LCAS/LCMAPS configuration options for pilot jobs

John White, Helsinki Institute of Physics.
David Groep, NIKHEF.

www.eu-egee.org

Information Society

Background

- "Pilot jobs" need to perform identity switch on the WN.
  - Mapping decision to be determined by LCMAPS.
- **In order to run glexec on the WN (short time scale), either:**
  1. Run LCAS/LCMAPS at a site-central location.
     - Single point of failure for site.
     - Scalability not guaranteed.
     - Not certified for NFS.
     - MUCH testing needed.
     - Consistent mapping decisions.
  2. Run LCAS/LCMAPS on each WN;
     - More durable. No single point of failure
     - Possible clash of FQAN mappings over WNs.
     - Synchronization of mapping configurations.
- **Eventually use a Site-Central Authorization Service.**
  - Technically best solution.
  - Consistent, synchronized mapping decisions.
  - Authorizes user credentials. Returns mapped uid/gid.

LCAS/LCMAPS Config Options

1. **Run LCAS/LCMAPS on each WN with centralized configuration.**
   - Shared home directory system (NFS?).
     - Provide the following files:
     - /etc/grid-security/gridmapdir
     - /etc/grid-security/grid-mapfile
     - /etc/grid-security/groupmapfile
   - Based on experience (NIKHEF) should scale to O(1000) nodes.
2. **Run LCAS/LCMAPS on each WN with local configurations. (preferred)**
   - Use WN-local, VO-agnostic, generic pool accounts. eg. pool0000 to pool0032.
   - WN-local home directory. Quick recycling.
   - No need for shared gridmapdir etc
   - Job cannot "escape" from your WN, pilot job contained, traceable.
   - Need to synchronize the LCAS/LCMAPS configs. YAIM or Quattor.

Logging

- glexec will log the uid switch based on FQAN.
- Ideally would use the (remote) syslog facility.
  (glexec $\geq$ 0.5.25)
- **In option 1:**
  - Will log FQAN/uid/gid switches from central configuration.
  - Consistent logging if syslogging centrally.
- **In option 2 .**
  - Will log FQAN/uid/gid switches within WN user-space.
  - Some WN-specific info also needed if syslogging centrally.
  - Otherwise a clash of uids from various WNs is possible.
    - syslog collector includes the machine name + timestamp.
      Alleviates problem
    - Accounting on pilot FQAN/uid:
      VOs are responsible, in their framework, for tracking the individual users.

Summary

- SCAS client code has been completed.
  - Being developer-tested and being tested by OSG.
- SCAS server code is written.
  - Important site-central code. Still being developer tested.
  - Interoperability library now stable (beta version).
- SCAS delivery dates. As of April 29th 2008.
  - Second week of May to end of June 2008.
  - Must be certified at the level of a WMS/CE.
- For short-term tests of the glexec-on-WN. We suggest:
  - Running glexec/LCAS/LCMAPS on the WN. With:
    - Centrally-maintained LCAS/LCMAPS configuration.
    - LCAS/LCMAPS configuration on WN. Maintained through central service.

GUMS vs. LCAS/LCMAPS

- From in-person discussions at the San Diego MWSG:
  - `http://docs.google.com/View?docid=dmqrsvn_5wvp7sx`
- Other "points" that GUMS assumes:
  - Create a pre-created dedicated account at every site for each possible combination of VO user/FQAN combination.
  - grid-mapfile synchronization delay, up to 6 hours for sites to 'recognise' your new vo membership/roles/groups.
  - No validation of the VOMS attributes.
  - Full-scale database with web management tools to do a simple account mapping.

Issues

- Is it transparent for CE/Batch system with shared home dirs?
  - The target uid and the pilot job uid are completely de-coupled.
  - Anything done as the target uid is transparent.
  - The batch system will only see the pilot job.
- Actions in the pilot do no affect the batch system, except:
  - How to completely kill the job:
    - gLExec will preserve the process tree and parent ID, the children, even if setuid'ed, will be killed.
  - How to do accounting:
    - gLExec will accrue CPU/wall time for it's children and report that back in it's "own" execution time at completion.
- Pilot job handles the "payload uid" scratch area ($TMPDIR).
  - This can be done with repeated invocations of gLExec.
  - Where it is guaranteed to give the same mapping on subsequent invocations.
  - A standard pattern will be documented.
- In summary: the batch system will not notice the difference between local or shared $HOME for the target uids.