

# Web Proxy Auto Discovery for WLCG

Monday, 10 October 2016 11:30 (15 minutes)

All four of the LHC experiments depend on web proxies (that is, squids) at each grid site in order 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 proxy 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. This paper describes a new Worldwide LHC Computing Grid (WLCG) system for discovering the addresses of web proxies that 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) files. Both the Frontier and CVMFS clients support this standard. The input into the WLCG system comes from squids registered by sites in the Grid Configuration Database (GOCDB) and the OSG Information Management (OIM) system, combined with some exceptions manually configured by people from ATLAS and CMS who participate in WLCG squid monitoring. Central WPAD servers at CERN respond to http requests from grid nodes all over the world with a PAC file describing how grid jobs can find their web proxies, based on IP addresses matched in 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 WPAD servers at CERN. The CERN WPAD servers additionally support requests from jobs running at non-grid sites (particularly for LHC@Home) which it directs to the nearest publicly accessible web proxy servers. The responses to those requests are based on a separate database that maps IP addresses to longitude and latitude.

## Secondary Keyword (Optional)

Distributed data handling

## Primary Keyword (Mandatory)

Distributed workload management

## Tertiary Keyword (Optional)

**Primary author:** DYKSTRA, Dave (Fermi National Accelerator Lab. (US))

**Co-authors:** DEWHURST, Alastair (STFC - Rutherford Appleton Lab. (GB)); DE SALVO, Alessandro (Universita e INFN, Roma I (IT)); BLUMENFELD, Barry Jay (Johns Hopkins University (US)); BLOMER, Jakob (CERN); VERGUILLOV, Vassil (Bulgarian Academy of Sciences (BG))

**Presenter:** DYKSTRA, Dave (Fermi National Accelerator Lab. (US))

**Session Classification:** Track 3: Distributed Computing

**Track Classification:** Track 3: Distributed Computing