

This is Dave's proposed standards for Http Proxy Discovery in the WLCG. See SquidMonitoringTFInfoSystem as background; from the information system's viewpoint we're talking about the "Worker Node View" here. See also the task force's references.

From a grid job's perspective, here's the proposed standards:

1. If the \$http_proxy environment variable is set, use it. It MAY refer to a round-robin of multiple proxies, and if so the application SHOULD try all of them if any fail.
2. Else if \$PAC_URLS is set, use it as a semicolon-separated list of web Proxy Auto-Config (PAC) file URLs. The URLs MUST start with either file:// or http://. Try them in order and if one of them successfully returns a file, parse it and use the list of returned proxies. If it is an http:// URL, it SHOULD refer to a round-robin of at least 2 servers for reliability, and if so the application SHOULD try them all if any fail.
3. Else try reading http://wpad/wpad.dat, and if it returns a file, parse it as a PAC file.
4. Else try reading http://wlcg-wpad.cern.ch/wpad.dat and parse it as a PAC file.

For an application that wants to download a file this can be easily implemented by using `pacwget`. If \$PAC_URLS is not set then set it to "http://wpad/wpad.dat" and in any case add ";http://wlcg-wpad.cern.ch/wpad.dat". Frontier & CVMFS are both being enhanced to support these Proxy Auto-Config files, although existing sites aren't expected to convert their current configurations to use them.

Here's the plan for the wlcg-wpad.cern.ch servers:

1. Look up registered squids in OIM and GOCDB.
2. For those relatively few sites that use different names (for example if they are on private networks), convert those registered squids to the "Worker Node View" through simple translation files maintained by hand on those servers by WLCG Squid operations personnel.
3. Automatically generate wpad.dat files corresponding to the GeoIP "ISP"s of the registered squids' addresses. To do this will also need to consult translation files from wlcg-squid-monitor.cern.ch that distinguish between different applications for registered squids, for the relatively few sites that distinguish.
4. Have the apache server look up a requesting IP address and return matching wpad.dat files based on GeoIP "ISP" matches with squids.
5. Enhance existing experiment-specific SUM tests from ATLAS & CMS to audit Frontier configuration information with a grid job's perspective above.
6. Add a SUM test to audit CVMFS configuration information.

Large sites that support opportunistic use SHOULD have a separate set of squids strictly for opportunistic use, separate from their production squids. Large sites SHOULD also run a pair of their own wpad servers.

This topic: LCG > HttpProxyDiscoveryProposal

Topic revision: r6 - 2016-02-17 - DaveDykstra



Copyright &© 2008-2024 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback