

The Open High Throughput Computing Content Delivery Network

Dave Dykstra^{1,1} , Brian Bockelman² , Jakob Blomer³  and Laurence Field⁴ 

¹Scientific Computing Division, Fermilab, Batavia, IL, USA

²University of Nebraska-Lincoln, Lincoln, NE, USA

³EP-SFT Department, CERN, Geneva, Switzerland

⁴IT-CM-IS Department, CERN, Geneva, Switzerland

Abstract. LHC experiments make extensive use of web proxy caches, especially for software distribution via the CernVM File System and for conditions data via the Frontier Distributed Database Caching system. Since many jobs read the same data, cache hit rates are high and hence most of the traffic flows efficiently over Local Area Networks. However, it is not always possible to have local web caches, particularly for opportunistic cases where experiments have little control over site services. The Open High Throughput Computing (HTC) Content Delivery Network (CDN), openhtc.io, aims to address this by using web proxy caches from a commercial CDN provider. Cloudflare provides a simple interface for registering DNS aliases of any web server and does reverse proxy web caching on those aliases. The openhtc.io domain is hosted on Cloudflare's free tier CDN which has no bandwidth limit and makes use of data centers throughout the world, so the average performance for clients is much improved compared to reading from CERN or a Tier 1. The load on WLCG servers is also significantly reduced. WLCG Web Proxy Auto Discovery is used to select local web caches when they are available and otherwise select openhtc.io caching. This paper describes the Open HTC CDN in detail and provides initial results from its use for LHC@Home and USCMS opportunistic computing.

1 Introduction

A Content Delivery Network (CDN) is a geographically distributed set of caching web proxy servers designed to improve performance and availability of web services. LHC experiments make extensive use of the WLCG CDN, which is made up of general purpose Squid web proxy cache servers installed at WLCG sites. The primary applications that use the WLCG CDN are the Frontier Distributed Database Caching system [1] (referred to as just Frontier below) and the CernVM File System [2] (CVMFS). Having those cache servers at each site is important for low-latency and high performance with many queries and clients at those sites, but sometimes sites are very small and can't justify having their own caches or are highly distributed such as with the volunteer computing LHC@Home [3] project. There

1 Corresponding author: dwd@fnal.gov

are also opportunistic use cases where experiments have no control over the site services and so need to use off-site caching. This paper describes the new Open High Throughput Computing (HTC) CDN for those cases where on-site web proxy caches are not available.

2 Free Tier of a Commercial CDN

There are many commercial CDN providers that provide web caching on the internet. There is one provider – Cloudflare [4] – that has a free tier with unlimited bandwidth and works with the Frontier and CVMFS applications. Cloudflare provides an easy to use web interface for administration and works with any internet domain as long as they are allowed to host the domain on their DNS servers. Then the administrator can create DNS records and choose whether each name is cached or not cached. Cloudflare includes many useful features in the free tier, including Distributed Denial of Service protections, IPv6 support, support for DNSSEC, and support for https. They have data centers distributed all over the world, and the free tier includes use of all of them except their data centers in China.

Cloudflare uses a small number of IP addresses for all services that they cache, and automatically route requests to those addresses to their nearest data centers by use of AnyCast [5]. The addresses are shared with multiple domains, and requests not yet cached are directed from the data centers to the correct origin server based on the http ‘Host’ header or https Server Name Indication.

Note that AnyCast is not something that could be used by the WLCG CDN to locate Squid proxies at WLCG sites, because it requires one administrative entity to control the distributed network and requires a lot of work to negotiate with all the local Internet Service Providers. Instead, the Frontier and CVMFS clients have additional functionality to locate web proxy servers (for example using Web Proxy Auto Discovery [6]) and to robustly try more than one if one fails and another one is available.

3 OpenHTC.io

Since the Cloudflare CDN works by hosting domain names, the domain name “openhtc.io” was registered for the purpose of hosting the Open HTC CDN on the Cloudflare free tier. Provision was made for the longevity of the domain, with 3 people given administrative access. Aliases have been registered for CVMFS stratum 1 servers (for example s1fnal-cvmfs.openhtc.io and s1cern-cvmfs.openhtc.io) and for CMS frontier servers (for example cms-frontier.openhtc.io). The free tier includes up to three “Page Rules” for customization, and two (“Cache Everything” and “Respect server expiration headers”) are needed to work correctly with CVMFS and Frontier. The introductory home page at <https://openhtc.io> offers to register aliases for other appropriate HTC web services as well.

4 Use Cases

The Open HTC CDN is not intended for use by the majority of the WLCG, but there are a number of important use cases for it that are itemized in this section.

4.1 Volunteer computing

The CVMFS openhtc.io aliases are now in production use for all LHC@Home jobs, and the CMS Frontier openhtc.io aliases are used for CMS jobs running on LHC@Home. The jobs

also use the WLCG Web Proxy Auto Discovery (WPAD) so if they run at WLCG sites they use local Squids in addition to Cloudflare caches.

4.2 Opportunistic jobs

The Frontier `openhtc.io` aliases are also in production use by CMS opportunistic jobs running on the Open Science Grid (OSG). They use WLCG WPAD too and for the most part they run at grid sites where local Squids are available, so they use those Squids which then read from `openhtc.io`. If they ran somewhere where there was no local Squids they would use only `openhtc.io` caches. Also, there was a case where the local Squids at one large opportunistic site were failing, so for a weekend all of the Frontier accesses on the site were directed to use the `openhtc.io` aliases without Squids and did so without any problems noticed. Normally WLCG WPAD includes backup proxies at CERN and Fermilab for sites that it recognizes, for the purpose of detecting when the local Squids fail, but for that weekend those backup proxies were intentionally disabled in order to test how well `openhtc.io` would operate and to improve performance at the site.

4.3 CVMFS configurations with no proxy

The OSG CVMFS configuration repository is set up to switch to `openhtc.io` aliases for CVMFS stratum 1 servers whenever the CVMFS client configuration indicates that it has no proxies available; that is, it has `CVMFS_HTTP_PROXY` set to "DIRECT".

4.4 CernVM

The CernVM CVMFS configuration is planned to do something similar to OSG: it will select `openhtc.io` stratum 1s when Web Proxy Auto Discovery is selected but with no proxy as a fallback, that is, when `CVMFS_HTTP_PROXY` is set to "auto;DIRECT". That will be the default setting of `CVMFS_HTTP_PROXY` on CernVM, and a CernVM-specific version of the WLCG WPAD service will be configured to provide grid site squids if they're available and otherwise send requests directly to `openhtc.io`. However, in order to prevent large numbers of clients in a single organization (for example a commercial cloud provider) from using `openhtc.io` instead of setting up their own squids, WPAD will keep track of the number of requests and if there are too many in too short of a period of time from one organization, it will direct the requests instead to monitored backup proxies at CERN and Fermilab.

4.5 Open data releases

The LHC OpenData releases are based on CernVM, so when the CernVM version is updated for those releases, they will automatically use the `openhtc.io` aliases.

5 Performance

Because the Cloudflare servers are located so widely around the world, after the caches are loaded the latency of each request on average for LHC@Home jobs is much reduced compared to the previous method of using Squid proxy servers at CERN and Fermilab.

Figure 1 shows a plot of CMS job times up to the first processed event, in seconds, comparing the same LHC@Home clients without Cloudflare and with Cloudflare. The CVMFS caches were empty before starting in all the plotted cases. The measurements include both CVMFS

and Frontier accesses. The average speedup using Cloudflare openhtc.io aliases was 2.2 times, with an average savings of 295 seconds which is almost 5 minutes.

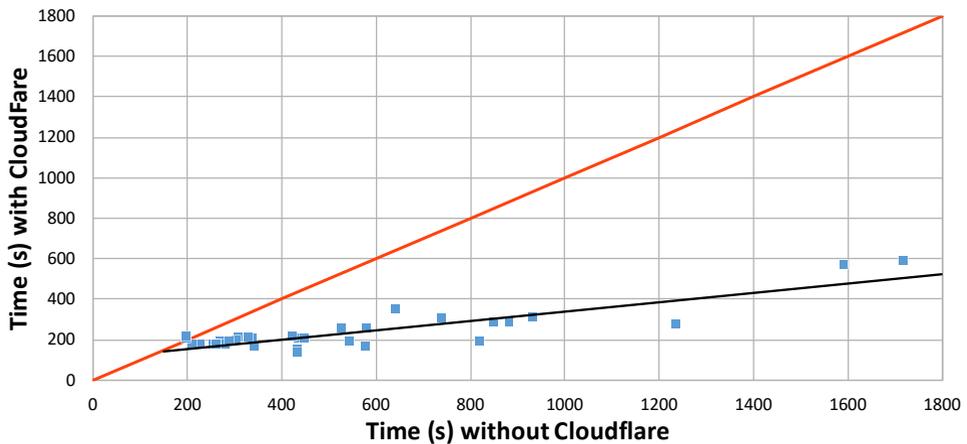


Figure 1. CMS job start times to the first event without vs. with Cloudflare

6 Limitations

There are a number of limitations with the approach described in this paper:

1. It uses more WAN bandwidth and has higher latency than local Squids. This is why we only want to use it when local Squids are not an option or the number of clients in a LAN is very small.
2. The monitoring is much less detailed than what we have with WLCG Squids. There are only aggregates of the total number of requests and total bandwidth, with hourly as the smallest granularity, for the whole domain. There is also a list of top requesting countries for the last 24 hours.
3. Unlike Squid, Cloudflare caches to not support the http If-Modified-Since feature. At one time this feature was very important for reducing Frontier infrastructure load [7], but it isn't so important anymore for CMS; CMS changed their conditions data system to index data blobs by a hash of their contents, so most of the data never changes and is cached indefinitely. If-Modified-Since is still used by the WLCG infrastructure, but it is disabled for Cloudflare by a configuration in the Squids running on the Frontier servers that removes the http Last-Modified header when a request comes from a Cloudflare server. CVMFS does not use If-Modified-Since because almost all of its data is cacheable indefinitely.
4. Cloudflare forwards more requests to the origin servers than Squids do, but it is still much fewer than the number of clients so it is good enough.
5. Cloudflare only caches http on ports 80 and 8080. Frontier clients have always been configured to contact Frontier servers on port 8000, but the CMS Frontier servers were changed to support both ports 8000 and 8080. CVMFS stratum 1s always supported ports 80 and 8000, so for openhtc.io we use port 80. Cloudflare also caches https on port 443, but we don't take advantage of that at this time.
6. The Cloudflare Terms of Service at one time [8] said that they might disable a domain if it served things not intended for web browsers and if they noticed it causing any adverse affects on their other customers. We communicated with them about it, described our service, and they said it was unlikely to cause any adverse

effects. Since that time the Terms of Service have changed [9] and they no longer mention this restriction.

6 Risk Mitigation

If Cloudflare did disable the openhtc.io domain for any reason there is not another equivalent service provider at this time, but we could avoid service disruption by hosting the domain on any DNS service provider, use ordinary uncached DNS aliases, and update Web Proxy Auto Discovery to list Squids again. The performance would be reduced, but it would continue to function without having to change client configurations.

7 Conclusions

The Open HTC CDN is an effective method of improving performance for jobs that run where no local Squids are available, and it has only the low costs of development and support. It is still preferable to use local Squids whenever possible. If the commercial service were ever to become unavailable, there is an effective recovery plan to continue service at reduced performance.

Acknowledgements

This document was prepared by CMS using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359.

References

1. D. Dykstra, J. Phys. Conf. Ser. **331**, 042008 (2011)
2. J. Blomer, et. al., J. Phys. Conf. Ser. **396**, 052013 (2012)
3. J. Barranco, et. al., Open Eng., **7**, Issue 1, 379-393 (2017)
4. Cloudflare : <https://cloudflare.com>
5. Anycast : <https://blog.cloudflare.com/a-brief-anycast-primer>
6. D. Dykstra, et. al., J. Phys. Conf. Ser. **898**. 052043 (2017)
7. D. Dykstra and L. Leuking, J. Phys. Conf. Ser. **219**, 072034 (2010)
8. Previous TOS : <https://web.archive.org/web/20180104152120/cloudflare.com/terms>
9. Current TOS : <https://web.archive.org/web/20181011084248/cloudflare.com/terms>