Case Study

Newspaper Publisher Gains Control Over Content Theft with Netacea Bot Protection

Highlights

24%
reduction in AWS costs

Potential savings of
£15m
by blocking bad bots

Customer Profile

- One of the largest media groups and news publishers in the world
- Operates several of the UK's biggest news websites

Results

- Paywall bypass scrapers and GenAI training bots identified and mitigated
- 24% reduction in AWS costs by blocking scraper traffic on flagship news site
- Commercial licencing of scraper traffic enabled, opening new revenue streams
- Potential savings of £15m annually by preventing scraping attacks

The Challenge

Content theft has emerged as a critical concern for media and news publishers. The rapid rise of generative AI (GenAI) has disrupted traditional revenue models, as automated systems increasingly scrape content from news websites without consent.

These systems generate outputs based on stolen articles, editorials, and images, undermining the core revenue streams of publishers – advertising and paid subscriptions.

While this disruption presents challenges, it also opens new opportunities. Media organizations can establish financial agreements with GenAI businesses to license content. This has the potential to open new and highly valuable revenue streams. However, this strategy is only viable if publishers can identify scraper traffic and block unauthorized access effectively.

NETACEA

One of the world's largest media publishers faced a growing problem with scraper bot activity across their extensive network of news websites. These bots facilitated content theft in multiple ways:

Bypassing paywalls to access subscriber -only content.

Extracting trainable data for GenAI models without permission.

Increasing infrastructure costs by consuming server resources to deliver content to automated systems.

Despite having a rudimentary Web Application Firewall (WAF) in place, the client struggled with:

Limited visibility into bot traffic.

Inability to counter sophisticated, evolving bot tactics.

High utilization of their AWS infrastructure by bot traffic.

> The executive leadership team prioritized solving this issue.
> The organization needed better visibility, control, and protection to combat content theft effectively, without having to manually block bots themselves.

The Solution

To tackle the issue head-on, the publisher partnered with Netacea, a leader in bot protection for a fully managed bot protection service.

STEP 01

Understanding the Bot Threat Landscape

> Netacea's Threat Intel Center conducted extensive research across hidden marketplaces and forums to uncover content scraping configurations targeting the client's websites.

STEP 02

Data-Driven Detection

> Netacea seamlessly integrated with the client's CloudFront CDN, ingesting web log data from key news sites. Using server-side analysis and machine learning algorithms, Netacea analyzed each web request's intent to differentiate between benign visitors and malicious bots.

> This highly scalable and adaptive approach made it impossible for attackers to detect and bypass Netacea's defenses.

STEP 03

Automated Identification and Response

The detection models quickly identified that 24% of all website traffic came from malicious bots. Further investigation revealed advanced techniques employed by attackers, including:

- Rotating IP addresses to evade detection.
- Using spoofed user agents to mimic human visitors.

Netacea's machine learning algorithms automatically adapted to counter these evolving tactics, requiring no manual intervention from the publisher's internal teams.

The Outcome

Netacea's Bot Protection is now fully integrated across the publisher's news websites, delivering tangible results:

Reduced Infrastructure Costs

The publisher's infrastructure team reported a 24% drop in AWS usage, saving operational expenses.

Visibility into Bot Traffic

Through the Netacea portal, the organization can now see and analyze every bot attempting to scrape their content.

Commercial Licencing of Ethical Bot Traffic

While malicious bots are automatically blocked, ethical scrapers that declare their identity can now be approached with financial agreements to license content access.

> With Netacea Bot Protection, the publisher is not only addressing content theft effectively but also establishing new revenue opportunities in the era of generative AI.

By investing in robust bot protection technology, this organization has safeguarded its content assets, protected revenue streams, and positioned itself as a leader in managing the challenges posed by GenAI-driven content scraping.