
THE BENEFITS OF
BETTER
CDN
MONITORING
AND HOW TO GET THERE
GlobalDots

The Transformative power of CDN monitoring

As a business, If you wish to scale your service or website, a CDN has become a non-negotiable component of your infrastructure. Between 2020 and 2025, the cloud CDN market is forecast to grow 28% and will have a market size of $49.6 billion. CDNs have moved from a tool for big tech to a necessity for every company who is looking to handle the scale that the internet offers.

It should not come as a surprise, as CDNs have multiple benefits. Some of the main benefits of CDNs, resulting from their distributed nature are:

- Improved website speed and performance
- Reduced server load time, bringing down bandwidth and delivery costs
- Paramount website security, especially against DDoS attempts
- High-quality data about users that can be used for analytics and accurate segmentation of your audience

Once a CDN is in place though, a lot of businesses don't take the next step of monitoring their new addition to their product. Within the CDN, there is a wealth of information. Everything from HTTP request times and user journeys through to load profiles that can help operations teams develop a deep understanding of how their customers use their site. In fact, the entire reality of the website, as users actually experience it, occurs in the CDN realm.

the entire reality of the website, as users actually experience it, occurs in the CDN realm

The misconception of the value of CDN logs (and why so many companies fail to use it)

While CDN logs contain tremendous value - many companies do not monitor them. In fact, there are a couple of good reasons why not do so:

- CDN logs come in great volume, while many/most of them considered useless
- It's expensive to store, and complicated to manage, index, and analyze
- Major CDN providers don't make it easier, not paying much attention to the issue, leaving only the more sophisticated players to solve it by themselves

When you're ready to take advantage of some of these great benefits to analyzing your CDN logs, you will quickly run into a few problems. These problems range from the commercial to the technical.

1. CDN logs are often plaintext and not in a consumable format

CDN logs are usually basic text web access logs, rather than JSON. This means that parsing them can be difficult. Web access logs don't necessarily follow a modern data structure, and instead follow standards like the W3C Extended log file format.

127.0.0.1 username [10/Oct/2021:13:55:36 +0000] "GET /my-page HTTP/2.0" 200 150 1289

Web Access Log Sample

{ "ip": "127.0.0.1",
"username": "username",
"timestamp": "10/Oct/2021:13:55:36 +0000",
"method": "GET",
"url": "/my-page",
"httpVersion": "HTTP/2.0",
"responseStatus": 200,
"latency": 150,
"requestBodySizeBytes": 1289 }

The same access log in JSON

This means that if you decide to ingest your CDN logs into a 3rd party, you need to ensure that they are able to process unstructured logs and convert them into key-value pairs that can be queried and understood.

2. A lot of data to process

Companies deploy CDNs to handle high volumes of traffic and CDN usage is increasing globally, year on year. Each of these requests need to be logged, ingested, parsed, analysed and visualised. This adds up a huge amount needed to get through all of that information. When you're trying to gain insights from your CDN data, you need to find a solution that can consume vast quantities of information with minimum latency and provide insights right away.

3. High storage costs

CDNs generate a huge volume of logs. The size of the logs is proportional to the amount of traffic you receive, but if you're using a CDN, the chances are you're expecting a lot of traffic. You may need to store logs in the short term for operational reasons, but you may need to store your logs for regulatory requirements, such as GDPR.

The period of time you need to store your logs will depend on your requirements, but in some cases, it can be over a year. This means you're storing huge volumes of data for long periods of time.

When you're looking at a logging provider, you should search for a provider that supports tiered pricing, to enable you to optimise your storage.

And yet, in recent years CDN-expert-service-players, along with monitoring and observability innovators, had been working together to develop smart, cost-effective solutions to allow businesses to act on these logs, gain huge business value and a significant competitive advantage.

The great business value of your CDN logs

We're going to tag each benefit with labels, indicating which aspect of your business they will support. These are:

REVENUE - Helping your business to strategically utilise CDN observability data to drive revenue generating changes

SECURITY - Improving your security posture by giving you new insights into how your system is behaving

OPERATIONAL - Aiding your operational success, by helping you to improve uptime, optimise server resources, compress content and more.

These three dimensions offer a clear insight into the value of CDN monitoring, and make a compelling argument for why all CDN users should prioritise better CDN monitoring to access the value, locked away inside their CDN solution.

Ready to harness your CDN logs to boost speed and security?

Book a Demo

1. Understand site performance by page

REVENUE SECURITY OPERATIONAL

Site performance is a key metric for your whole business. While it may seem like a solely operational concern, performance metrics, like average and median latency, are indicators across all three of our categories.

So how do you visualise your page latency?

The easiest way to track your site latency is to render the values in your logs out as a line graph. This is going to involve a couple of key steps.

1. Parse your logs and turn them into metrics

Your access logs from your CDN solution will contain a latency metric. It looks something like this:

> 127.0.0.1 username [10/Oct/2021:13:55:36 +0000] "GET /my-page HTTP/2.0" 200 150 1289

Your first challenge is to extract the target page and the latency from that log, both highlighted in green. For this, you can either utilise regex parsing or use a 3rd party service that can process and structure this data for you.

2. Render in a table

Once you've got it, the simplest and easiest way to track your data is to then render it out in a table, showing the average, median and 95th percentile latency for each of your pages.

This will give you a clear insight into which pages are performing the best and which are slowing down. Making use of the percentile also gives you rapid feedback on when these statistics suddenly change.

Target page	Average Latency MS	50th percentile of Latency MS	95th percentile of json.http_resp_took_ms.numeric
/cart/checkout	187,854	98,286	456
/	50,634	23	163,2
/product	46,2	18	180,306
/cart	45,67	23	170,058
/_healthz	0,069	0	0

Site performance is key for conversion

25% of users will leave a site if it takes longer than 4 seconds to load. 46% of users won't come back to a slow site. Your CDN is in place to ensure that your site loads instantaneously, inline with what consumers expect. Your CDN monitoring and the insights it delivers are key to understanding the true experience your customers are having, when they navigate your site.

But I thought a CDN would mean every page loads quickly?

A CDN will certainly improve the performance of your site, but cache misses are more common than you think. Even when you get a cache hit, you don't know whether it's from memory, drive, or after a series of network hops inside your CDN.

Without a deep insight into the performance of your CDN, you have no way of knowing if your CDN is providing you with the service that you're paying for

Without a deep insight into the performance of your CDN, you have no way of knowing if your CDN is providing you with the service that you're paying for. Site speed is a key factor in conversion, so you need to make sure you've got a clear view of how long it truly takes for your customers to see your page, when they navigate to your site through your CDN. This is given clearly by the latency in milliseconds, present in almost all CDN logs. By visualizing the latency in your CDN logs, you're dealing with a very accurate measure, which is the time it took for your CDN to fulfill the HTTP response.

Ready to harness your CDN logs to boost speed and security?

Book a Demo

Knowing your site hotspots means you can

Each page on your site will also make requests through to APIs that are hosted on your backend. This means that even though your CDN is handling your web page traffic, it may still have an impact on the kind of traffic your backend services experience because this sort of data might only have a short cache life.

Having a clear understanding of which pages are being used the most, means you can scale the services on which those pages depend, and make sure that both the web traffic and the backend data needed to display the page are being loaded in record time.

20% of DDoS incidents in 2020 were coupled with other attacks - a DDoS attack is often the opening salvo of a more complex scheme

Your slow pages might be a vulnerability

DDoS attacks aim to consume all of the resources on your site. They can cost small companies thousands of dollars and result in prolonged outages of your site, and potentially can open the door for other, more impactful attacks that cause leaks of your company's personal information. Many DDoS attacks utilise multiple attack vectors at once to deepen the impact - in fact, 20% of DDoS attacks in 2020 were coupled with other attacks. This means that a DDoS attack is often the opening salvo of a more complex, sophisticated attack.

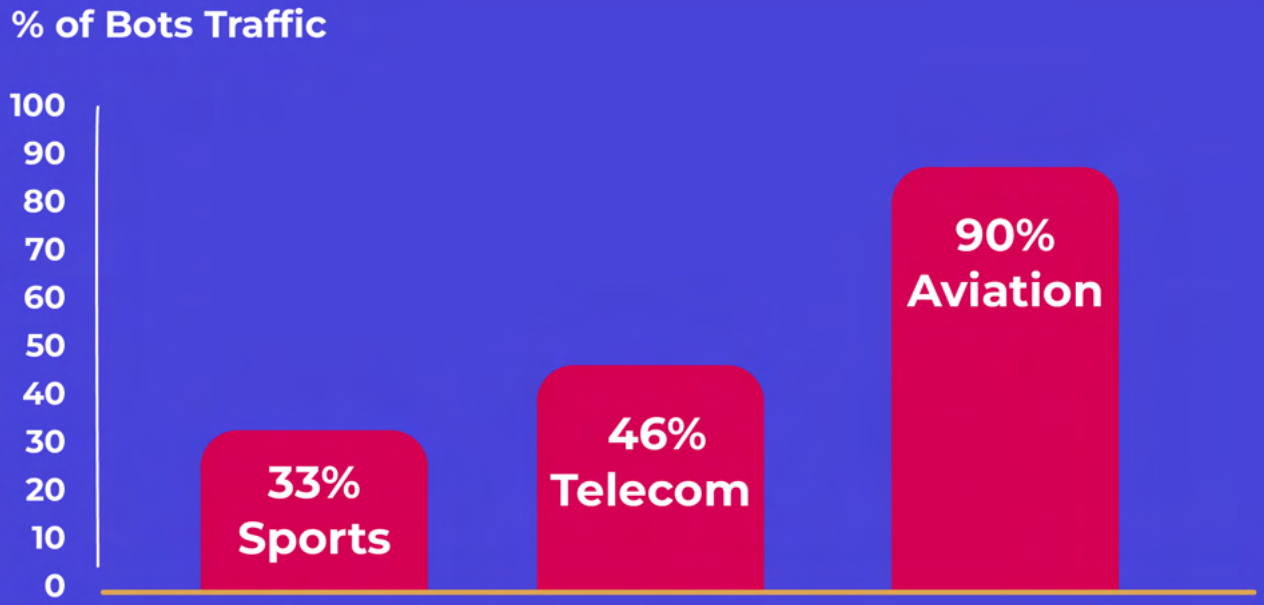
A slow loading web page might be an indication that this page requires a lot of processing power to load. For example, if the shopping cart on an eCommerce site is taking a long time to load, it may be due to a slow running service running in the backend.

This is a signal to a hacker that if they target a DDoS or slow DDoS attack on your slow page, they may be able to take out your site more quickly.

What should I look for when assessing my vulnerability risk?

Your CDN data includes latencies that you can monitor. A sharp increase in your HTTP request latency may indicate that someone is attempting a DDoS attack. You can expect request latencies of between 1 and 100 milliseconds to be within the bounds of normal. Higher may be okay, it depends on your system, but most CDNs boast consistent speeds of below 100ms.

Secondly, you should understand how much of your traffic is true, and how much is driven by bots. Bots contributed 40% of online traffic in 2020, with 25.6% of that being "bad bots". Detecting bot traffic involves understanding what your normal traffic looks like, and spotting changes in metrics like sudden spikes in traffic from an unknown region, sudden shifts in latency metrics, very low session durations and more. This is important, because when we've worked with clients, we've found that permissive bot policies allow these "bad bots" to drive up your hosting costs and can ultimately lead to outages.



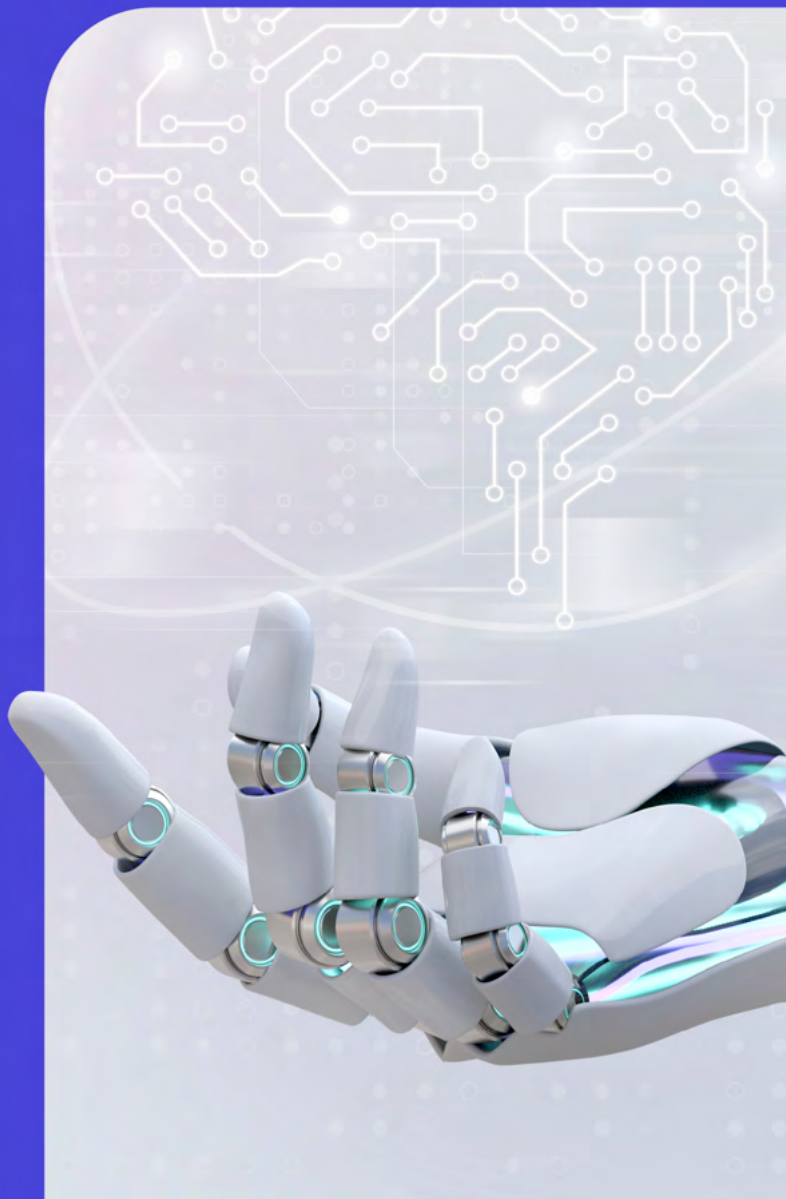
It's important to understand how many bots you should expect on your site. This depends on where and how you do business. For example, an international airline with destinations in all continents, relying heavily on Online Travel Agents (OTAs), will experience a higher bot traffic share, compared to a local airline. In fact, based on our experience with clients with whom we have partnered, it's possible to see 90% (good and bad) bot traffic of this kind, hitting a site in the aviation industry. Other industries include Telecoms (45.7% are bad bots) and Sports (33% are bad bots). This bot traffic needs to be closely monitored, to ensure that it isn't having an adverse impact on the customer experience on your site, but also that you're not paying huge hosting costs to support the activity of bad bots.

Ready to harness your CDN logs to boost speed and security?

Book a Demo

And don't forget to measure your latency percentiles. When you're dealing with latencies, most people simply graph the average or the median. This is useful, but it doesn't give you the instant feedback you need when something has drastically changed in your site traffic. Tracking the 95th and, in some cases, the 99th percentile will tell you the second that your outlier data has changed. These few seconds may be the difference between dealing with a full blown DDoS attack and preventing it in its infancy.

These few seconds may be the difference between dealing with a full blown DDoS attack and preventing it in its infancy



2. Understanding the customer journey through your site

REVENUE SECURITY

While it's important to understand the kind of traffic each page is experiencing, a clear grasp of customer journey has powerful benefits for both your security and your revenue interests. Your customer journey is the pages that a customer commonly travels through, as they move from browsing to purchasing. This information is available in the CDN logs in the form of requested endpoints and session IDs or IP addresses.

How do you map your customer journey?

Tracking customer journeys through your site is not an easy challenge. The problem is chaining together a single session. The process can be broken down into simple steps, but unfortunately these are quite manual:

1. Use a value, like IP address or one of the HTTP headers, to track a single session in a given time frame

2. Group all sessions together on IP address, so you've got a list of all requests made by that IP address in a short time frame (say, one hour)

3. Count the number of journeys that have the same (or similar) pages

When you've got some data to work with, the most common approach is to render your information on a customer journey map. A customer journey map is designed to help you understand how your customer is truly interacting with your site.

Some challenges with working out the customer journey

The big challenge you're going to have is that users typically browse around in random ways, before eventually buying. This means that two users may request a lot of different resources but actually they're on the same journey. For example, eCommerce sites report that 75% of their queries, every month, are new. This means different resources, but ostensibly the same customer journey. Your challenge is to work out what type of resource the consumer is interested in, so you can group similar shopping experiences together on your site.

Your second challenge is going to be crunching all of this data. Once you've pulled out IP address and page from each of your logs, perhaps using regex, then you may wish to use something like Tableau to visualise the data that you've got, so you can understand your trends.

Your customer journey is your baseline

Anomaly detection is a cutting edge cyber security technique. It utilises machine learning to detect when something happens out of the ordinary. The real challenge with anomaly detection, however, isn't getting the machine learning capability, which can be bought off the shelf. The real challenge with machine learning is gathering the necessary training data. Machine learning algorithms typically require a lot of data to achieve an acceptable level of accuracy.

Your CDN logs contain the baseline for normal usage of your site, and so it is a repository of training data for your machine learning algorithms. A combination of IP addresses, URLs, request payloads and latencies can give you a good picture of usage. An anomaly detection algorithm, trained with your CDN logs, will be able to detect traffic patterns that differ from the normal user journey, and give you an edge on attackers, before they've managed to mount their attack.

Your CDN logs contain the baseline for normal usage of your site, and so it is a repository of training data for your machine learning algorithms

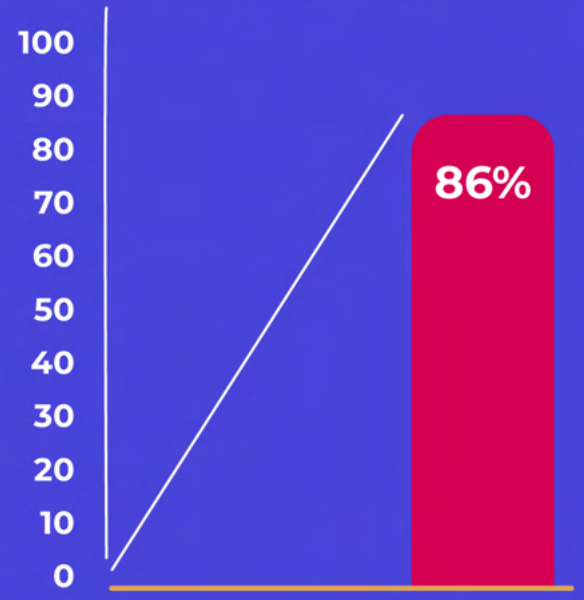
Make the journey as smooth as possible

Once you know the path your user is taking through your site, you can optimise your website for that path. By optimising this journey, you're ensuring that most of your customers have a smooth journey from browsing to buying.

Ready to harness your CDN logs to boost speed and security?

Book a Demo

The impact of great UX



A great user journey can have a serious impact. 86% of customers are willing to pay more for a great user experience. This translates into a direct revenue stream that you may be missing out on, because you haven't yet built a map of the journey your customers take through the site, and then optimised your site for that journey.

3. Detailed understanding of Regional traffic

REVENUE SECURITY

CDNs instantly give your site a global presence, and with powerful edge caching technology, can deliver content that would typically take seconds in under 100ms. CDNs also commonly log the IP address of the CDN edge node that handled the request, meaning they can provide geolocation data on where your customers are coming from.

What is the best way to visualise regional traffic?

Most observability platforms offer a map view that will allow you to plot geolocation data onto a world map, so you can clearly visualise where your traffic is coming from.

The best place to get geolocation data is to use the IP address. IP addresses aren't perfect for geolocation, because they can be spoofed or users can utilise a VPN solution, but across your whole dataset, they'll give you a clear trend. In order to get geolocation information from your IP addresses, you'll need to either enrich your logs or make use of a geolocation API that will convert your IP addresses into coordinates that can be mapped onto your world map.

Some regions are hotspots for security attacks

In 2021, an estimated 74% of proceeds from ransomware went to Russian hackers. Earlier in 2021, China was accused of being responsible for an unprecedented level of hacking attempts. More and more countries are engaging in cybercrime at the state level. It is estimated that by 2025, the global cost of cybercrime will exceed $10.5 trillion.

By tracking the countries that are engaging with your site, a sudden shift in regional requests may be the first sign of a mounting attack against your service. For established companies with stable markets, a sudden change in their user geography may indicate that a large portion of requests have suddenly arrived from somewhere new. This may be perfectly legitimate usage, but it can form part of a chorus of metrics that can give you early warning of a cyber attack.

a sudden shift in regional requests may be the first sign of a mounting attack against your service

Recent innovation makes CDN logs monitoring easy and cost-effective

There is tremendous value, hidden away inside your CDN. A robust monitoring platform will help you to prise out key insights that will help you to strengthen your security posture, grow your revenue streams and achieve operational excellence. While traditional solutions simply couldn't cut it, recent developments tackled the problem and now enable a simple and effective way to leverage your logs, while revamping pricing models to enable a cost-effective monitoring operation, dropping costs by 40-70%.

About GlobalDots

GlobalDots, a 20-year CDN world expert and a premier partner of all major CDN providers, decided to tackle the CDN monitoring challenge. By collaborating with some of the most recognized Monitoring & Observability innovators, it now offers centralized monitoring platforms to store, analyze, and monitor CDN logs for web-based companies. We empower our customers to simply and affordably gain access to key data that they would have otherwise missed, enabling better decision making and strategizing. Some benefits include faster incident resolution time, reduction in development time, a 99 percent decrease in errors, and more.

Ready to harness your CDN logs to boost speed and security?

Book a Demo

GlobalDots

Learn more:

cdn-monitoring.globaldots.com

follow us