SOTI

10 YEARS
OF SECURITY INSIGHT

10권, 3호

수익성을 떨어뜨리는 스크레이핑:

웹 스크레이퍼가 이커머스에 미치는 영향

Akamai

인터넷 보안 현황

목차

봇이 전체 웹 트래픽의 절반 이상을 발생시킨다는 사실을 알고 계셨나요? 특히 웹 애플리케이션과 자산으로 매출을 창출하는 커머스 업계는 리스크가 높은 봇 트래픽의 영향을 가장 많이 받아왔습니다(그림 1). 봇이 진화하고 있다는 소식은 자주 들리지만, 웹 스크레이퍼 봇은 다른 종류의 봇과 달리 표면 아래에 숨겨진 경제적 영향이 크기 때문에 오늘날 이커머스 중심 기업의 관심을 끌고 있습니다. 또한, 스크레이퍼 봇은 AI(Artificial Intelligence) 봇넷과 헤드리스 브라우저 기술의 등장으로 인해 탐지하기가 극도로 어려워졌습니다. 예를 들어, Akamai의 이커머스 고객사 중 한 곳은 리스크가 높은 트래픽의 99%를 차단했는데 이 트래픽이 스크레이퍼 봇에서 발생하는지 몰랐습니다.

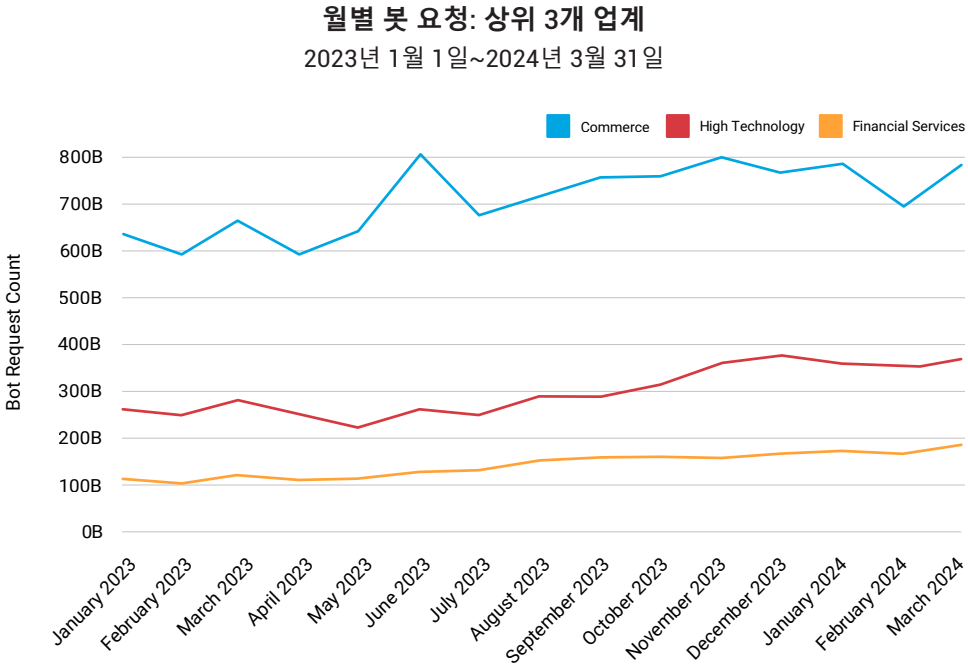

그림 1: 커머스는 봇 요청이 가장 많은 업계로, 커머스 업계의 글로벌 봇 트래픽이 2023년 초부터 2024년 1분기까지 증가한 것으로 관측되었습니다.

따라서 이번 SOTI(State of the Internet) 보고서에서는 이러한 봇과 그 운영자의 진화 및 전문화에 집중합니다. 봇은 오래전부터 존재해 왔으며 여러 그룹은 범죄 공격, 사기 계획, 경쟁업체 정보 획득을 위해 지속적으로 봇을 활용합니다. 최근에는 모든 봇의 사용이 증가하고 스크레이퍼 봇이 비즈니스에 미치는 부정적인 영향이 증가하는 트렌드를 보입니다. 이 보고서는 기술적 인사이트와 공격 방법론을 공유해 이 문제에 대한 커머스 업계 전반의 인식을 제고하고자 작성되었습니다.

봇: 정상 봇, 악성 봇, 유해한 봇

모든 대표적인 이커머스 기업은 봇으로 인해 어려움을 겪고 있습니다. 봇은 지속적으로 진화하고 있으며 달성하고자 하는 목표에 맞춰 전문화되고 있습니다. 커머스 업계에는 다양한 작업을 수행하는 수많은 종류의 봇이 존재합니다. 쉽게 생각하자면 정상 봇, 악성 봇, 회색 봇의 3개 그룹으로 나눌 수 있습니다. 정상 봇은 고객이 사이트를 찾게 도와주지만, 악성 봇은 악의적인 목적을 위해 사이트를 스크레이핑합니다. 회색 봇(지속적으로 핑을 하는 파트너 봇, 자주 호출하는 기타 프로그램 API 등)은 정상적이고 실제로는 정상 봇의 하위 범주이지만 소음을 발생시킵니다.

따라서 사용자의 기본적인 질문에 답하고 더 정확한 검색 결과를 보여주는 웹사이트 콘텐츠를 제공하는 등 긍정적인 영향을 미칠 수 있는 유용한 챗봇과 검색 엔진 봇을 고려했을 때, Akamai는 이러한 종류의 봇을 최적화하면서 IT 비용을 절감하고자 합니다. 고객 계정에 무단으로 접속해 계정을 탈취하는 크리덴셜 스터핑 봇과 같은 유해한 봇의 경우, Akamai는 전반적인 고객 경험에 영향을 주지 않으면서 예방 조치를 취하고자 합니다. 최근 특히 문제가 되는 봇 종류는 매출 감소, 충성도 저하, 비용 증가를 발생시키는 웹 스크레이퍼 봇입니다.

인터넷 웹사이트의 데이터와 콘텐츠를 직접 가져오기 위해 사용하는 봇넷인 스크레이퍼 봇은 독특합니다. 다른 봇과 다르게 작동하고 비즈니스에 미치는 영향과 탐지 방식이 다르기 때문에 주의가 요구됩니다. 또한, 웹 스크레이퍼는 기업과 운영자가 수집한 정보로 수익을 창출하는 방식에 따라 사용 사례가 다양하다는 점에서 다면적입니다. 특정 목표와 관계없이 스크레이퍼는 매출 감소, IT 비용 증가, 전반적인 고객 경험 저하를 야기합니다.

이번 SOTI 보고서에서는 스크레이핑이 이커머스 전반에 미치는 영향을 살펴보고 비즈니스 소유자(디지털, 마케팅, 브랜드, 재무, 리스크, 보안 등)가 악성 스크레이퍼를 차단하는 데 공동의 관심을 가져야 하는 이유를 살펴봅니다. 이러한 영향을 더 잘 이해하려면 웹 스크레이퍼 봇이 진화한 이유, 사용 목적, 작동 방식, 영향, 커머스 기업이 취할 수 있는 조치에 대한 전체 그림을 살펴보는 것이 중요합니다.

보고서의 핵심 인사이트

웹 스크레이핑은 단순한 사기나 보안 문제가 아니라 비즈니스 문제이기도 합니다. 스크레이퍼 봇은 매출, 경쟁력, 브랜드 정체성(BI), 고객 경험, 인프라 비용, 디지털 경험 등 기업의 여러 측면에 부정적인 영향을 미칩니다.

Akamai의 사례 리서치에 따르면, 전체 트래픽 활동의 42.1%가 봇에서 발생했으며, 이 중 65.3%가 악성 봇에서 발생한 트래픽이었습니다. 그리고 악성 봇 트래픽의 총 63.1%는 고급 기법을 사용했습니다.

헤드리스 브라우저 기술은 스크레이퍼 환경을 변화시켰으며, 이러한 종류의 봇 활동을 관리하려면 자바스크립트 기반의 방어 조치보다 더 정교한 접근 방식을 취해야 합니다.

스크레이핑의 의도가 악의적이든 유익하든 상관없이 그로 인해 기업이 직면하는 기술적 영향에는 웹사이트 성능 저하, 사이트 지표 오염, 피싱 사이트의 감염된 인증정보 공격, 컴퓨팅 비용 증가 등이 포함됩니다.

웹사이트에 사람, 기본 봇 또는 정교한 봇 트래픽이 발생하는지 파악하려면 다양한 트래픽 패턴을 관찰하고 이해하는 것이 중요합니다. 패턴은 주기적, 간헐적, 지속적 등 다양합니다.

정상 봇과 악성 봇 비교

기본 개념부터 시작해 봅시다. '로봇'의 줄임말인 봇은 자동화된 작업을 사람보다 더 빠르고 정확하게 수행할 수 있는 컴퓨터 프로그램입니다. 봇의 다양한 역할과 종류는 크게 정상 봇과 악성 봇의 2개 범주로 나뉩니다(그림 2). 회색 봇은 정상 봇의 하위 범주이지만 비교를 단순화하기 위해 지금은 정상 봇과 통합하겠습니다.

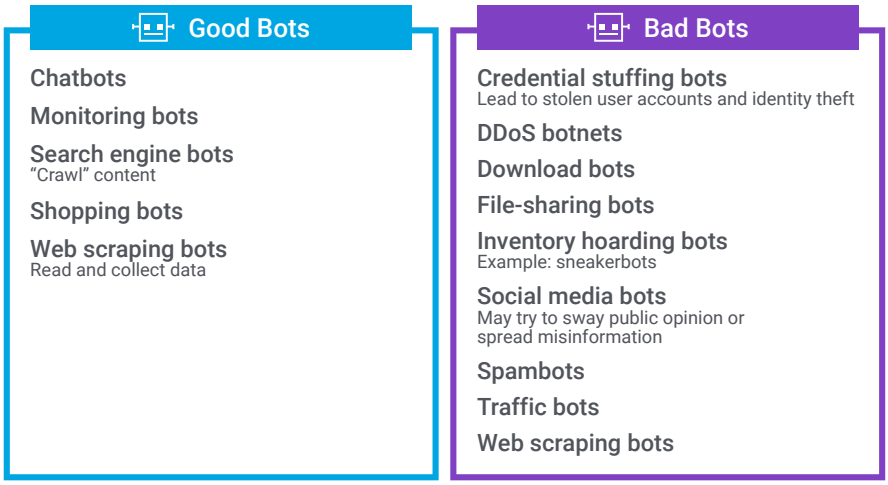

그림 2: 정상 봇과 악성 봇 비교표(예시 포함)

정상 봇은 툴과 서비스를 제공하는 데 도움이 되는 유용한 봇이지만, 악성 봇은 사이버 범죄자나 사기꾼이 악의적인 의도를 가지고 사용하는 경우가 많습니다. 이런 악성 봇의 예로는 온라인에서 사람의 행동을 모방해 웹사이트의 클릭 수와 트래픽을 늘리는(광고 사기 등) 트래픽 봇이 있습니다.

웹 스크레이핑 봇은 정상 봇과 악성 봇 범주에 모두 포함되며, 기업이 수집한 정보를 사용하는 방식에 따라 구분합니다. 이제 세계 최대 리테일 기업 및 이커머스 브랜드 중 일부가 직면한 스크레이퍼 봇의 좋은 영향과 나쁜 영향과 관련된 다양한 사용 사례를 자세히 살펴보겠습니다.

스크레이핑 기초

웹 스크레이핑은 이커머스 기업에서 일반적으로 사용됩니다. 예를 들어, 여행 및 숙박 업계에서는 여행 애그리게이터들이 호텔 및 항공사 파트너로부터 동적 콘텐츠를 스크레이핑해 예약 가능 여부와 가격을 최신 상태로 유지합니다. 이러한 종류의 스크레이핑이 예상되며, 기업은 일반적인 봇 제어를 사용해 실제 사용자가 예약을 원하는 시간대에 스크레이퍼의 속도를 조절합니다. 또한, 기업은 데이터 추출 서비스 공급업체를 통해 경쟁업체로부터 리드 및 기타 관련 정보를 수집하기도 합니다. 게다가 스크레이핑 봇은 데이터를 분석하고 트렌드를 파악하는 데 사용할 수 있습니다. 스크레이핑은 온라인 제품 및 서비스를 개선하고 잠재적인 소비자가 검색 엔진 등을 통해 기업 제품을 더 쉽게 찾도록 사이트를 검토하는 데에도 유용할 수 있습니다. 이러한 모든 작업은 기업이 경쟁 우위를 확보하는 데 도움이 될 수 있습니다. 하지만 많은 기업이 그다지 바람직하지 않은 이유로 스크레이퍼를 사용하고 있다는 사실을 부인할 수 없습니다.

스크레이핑에 직면한 고객들

안타깝게도 피싱 사기의 희생양이 된 소비자에 대한 이야기가 종종 들립니다. 이 경우 스크레이퍼 봇이 제품 이미지, 설명, 가격 정보를 수집해 인증정보나 신용 카드 정보를 훔치기 위한 위조된 상점 또는 피싱 사이트를 만드는 데 사용되었을 수 있습니다. 이러한 피싱 및 위조 사이트는 브랜드 사칭의 한 형태로, 피해 기업의 지적 재산이 잠재 고객과의 신뢰를 구축하는 데 사용됩니다.

세계 최대 규모의 이커머스 브랜드 중 일부는 브랜드 사칭 캠페인의 일환으로 위조 사이트, 피싱 캠페인, 회사 웹 데이터 도난으로 인해 피해를 보았습니다(그림 3). 안타깝게도 피싱 사이트가 성공하면 정상적인 브랜드는 고객 신뢰와 충성도를 잃게 되어 그 여파를 고스란히 떠안습니다.

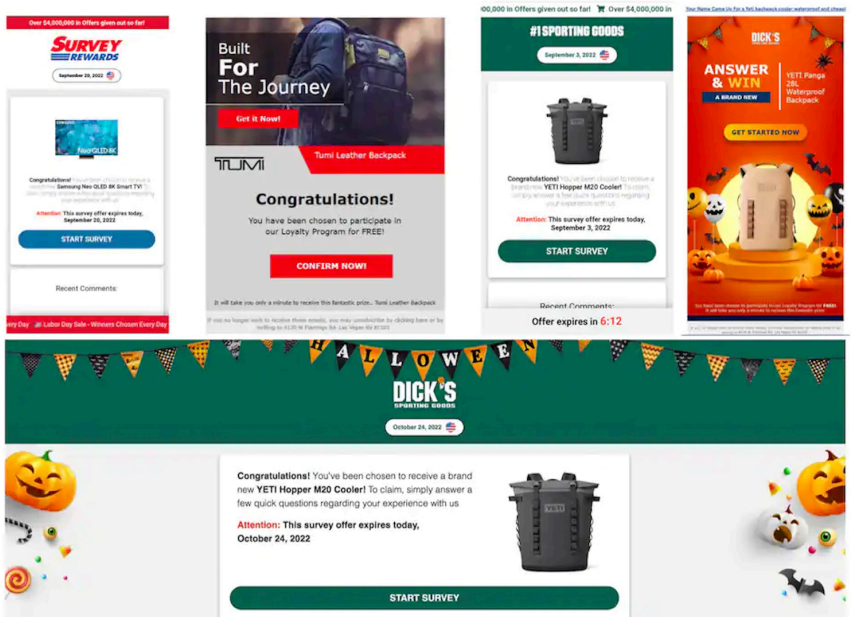

그림 3: 브랜드 사칭의 피해를 본 주요 이커머스 기업의 예

스캘핑은 웹 스크레이핑으로 인해 발생할 수도 있는데, 스캘퍼는 사이트에서 구매 가능한 제품을 스크레이핑해 정상적인 고객이 기회를 얻기 전에 구매할 수 있기 때문입니다(그림 4).

스크레이퍼 사용 사례

콘텐츠 스크레이핑으로 돈을 버는 공격자

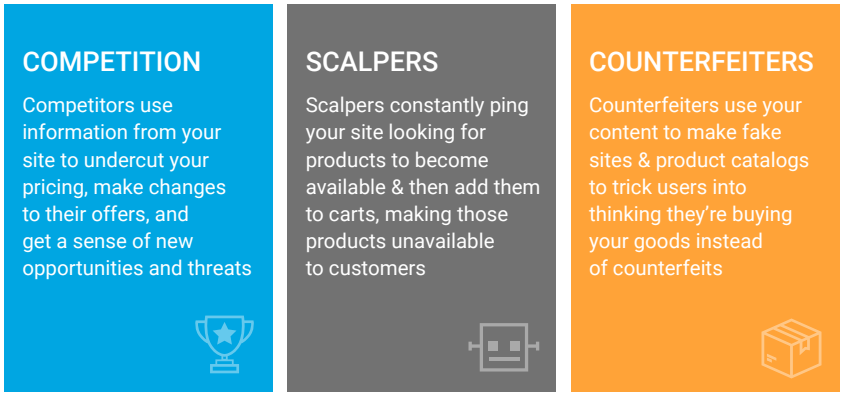

그림 4: 스크레이퍼 사용 사례

이러한 종류의 유해한 스크레이핑 활동을 수행하는 공격자는 자신의 악의적인 목적이 피해자에게 미치는 영향을 알고 있습니다. 여기에는 경쟁업체 정보 및 스파이 활동, 재고 비축 및 스크레이핑, 위조 및 모조품 사이트·상품, 미디어 사이트 스크레이핑 및 재게시 등의 부정적 영향이 포함됩니다(표 1). 그리고 스크레이퍼 봇의 사용을 명시적으로 금지하는 현행 법률은 없습니다.

영향	설명
경쟁을 위한 정보 수집·스파이 활동	경쟁업체는 한 기업의 사이트 정보를 이용해 가격을 낮추고, 오퍼를 변경하고, 새로운 기회와 위협을 파악합니다.
재고 비축 및 스크레이핑	스캘퍼는 표적 사이트를 지속적으로 핑해 구매할 수 있는 제품을 찾은 다음 장바구니에 추가해 실제 고객이 해당 제품을 사용할 수 없게 만듭니다.
위조 및 사칭 사이트·상품	위조범들은 스크레이핑한 콘텐츠를 사용해 가짜 사이트와 제품 카탈로그를 만들어 사용자가 위조품이 아닌 정상적인 제품을 구매한다고 속입니다.
미디어 사이트 스크레이핑 및 재게시	공격자는 뉴스 기사, 블로그, 기타 콘텐츠를 스크레이핑해 자신의 사이트에 게시함으로써 원래 기업의 방문자와 잠재적인 광고 매출을 잃게 만들 수 있습니다. 광고비는 사이트 방문자 수/잠재 고객을 기반으로 하는 경우가 많으므로 방문자 수가 줄면 미디어 사이트가 높은 광고비로 얻을 수 있었던 매출을 잃게 됩니다.

표 1: 웹 스크레이퍼로 인한 의도적인 부정적 영향

웹 스크레이핑의 일반적인 부작용

웹 스크레이핑의 의도와 상관없이 기업은 웹 스크레이핑의 부작용으로 인한 비용을 감당해야 합니다. 유익한 스크레이핑 서비스에 비용을 지불하는 기업들도 있지만, 스크레이핑을 당하는 기업은 자체적으로 비용을 부담해야 합니다. 여기에는 안티봇 솔루션에 대한 비용과 사이트 성능 저하 및 주요 지표 오염으로 인한 부정적인 경제적 영향이 포함됩니다(표 2).

영향	설명
봇 트래픽을 처리하기 위한 서버, CDN, 클라우드 비용 증가	이는 매출에 영향을 미치고 경쟁업체, 공격자, 위조범의 콘텐츠 사용으로 인한 평판 손실을 초래합니다.
사이트 성능 저하	스크레이퍼 봇은 중지될 때까지 지속적으로 실행되기 때문에 기업이 원치 않는 봇 트래픽을 처리하고 사이트 속도 및 앱 성능 저하와 같은 사용자 경험 저하로 인해 서버 및 전송 비용이 증가합니다.
주요 지표 오염	탐지되지 않은 봇 활동은 비즈니스팀이 제품 포지셔닝 전략 및 마케팅 캠페인과 같은 투자 결정을 내릴 때 의존하는 사이트 전환과 같은 주요 지표를 심각하게 왜곡합니다.

표 2: 웹 스크레이퍼로 인한 의도치 않은 부정적 영향

스크레이핑의 활용: 써드파티 웹 스크레이핑 서비스

앞서 언급했듯이 웹 스크레이퍼 봇은 좋은 목적이나 나쁜 목적으로 사용될 수 있습니다. 악성 봇으로 알려져 정당하게 차단되는 크리덴셜 스터핑 공격에 사용되는 봇과 달리, 정상적인 웹 스크레이핑 봇을 제공하는 회사도 있습니다. 많은 기업이 이러한 써드파티 웹 스크레이핑 서비스를 사용해 데이터를 추출하고 자사에 제공하는데, 이는 특히 경쟁이 치열한 마케팅 분야에서 유용할 수 있습니다.

다양한 종류의 웹 스크레이핑 및 데이터 추출 서비스를 제공하는 수십 개의 기업이 있으며, 심지어 이를 홍보하는 콘퍼런스도 있습니다. 예를 들어, Bright Data는 기업이 데이터 스크레이핑 방법을 배우도록 봇 탐지 회피에 관한 전문가가 한자리에 모이는 ScrapeCon이라는 콘퍼런스를 개최합니다. 표 3에는 써드파티 웹 스크레이핑 회사가 제공할 수 있는 서비스 수준의 예시가 나와 있습니다.

서비스 수준 1	프록시 서비스가 스크레이핑의 일부가 될 수 있으며 데이터 센터의 모바일 IP 및 거주지 주소가 포함될 수 있는 인프라를 제공할 수 있습니다.
서비스 수준 2	이 두 번째 수준에는 고객의 데이터 과학 팀원이 유용한 인텔리전스를 추출하여 비즈니스 의사 결정을 내리는 데 쉽게 사용하도록 데이터를 정리하고 구조화하는 자동화된 데이터 추출도 포함될 수 있습니다.
서비스 수준 3	가장 높은 레벨은 실제 비즈니스 인텔리전스 자체의 추출을 추가해 비즈니스의 의사 결정 프로세스를 더 향상시킬 수 있습니다. 이를 'AI 봇넷'이라고 합니다.

표 3: 써드파티 웹 스크레이핑 회사가 제공하는 다양한 서비스 수준

고객은 가장 기본적인 서비스부터 고급 서비스까지 서비스 수준과 데이터 수집 빈도를 선택할 수 있으며, 대상을 지정할 수도 있습니다. 종종 제공되는 서비스 수준 또는 선택되는 봇넷은 극복해야 하는 보안 수준에 따라 달라집니다. 더 기본적인 봇넷은 트래픽 부하를 분산하는 데이터 센터에 위치한 수천 개의 프록시 서버를 통해 고급 스크립트를 이용해 데이터를 수집할 수 있습니다. 보안 수준이 가장 기본적인 경우, 봇넷은 이 기술을 사용해 보안 인프라의 봇 관리 방어 및 웹 애플리케이션 방화벽을 통과할 수 있습니다.

그러나 보안 수준이 더 높은 경우에는 헤드리스 브라우저 공격과 같은 더 정교한 스크레이핑 접근 방식이 필요할 수 있습니다. 이는 스크레이핑이 선한 의도를 가진 행위자에 의해 수행되든 악한 의도를 가진 행위자에 의해 수행되든 마찬가지입니다. 그리고 저렴하지 않습니다. 일반적으로 기본 서비스 수준 대비 더 정교한 인프라로 인해 훨씬 더 큰 비용이 발생하기 때문입니다. 고급 방어에는 챌린지 기술(CAPTCHA, 작업 증명 등), 클라이언트 측 핑거프린트 평가를 위해 설계된 여러 탐지 레이어, 하이퍼텍스트 전송 프로토콜(HTTP), 전송 계층 보안(TLS) 특성 분석이 포함될 수 있습니다.

AI 봇넷의 스크레이핑 프로세스

기본 웹 스크레이퍼는 스크레이핑 기술이 더 일관적일 수 있지만, AI 봇넷은 형식이나 위치의 일관성이 떨어지는 비정형 데이터와 콘텐츠를 발견하고 스크레이핑할 수 있습니다. 또한, AI 봇넷은 실제 비즈니스 인텔리전스를 사용해 의사 결정 프로세스를 개선할 수 있습니다. 표 3, 서비스 수준 3에 언급된 정교한 AI 봇넷은 데이터를 스크레이핑하는 3단계 프로세스를 가지고 있습니다. AI 봇넷은 데이터를 수집하고 추출한 다음 처리하는 방식으로 작동합니다(그림 5).

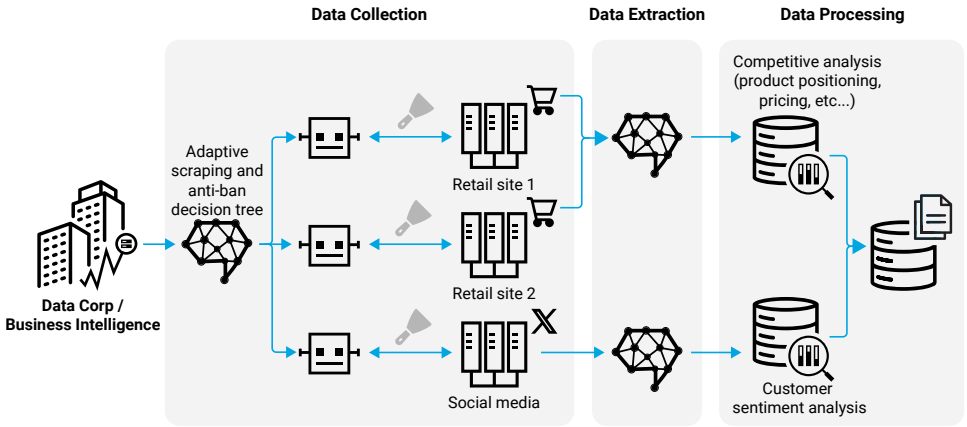

그림 5: AI 봇넷과 그 3단계 프로세스의 표현

이 3단계를 좀 더 자세히 살펴보고 각 단계가 무엇을 포함하는지 알아보겠습니다.

데이터 수집

웹 스크레이핑은 웹사이트 또는 웹사이트에서 추출한 데이터를 정리해 기업이 원하는 대로 적용하고 분석할 수 있는 새로운 데이터 세트를 생성할 수 있게 합니다. 그리고 이것은 데이터 수집에서 시작됩니다.

데이터 수집이 신속하고 원활하게 작동하려면 '안티 밴' 또는 '안티 봇 탐지' 기술과 결합된 적응형 스크레이핑으로 구성되어야 합니다. 이러한 기술은 의사 결정 트리로 설정되어 있을 수 있는 모든 보안 기능의 다양한 측면을 탐지합니다. 여기서 중요한 것은 안정성입니다. 봇 방어에는 자바스크립트 핑거프린팅, HTTP 및 TLS 핑거프린팅(HTTP 헤더 및 TLS 커넥션 평가), 인터넷 프로토콜(IP) 평판 탐지가 포함될 수 있습니다(그림 6). 이러한 워크플로우 중 일부에는 특히 성공률에 대한 통계 수집과 쿠키 전략, HTTP 헤더 및 TLS 매개변수 조정 그리고 자바스크립트 핑거프린팅 코드 평가 시 머신 러닝(ML)이 포함될 수 있습니다. 헤드리스 브라우저의 역할도 여기에 포함됩니다.

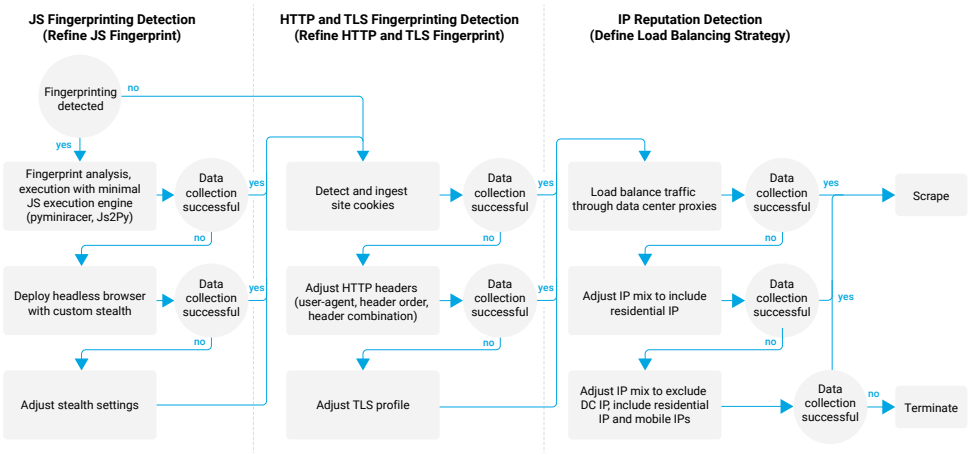

그림 6: 데이터 수집을 시도할 때 자바스크립트 핑거프린팅, HTTP 및 TLS 핑거프린팅, IP 평판 탐지를 피하려 하는 안티 봇 탐지 의사 결정 트리

헤드가 없는 브라우저

헤드리스 브라우저는 그래픽 사용자 인터페이스(GUI)가 없는 웹 브라우저입니다. 즉, 헤드리스 브라우저가 표시되는 웹 페이지와 사람이 직접 상호작용을 할 수 없으며, 대신 명령줄 인터페이스(CLI) 또는 네트워크 통신을 통해 브라우저가 실행됩니다. 인기 있는 오픈 소스 헤드리스 브라우저인 Selenium의 경우 자동화되어 있으며 웹 스크레이핑에 널리 사용됩니다. 이는 동적 콘텐츠를 스크레이핑하려는 검색자에게 매우 유용할 수 있습니다.

헤드리스 브라우저를 사용하면 스크린샷과 웹사이트 코드를 효율적으로 복사하고 전체 페이지를 렌더링하지 않고도 선택한 데이터를 추출할 수 있습니다. 그러나 헤드리스 브라우저 공격은 비용이 많이 들며, 공격자가 남긴 핑거프린트를 통해 탐지될 수 있습니다. 그러나 다른 정교한 인프라에 대한 비용은 헤드리스 브라우저의 비용과 비슷합니다. 즉, 일반적으로 높습니다.

데이터 추출 및 데이터 처리

추출된 정보는 일반적으로 HTML과 JSON 콘텐츠로 구성됩니다. 추출된 모든 데이터 중 일부만 분석에 도움이 될 수 있습니다. 예를 들어, 경쟁 분석에는 일반적으로 가격, 할인, 재고, 제품 SKU 번호, 카테고리, 설명이 포함됩니다. 필수 정보는 다양한 구조와 데이터 형식으로 학습된 ML 모델이 자동으로 추출해 인식할 수 있습니다. 이렇게 하면 데이터를 수동으로 추출하기 위해 수행해야 하는 모든 추가 처리 작업을 피할 수 있으며 HTML 및 JSON 콘텐츠 코드 구조를 학습할 필요가 없습니다. 또한, 사이트 디자인이 변화함에 따라 콘텐츠 코드 구조가 바뀔 수 있습니다. 분석 범위에 여러 웹사이트가 포함된 경우 처리를 위해 추가적인 머신 러닝 로직도 필요합니다.

사례 연구: 웹 스크레이핑 탐지 솔루션의 장점

Akamai 연구원들은 스크레이핑 활동을 탐지하는 웹 스크레이핑 솔루션을 사용하는 이커머스 고객 중 일부를 대상으로 한 주 동안의 트래픽 활동 내용을 관찰했습니다. 샘플 규모는 약 69억 건의 요청에 달합니다. 이 분석에서는 HTML과 AJAX 요청만 고려했습니다. 봇 대부분은 정적 콘텐츠를 요청하지 않기 때문에 정적 콘텐츠(이미지, 자바스크립트, 스타일시트)는 분석에 포함되지 않았으며, 이는 데이터의 불필요한 부풀리기를 방지하는 데 도움이 되었습니다.

전체 활동은 Akamai Content Protector를 통해 49.3%의 리스크가 낮은 사용자 트래픽, 42.1%의 봇 트래픽(리스크가 높은 악성 봇 27.5%, 정상 봇 14.6%), 8.7%의 리스크가 중간 수준인 미분류 트래픽으로 분류했습니다(그림 7).

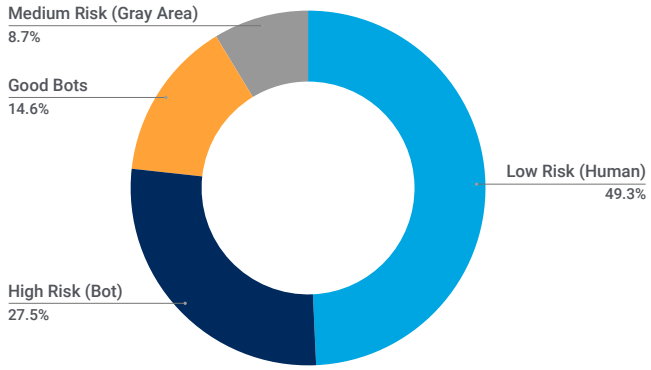

그림 7: 트래픽 활동 분류

그림 8은 봇에서 발생한 트래픽의 42.1% 중 65.3%는 악성 봇으로 간주되는 스크레이퍼에서 발생했으며, 나머지 34.7%는 웹 검색 엔진, SEO, 소셜 미디어, 온라인 광고 등 정상 봇으로 분류되는 스크레이퍼에서 발생했음을 보여줍니다.

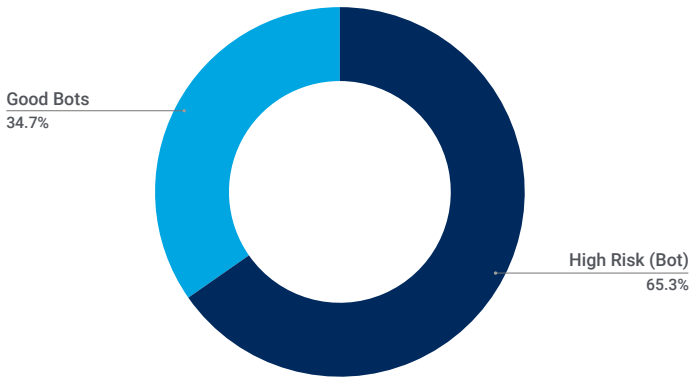

그림 8: 정상 봇 트래픽과 악성 봇 트래픽 비교

전체 봇 트래픽의 65.3%에 기여하는 리스크가 높은 악성 봇의 정교함 수준도 측정했습니다. 이 트래픽의 37%는 간소한 스테이트리스 방식으로 쉽게 탐지할 수 있는 기본 스크립트 봇넷에서 발생했으며, 47.6%는 ML을 사용한 고급 스테이트풀 탐지 방법이 필요한 고급 스크립트 봇넷에서, 15.5%는 고급 자바스크립트 핑거프린팅 및 스테이트풀 탐지 방법이 필요한 헤드리스 브라우저에서 발생했습니다(그림 9).

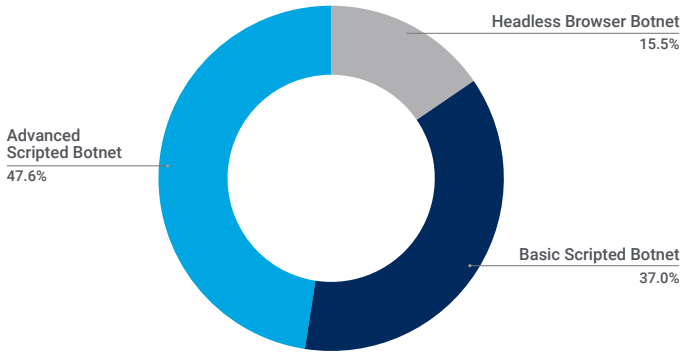

그림 9: 정교함에 따른 악성 봇 트래픽 분포(반올림으로 인해 합계가 100%가 되지 않음)

이 데이터를 통해 악성 봇 스크레이퍼가 정상 봇 스크레이퍼보다 훨씬 더 많으며, 전체 트래픽의 절반 가까이가 봇으로 구성되어 있고, 고급 스크립트 봇넷이 가장 많은 악성 봇 트래픽(47.6%)을 생성한다는 것을 알 수 있습니다.

이러한 봇에 대한 방어 기능이 마련되고 스크레이퍼가 제거되면 웹사이트 활동이 훨씬 빠르고 효율적으로 실행되며 사이트 지표가 더 깔끔하게 표시될 것입니다. 그리고 이러한 결과는 더 나은 사용자/고객 경험으로 이어질 것입니다. 그림 10에서 볼 수 있듯이, 방어 기능이 활성화된 후 리스크가 높은 봇 요청의 수가 많이 감소했습니다.

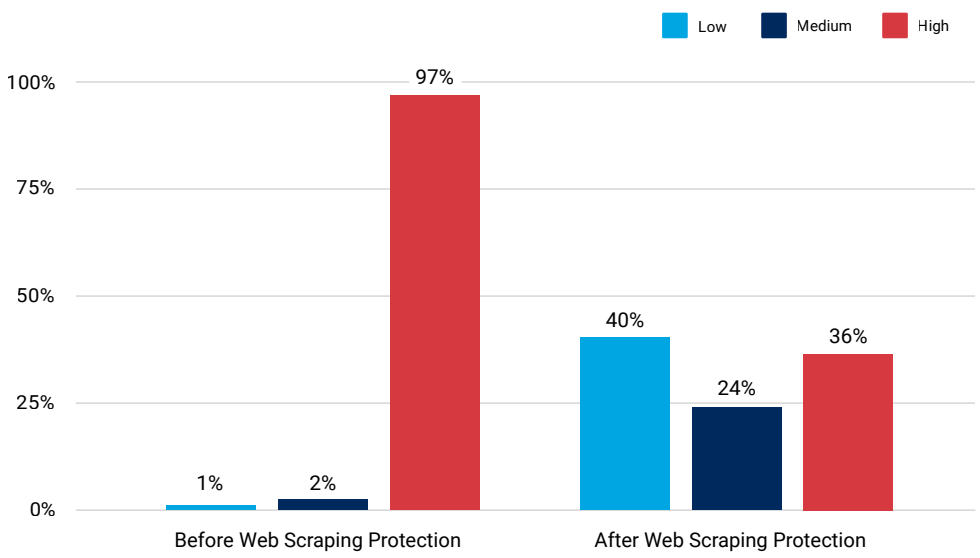

그림 10: Content Protector로 방어하기 전과 후의 리스크 수준

보호 및 방어

이 섹션에서는 웹 스크레이퍼를 탐지하는 데 중요한 몇 가지 지표와 이에 대한 방어 조치를 제공할 수 있는 툴에 대한 정보를 제공합니다.

기본 스크레이퍼 탐지

정교한 스크레이퍼는 탐지하기 어려울 수 있지만, 봇 관리 솔루션은 모든 종류의 침입형 스크레이퍼가 수집하는 데이터를 방어할 수 있으며 특히 다음의 특징을 찾아내어 간소한 웹 스크레이퍼 봇을 탐지할 수 있습니다.

- 구형 브라우저 및 OS 버전을 광고하는 요청
- HTTP 헤더 서명의 비정상
- 더 일반적인 HTTP v2 또는 새롭게 등장한 HTTP v3 대신 이전 버전의 HTTP(v1.1 등)를 사용하는 경우
- 수천 개의 클라우드 서비스/데이터 센터에서 발생하는 요청

고급 스크레이퍼 탐지

고급 스크레이퍼에서는 위 목록의 특성 중 어느 것도 관찰할 수 없습니다. 따라서 더 정교한 스크레이퍼에 관한 몇 가지 특징을 소개합니다.

- 최신 브라우저 및 OS 버전에서 오는 요청
- HTTP 헤더 세트가 정상적인 브라우저와 동일하게 보임
- HTTP v2 사용
- 수십만 개의 주거 및 모바일 IP 주소에서 오는 요청

트래픽 패턴 식별

웹사이트에 발생하는 트래픽의 종류가 사람(그림 11), 기본 봇(그림 12), 정교한 봇(그림 13)인지를 식별할 수 있는 몇 가지 주요 지표가 있습니다.

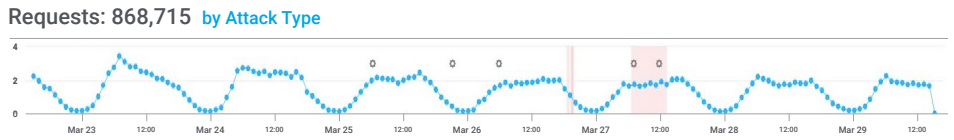

그림 11: 일반적으로 주기적인 활동 주기를 보이는 정상적인 사용자 트래픽

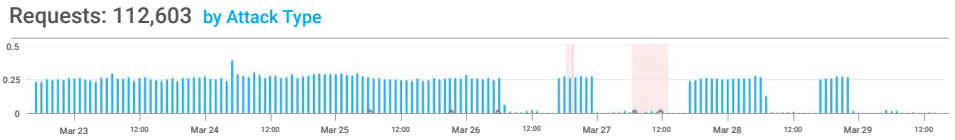

그림 12: 규칙적인 활동과 간헐적인 휴식 시간을 가지는 일반적인 봇 트래픽

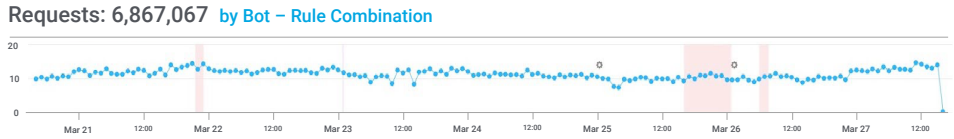

그림 13: 낮과 밤에 지속적으로 나타나는 더 정교한 봇의 트래픽

부하 분산 전략은 약하지만 핑거프린트 전략은 정교한(또는 그 반대의) 봇넷도 그 중간 어딘가에 있는 경우가 많습니다. 그러나 고급 봇넷은 매우 정교하여 완벽한 핑거프린트를 보유하거나 정상적인 사용자 트래픽 패턴을 재현할 수 있습니다.

이러한 스크레이퍼 봇을 경계하는 것 외에도 콘텐츠 보호기와 같이 웹 스크레이핑을 방지하는 툴을 사용하면 스크레이퍼가 득실거리는 거친 바다에서 특별한 이점을 누리며 순조롭게 항해할 수 있습니다. 다음과 같은 장점이 있습니다.

- 전환율 향상 및 IT 비용 절감
- 더 정확한 지표를 통해 더 나은 투자 결정을 내리고 매출 증가 뒷받침
- 가격 책정에 대한 부담 감소, 경쟁업체의 가격 인하로 인한 매출 감소 효과
- 원하는 상품에 접속한 고객의 만족도 제고, 고객이 인기 상품을 확보한 후 장바구니에 제품을 추가할 때 업셀 기회로 인한 매출 증가
- 정품 판매자가 판매하는 정상적인 상품으로 보이는 저품질 가품으로부터 고객을 보호해 브랜드 평판 유지
- 제품 매출 및 고객 충성도 유지
- 광고 수익 증가 및 보호
- 잠재 고객 및 사이트 방문자 유지

컴플라이언스 고려 사항

결제 카드 산업 데이터 보안 표준(PCI DSS) v4.0이 현재 시행 중이며, 여전히 기업에 영향을 미치고 있는 위협 트렌드에 의해 많은 변화가 발생하고 있습니다. 이러한 공격에 대응하기 위해서는 가시성이 핵심입니다. 기존의 자바스크립트 환경이든 변환을 용이하게 하는 데 사용되는 API든, 이러한 공격을 신속하게 탐지하고 해결하는 것이 중요합니다.

또한, 거버넌스 기능이 추가된 새로운 NIST Cybersecurity Framework 버전 2.0에서 새로운 컴플라이언스 트렌드를 확인할 수 있습니다. NIST는 여러 정부 규정의 기반이 되는 경향이 있으며 많은 상용 사이버 보안 프레임워크에 영향을 미치고 있습니다. 따라서 지금이 바로 새로운 가이드를 검토해 정책을 업데이트하거나 현재 문서를 매핑해 부족한 부분을 파악할 좋은 시기입니다.

상장 기업과 일반회계기준(GAAP)을 사용하는 기업의 경우, 또 다른 컴플라이언스 영역은 사이버 보안 중대성입니다. 중대한 리스크와 위협을 정의하려면 경영진 전반의 협력이 필요합니다. 랜섬웨어와 같은 중대한 위협을 식별한 후에는 마이크로세그멘테이션과 같은 방어 조치를 매핑해야 합니다. 위기 관리 계획에 공개 일정이 포함되어 있는지 확인하고 Security and Exchange Commission Cyber Incident Form 8-K를 제출해야 하는 최악의 시나리오에 대비한 플레이북을 준비하세요.

결론

이 보고서를 통해 기업에 부정적인 경제적 영향을 미칠 수 있는 영역에 대한 인사이트를 얻을 수 있기를 바랍니다. 봇이 사이트에 미치는 영향이 점점 더 커지고 있기 때문에 유익한 봇을 최적화하고 악성 봇을 방어하며 전반적인 고객 경험에서 발생하는 불편을 최소화해야 합니다. 이는 비즈니스에 영향을 미치는 보안 문제입니다. 모든 보안 문제와 마찬가지로 첫 번째 단계는 가시성을 확보하는 것이고, 두 번째 단계는 영향을 분석하는 것이며, 마지막 단계는 적절한 보안 제어를 구현할 수 있도록 리스크와 매출에 대한 ROI를 정하는 것입니다.

보이지 않는 것은 보호할 수 없으므로 지금이 바로 가시성이 부족한 부분을 파악할 때입니다. 이를 위해서는 사이트에서 웹 스크레이핑 활동의 수준과 그 의도를 파악해야 합니다. 정상 봇과 악성 봇이 모두 봇 환경을 구성하며, 스크레이퍼 봇은 용도에 따라 두 범주 모두에 속합니다. 유익한 스크레이퍼 봇과 유해한 스크레이퍼 봇 사이의 경계가 모호할 수 있지만, 봇의 정교함(헤드리스 브라우저 공격을 수행하는 웹 스크레이퍼 등)은 계속 진화하고 있습니다. 이 모든 것은 웹 스크레이퍼 봇이 이커머스 기업의 IT 비용과 고객 경험에 미치는 막대한 영향과 함께 발생합니다. 따라서 봇의 활동과 사이트에 미치는 영향을 분석할 수 있는 툴을 확보하는 것이 중요합니다.

사이트에서 범죄 비즈니스 모델을 실행하고 멤버십 포인트 현금화, 사기 주문, 반품 사기 등 다양한 악성 활동을 하는 공격자는 누구도 원치 않을 것입니다. 티켓 봇이 기간 한정 이벤트 티켓을 구매하거나 쇼핑 봇이 인기 제품을 구매하는 것도 원치 않을 것입니다. 봇은 특별 상품을 이용해 신규 계좌 개설을 악용하는 데 쓰일 수 있으며, 이는 캠페인 분석 및 비용에 영향을 미칩니다. 대형 분산 서비스 거부(DDoS) 봇넷은 웹 애플리케이션을 마비시켜 사용자 경험을 저하시키거나 주문이나 예약을 할 수 없게 만들어 매출 손실과 고객 불만을 초래할 수 있습니다. 심지어 봇은 온라인에서 사람의 행동을 모방해 웹사이트의 클릭과 트래픽을 증가시켜 신중하게 제작된 디지털 경험의 마케팅 및 성과 애널리틱스를 왜곡할 수 있습니다. 누구도 이런 일은 원치 않을 것입니다.

앞서 언급했듯이 전 세계 커머스 웹 트래픽의 절반 이상이 봇으로 구성되며 봇 트래픽 수준은 계속 증가합니다. 이 보고서의 인사이트와 조언은 웹 스크레이핑 방어와 콘텐츠 보안 기능이 포함된 Akamai의 보안 플랫폼을 기반으로 합니다. Akamai는 많은 이커머스 리더와 파트너십을 맺고 있기 때문에 기업이 고객을 가장 효과적으로 보호하기 위해 사용할 수 있는 보호 및 방어 조치를 공유하고자 합니다. Akamai는 웹 스크레이퍼 봇의 사용, 서비스 수준 선택, 사용 가능한 종류의 증가를 예상합니다. 따라서 기업의 리스크 체계를 지속적으로 평가하고 현재 보안 관리가 경영진의 리스크 성향에 부합하는지 판단하는 것이 필요합니다.

Akamai의 보안 리서치 허브에서 최신 리서치를 확인하세요.

방법론

Content Protector 데이터

이 데이터 샘플은 Content Protector 툴이 모니터링하는 트래픽에 할당하는 리스크 수준 분류에 대해 설명합니다. 이러한 분류는 봇 스크레이핑 활동을 탐지하고 정상 봇인지 악성 봇인지 판단하는 데 모두 사용됩니다. 봇 대부분은 정적 콘텐츠를 요청하지 않기 때문에 이 분석에서는 불필요하게 데이터가 부풀려지는 것을 방지하고자 HTML 및 AJAX 요청만 고려했습니다.

이 데이터 샘플은 2024년 4월 12일부터 4월 19일까지 일주일의 기간을 포괄합니다. 총 샘플 규모는 65억 건 이상의 요청으로 구성되었습니다.

봇 공격

이 데이터는 웹 애플리케이션 방화벽(WAF)과 봇 관리 툴을 통해 관측된 트래픽에 대한 애플리케이션 레이어 알림을 설명합니다. 보호하고 있는 웹사이트, 애플리케이션 또는 API에 대한 요청에서 악성 페이로드가 탐지되면 봇 알림을 트리거합니다. 이런 봇 알림은 악성 봇과 정상 봇 모두에 의해 트리거될 수 있습니다. 알림이 트리거됐다고 해서 공격이 성공한 것은 아닙니다. 이러한 제품은 높은 수준의 사용자 맞춤화가 가능하지만 Akamai가 여기에 제공한 데이터는 프로퍼티의 맞춤형 설정을 고려하지 않는 방식으로 수집했습니다. 이 데이터는 130여 개국 약 1300개 네트워크, 4000개 이상의 위치, 약 34만 대의 서버로 구성된 Akamai Connected Cloud에서 탐지된 보안 이벤트를 분석하는 내부 툴에서 추출한 것입니다. Akamai 보안 팀은 매달 페타바이트 규모의 데이터를 활용하여 공격을 연구하고, 악성 행동을 식별하며, Akamai 솔루션에 인텔리전스를 추가합니다.

이 데이터는 2023년 1월 1일부터 2024년 3월 31일까지 15개월의 기간을 포괄합니다.

저자 소개

편집장

랜스 로즈(Lance Rhodes)

편집 및 작성

데이비드 세네칼(David Senecal)
마리아 블라삭(Maria Vlasak)

검토 및 주제별 기여

미치 메인(Mitch Mayne)
수잔 맥레이놀즈(Susan McReynolds)
크리스틴 로스(Christine Ross)
바데트 트리비(Badette Tribbey)
스티브 윈터펠드(Steve Winterfeld)

데이터 분석

첼시 터틀(Chelsea Tuttle)

홍보 자료

애니 브런홀츨(Annie Brunholzl)

마케팅 및 출판

조지나 모랄레스(Georgina Morales)
에밀리 스핑크스(Emily Spinks)

인터넷/보안 현황 보고서

지난 보고서를 읽고 Akamai의 다음 인터넷 보안 현황 보고서를 확인하세요. **akamai.com/soti**

Akamai 위협 연구팀 자세히 살펴보기

akamai.com/security-research에서 최신 위협 인텔리전스 분석, 보안 보고서, 사이버 보안 연구 내용을 확인하세요.

이 보고서의 데이터 확인

이 보고서에 참조로 사용된 그래프와 차트의 고품질 버전을 확인하세요. Akamai가 제공한 소스라는 점이 정식으로 인정되고 Akamai 로고가 보존되는 경우 이러한 이미지를 무료로 사용 및 참조할 수 있습니다. **akamai.com/soti data**

Akamai 솔루션 자세히 알아보기

웹 스크레이퍼를 탐지하고 방어하는 Akamai 솔루션에 대한 자세한 내용은 **Content Protector** 페이지를 참조하시기 바랍니다.