SOTI

10 YEARS
OF SECURITY INSIGHT

VOLUME 10
NUMÉRO 03

Protégez vos données et vos revenus

Impact des extracteurs Web sur le commerce électronique

Akamai

État des lieux d'Internet/Sécurité

Table des matières

Saviez-vous que les bots génèrent plus de la moitié de la totalité du trafic Web ? Le segment de marché du commerce, en particulier, avec sa dépendance à l'égard des applications et des actifs Web générateurs de revenus, a été le plus touché par le trafic de bots à haut risque (Figure 1). Et bien qu'on entende souvent dire que les bots évoluent, les bots d'extraction Web sont ceux qui attirent l'attention des entreprises de commerce électronique aujourd'hui, car leurs impacts économiques, souvent tapis sous la surface, diffèrent de ceux d'autres types de bots. La détection des bots d'extraction est également devenue beaucoup plus difficile en raison de la montée en puissance des botnets d'intelligence artificielle (IA) et des technologies de navigateur sans interface, qui les rendent extrêmement évasifs. Par exemple, l'un des clients d'Akamai spécialisé dans le commerce électronique a vu s'arrêter 99 % du trafic à haut risque dont il ignorait qu'il provenait de bots d'extraction.

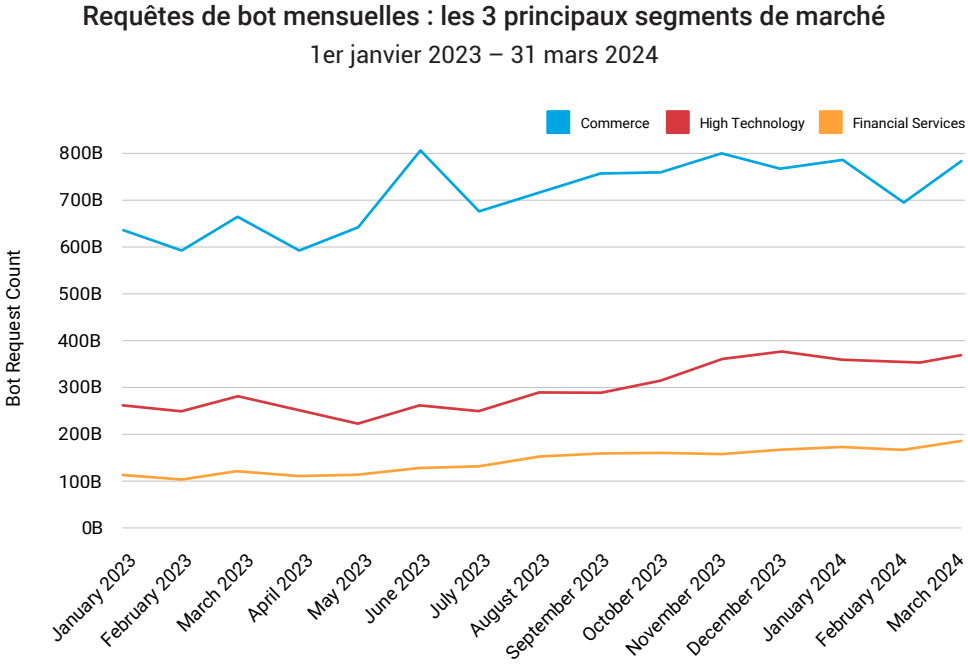


Figure 1 : Le commerce arrive en tête des segments de marché pour les demandes de bots, et enregistre une augmentation du trafic mondial de bots de début 2023 jusqu'au premier trimestre 2024

C'est pourquoi, dans ce rapport État des lieux d'Internet, nous nous concentrons sur l'évolution et la spécialisation de ces bots et de leurs opérateurs. Bien que les bots existent depuis un certain temps, nous continuons à voir leur application par divers groupes pour permettre des attaques criminelles, des stratagèmes de fraude et l'obtention d'informations concurrentielles. Récemment, nous avons constaté une tendance à l'augmentation de l'utilisation de tous les bots ainsi qu'un accroissement de l'impact négatif des bots d'extraction sur les entreprises. Ce rapport est conçu pour partager à la fois des informations techniques et des méthodes d'attaque afin de sensibiliser à ce problème croissant dans le secteur du commerce.

Bots : le Bon, la Brute et le Truand

Toutes les grandes organisations axées sur le commerce électronique sont confrontées à des bots qui évoluent continuellement et se spécialisent de plus en plus en fonction de ce qu'ils cherchent à accomplir. Dans le segment de marché du commerce, il existe une grande variété de types de bots qui effectuent de nombreuses tâches différentes. Ces différents types peuvent être facilement classés en trois groupes : les inoffensifs, les malveillants et les « gris ». Les bots inoffensifs aident les clients à trouver votre site. Les bots malveillants extraient votre site à des fins malveillantes. Enfin, bien qu'ils soient légitimes, les bots « gris » ont tendance à être bruyants ; ils constituent une sous-catégorie des bots inoffensifs (par exemple, les bots partenaires qui envoient constamment des messages et d'autres API de programmes qui passent des appels fréquents).

Ainsi, lorsque nous pensons aux chatbots utiles et aux bots de moteur de recherche qui peuvent avoir des impacts bénéfiques, comme répondre aux questions de base des utilisateurs et fournir un contenu de site Web qui renvoie des résultats de recherche plus précis, notre objectif est d'optimiser ces types de bots tout en limitant les coûts informatiques. En ce qui concerne les bots nuisibles, tels que les bots de credential stuffing qui tentent d'obtenir un accès non autorisé au compte d'un client, entraînant ainsi un piratage du compte, notre but est de prendre des mesures préventives sans pour autant nuire à l'expérience globale du client. Enfin, un type de bot qui a récemment fait son apparition devient particulièrement problématique dans la mesure où celui-ci diminue les revenus, tout en réduisant la fidélité et en augmentant les coûts : les bots d'extraction de site Web.

Ces bots, qui sont utilisés pour extraire directement des données et du contenu de sites Web sur l'Internet, sont uniques. Ils exigent notre attention, car ils fonctionnent différemment et leurs impacts commerciaux et leurs détections varient par rapport à ceux des autres bots. Les extracteurs Web ont également de multiples facettes, dans la mesure où leurs cas d'utilisation varient en fonction de la manière dont les entreprises et les opérateurs monétisent les informations collectées par ces bots. Quel que soit l'objectif, ces extracteurs font perdre des revenus, augmentent les coûts informatiques et détériorent l'expérience globale des clients.

Dans ce rapport État des lieux d'Internet, nous examinons les impacts de l'extraction sur le commerce électronique et les raisons pour lesquelles les propriétaires d'entreprises (qui pensent digital, marketing, marque, finance, risque et sécurité) devraient collaborer pour stopper les extracteurs abusifs. Pour mieux comprendre ces impacts, il est essentiel de savoir pourquoi les bots d'extraction évoluent, mais aussi d'avoir un aperçu global de leur utilisation, de leur mode de fonctionnement, de leurs impacts et de ce que les entreprises commerciales peuvent faire à leur sujet.

Principales conclusions du rapport

- L'extraction Web n'est pas seulement un problème de fraude ou de sécurité, c'est aussi un problème commercial. Les bots d'extraction ont un effet négatif sur de nombreuses facettes de l'organisation, notamment sur les revenus, l'avantage concurrentiel, l'identité de marque, l'expérience client, les coûts d'infrastructure et l'expérience digitale, pour n'en citer que quelques-unes.

- Selon une étude de cas d'Akamai, 42,1 % de l'activité globale du trafic provenait de bots, dont 65,3 % de bots malveillants. Au total, 63,1 % de ce trafic de bots malveillants utilisaient des techniques avancées.

- La technologie des navigateurs sans interface a changé le paysage des extracteurs, nécessitant l'adoption d'une approche de gestion de ce type d'activité de bots plus sophistiquée que d'autres mesures d'atténuation reposant sur JavaScript.

- Les impacts techniques auxquels les entreprises sont confrontées en raison de l'extraction, que l'extraction ait été effectuée avec des intentions malveillantes ou bénéfiques, incluent la dégradation des performances des sites Web, la pollution des métriques du site, les attaques d'informations d'identification compromises provenant de sites d'hameçonnage, l'augmentation des coûts de calcul et plus encore.

- Il est important d'observer et de comprendre les différents modèles de trafic afin de déterminer si un site Web est victime d'un trafic humain, d'un bot de base ou d'un bot sophistiqué. Ces modèles peuvent être circadiens, intermittents ou continus.

Bots inoffensifs et bots malveillants

Commençons par le commencement : un bot, abréviation de « robot », est un programme informatique conçu pour effectuer des tâches automatisées plus rapidement et avec plus de précision qu'un humain. Les différents rôles et types de bots se répartissent en deux grandes catégories : les bots inoffensifs et les bots malveillants (Figure 2). Les bots gris sont une sous-catégorie des bots inoffensifs, mais nous les considèrerons comme des bots inoffensifs pour simplifier la comparaison.

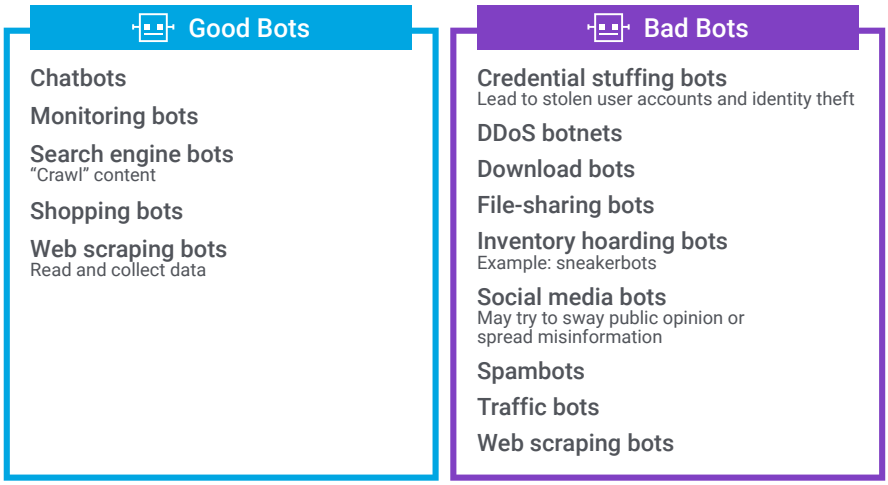
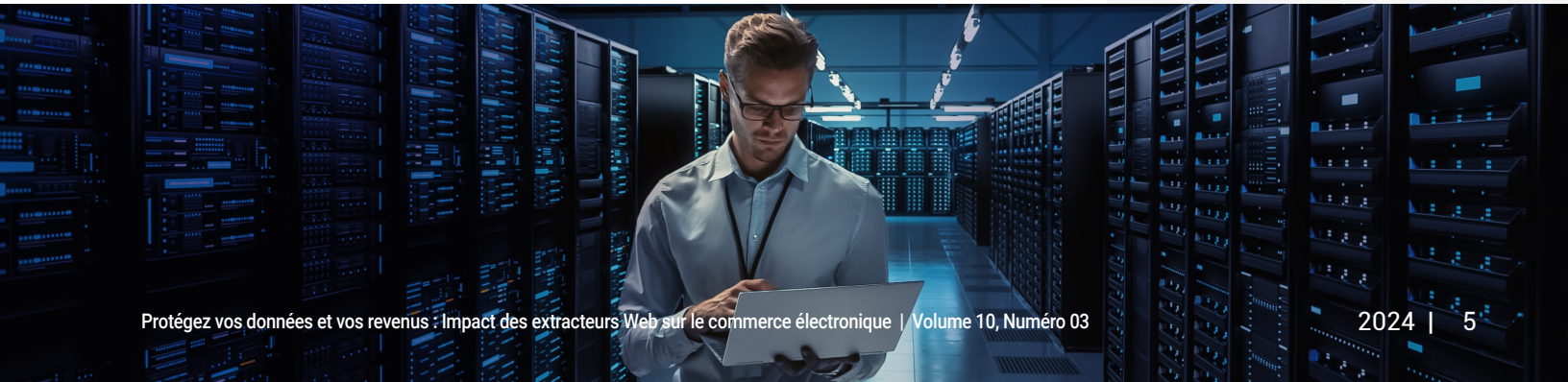


Figure 2 : Comparaison juxtaposée, avec des exemples, entre les bots inoffensifs et les bots malveillants

Les bots inoffensifs sont des bots utiles qui aident à fournir des outils et des services, tandis que les bots malveillants, comme leur nom l'indique, sont souvent utilisés avec des intentions malveillantes par les cybercriminels et les fraudeurs. Par exemple, un bot de trafic qui imite le comportement humain en ligne pour augmenter le nombre de clics et le trafic sur un site Web (c'est-à-dire commettre une fraude publicitaire).

Les bots d'extraction Web appartiennent à la fois à la catégorie des bots inoffensifs et malveillants. La distinction est liée à la manière dont les organisations utilisent les informations collectées par ces bots. Nous allons maintenant nous intéresser de plus près à divers cas d'utilisation associés aux bons et aux mauvais effets des bots d'extraction, auxquels sont confrontés certains des plus grands détaillants et marques de commerce électronique au monde.



Extraction 101

L'extraction Web est couramment utilisée par les entreprises de commerce électronique. Dans les secteurs des voyages et de l'hôtellerie, par exemple, les agrégateurs de voyages extraient le contenu dynamique de leurs partenaires hôteliers et aériens pour rester à jour sur les disponibilités et les prix. Ce type d'extraction est attendu, et les entreprises utilisent des contrôles de bots courants pour limiter les extracteurs pendant les périodes de la journée où les utilisateurs réels cherchent à faire une réservation. Les organisations font également appel à des fournisseurs de services d'extraction de données pour recueillir des pistes et d'autres informations connexes auprès de leurs concurrents. Dans le même temps, les bots d'extraction peuvent être utilisés pour analyser les données et identifier les tendances. L'extraction peut également s'avérer utile pour la révision de sites afin d'améliorer les offres et les services en ligne et de permettre aux clients potentiels de trouver plus facilement les produits de l'entreprise, par exemple via un moteur de recherche. Toutes ces actions peuvent aider les entreprises à acquérir un avantage concurrentiel. Cependant, il est indéniable que de nombreuses entités utilisent des bots d'extraction pour des raisons moins louables.

L'extraction évolue et les clients le remarquent

Malheureusement, nous entendons souvent parler d'utilisateurs qui ont été victimes d'escroqueries par hameçonnage. Dans ce cas, des bots d'extraction ont pu être utilisés pour s'emparer des images de produits, des descriptions et des informations sur les prix pour créer des vitrines contrefaites ou des sites d'hameçonnage visant à voler des identifiants ou des informations de carte de crédit. Ces sites d'hameçonnage ou de contrefaçon sont une forme d'usurpation de marque, dans laquelle la propriété intellectuelle des organisations victimes est utilisée pour établir une relation de confiance avec des clients potentiels.

Certaines des plus grandes marques de commerce électronique au monde ont été victimes de sites contrefaits, de campagnes d'hameçonnage et de vol de données Web de l'entreprise dans le cadre de campagnes d'usurpation d'identité (Figure 3). Malheureusement, lorsque les sites d'hameçonnage réussissent, les marques légitimes sont les victimes directes des retombées de la perte de confiance et de fidélité des clients.

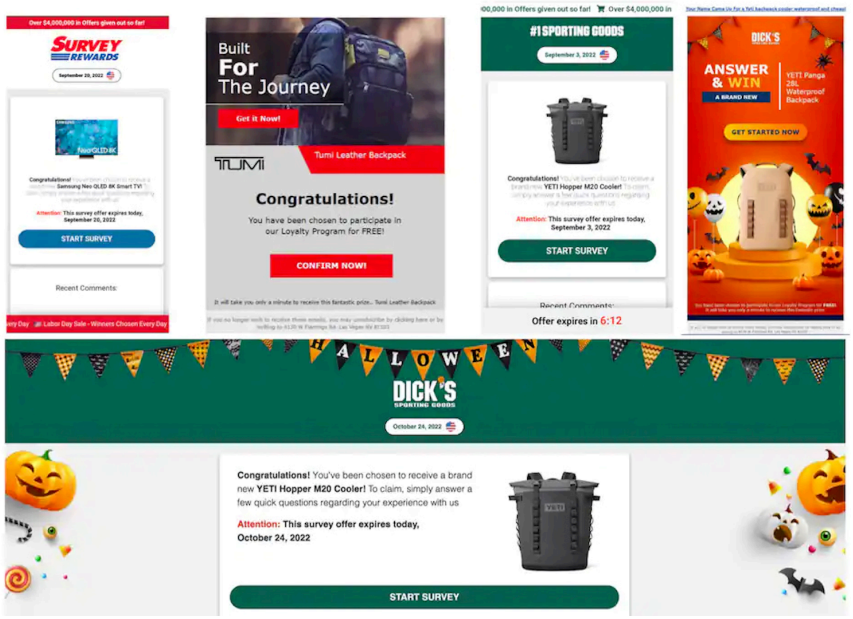


Figure 3 : Exemple de grandes entreprises de commerce électronique victimes d'usurpation d'identité de marque

La revente peut également être liée à l'extraction Web, dans la mesure où les revendeurs peuvent rechercher sur un site les produits disponibles et les acheter avant que les clients légitimes n'aient pu le faire (Figure 4).

Figure 4 : Cas d'utilisation d'extracteurs

Les acteurs malveillants qui se livrent à ce type d'activités d'extraction nuisibles sont conscients des effets que leurs objectifs malveillants ont sur les victimes. Parmi ceux-ci, on retrouve notamment les effets négatifs de la veille concurrentielle/l'espionnage, de l'accumulation/l'extraction des stocks, de la contrefaçon/création de sites et de marchandises factices, ainsi que de l'extraction et de la rediffusion de sites de médias (Tableau 1). Malheureusement il n'existe aucune loi interdisant explicitement l'utilisation de bots d'extraction.

Impact	Description
Veille concurrentielle/ Espionnage	Les concurrents utilisent les informations du site d'une organisation pour réduire les prix, apporter des modifications à ses offres et avoir une idée des nouvelles opportunités et menaces.
Accumulation/Extraction des stocks	Les bots de revente fouillent constamment les sites ciblés à la recherche de produits disponibles, puis les ajoutent aux paniers, rendant ces produits indisponibles pour les vrais clients.
Contrefaçons et sites/ marchandises factices	Des faussaires utilisent le contenu extrait pour créer de faux sites et de faux catalogues de produits pour tromper les utilisateurs en leur faisant croire qu'ils achètent des biens légitimes au lieu de contrefaçons.
Extraction et rediffusion de sites de médias	Les attaquants peuvent extraire des articles d'actualité, des blogs et d'autres contenus et les placer sur leurs propres sites, entraînant une perte de visiteurs et de revenus publicitaires potentiels pour l'organisation d'origine. Les contrats publicitaires sont souvent basés sur le nombre de visiteurs ou la visibilité du site, de sorte qu'une diminution du nombre de visiteurs signifie que le site de médias perd les revenus qu'il aurait obtenus grâce à des contrats publicitaires plus élevés.

Tableau 1 : Impacts négatifs intentionnels des extracteurs Web

Les effets secondaires généraux de l'extraction Web

Quelle que soit l'intention de l'extraction Web, les organisations doivent faire face aux dépenses liées à ses effets secondaires. Certaines entreprises paient pour des services d'extraction bénéfiques, mais les entreprises qui font l'objet de cette extraction sont elles-mêmes exposées à des coûts. Parmi ces coûts figurent notamment des dépenses liées aux solutions anti-bots et des impacts économiques négatifs de la dégradation des performances du site et de la pollution des indicateurs clés (Tableau 2).

Impact	Description
Augmentation des coûts des serveurs, du réseau de diffusion de contenu (CDN) et du cloud pour desservir le trafic des bots	Cela a un impact sur les revenus et affecte la réputation liée à l'utilisation du contenu par les concurrents, les attaquants et les faussaires.
Dégradation des performances du site	Dans la mesure où les bots d'extraction fonctionnent en continu jusqu'à ce qu'on les arrête, ils augmentent les coûts de serveur et de diffusion, car les organisations assument le trafic indésirable des bots et souffrent d'une dégradation de l'expérience utilisateur en raison de performances plus lentes du site et de l'application.
Pollution des indicateurs clés	Les activités de bots non détectées faussent considérablement les indicateurs clés, tels que la conversion sur le site, sur lesquels les équipes commerciales s'appuient pour prendre des décisions d'investissement, comme les stratégies de positionnement des produits et les campagnes de marketing.

Tableau 2 : Impacts négatifs involontaires des extracteurs Web

Extractions à louer : services tiers d'extraction Web

Comme nous l'avons mentionné, les bots d'extraction Web peuvent être utilisés à bon ou à mauvais escient. Contrairement aux bots utilisés pour les attaques de credential stuffing, qui sont des bots malveillants connus et donc bloqués à juste titre, il existe des entreprises qui proposent des bots d'extraction Web légitimes. De nombreuses organisations utilisent ces services d'extraction Web tiers pour extraire et fournir des données à leur propre organisation, ce qui peut être bénéfique, en particulier dans le monde du marketing concurrentiel.

Il existe des dizaines d'entreprises qui fournissent différents types de services d'extraction Web/de données, et des conférences en font même la promotion. Par exemple, Bright Data organise une conférence intitulée ScrapeCon, qui rassemble des experts du contournement de la détection des bots, afin de permettre aux entreprises d'apprendre à extraire des données. Le tableau 3 présente des exemples de niveaux de services pouvant être fournis par des sociétés tierces spécialisées dans l'extraction Web.

Niveau de service 1	Les services proxy peuvent faire partie de l'infrastructure d'extraction et d'offre qui pourrait inclure les adresses IP mobiles et résidentielles des centres de données.
Niveau de service 2	Ce deuxième niveau peut également inclure l'extraction automatisée de données qui nettoie et structure les données pour faciliter leur utilisation par les membres de l'équipe de science des données du client, qui en extraient des informations nécessaires pour orienter les décisions de l'entreprise.
Niveau de service 3	Le niveau le plus élevé peut ajouter l'extraction de renseignements commerciaux proprement dits, qui peuvent encore améliorer davantage le processus de prise de décision des entreprises. C'est ce que l'on appelle les « botnets d'IA ».

Tableau 3 : Différents niveaux de services fournis par des sociétés tierces spécialisées dans l'extraction Web

Les clients peuvent choisir n'importe lequel de ces niveaux de service, du plus basique au plus avancé, ainsi que la fréquence de la collecte de données, mais aussi préciser leurs cibles. Souvent, le niveau de service fourni, ou le botnet choisi, dépend du niveau de protection qu'ils doivent surmonter. Un botnet basique peut collecter des données par le biais d'un script avancé avec quelques milliers de serveurs proxy situés dans des centres de données qui équilibrent la charge de trafic. Si la protection est suffisante, le botnet peut utiliser cette technique pour passer à travers les défenses de gestion du bot et le Web Application Firewall (pare-feu d'application Web) de l'infrastructure de sécurité.

En revanche, si la protection est plus avancée, une approche plus sophistiquée de l'extraction, telle qu'une attaque par navigateur sans interface, peut s'avérer nécessaire. Ceci est vrai, que l'extraction soit effectuée par un acteur malveillant ou non. De plus, cela a un coût, car les entreprises vont devoir prendre en charge des frais qui sont généralement beaucoup plus élevés pour une infrastructure plus sophistiquée que pour le niveau de service de base. Une défense avancée peut inclure des technologies de défi (comme CAPTCHA ou preuve de travail), plusieurs couches de détection conçues pour l'évaluation des empreintes digitales côté client, et une analyse des caractéristiques Hypertext Transfer Protocol (HTTP) et Transport Layer Security (TLS).

Processus d'extraction des botnets d'IA

Bien que les techniques d'extraction des bots d'extraction Web de base soient plus cohérentes, les botnets d'IA ont la capacité de découvrir et d'extraire des données et du contenu non structurés qui sont dans un format ou un emplacement moins cohérent. En outre, ils peuvent utiliser la veille stratégique réelle pour améliorer le processus de prise de décision. Les botnets d'IA sophistiqués, mentionnés dans le tableau 3, de niveau de service 3, suivent un processus en trois étapes pour extraire les données. Ils collectent, extraient et traitent les données (Figure 5).

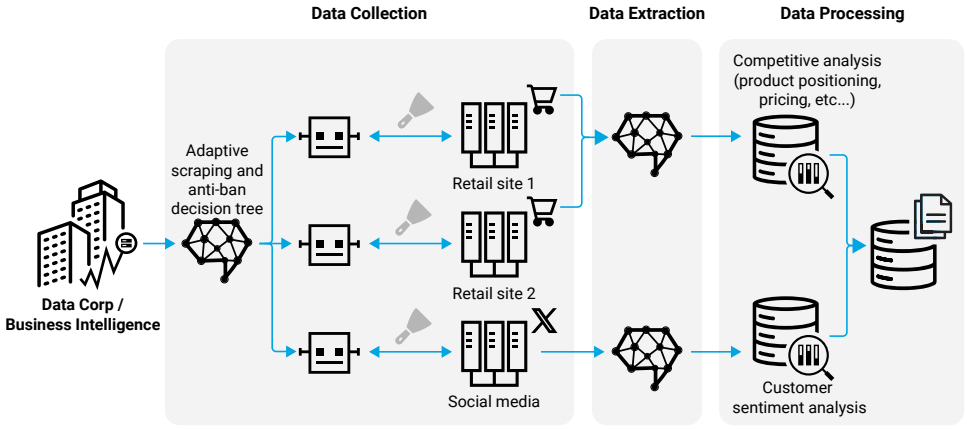
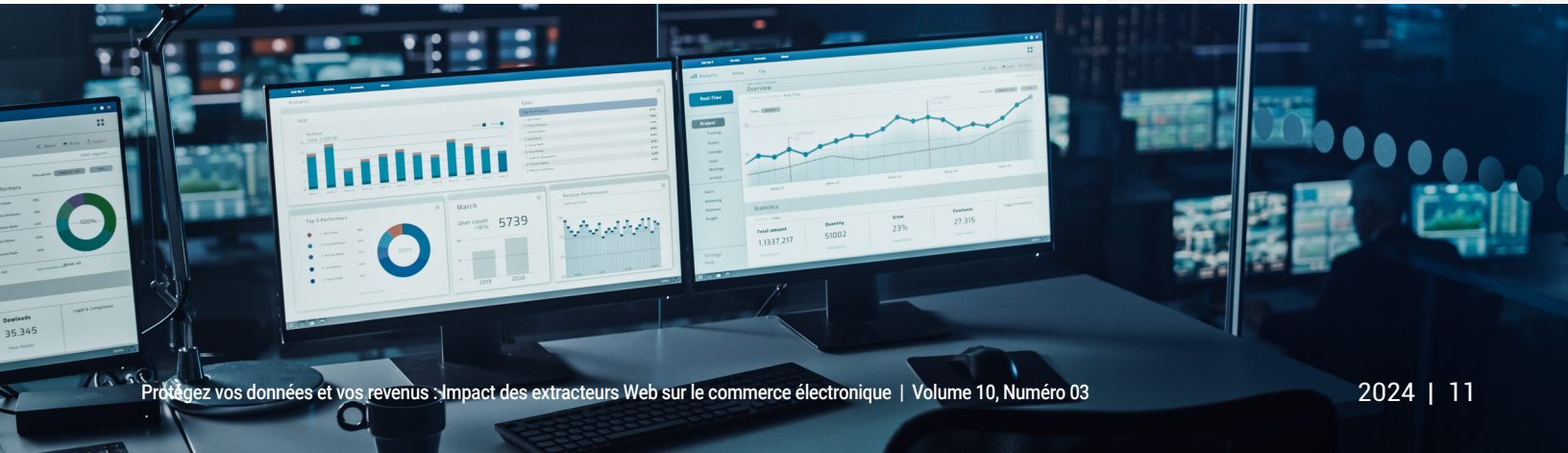


Figure 5 : Représentation d'un botnet d'IA et de son processus en trois étapes

Examinons ces trois étapes plus en détail pour mieux comprendre ce qu'elles impliquent.

Collecte de données

L'extraction Web implique l'organisation des données extraites d'un ou de plusieurs sites Web, afin que les organisations puissent produire de nouveaux ensembles de données qui peuvent être appliqués et analysés comme elles l'entendent. Et cela commence par la collecte des données.



La collecte de données consiste en une extraction adaptative combinée à des technologies « anti-interdiction » ou de « détection anti-bots » pour fonctionner rapidement et sans encombre. Ces technologies sont conçues comme des arbres de décision pour détecter les différents aspects des protections qui peuvent être en place. La résilience est ici le mot d'ordre. La protection des bots peut inclure l'empreinte JavaScript, l'empreinte HTTP et TLS (évaluation des en-têtes HTTP et de la négociation TLS) et la détection de la réputation du protocole Internet (IP) (Figure 6). Certains de ces flux de travail peuvent inclure l'apprentissage automatique, notamment lors de la collecte de statistiques sur le taux de réussite, l'adaptation à la stratégie des cookies, l'en-tête HTTP et les paramètres TLS, et l'évaluation du code d'empreinte JavaScript. C'est également à ce stade qu'un navigateur sans interface peut entrer en jeu.

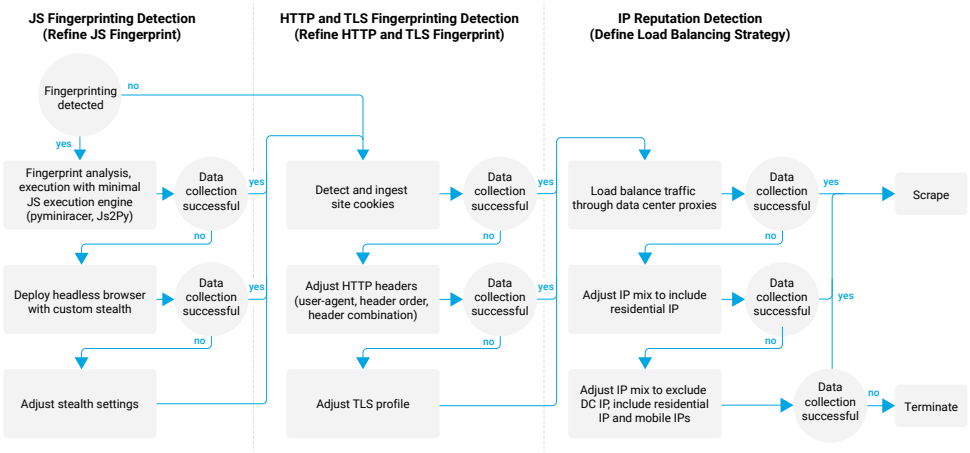


Figure 6 : Lorsqu'il tente de collecter des données, cet arbre de décision de détection anti-bots tente d'éviter l'empreinte JavaScript, l'empreinte HTTP et TLS et la détection de la réputation IP

Navigateur sans interface

Un navigateur sans interface est un navigateur Web qui ne possède pas d'interface utilisateur graphique. Par conséquent, cela signifie que les humains ne peuvent pas interagir directement avec la page Web sur laquelle le navigateur sans interface apparaît, et que le navigateur est exécuté via une interface de ligne de commande ou par une communication réseau. Dans le cas de Selenium, un navigateur sans interface open source populaire, le navigateur est automatisé et largement utilisé pour l'extraction Web. Cette caractéristique peut d'ailleurs s'avérer très utile pour les chercheurs de données qui tentent de récupérer du contenu dynamique.

Les navigateurs sans interface peuvent également permettre de copier efficacement des captures d'écran et le code du site Web, mais aussi d'extraire les données choisies sans rendre la page entière. Cependant, les attaques par navigateur sans interface sont coûteuses et peuvent parfois encore être détectées à cause des empreintes digitales qu'elles laissent derrière elles. Les dépenses liées à d'autres infrastructures sophistiquées sont toutefois similaires à celles des navigateurs sans interface, c'est-à-dire généralement élevées.

Extraction et traitement des données

Les informations extraites sont généralement constituées de contenu HTML et JSON. Parmi toutes les données extraites, seule une fraction peut être utile à l'analyse. Par exemple, l'analyse de la concurrence porte généralement sur les prix, les remises, les stocks et les numéros d'UGS, les catégories et les descriptions des produits. Les informations essentielles peuvent être extraites automatiquement par des modèles d'apprentissage automatique qui peuvent être entraînés à reconnaître de multiples structures et formats de données. Cela permet ainsi d'éviter le travail de traitement supplémentaire nécessaire pour extraire manuellement les données et d'avoir à étudier la structure du code de contenu HTML et JSON. De plus, il est également important de garder en tête que la structure du code du contenu peut changer au fur et à mesure que la conception du site évolue. Une logique d'apprentissage automatique supplémentaire est également nécessaire pour le traitement si plusieurs sites Web sont impliqués dans l'analyse.

Étude de cas : avantages des solutions de détection d'extraction Web

Les chercheurs d'Akamai ont observé un sous-ensemble de clients du commerce électronique protégés par une solution d'extraction Web qui détectait les activités d'extraction, et ont examiné la répartition de l'activité du trafic pendant une semaine. Cela représente un échantillon d'environ 6,9 milliards de requêtes. L'analyse n'a pris en compte que les requêtes HTML et AJAX. Le contenu statique (images, JavaScript, feuilles de style) n'a pas été inclus dans l'analyse dans la mesure où la plupart des bots ne requièrent pas de contenu statique ; dans le même temps, cette omission a également permis d'éviter de gonfler inutilement les données.

L'activité globale a été classée par Akamai Content Protector et comprenait 49,3 % de trafic humain à faible risque, 42,1 % de trafic de bots (27,5 % de bots malveillants à haut risque et 14,6 % de bots inoffensifs) et 8,7 % de trafic à risque moyen non classé (Figure 7).

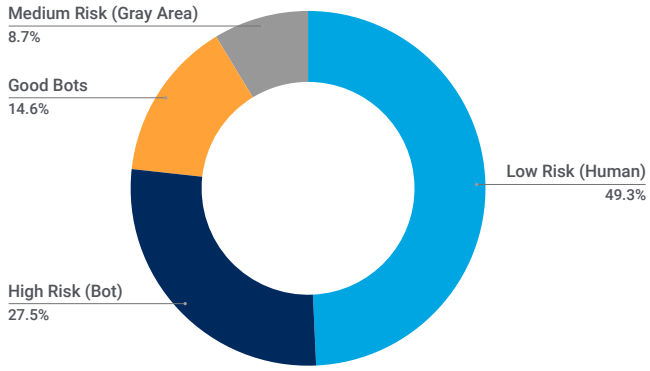


Figure 7 : Classification des activités de trafic

La figure 8 montre que sur les 42,1 % de trafic provenant de bots, 65,3 % proviennent de bots d'extraction considérés comme des bots malveillants, et les 34,7 % restants de bots d'extraction classés comme des bots inoffensifs (par exemple, moteurs de recherche sur le Web, SEO, réseaux sociaux et publicité en ligne).

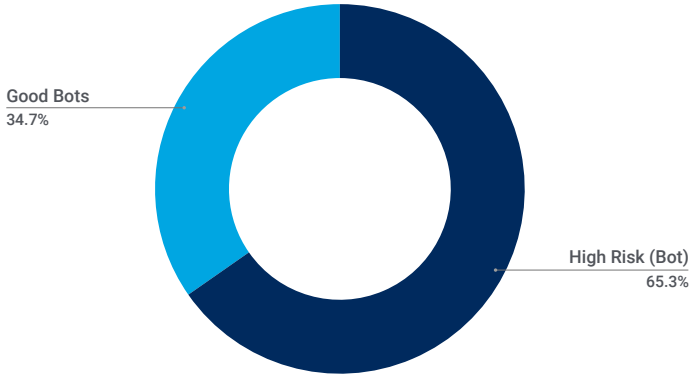


Figure 8 : Trafic de bots inoffensifs vs. trafic de bots malveillants

Les niveaux de sophistication ont également été mesurés pour les bots malveillants à haut risque ayant contribué à 65,3 % de l'ensemble du trafic de bots. 37 % de ce trafic provenait de botnets scriptés de base faciles à détecter avec des méthodes simples sans état, 47,6 % provenaient de botnets scriptés plus avancés qui nécessitent des méthodes de détection avec état plus avancées utilisant l'apprentissage automatique, et 15,5 % provenaient de navigateurs sans interface qui nécessitent des méthodes avancées d'empreintes JavaScript et de détection avec état (Figure 9).

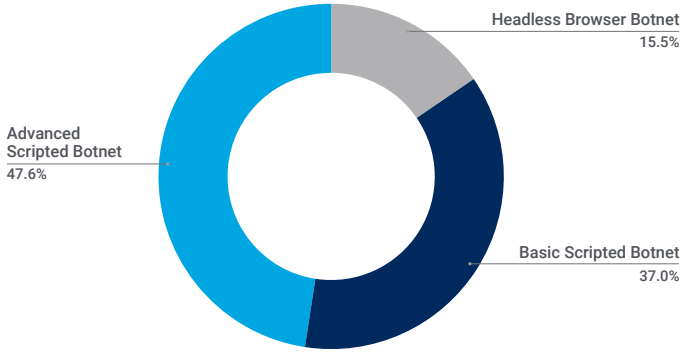
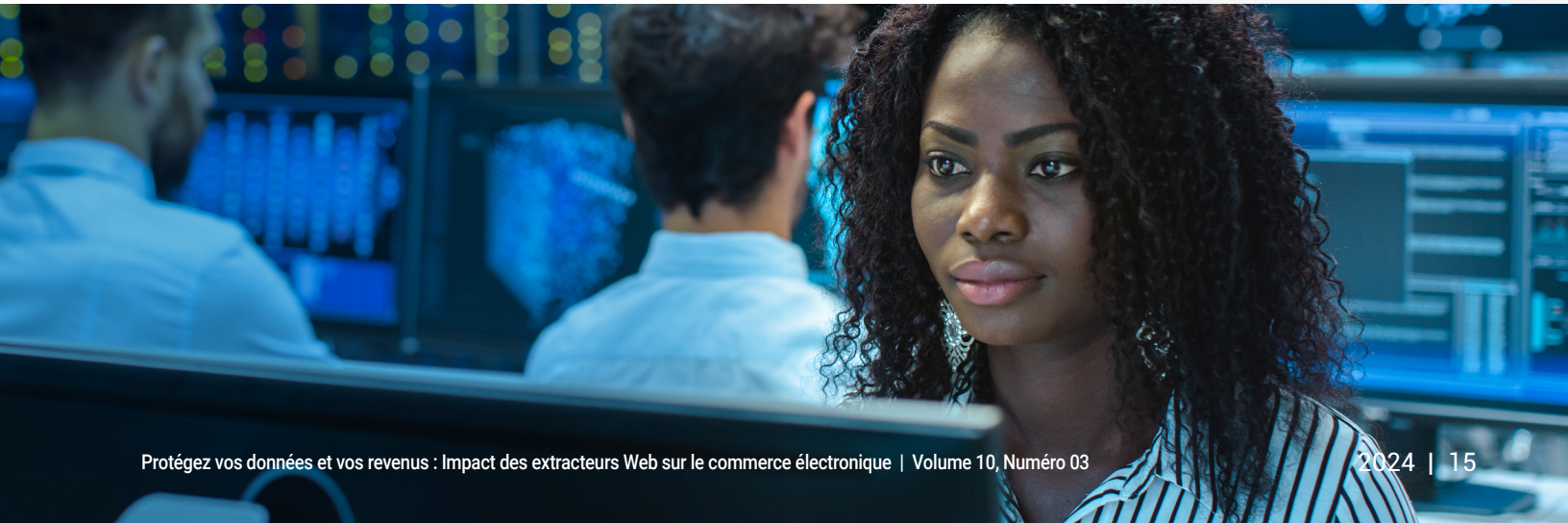


Figure 9 : Répartition du trafic des bots malveillants en fonction de leur degré de sophistication (les totaux ne sont pas égaux à 100 % en raison des arrondis)

Ces données montrent que les bots d'extraction malveillants sont nettement plus nombreux que les bots d'extraction inoffensifs et que près de la moitié du trafic global est constitué de bots, les botnets scriptés avancés produisant le plus de trafic de bots malveillants (47,6 %).

L'activité du site Web sera beaucoup plus rapide et efficace, et les indicateurs de celui-ci seront plus faciles à lire, une fois que les défenses contre ces bots auront été mises en place et que les bots d'extraction auront été supprimés. À terme, ces résultats se traduiront par une meilleure expérience pour les utilisateurs/clients. Comme le montre la figure 10, le nombre de requêtes de bots à haut risque a considérablement diminué une fois que les mesures d'atténuation ont été mises en place.



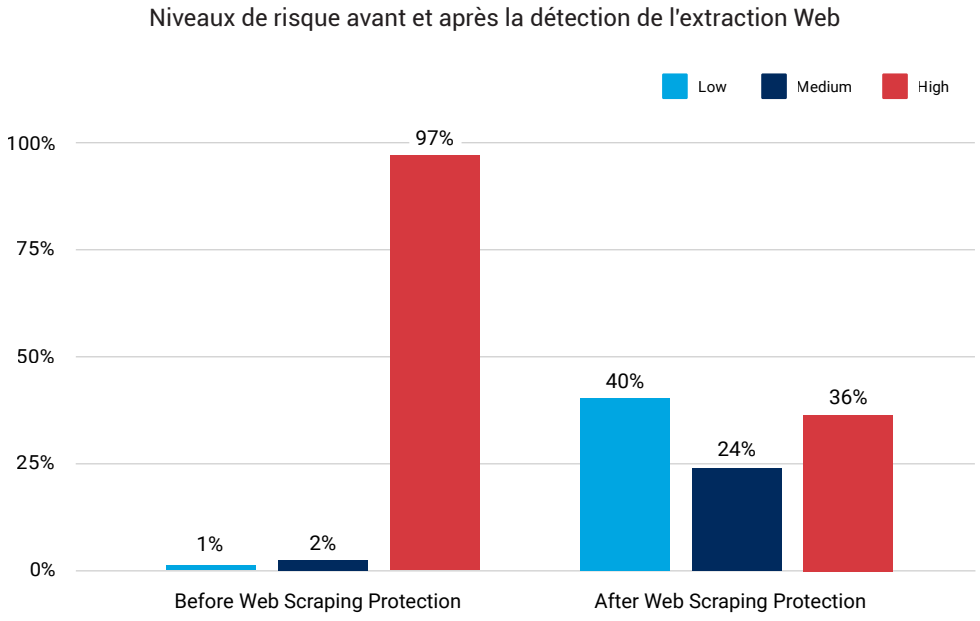


Figure 10 : Niveaux de risque avant et après atténuation avec Content Protector

Sauvegarde et atténuation des risques

Cette section fournit quelques indicateurs cruciaux pour la détection des bots d'extraction Web, ainsi que des informations sur les outils qui peuvent fournir des mesures défensives contre ces derniers.

Détection des bots d'extraction de base

Bien que les bots d'extraction sophistiqués puissent être difficiles à détecter, les solutions de gestion des bots peuvent empêcher la collecte de données par toutes sortes de bots d'extraction intrusifs et notamment rechercher les caractéristiques suivantes pour détecter les bots d'extraction Web les plus simples :

- Demandes annonçant des navigateurs et des versions de système d'exploitation plus anciens
- Anomalies dans la signature de l'en-tête HTTP
- Utilisation d'anciennes versions de HTTP (par exemple, v1.1) au lieu de HTTP v2, plus courant, ou du HTTP v3 émergent
- Demandes provenant de milliers de services cloud/centres de données

Détection de bots d'extraction plus avancés

Aucune des caractéristiques de la liste ci-dessus ne sera observable pour les bots d'extraction plus avancés. Voici donc quelques caractéristiques de bots d'extraction plus sophistiqués :

- Demandes provenant de la dernière version du navigateur et du système d'exploitation
- L'ensemble d'en-têtes HTTP est identique à celui du navigateur légitime
- Utilisation de HTTP v2
- Demandes provenant de centaines de milliers d'adresses IP résidentielles et mobiles

Identification des modèles de trafic

Certains indicateurs clés permettent d'identifier si le type de trafic d'un site Web est humain (Figure 11) ou dû à un bot de base (Figure 12) ou un bot sophistiqué (Figure 13).

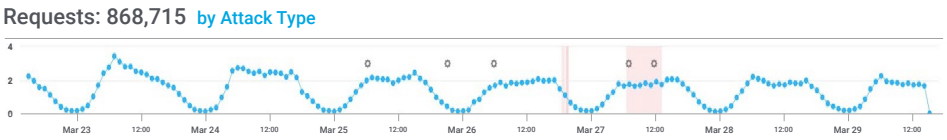


Figure 11 : Le trafic utilisateur légitime présente généralement un cycle d'activité circadien

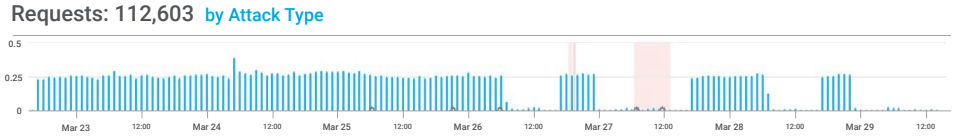


Figure 12 : Le trafic typique des bots montre une activité régulière avec des pauses occasionnelles

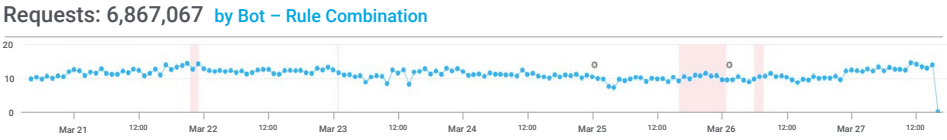


Figure 13 : Des bots plus sophistiqués affichent du trafic en continu jour et nuit

Nous constatons aussi souvent des botnets qui oscillent entre les trois et adoptent une stratégie d'équilibrage de charge faible, mais une stratégie d'empreinte sophistiquée (ou vice versa). Cependant, les botnets les plus avancés peuvent parfois être si sophistiqués qu'ils peuvent passer comme ayant une empreinte digitale parfaite ou même reproduire un modèle de trafic d'utilisateur légitime.

En plus d'être à l'affût de ces bots d'extraction, les outils de protection contre l'extraction Web, tels que Content Protector, peuvent offrir des avantages particuliers et une navigation plus fluide dans les environnements agités infestés de bots d'extraction. Les avantages peuvent inclure :

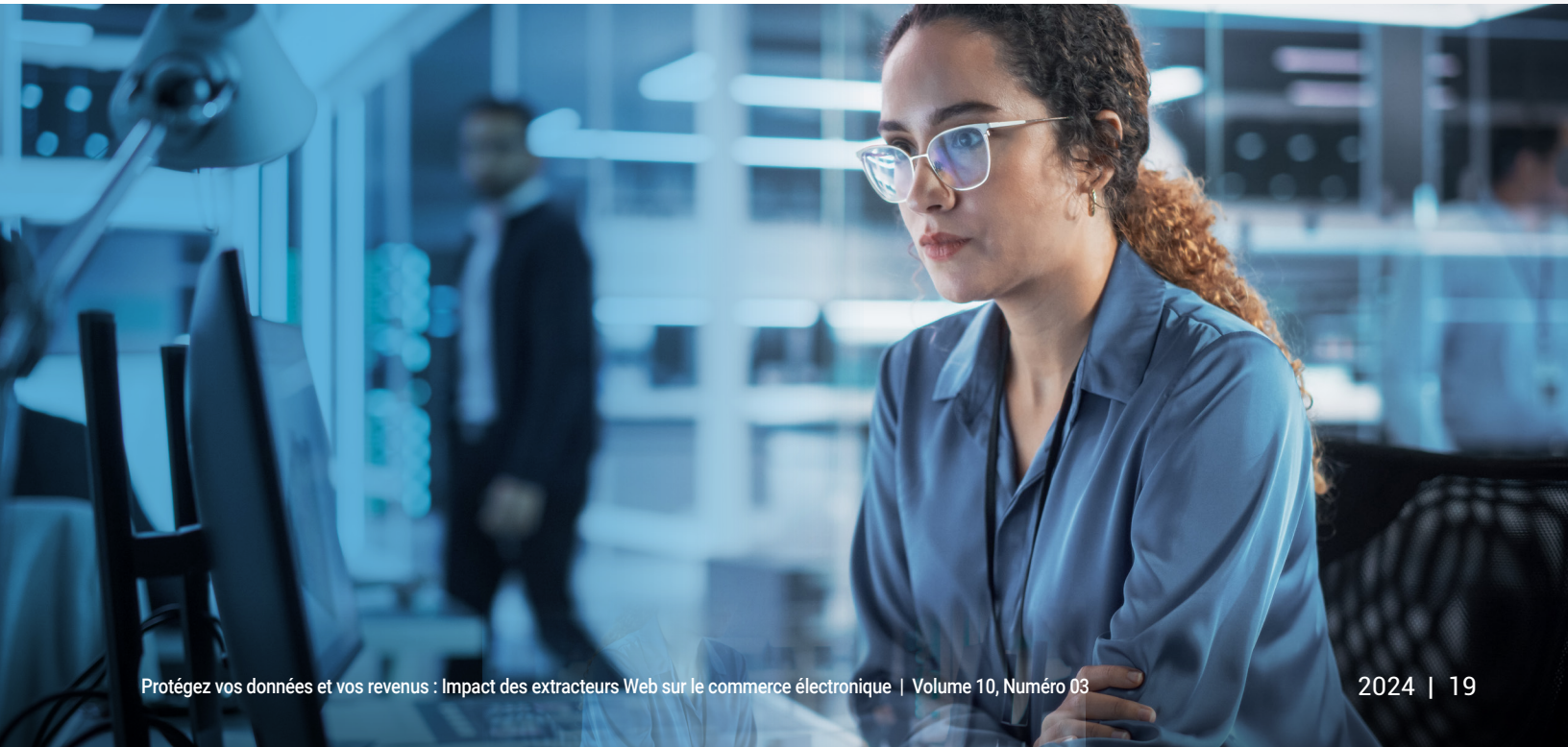
- des taux de conversion plus élevés et des coûts informatiques réduits ;
- des mesures plus précises, qui peuvent conduire à de meilleures décisions d'investissement et à une augmentation du chiffre d'affaires ;
- une réduction de la pression sur les prix, qui peut se traduire par des ventes épargnées par la concurrence ;
- des clients satisfaits de pouvoir accéder aux produits souhaités, et une augmentation des revenus provenant des opportunités de vente incitative lorsque les utilisateurs ajoutent des produits supplémentaires à leur panier une fois qu'ils ont sécurisé l'article convoité ;
- la préservation de la réputation de la marque, dans la mesure où les clients sont protégés contre les contrefaçons de mauvaise qualité qu'ils pensent être des produits légitimes du vendeur d'origine ;
- le maintien du chiffre d'affaires généré par les produits et la fidélisation des clients ;
- l'augmentation/la protection des revenus publicitaires ;
- la fidélisation du public et des visiteurs du site.

Considérations relatives à la conformité

La norme de sécurité des données de l'industrie des cartes de paiement (PCI DSS) v4.0 est maintenant en vigueur, et bon nombre des changements ont été motivés par des tendances de menaces qui ont encore un impact sur les entreprises. La visibilité est essentielle pour faire face à ces attaques. Qu'elles visent votre environnement JavaScript historique ou les API utilisées pour faciliter la transformation, il est essentiel de détecter rapidement ces attaques et d'y remédier.

Nous observons également des tendances émergentes en matière de conformité dans la nouvelle version 2.0 du cadre de cybersécurité du NIST, qui a ajouté une fonction de gouvernance. Le NIST constitue une base pour un certain nombre de réglementations gouvernementales et est présent dans de nombreux cadres de cybersécurité commerciaux. C'est donc le moment idéal pour examiner les nouvelles directives et les utiliser pour mettre à jour vos politiques ou pour cartographier votre documentation actuelle afin de repérer les lacunes.

Pour les entreprises cotées en bourse et celles qui utilisent des principes comptables généralement acceptés (PCGR), un autre domaine de conformité est l'importance relative à la cybersécurité. La nécessité de définir les risques et les menaces importants implique la collaboration de toute l'équipe de direction. Une fois que vous avez identifié les menaces importantes (comme les ransomwares), vous devez définir des mesures d'atténuation (comme la microsegmentation). Veillez à ce que vos plans de gestion de crise tiennent compte des délais de divulgation et à ce que vous disposiez d'un cahier des charges pour le pire des scénarios, pour lequel vous devriez déposer un formulaire 8-K relatif à un incident cybernétique auprès de la Security and Exchange Commission (Commission de la sécurité et des échanges).



Conclusion

Nous espérons que ce rapport vous aura donné un aperçu d'un domaine qui pourrait avoir un impact économique négatif sur votre organisation. Les bots affectent vos sites, et ce de manière de plus en plus importante, et il est essentiel d'optimiser les bots bénéfiques, d'atténuer les bots malveillants et de garantir une expérience client globale fluide. Il s'agit d'un problème de sécurité qui a des répercussions sur l'entreprise. Comme pour tous les problèmes de sécurité, la première étape consiste à gagner en visibilité, la deuxième à analyser l'impact et la dernière à déterminer le retour sur investissement pour les risques et les revenus afin de mettre en œuvre les contrôles de sécurité appropriés.

Vous ne pouvez pas protéger ce que vous ne pouvez pas identifier, le moment est donc venu de déterminer où se situent vos lacunes en matière de visibilité. Pour ce faire, vous devez déterminer le niveau d'activité de l'extraction Web sur vos sites ainsi que son intention. Le paysage des bots est composé de bots inoffensifs et malveillants, et en fonction de leur utilisation, les bots d'extraction appartiennent aux deux catégories. Bien que la frontière entre les extracteurs bénéfiques et les nuisibles puisse être floue, l'évolution de la sophistication des bots (par exemple, les bots d'extraction Web menant des attaques de navigateur sans interface) se poursuit. Tout cela s'accompagne d'un impact considérable des bots d'extraction sur les coûts informatiques et sur l'expérience des clients des entités de commerce électronique. Il est primordial de veiller à ce que vous disposiez des outils nécessaires pour analyser l'activité des bots et leur impact sur votre site.

L'objectif est de bloquer les attaquants afin qu'ils ne puissent pas exécuter leur modèle commercial criminel sur vos sites et commettre toute une série d'activités malveillantes, comme encaisser des points de fidélité, passer des commandes frauduleuses ou même effectuer des retours frauduleux. Le but est aussi d'empêcher les bots de billetterie d'acheter des places pour des événements limités ou les bots d'achat de commander des produits à la mode. Les bots peuvent être utilisés pour faciliter l'ouverture abusive de nouveaux comptes en profitant d'offres spéciales, ce qui a une incidence sur l'analyse et les coûts de la campagne. Les grands botnets de déni de service distribué (DDoS) peuvent submerger les applications Web et provoquer une mauvaise expérience utilisateur ou l'impossibilité de passer des commandes ou de faire des réservations, ce qui entraîne des pertes de revenus et nuit aux clients. Les bots peuvent également imiter le comportement humain en ligne pour augmenter les clics et le trafic sur un site Web, faussant ainsi les analyses marketing et de performance d'expériences digitales soigneusement élaborées. Vous ne voulez certainement rien de tout cela.

Comme nous l'avons indiqué précédemment, plus de la moitié du trafic commercial mondial sur le Web est constitué de bots, et les niveaux de trafic des bots continuent d'augmenter. Akamai a basé les idées et les conseils de ce rapport sur notre plateforme de sécurité, qui comprend la protection du contenu et la défense contre l'extraction Web. Nous travaillons en partenariat avec de nombreux leaders du commerce électronique, et nous avons donc voulu partager les mesures de protection et d'atténuation que les entreprises peuvent utiliser pour protéger au mieux leurs clients. Nous prévoyons une augmentation de l'utilisation, des choix de niveau de service et des types de bots d'extraction disponibles. Il est donc nécessaire d'évaluer continuellement la posture de risque de votre entreprise et de déterminer si vos contrôles de sécurité actuels répondent à l'appétence pour le risque de vos dirigeants.

Tenez-vous au courant de nos dernières recherches en consultant notre centre de recherche sur la sécurité.

Méthodologies

Données Content Protector

Cet échantillon de données décrit les classifications de niveau de risque que notre outil Content Protector attribue au trafic qu'il surveille. Ces classifications sont utilisées pour détecter les activités d'extraction des bots et pour déterminer si nous avons affaire à un bot inoffensif ou malveillant. Dans la mesure où la plupart des bots ne requièrent pas de contenu statique, cette analyse n'a pris en compte que les requêtes HTML et AJAX afin de ne pas gonfler inutilement les données.

Cet échantillon de données couvrait une période d'une semaine allant du 12 au 19 avril 2024. La taille totale de notre échantillon comprenait plus de 6,5 milliards de requêtes.

Attaques de bots

Ces données décrivent les alertes de la couche applicative sur le trafic vu à travers notre pare-feu d'application Web (WAF) et notre outil de gestion des bots. Les alertes de bots sont déclenchées lorsque nous détectons une charge utile de bot dans une requête adressée à un site Web, à une application ou à une API protégé(e). Ces alertes de bot peuvent être déclenchées par des bots malveillants et inoffensifs. En revanche, elles n'indiquent pas si ces attaques sont fructueuses. Bien que ces produits permettent un haut niveau de personnalisation, nous avons recueilli les données présentées ici d'une manière qui ne tient pas compte des configurations personnalisées des propriétés protégées. Les données sont issues d'un outil interne d'analyse des événements de sécurité détectés sur Akamai Connected Cloud, un réseau d'environ 340 000 serveurs répartis sur plus de 4 000 sites et environ 1 300 réseaux dans plus de 130 pays. Nos équipes de sécurité utilisent ces données, qui se mesurent en pétaoctets par mois, pour étudier les attaques, signaler des comportements malveillants et fournir des informations supplémentaires aux solutions Akamai.

Ces données couvraient une période de 15 mois, du 1er janvier 2023 au 31 mars 2024.

Akamai

Crédits

Rédacteur en chef

Lance Rhodes

Édition et rédaction

David Senecal
Maria Vlasak

Révision et expertise

Mitch Mayne
Susan McReynolds
Christine Ross
Badette Tribbey
Steve Winterfeld

Analyse des données

Chelsea Tuttle

Documents promotionnels

Annie Brunholzl

Marketing et publication

Georgina Morales
Emily Spinks

Autres rapports État des lieux d'Internet/Sécurité

Lisez les numéros précédents et surveillez les prochaines parutions du célèbre rapport État des lieux d'Internet/Sécurité d'Akamai, **akamai.com/soti**

D'autres recherches sur les menaces d'Akamai

Tenez-vous au courant des dernières analyses d'informations sur les menaces, des rapports de sécurité et des recherches sur la cybersécurité sur **akamai.com/threatresearch**

Accéder aux données de ce rapport

Consultez des versions de haute qualité des graphiques et des tableaux référencés dans ce rapport. Ces images sont libres d'utilisation et de référence, à condition qu'Akamai soit dûment crédité en tant que source et que le logo Akamai soit conservé. **akamai.com/sotidata**

En savoir plus sur les solutions Akamai

Pour en savoir plus sur les solutions Akamai de détection et de protection contre les bots d'extraction Web, consultez notre **page Content Protector**.

Akamai

Akamai protège votre expérience client, votre personnel, vos systèmes et vos données en vous aidant à intégrer la sécurité dans tout ce que vous créez, dans toutes vos conceptions et diffusions. La visibilité de notre plateforme sur les menaces mondiales nous aide à adapter et à faire évoluer votre stratégie de sécurité (activer le Zero Trust, bloquer les ransomwares, sécuriser les applications et les API, ou lutter contre les attaques DDoS), vous donnant ainsi la confiance nécessaire pour innover, prospérer et ouvrir le champ des possibles. Pour en savoir plus sur les solutions de Cloud Computing, de sécurité et de diffusion de contenu d'Akamai, rendez-vous sur akamai.com et akamai.com/blog, ou abonnez-vous à Akamai Technologies sur X (anciennement Twitter) et LinkedIn. Publication : 06/24.