Measuring CDN Performance, Availability and Scalability

Abhishek Tiwari

When it comes to select a Content Delivery Networks (CDN) there are many factors you need to consider such as performance, availability, scalability, reliability, pricing model, features, SLA etc. In my opinion performance and pricing are key in the decision making process.

Performance

CDN performance is measured in terms of speed of DNS resolution and content delivery. Normally performance measurement includes both,

1. How quickly CDN's DNS system supply the address of the best CDN content server? DNS resolution is based on latency, geo-graphic location or price class. It has been reported that many CDNs employ a huge number of DNS servers that are co-located with their content servers.
2. How quickly content is delivered from CDN server to client?

Quantifying performance of a CDN at large scale requires the ability to either capture traffic on the CDN's servers or control a large number of globally distributed clients.

For CDN performance testing there are not any standard protocols to measure the performance across CDN. Each CDN vendors tends to use different methodologies to measure performance. Given that underlaying architecture and network distribution is neither same nor very visible so it is really hard to compare CDN vendors in terms of performance. Geographic location of CDN data centers plays a big role in performance measurements.

Having said that, CDN performance testing can be segment based on delivery path and main Internet bottlenecks

- First mile testing (CDN server level)
- Backbone testing (Internet hubs)
- Last mile testing

Backbone testing

Most of the CDN providers focus mainly on Backbone testing. Backbone tests measure how fast your content (website, video etc.) is loading at major Internet hubs. Backbone testing does not cover performance end-to-end and it ignores the performance impact from Internet hubs to actual user device. Backbone testing results are often misleading as they are not measuring the true end-user experience. Often CDN vendors prefers this as it costs less, takes less time and easy to execute.

Last mile testing

Real end-user experience and CDN performance can be measured using Last mile testing. Despite shortcomings Last mile measurements are more realistic as they measure how quickly your content is delivered from CDN servers to end-user device. Because it incorporates last leg of connectivity service required to deliver content, Last mile testing provides more accurate picture on end-user experience.

There are two ways to perform Last mile testing:

1. Use independent testing company such as Gomez to run Last mile tests. Gomez collects performance data from a large number of real customers. Companies likes Gomez help CDN vendors to measure their own latency for delivering objects to real-world users located around the globe.
2. Use a tracking code on client side to collect real latency data from actual users. This is often used by CDN customers to measure actual performance they are getting from CDN providers.

Availability

Although speed of content delivery is quite important, CDN availability and uptime can not be ignored. For availability and uptime evaluation, both cluster availability and server availability are measured where "cluster" refers to a collection of CDN servers in the same location. As you may expect that clusters have better availability than individual content servers. Often CDN vendors employ auto-scaling techniques which provisions new content servers during high-load conditions.

Scalability

Load testing is a way to measure scalability and elasticity of a CDN service. Unfortunately traditional load testing methods don't work well with CDNs. For instance if you test by sending requests from a single client to just one of the IP addresses that DNS returns, your requests may resolve to a single server in one CDN edge location. Often load testing is not very straightforward as each CDN network uses different load handling techniques. Especially due to auto-scaling, clustering of content servers and geo-location based DNS resolution - it is not easy to use something like Apache Benchmark. Moreover IP address range also keeps changing due to adding and removing of the content server as part of auto-scaling.

To perform load testing with CDNs one must,

1. Send client requests from multiple geographic regions (Load testing at edge location level based on DNS resolution).

2. For each geographic region, use multiple measurement clients with each client making an independent DNS request so that each client will receive a different set of IP addresses from DNS (Load testing at cluster level).

3. Now for each client that is making requests, spread your client requests across the set of IP addresses that are returned by DNS, which ensures that the load is distributed across multiple servers (Load testing at server level).

Reliability

Packet-loss measurements are used to determine the reliability of a CDN. High reliability indicates that a CDN incurs less packet loss and is always available to the clients.

References:

- Last mile
- Measuring and Evaluating Large-Scale CDNs
- Load Testing CloudFront
- A Taxonomy and Survey of Content Delivery Networks
- Comparing CDN Performance: Amazon CloudFront's Last Mile Testing Results