Update on the WLCG Squid Monitoring Task Force Squid Configuration Information

WLCG Operations Coordination Meeting
24 January 2013

Dave Dykstra
dwd@fnal.gov

Squid Configuration Information

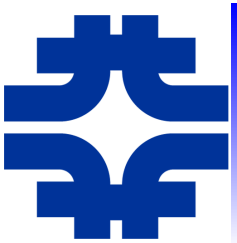
- We have been debating my comprehensive proposal for handling squid configuration information
  - https://twiki.cern.ch/twiki/bin/view/LCG/SquidMonitoringTFInfoSystem
- There is general agreement that GOCDB & OIM should contain a public name for each squid service
  - Multiple squids in a service in round-robin DNS
  - Maintained by site administrators
  - Multiple services per site supported
  - Enables marking services down for maintenance
  - Enables SAM test of the service
  - Debating whether should include port number 3128
  - Debating whether should indicate type (Frontier or CVMFS)

MRTG Monitoring information

- MRTG monitoring configuration information isn't always the same as what will be stored in GOCDB & OIM
  - Need to know order of names for multiple machines
  - Sometimes use different port or multi-squid per machine
  - Sometimes monitor extra squids (e.g. reverse proxies)
- There is general agreement that configuration differences from GOCDB & OIM for MRTG monitoring will be stored on the monitoring server only
  - This is already the way it is done
  - Maintained by hand by each VO operations personnel
  - Will use a simple common file format for differences (now done differently between CMS & ATLAS)
  - Will audit against DNS entries

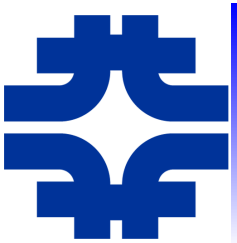
Worker Node information

- Debating whether Worker Node information is in-scope
  - Getting WLCG agreement on this is a main goal of mine
  - There are many cases already using auto-discovery (e.g. CernVM, GlideinWMS, & ATLAS site configurator)
  - Becoming more important as more applications use proxies
  - The ways that this has been done so far are all different and none are good, complete solutions
- Again, Worker Node configuration information sometimes differs from GOCDB & OIM
  - Sometimes use private network access to squids
  - Sometimes use different port (that might already be there)
  - If multiple squid services, which one to use?



Web Proxy Auto Discovery

- It turns out there's an internet standard for this called Web Proxy Auto Discovery (WPAD)
  - Clients load javascript subset called Proxy Auto Config (PAC) from http://wpad/wpad.dat
  - Translates client IP address to proxy list
  - Open source library available to parse
  - Both frontier client and CVMFS client plan to support



PAC file example

```
function FindProxyForURL(url, host) {
    if (isInNet(host,”135.225.160.0”,”255.255.160.0”) &&
        shExpMatch(url,”*/cmsfrontier*”))
          return “PROXY http://cmsfrontier.fnal.gov:3128; DIRECT”;
   else
          return “PROXY http://squid.fnal.gov:3128; DIRECT”;
}
```

WLCG WPAD service

- I propose providing http://wlcg-wpad.cern.ch/wpad.dat that supplies a valid PAC file for all WLCG sites
  - Capable of supporting queries from all WLCG worker nodes but most will continue to use existing methods to configure Frontier & CVMFS
  - Store differences from GOCDB & OIM only on this server
  - Maintained by same people maintaining squid monitor
  - SAM audits of existing Frontier & CVMFS configs to ensure correctness, and other applications will use this service
  - Large sites may provide their own at http://wpad/wpad.dat which will override central service
  - This would be the only service providing Worker Node squid proxy configuration information