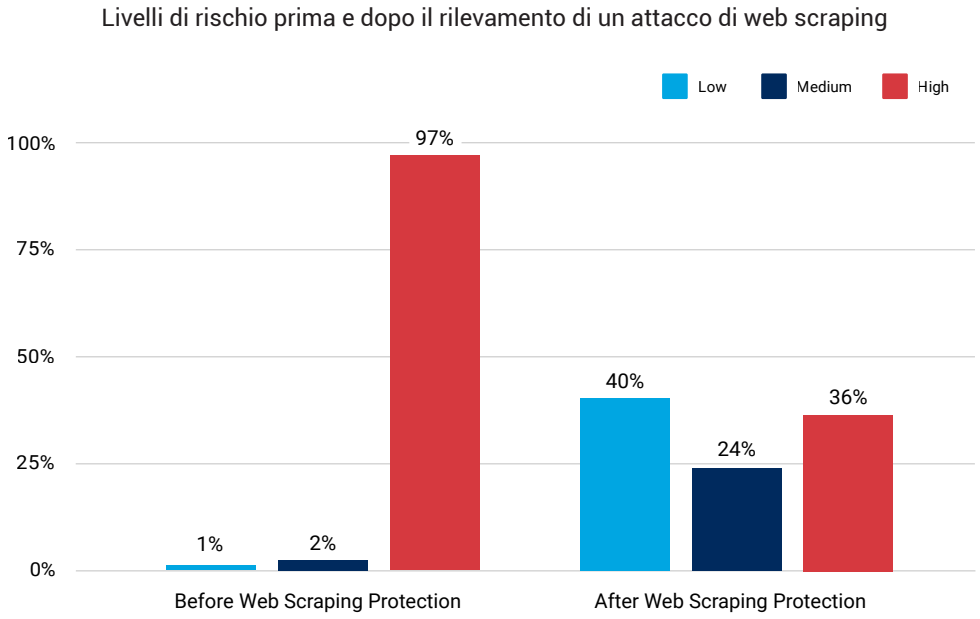


Figura 10. Livelli di rischio prima e dopo la mitigazione con Content Protector

Salvaguardia e mitigazione

Questa sezione fornisce alcune indicazioni fondamentali per il rilevamento dei web scraper e informazioni sugli strumenti utili per difendersi da queste minacce.

Rilevamento degli scraper standard

Anche se gli scraper sofisticati possono risultare difficili da rilevare, le soluzioni di gestione dei bot possono difendere dalla raccolta di dati effettuata da parte di ogni tipo di scraper intrusivo e possono cercare in particolar modo le seguenti caratteristiche per rilevare i bot dei web scraper più semplici:

- Richieste che pubblicizzano versioni precedenti di browser e sistemi operativi
- Anomalie nella firma dell'intestazione HTTP
- Utilizzo della versione precedente del protocollo HTTP (ad es., la versione 1.1) invece della più comune versione 2 o della nuova versione di questo protocollo
- Richieste che provengono da migliaia di servizi cloud/data center

Rilevamento degli scraper più avanzati

Nessuna delle caratteristiche riportate nell'elenco qui sopra sarà presente negli scraper più avanzati. Ecco alcune caratteristiche degli scraper più avanzati:

- Richieste che provengono dalle ultime versioni di browser e sistemi operativi
- Il set delle intestazioni HTTP sembra identico a quello del browser legittimo
- Utilizzo del protocollo HTTP v2
- Richieste che provengono da centinaia di migliaia di indirizzi IP residenziali e mobili

Identificazione dei modelli di traffico

Ecco alcuni indicatori chiave che consentono di identificare se il traffico di un sito web proviene da utenti (Figura 11), da bot semplici (Figura 12) o da bot sofisticati (Figura 13).

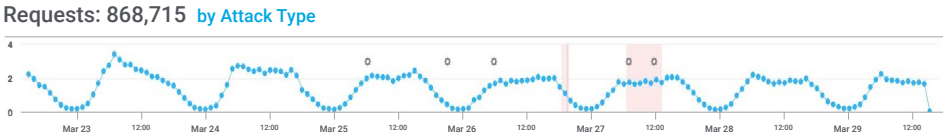


Figura 11. Traffico di utenti legittimi, che mostra, di solito, un ciclo di attività circadiano

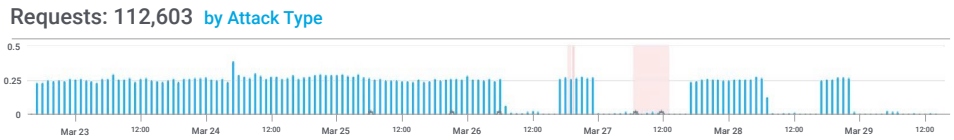


Figura 12. Traffico tipico di bot, che mostra attività regolari con interruzioni occasionali

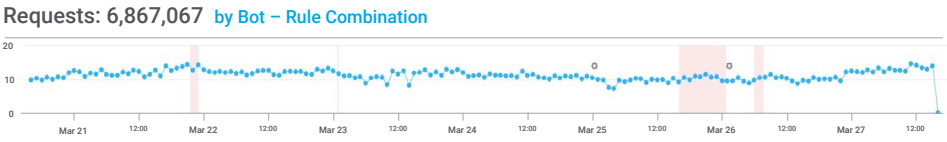


Figura 13. Bot più sofisticati, che mostrano il traffico continuamente, di giorno e di notte

Spesso, osserviamo anche botnet intermedie, ossia con una strategia di bilanciamento del carico debole, ma con una strategia di fingerprinting sofisticata (o viceversa). Tuttavia, le botnet più avanzate possono essere talmente sofisticate da presentare un'impronta digitale perfetta o da riuscire persino a riprodurre modelli di traffico di utenti legittimi.

Oltre a cercare questi bot scraper, gli strumenti che proteggono dal web scraping, come Content Protector, possono offrire vantaggi speciali e una navigazione più agevole nelle agitate acque infestate dagli scraper. I vantaggi includono:

- Tassi di conversione più elevati e costi IT ridotti
- Metriche più accurate, che possono condurre a migliori decisioni in termini di investimenti e favorire l'incremento dei profitti
- Ridotta pressione sui prezzi, che può evitare l'abbassamento dei prezzi da parte della concorrenza
- Soddisfazione dei clienti che possono accedere ai prodotti desiderati e aumento dei ricavi derivante dalle opportunità di upselling sui clienti che aggiungono ulteriori prodotti al carrello dopo essersi assicurati l'articolo desiderato
- Reputazione del brand preservata e clienti protetti da contraffazioni di scarsa qualità scambiate per beni forniti dal venditore originale
- Mantenimento dei ricavi dei prodotti e della fedeltà dei clienti
- Aumento/Protezione dei ricavi pubblicitari
- Conservazione degli utenti e dei visitatori del sito

Considerazioni sulla conformità

Lo standard Payment Card Industry Data Security Standard (PCI DSS) v4.0 è entrato in vigore, ed è stato determinato soprattutto da una serie di minacce che continuano a interessare le aziende. La visibilità è la chiave per affrontare questi attacchi. Che vengano sferrati contro l'ambiente JavaScript storico o contro le API utilizzate per facilitare la trasformazione, è fondamentale rilevare e mitigare rapidamente questi attacchi.

Osserviamo anche tendenze di conformità emergenti nel nuovo NIST Cybersecurity Framework versione 2.0, che ha aggiunto una funzione di governance. Il NIST tende a fungere da base per numerose normative governative estendendosi a molti modelli di cybersicurezza commerciale. Ora è il momento giusto per esaminare le nuove linee guida e usarle per aggiornare le proprie policy o confrontarle con la propria documentazione attuale per verificare le aree in cui esistono eventuali falle.

Per le società quotate in Borsa e per le aziende che usano principi contabili generalmente accettati (GAAP), un'altra area relativa alla conformità è rappresentata dagli aspetti materiali della cybersicurezza. La necessità di definire i rischi e le minacce materiali richiede la collaborazione dei dirigenti aziendali. Una volta identificate le minacce materiali (come i ransomware), è necessario adottare gli opportuni sistemi di mitigazione (come la microsegmentazione). È consigliabile assicurarsi che i piani di gestione delle crisi rispettino la cronologia delle divulgazioni e predisporre un playbook in previsione degli scenari peggiori, nel qual caso bisogna compilare il modulo Cyber Incident Form 8-K della Security and Exchange Commission.

Conclusione

Speriamo che questo rapporto possa fornirvi informazioni approfondite su un'area che potrebbe influire negativamente da un punto di vista economico sulla vostra organizzazione. I bot influiscono sui siti in modo sempre più massiccio, pertanto è importante ottimizzare i bot legittimi, mitigare i bot dannosi e garantire eccellenti customer experience complessive. Si tratta di un problema di sicurezza con un impatto sulle aziende. Come per tutti i problemi di sicurezza, il primo passo è ottenere visibilità, il secondo passo è analizzare l'impatto e l'ultimo passo è stabilire il ROI per i rischi e i ricavi in modo da poter implementare i controlli di sicurezza appropriati.

È impossibile proteggersi da ciò che non si vede. Ecco perché questo è il momento giusto per stabilire se esistono lacune di visibilità nel vostro sistema. Per farlo, dovete stabilire il livello dell'attività di web scraping sul vostro sito e il suo scopo. Lo scenario dei bot è costituito da bot legittimi e bot dannosi, mentre i bot scraper appartengono ad entrambe le categorie, a seconda del loro utilizzo. Anche se il confine tra i bot scraper legittimi e quelli dannosi non è ben definito, l'evoluzione della sofisticazione dei bot (ossia, i web scraper che sferrano attacchi basati sui browser headless) continua. Al contempo, l'impatto esercitato dai bot dei web scraper sulle società di e-commerce in termini di costi dell'IT e di customer experience resta immenso. È fondamentale assicurarsi di aver messo in atto gli strumenti necessari per analizzare le attività dei bot e il loro impatto sul vostro sito.

Ciò che volete evitare è trovarvi di fronte a criminali che sferrano i loro attacchi contro il vostro sito e commettono una serie di attività dannose, come riscattare i punti fedeltà, effettuare ordini fraudolenti o persino commettere frodi sui resi. Inoltre, sicuramente volete evitare che i biglietti di eventi limitati o i prodotti più richiesti vengano acquistati dai bot. I bot possono essere utilizzati per facilitare l'abuso di apertura di nuovi account usufruendo di offerte speciali, che influiscono sui costi e sull'analisi delle campagne. Le grandi botnet DDoS (Distributed Denial-of-Service) possono sovraccaricare le applicazioni web e causare scarse user experience oppure bloccare ordini o prenotazioni, determinando perdite di fatturato e problemi ai clienti. I bot possono persino imitare il comportamento umano online per aumentare il numero di clic e il traffico su un sito web, alterando l'analisi del marketing e delle performance di experience digitali accuratamente create. Sicuramente non volete tutto ciò.

Come abbiamo osservato in precedenza, più della metà del traffico web del commercio a livello globale è costituito dai bot e i livelli del traffico dei bot continuano a salire. Le informazioni e i consigli forniti da Akamai in questo rapporto sono basati sulla sua piattaforma di sicurezza, che include la funzionalità di protezione dei contenuti con una difesa dal web scraping. Collaborando con molte aziende leader nel settore dell'e-commerce, desideriamo condividere consigli sui metodi di salvaguardia e mitigazione perché le aziende possano adottarli per proteggere al meglio i loro clienti. Abbiamo notato un incremento nell'utilizzo, nelle opzioni dei livelli di servizio e nei tipi di bot di web scraper disponibili. Pertanto, è necessario valutare continuamente il livello di rischio della vostra azienda per stabilire se gli attuali controlli di sicurezza sono in grado di soddisfare la propensione al rischio della leadership.

Restate aggiornati sulla nostra ultima ricerca consultando il nostro Security Research Hub.

Metodologie

Dati di Content Protector

Questo campione di dati descrive le classificazioni dei livelli di rischio assegnate dal nostro strumento Content Protector al traffico monitorato. Queste classificazioni vengono utilizzate per rilevare le attività di scraping condotte da bot e per stabilire se si tratta di bot legittimi o dannosi. Poiché la maggior parte dei bot non richiede contenuti statici, questa analisi ha considerato solo le richieste HTML e AJAX per evitare di gonfiare inutilmente i dati.

Questo campione di dati ha riguardato un periodo di una settimana, dal 12 aprile al 19 aprile 2024. Il nostro campione totale di dati era costituito da oltre 6,5 miliardi di richieste.

Attacchi bot

Questi dati descrivono gli avvisi a livello di applicazione relativi al traffico osservato tramite la nostra soluzione WAF (Web Application Firewall) e lo strumento di gestione dei bot. Gli avvisi sui bot vengono attivati quando si rileva un payload bot all'interno di una richiesta a un sito web, un'applicazione o un'API protetta. Questi avvisi possono essere attivati sia da bot dannosi che non dannosi, e non indicano la corretta riuscita della violazione. Sebbene questi prodotti consentano un alto livello di personalizzazione, i dati qui presentati sono stati raccolti senza prendere in considerazione le configurazioni personalizzate delle proprietà protette. I dati sono stati ricavati da uno strumento interno per l'analisi degli eventi di sicurezza rilevati sull'Akamai Connected Cloud, una rete di circa 340.000 server in più di 4.000 sedi su quasi 1.300 reti in oltre 130 paesi. I nostri team addetti alla sicurezza utilizzano questi dati, misurati in petabyte al mese, per effettuare ricerche sugli attacchi, segnalare comportamenti dannosi e includere ulteriori informazioni nelle soluzioni Akamai.

Questi dati hanno riguardato un periodo di 15 mesi, dal 1° gennaio 2023 al 31 marzo 2024.

Riconoscimenti

Redattore capo

Lance Rhodes

Editoria e stesura

David Senecal
Maria Vlasak

Revisione e contributi di esperti del settore

Mitch Mayne
Susan McReynolds
Christine Ross
Badette Tribbey
Steve Winterfeld

Analisi dei dati

Chelsea Tuttle

Materiali promozionali

Annie Brunholzl

Marketing ed editoria

Georgina Morales
Emily Spinks

Altri rapporti sullo stato di Internet - Security

Leggete i numeri precedenti e guardate le prossime uscite degli acclamati rapporti sullo stato di Internet - Security di Akamai sul sito **akamai.com/soti**

Ulteriori informazioni sulla ricerca delle minacce Akamai

Restate aggiornati con le ultime intelligence sulle minacce, rapporti sulla sicurezza e ricerche sulla cybersicurezza. **akamai.com/security-research**

Accesso ai dati del rapporto

Potete visualizzare i grafici e i diagrammi citati in questo rapporto in versioni di alta qualità. L'utilizzo e la consultazione di queste immagini sono forniti a scopo gratuito, purché Akamai venga debitamente citata come fonte e venga conservato il logo dell'azienda: **akamai.com/sotidata**

Ulteriori informazioni sulle soluzioni di Akamai

Per ulteriori informazioni sulle soluzioni di Akamai per il rilevamento e la protezione dai web scraper, visitate la nostra **pagina su Content Protector**.

Data di pubblicazione: 06/24.