User traceability and log analysis tools

Eygene Ryabinkin, RRC Kurchatov Institute
Giuseppe Misurelli, INFN-CNAF (speaker)
Daniel Kouril, CESNET

EGEE09 Conference
September 22, 2009
Barcelona

www.eu-egee.org

e-infrastructure

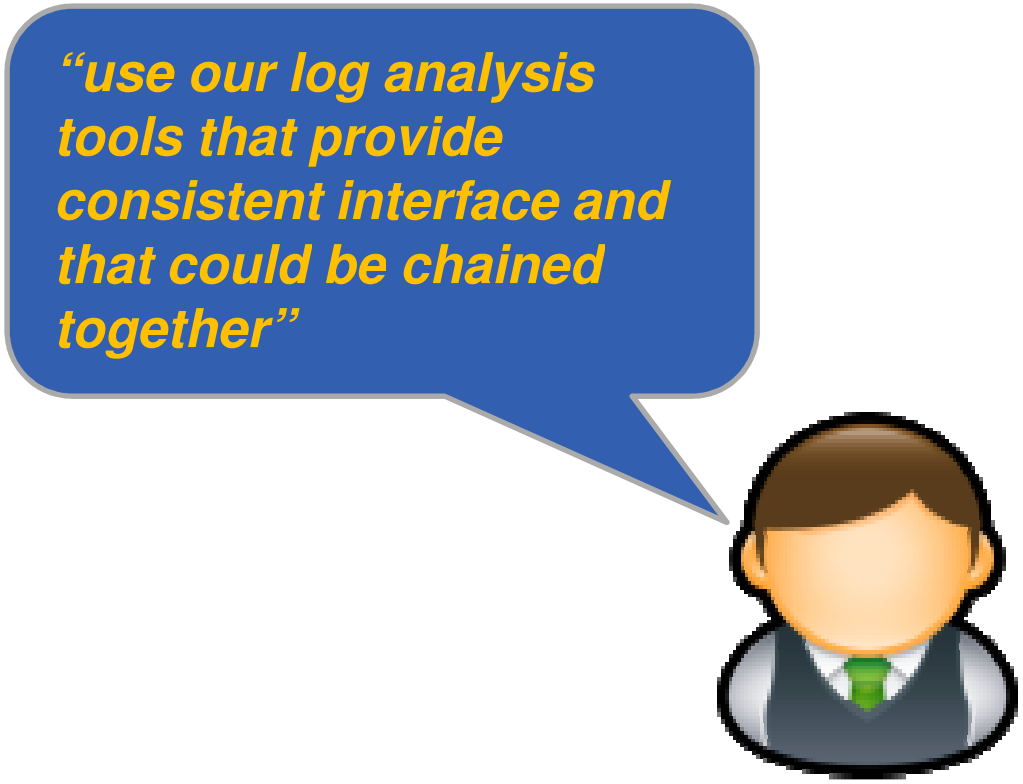
Outline

- **Log analysis**
  - How to figure out what's going on
  - How to relate the records at various sources
- **What OSCT has been provided so far**
  - gLite-LB tracer (`lbtrace`)
  - lcg-CE tracer (`dig-lcgce`)
- **How does OSCT distribute the tools**
  - RPMs available at?
  - Documentation
  - Support
  - Future plans

Log analysis: security officer viewpoint

Log analysis: OSCT suggested solution

- **The OSCT starting point:**
  - create at least some usable tools and show them to the public



gLite services currently traced

- **gLite-LB (`lbtrace`)**
- **lcg-CE (`dig-lcgce`)**

- **Why we started with these two services?**
  - they keep the majority of the information about user jobs
    - Not counting data transfer
    - CREAM CE? Not yet investigated, but in TODO list
  - easily relation between log records and relevant attributes to be searched
    - Job IDs
    - User DNs
    - VOMS attributes

gLite-LB tracer: `lbtrace`

- **Queries LB hosts about user jobs and bookkeeping information**
  - job hops through the different Grid services
- **Can list job records basing on simple query syntax**
  - `status eq done and dst eq <CE>`
- **Currently can interrogate only the live part of the LB database**
  - job records are periodically purged into the offline record library
- **Needs creation of indices on the LB backend**
  - easy but a bit annoying on production LB (services restart required)

gLite-LB tracer: usage example

- **Looking for information about jobs recorded by a given LB**
  - Security officer needs forensics on a specific job

```
#lbtrace -k host -H octopus.grid.kiae.ru list owner eq
'/DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=samoper/CN=582979/CN=Judit Novak'
and status eq done and destination eq snowpatch-
hep.westgrid.ca:2119/jobmanager-lcgpbs-ops

=== Job 1:
JobId: https://octopus.grid.kiae.ru:9000/B1uLHolwYwpTYh1uwudmjQ
Owner: /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=samoper/CN=582979/CN=Judit
Novak
Source: sam111.cern.ch
JobState: Done
StatusReason: Job terminated successfully
Destination: snowpatch-hep.westgrid.ca:2119/jobmanager-lcgpbs-ops CondorID: 664
GlobusID: [none]
PBSOwner: [none]
PBSNode: [none]
```

lcg-CE tracer: `dig-lcgce`

- **Uses jobmap records from the gatekeeper**
- **A SQL-like interface to the job records**
  - Understands conditional expressions that could be combined
- **Can invoke LRMS-specific tools to trace selected jobs down to the batch system logging layer**
  - Think <<tracejob>>
- **Can be used on central logging host**
  - Needs only jobmap files and installed Python

lcg-CE tracer: usage example

- **Digging for information related to a suspected user DN**
  - Site CSIRTs notified about malicious job submitted by a given DN

```
# dig-lcgce -s 20090901 -e 20090916 userDN eq '/C=IT/O=INFN/OU=Personal
Certificate/L=CNAF/CN=giuseppe misurelli'

{
'localUser': '18700',
'ceID': 'gridit-ce-001.cnaf.infn.it:2119/jobmanager-lcgpbs-cert',
'timestamp': '2009-09-07 14:00:21',
'userFQAN': ['/dteam/Role=NULL/Capability=NULL',
'/dteam/italy/Role=NULL/Capability=NULL', '/dteam/italy/INFN-CNAF/Role=NULL
/Capability=NULL'],
'userDN': '/C=IT/O=INFN/OU=Personal Certificate/L=CNAF/CN=giuseppe misurelli',
'jobID': 'https://lb009.cnaf.infn.it:9000/ZTHAJucuJpw4mgwKysV_2A',
'lrmsID': '118258.gridit-ce-001.cnaf.infn.it'
}
```

Distributing the tools

- **Sources live in SA1 subversion repository**
  - https://www.sysadmin.hep.ac.uk/svn/security
- **Packages will be provided by the SA1 repository**
  - https://twiki.cern.ch/twiki/bin/view/EGEE/EGEESA1PackageRepository
    - basically `yum install <<package_name>>`
- **Documentation and usage examples provided at the OSCT twiki web site**
  - https://twiki.cern.ch/twiki/bin/view/LCG/LogTracing
- **Packages install standard Unix man pages**
  - `man lbtrace`
  - `man dig-lcgce`

Support and future plans

- **Support and feature requests are handled through the SA1 Savannah section**
  - https://savannah.cern.ch/projects/sa1tools/

- **Future plans**
  - Use gLite Job Provenance to access data from offline LB records store
  - Add more SQL-like features useful for overview of the user activity
    - order by, count
  - Write tools for CREAM CE
  - Investigate the possibility for tracing the data movements over storage elements
  - Anything sensible requested by the end-users of the tools

Thanks for your attention

Questions, comments, feature requests?

You're welcome!