

Squid in OSG

Igor Sfiligoi
for the CDF VO

Product Description

- **Squid** is a Web proxy and cache
 - Supports several protocols, including HTTP and FTP
 - Can be used both for firewall traversal and for reducing the WAN traffic
- It is a very mature and stable product
 - Used in production environments all over the world
- <http://www.squid-cache.org>

Why CDF wants Squid

- Each CDF grid job has a 100+MB tar ball attached to it (user binaries)
- CDF grid jobs are heavily parallel
 - Each job cluster has hundreds of identical sections
 - only one tarball per cluster
- CDF loads tarballs from a central HTTP server
- A Squid cache can greatly reduce the network load and shorten startup times

Other users of Squid

- CDF is using FroNTier to access the calibration data
 - Lower bandwidth but more latency sensitive
 - Can greatly improve job startup
- ATLAS and CMS has also expressed interest in using Squid
 - My understanding is that mostly for FroNTier

Packaging and Configuration

- Squid does not rely on any other software
- It is distributed as part of VDT (1.3.11+)
- It can be used out of the box
 - Small configuration changes can be useful to better adapt to the available hardware and the size of the Grid pool
 - Memory cache size
 - Disk cache size

Required Resources

- Squid can be run on any machine that
 - Can be contacted from the worker nodes
 - Has outgoing connectivity to the WAN
- It uses a very small amount of CPU
 - $<10\%$ of a CPU
- It requires a reasonable amount of disk for caching
 - 40GB for a medium sized Grid pool

Service Location

- The most obvious candidates for a Squid server:
 - Storage Element
 - A dedicated machine on the border
- Installation on Compute Element not recommended
 - Apart for very small sites
- Need to decide how to publicize it
 - I would prefer an environment variable (\$OSG_SQUID)

Test and Validate (1)

- The service is considered tested after:
 - site level test (post-install):
 - no errors, no warnings
 - resource provider test:
 - Loading a set of tarballs via HTTP using Squid cache from all the available worker nodes
 - tarball size 150MB+-100MB, 20GB total
 - No errors should be seen
 - The Squid server should not overload the machine
 - Files should be loaded in the Squid cache just once

Test and Validate (2)

- The service is considered tested after:
 - application test:
 - CDF glideins start and run
 - No errors
 - others?

Service availability

- CDF would like every site to have a Squid server installed
 - We can survive without, but will heavily load the network and will waste CPU resources
- We are asking OSG to **strongly encourage** sites to install the service

Support and Documentation

- Part of VDT
- Official documentation at:
 - <http://www.squid-cache.org/>
- CDF support:
 - Barry Blumenfeld
 - Igor Sfiligoi

Pending Issues and Next Steps

- No known issues
- Would like to have it in OSG 0.6.0