Web Proxy Auto Discovery for the WLCG

D Dykstra[1], J Blomer[2], B Blumenfeld[3], A De Salvo[4], A Dewhurst[5], V Verguilov[6]

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

[2] EP-SFT Department, CERN, Geneva, Switzerland

[3] Johns Hopkins University, Baltimore, MD, USA

[4] Istituto Nazionale di Fisica Nucleare, Rome, Italy

[5] Rutherford Appleton Laboratory, Oxfordshire, England

[6] EP-UCM Department, CERN, Geneva, Switzerland

E-mail: dwd@fnal.gov, jblomer@cern.ch, bjb@jhu.edu, Alessandro.De.Salvo@cern.ch, Alastair.Dewhurst@cern.ch, Vassil.Verguilov@cern.ch

Abstract. All four of the LHC experiments depend on web proxies (that is, squids) at each grid site to support software distribution by the CernVM FileSystem (CVMFS). CMS and ATLAS also use web proxies for conditions data distributed through the Frontier Distributed Database caching system. ATLAS & CMS each have their own methods for their grid jobs to find out which web proxies to use for Frontier at each site, and CVMFS has a third method. Those diverse methods limit usability and flexibility, particularly for opportunistic use cases, where an experiment's jobs are run at sites that do not primarily support that experiment. This paper describes a new Worldwide LHC Computing Grid (WLCG) system for discovering the addresses of web proxies. The system is based on an internet standard called Web Proxy Auto Discovery (WPAD). WPAD is in turn based on another standard called Proxy Auto Configuration (PAC). Both the Frontier and CVMFS clients support this standard. The input into the WLCG system comes from squids registered in the ATLAS Grid Information System (AGIS) and CMS SITECONF files, cross-checked with squids registered by sites in the Grid Configuration Database (GOCDB) and the OSG Information Management (OIM) system, and combined with some exceptions manually configured by people from ATLAS and CMS who operate WLCG Squid monitoring. WPAD servers at CERN respond to http requests from grid nodes all over the world with a PAC file that lists available web proxies, based on IP addresses matched from a database that contains the IP address ranges registered to organizations. Large grid sites are encouraged to supply their own WPAD web servers for more flexibility, to avoid being affected by short term long distance network outages, and to offload the WLCG WPAD servers at CERN. The CERN WPAD servers additionally support requests from jobs running at non-grid sites (particularly for LHC@Home) which they direct to the nearest publicly accessible web proxy servers. The responses to those requests are geographically ordered based on a separate database that maps IP addresses to longitude and latitude.

1. Introduction

The Worldwide LHC Computing Grid (WLCG) has a Content Delivery Network (CDN) used primarily for delivery of conditions data via the Frontier Distributed Database caching system (Frontier) [1,2] and for delivery of software via the CernVM FileSystem (CVMFS) [3]. The WLCG CDN is based on general purpose web caching proxies, that is, the open source web tool called Squid [4]. The CDN was first deployed by CMS and ATLAS for Frontier. When CVMFS was introduced, its acceptance grew very quickly in large part because the Squid network was already deployed to many grid sites. There are many other smaller use cases that take advantage of the WLCG CDN squids, and

there would likely be many more if the squids were easier for applications to locate. Currently ATLAS and CMS each have their own incompatible methods of locating squids at the sites they manage, and CVMFS has yet another method. This makes it difficult for other applications to discover what squids they can use, and makes it difficult for ATLAS and CMS to run jobs opportunistically at each other's sites.

CDNs on the internet normally do not depend on clients that are smart enough to locate nearby general-purpose (also known as "forward") caching web proxies. Instead, the CDN servers, which are also typically caching proxies, are smart enough to locate the ultimate source of the data themselves (this type of proxy is known as a "reverse proxy"). In addition, DNS caching is often abused in these cases to resolve to different IP addresses for different geographical areas. Implementing it entirely on the server side can be made to work when all the cache servers in the CDN are controlled by one entity, but the grid model is distributed control. In addition, the cache hit rates are so high with the WLCG applications that it usually makes sense to deploy caches at the grid site premises, especially when they are as easy to maintain as a general-purpose proxy like Squid. There probably could be a way to distribute configurations that would work without decentralized control, but there is an easier modification to the existing design that is based on a de-facto internet standard: Web Proxy Auto Discovery [5].

This paper describes Web Proxy Auto Discovery and its design and implementation for the WLCG.

2. Web Proxy Auto Discovery standard
The Web Proxy Auto Discovery (WPAD) de-facto internet standard is supported by all major web browsers, often enabled by an option such as "Auto-detect proxy settings for this network." The standard is to try to read the URL http://wpad/wpad.dat, deliberately using a host name "wpad" that is not a fully qualified domain name so it will find the name in the local domain. If that URL can be found, it is interpreted as a Proxy Auto Config [6] file.

2.1. Proxy Auto Config standard
The Proxy Auto Config (PAC) standard is a subset of javascript. Specifically, it contains a javascript function called "FindProxyForURL" and returns a string that is a semi-colon separated list of proxy URLs preceded by the string "PROXY". For an example, see Figure 1.

```
function FindProxyForURL(url, host) {
  return "PROXY http://squid1.unl.edu:3128; PROXY http://squid2.unl.edu:3128";
}
```

Figure 1. Sample Proxy Auto Config file

The javascript code can include any amount of javascript conditional code, including defining additional functions, although the number of pre-defined functions available are very limited. The destination URL and host are provided, so different results can be returned based on those values. Also, there is a function available for reading the IP address of the source, so different results can be returned based on that information. The proxy URLs returned may include host names that are DNS round-robins with multiple IP addresses. If there is any kind of failure with one address, clients should try all the other addresses in a round-robin before moving on to the next URL in a list.

There is an open source library available that parses PAC files, called **pacparser** [7]. Both the CVMFS and Frontier clients use that library and have supported PAC files for two years or more. There is also an open source tool called **pacwget** [8] that has a command line interface similar to **wget** but understands how to locate proxies using a list of URLs to PAC files.

2.2. Extending WPAD for WLCG

The WPAD standard specifies reading from one URL, but the WLCG Http Proxy Discovery Task Force [9] has extended the standard so that clients read from up to four URLs as shown in Figure 2:

```
http://grid-wpad/wpad.dat
http://wpad/wpad.dat
http://wlcg-wpad.cern.ch/wpad.dat
http://wlcg-wpad.fnal.gov/wpad.dat
```

Figure 2. WLCG standard WPAD URLs

If any of the URLs successfully return an answer, the client stops looking and uses whatever results are returned from the FindProxyForURL function in the file. The first two on the list must be tried in that order, but the order of the last two may be swapped. Large sites are encouraged to provide their own grid-wpad or wpad services for additional flexibility, to reduce response latency, and to offload the wlcg-wpad servers. If a site does not provide their own WPAD service, then one of the shared wlcg-wpad servers at CERN or FNAL can respond with an answer. Details about how the wlcg-wpad service responds are in the next section.

3. wlcg-wpad service

The wlcg-wpad service is designed to recognize requests from jobs at WLCG grid sites and return a list of proxies that are available at that grid site. It does that by consolidating information from many different sources (details below), and matching the IP address of the requester to the IP address of a proxy using a GeoIP Organization database from a company called Maxmind [10].

3.1. wlcg-wpad cannot use a static PAC file

Although WPAD is designed to be able to be implemented in most cases as a static PAC file served from a web server, the wlcg-wpad service is implemented as an active python-based (WSGI [11]) web service that gives different answers depending on the incoming IP address. The primary reason why a single static file cannot be used for all of the WLCG is that multiple grid sites use private network addresses for the worker nodes, so the PAC function that supplies a source IP address cannot be reliably used to distinguish grid sites from each other. Instead, as grid jobs call out from a private network through a Network Address Translation (NAT) device to the wlcg-wpad service, a public IP address ends up being used, which the server can find. The public IP can then be used to determine the GeoIP Organization.
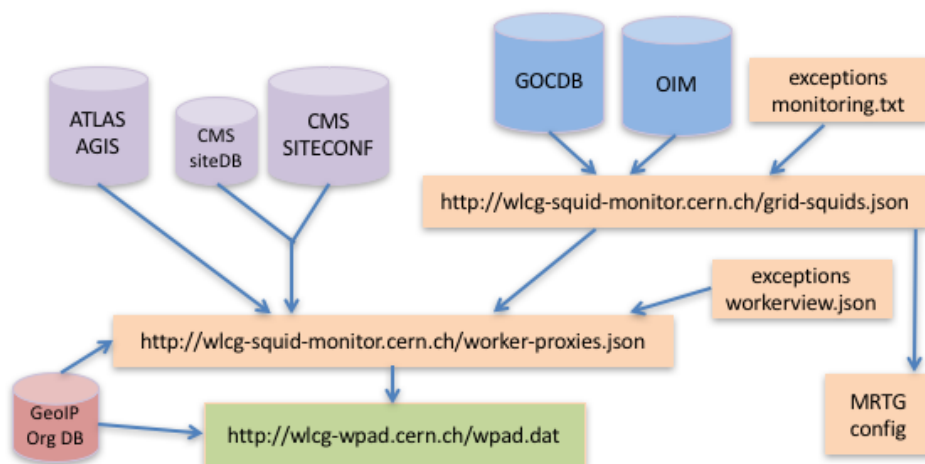
3.2. Information sources

Figure 3. wlcg-wpad service information flow

The wlcg-wpad service draws from a number of different information sources to determine what proxies can be used at grid sites. Figure 3 shows the sources and how they are related to each other.

The first sources of information are the databases where sites register information about the services they have, including squids. Those databases are the Grid Configuration Database (GOCDB) [12] for European Grid Initiative (EGI) sites and the OSG Information Management (OIM) [13] system for Open Science Grid (OSG) sites. Sites are asked to register the public fully qualified domain names of their squids in those databases. If there is more than one squid server providing one service, they are asked to be in a round-robin DNS under one name. This information plus modifications to the information in a monitoring.txt [14] exceptions file on the WLCG Squid monitoring machine (maintained by operators of the monitoring service) are combined to produce grid-squids.json [15]. That is used primarily as input to the WLCG MRTG monitoring service but it is also used by the wlcg-wpad service.

The information in grid-squids.json is helpful but not sufficient for the wlcg-wpad service, for two reasons:
1) Sites have not been asked to define a purpose in OIM or GOCDB for the squid services, so the services might not be ready for all the things that jobs might want to do with them. In particular, CMS and ATLAS Frontier place a much higher load on squids than CVMFS (because CVMFS has a per-node persistent cache and Frontier doesn't), so if a squid service were engineered for only CVMFS it could be overwhelmed by Frontier traffic.
2) OIM and GOCDB contain public names, but if a site runs jobs on a private network, the job needs to use the private network names of the squid services.

For those reasons, the primary sources for wlcg-wpad are the Frontier Squid registrations from CMS (SITECONF [16]) and ATLAS (AGIS [17]). The assumption is that if a site supports one of those two VOs and allows the other in opportunistically, the site would most likely want to share its squid service between the VOs as well. There is also a workerview.json [18] exceptions file on the WLCG Squid monitoring machine (again maintained by monitoring service operators) to modify the information from SITECONF and AGIS, and all the sources are combined to produce worker-proxies.json [19]. The CMS SiteDB [20] is also used to translate CMS site names into the site names used in OIM and GOCDB. The worker-proxies.json list is indexed by GeoIP Organization.

The wlcg-wpad service reads worker-proxies.json, and uses that to keep an in-memory list of proxies for GeoIP Organizations. When a request comes in, it looks up the client's GeoIP Organization and returns the list of proxies for that organization, when there are matching proxies available.

3.3. Private/public networks
SITECONF and AGIS have private network names for squids, which is good because those are what jobs need to use. However, a public name and address are needed to determine a GeoIP Organization. That is where the grid-squids.json information is used: if a squid host name is determined to be private (because it cannot be located in the DNS), then the site name is matched with the OIM or GOCDB site name, and a public address from grid-squids.json is used to locate the GeoIP Organization.

3.4. Multi-site GeoIP Organizations
Although generally each grid site maps to a unique GeoIP Organization, there are cases where multiple grid sites with different lists of proxies map to a shared GeoIP Organization. For those cases, a range of IP addresses assigned to each different site is required. GOCDB already contains this information, so it can be used to tell the sites apart. OIM does not contain the corresponding information, but most of the cases of multi-grid-site GeoIP Organizations are not in the OSG, so that does not make much difference.

3.5. Multiple squid services at a site

Sometimes sites have more than one squid service. In fact, sites are encouraged to run separate squid services for their most important VO(s) and for those VOs that run opportunistically, so no new opportunistic VO workflow can interfere with their most important jobs. There are two ways to handle this situation:

1) Sites can run their own http://grid-wpad/wpad.dat or http://wpad/wpad.dat services, and make those files be as complex as they need to direct different requests to different squid services.
2) The wlcg-wpad service supports shortcut aliases, defined in a configuration file, to associate destination URLs with VOs or services (e.g. atlas, cms, cvmfs, frontier). Then the workerview.json exceptions file can assign different lists of squids to different aliases.

3.6. Client-based load balancing

Both Frontier and CVMFS clients have configuration options that enable load balancing between a list of separate squid names. One limitation of the PAC format is that it does not fully support that. There is a supported approximation, however, and that is to use a javascript random number function to choose a different starting squid, and then list the other squids to try sequentially in case of failure. For example, Figure 4 shows the complete response from the wlcg-wpad service for a request from FNAL:

```
// For USCMS-FNAL-WC1; CMS: T1_US_FNAL
function FindProxyForURL(url, host) {
    if (shExpMatch(url, "*cmsfrontier*.cern.ch*") || shExpMatch(url, "*/cvmfs/cm
s*.cern.ch*")) {
        ran = Math.random();
        if (ran < 0.25) return "PROXY http://cmsfrontier2.fnal.gov:3128; PROXY h
ttp://cmsfrontier3.fnal.gov:3128; PROXY http://cmsfrontier4.fnal.gov:3128; PROXY
 http://cmsfrontier1.fnal.gov:3128";
        if (ran < 0.5) return "PROXY http://cmsfrontier3.fnal.gov:3128; PROXY ht
tp://cmsfrontier4.fnal.gov:3128; PROXY http://cmsfrontier1.fnal.gov:3128; PROXY
http://cmsfrontier2.fnal.gov:3128";
        if (ran < 0.75) return "PROXY http://cmsfrontier4.fnal.gov:3128; PROXY h
ttp://cmsfrontier1.fnal.gov:3128; PROXY http://cmsfrontier2.fnal.gov:3128; PROXY
 http://cmsfrontier3.fnal.gov:3128";
        return "PROXY http://cmsfrontier1.fnal.gov:3128; PROXY http://cmsfrontie
r2.fnal.gov:3128; PROXY http://cmsfrontier3.fnal.gov:3128; PROXY http://cmsfront
ier4.fnal.gov:3128";
    }
    return "PROXY http://squid.fnal.gov:3128";
}
```

Figure 4. Example of client-based load balancing approximation

This approximation is not as good as the load balancing that the CVMFS and Frontier clients do themselves, because this only selects a random starting squid once, when the clients first start. When the clients do their own load balancing, then they continue to balance the squids while the clients continue to run. For this reason, when using WPAD it is better to group squids together into a round-robin DNS service rather than listing them individually, so the clients can continue to balance the squids.

4. lhchomeproxy WPAD service
The LHC@Home [21] project enables individuals to run LHC jobs on their home computers. The users are not expected to supply their own squids, so squids that are open to the internet are made available. There is a http://lhchomeproxy.cern.ch/wpad.dat service that returns the list of available squids for LHC@Home, sorted by geographical distance between the client and the squids. This uses a

different database from Maxmind that maps IP addresses to longitude and latitude. The service is hosted on the same servers as wlcg-wpad.cern.ch, but they return a completely different response depending on the "Host" http header field which is set to the host name from the URL that the client requests. Currently, squids for this service are only available at CERN and FNAL, but there are plans to add more at other WLCG sites around the world. The service name and available squids are specified in a configuration file so they are easy to change.

5. cmsopsquid WPAD service

There is one more WPAD service, http://cmsopsquid.cern.ch/wpad.dat, also hosted on the same servers as wlcg-wpad and lhchomeproxy. This one is specifically for CMS opportunistic use. It is defined by configuration file to first try to look up a list of squids exactly like the wlcg-wpad service, but if that fails then it instead returns a geo-sorted list of squids available publicly for CMS opportunistic use.

6. Status and next steps

As of this writing, CMS is using the cmsopsquid WPAD service in production for opportunistic use cases on the OSG. LHC@Home is also using the lhchomeproxy WPAD service in production. Each service has been supplying a list of squids to use to approximately 80,000 to 150,000 jobs each week in January 2017.

The servers at CERN that supply the WPAD services are a pair of 10gbit/s physical machines that also provide four publicly available squid services on different TCP ports: the CMS opportunistic and LHC@Home squid services already mentioned, as well as backup squid services for CMS Frontier and for CVMFS. There are a similar pair of servers at FNAL supplying the same four squid services. The FNAL servers don't yet supply a mirror of the WPAD services, but that is planned to be added in the future.

So that anyone can easily see what response one of the WPAD services would give for any address, a "?ip=address" can be added to the end of the URL to find out what that response would be. For example, http://wlcg-wpad.cern.ch/?ip=131.225.152.161 shows the answer for FNAL.

Currently the WPAD services only support IPv4 addresses. IPv6 addresses are planned to be added sometime in 2017. Also, the IP address ranges of sites are not yet read from GOCDB but that is planned to be added soon.

In Infrastructure-as-a-Service cloud systems it is sometimes advantageous to dynamically create and track squid servers. A system called Shoal [22] is designed to manage that. Shoal supports the WPAD format for making the addresses of squids known to clients. To integrate with WPAD for WLCG, a Shoal server could be run in the cloud at the local name http://grid-wpad/wpad.dat. Alternatively, in the future it might be a good idea to run a Shoal server at CERN and have the WLCG WPAD service return the Shoal server's answers when requests come in from addresses matching squids tracked by the Shoal server.

7. Conclusions

There is now a standard method for any job to locate accessible web caching proxies when they run anywhere on the WLCG. The method isn't perfect, but it should continue to improve as usage increases.

References

[1]   Dykstra D and Lueking L 2010 Greatly improved cache update times for conditions data with Frontier/Squid *J. Phys.: Conf. Ser.* **219** 072034

[2]   Dykstra D 2011 Scaling HEP to Web Size with RESTful Protocols: The Frontier Example *J. Phys.: Conf. Ser.* **331** 042008

[3]   Blomer J, Buncic P, Charalampidis I, Harutyunyan A, Larsen D, and Meusel R 2012 Status and future perspectives of CernVM-FS *J. Phys.: Conf. Ser.* **396** 052013

[4]   Squid: http://www.squid-cache.org

[5]   WPAD: https://en.wikipedia.org/wiki/Web_Proxy_Autodiscovery_Protocol

[6]   PAC: http://findproxyforurl.com

[7]   pacparser: https://github.com/pacparser/pacparser

[8]   pacwget: https://github.com/pacwget/pacwget

[9]   WLCG task force: https://twiki.cern.ch/twiki/bin/view/LCG/HttpProxyDiscoveryTaskForce

[10]   Maxmind: http://maxmind.com

[11]   WSGI: https://en.wikipedia.org/wiki/Web_Server_Gateway_Interface

[12]   GOCDB: http://goc.egi.eu

[13]   OIM: https://oim.opensciencegrid.org

[14]   monitoring.txt: http://wlcg-squid-monitor.cern.ch/exceptions/monitoring.txt

[15]   grid-squids.json: http://wlcg-squid-monitor.cern.ch/grid-squids.json

[16]   SITECONF: In CVMFS, /cvmfs/cms.cern.ch/SITECONF

[17]   AGIS: In CVMFS, /cvmfs/atlas.cern.ch/repo/sw/local/etc/agis*

[18]   workerview.json: http://wlcg-squid-monitor.cern.ch/exceptions/workerview.json

[19]   worker-proxies.json: http://wlcg-squid-monitor.cern.ch/worker-proxies.json

[20]   SiteDB: https://cmsweb.cern.ch/sitedb/prod/sites

[21]   LHC@Home: https://lhcathome.cern.ch

[22]   Gable I 2014 Dynamic web cache publishing for IaaS clouds using Shoal *J. Phys.: Conf. Ser.* **513** 032035