Employing peer-to-peer software distribution in ALICE Grid Services to enable opportunistic use of OSG resources

R Jeff Porter & Iwona Sakrejda (LBNL)
Costin Grigoras, Latchezar Betev, Federico Carminati, Pablo Saiz (CERN)
Outline

- Introduction
  - ALICE Grid & AliEn
  - Traditional Software Deployment
- BitTorrent software deployment → AliTorrent
- Testing on OSG resources
- Performance characteristics
- Target multiple resources
- Summary
AliEn & the ALICE Grid

- ALICE Grid Facility
  - In production since ~2004
  - more than 80 sites
  - Small, central operations team
    - & site admins around the world

- AliEn: ‘Alice Environment’
  - Central Services at CERN
    - Task Queue, Job & File Catalog
    - Software management (build & deployment)
  - VO box site-specific operations
    - Job Agent submission
    - Software deployment
    - Resource monitoring
  - Grid Monitoring with MonALisa
  - Data Management
    - AliEn FileCatalog
    - Grid-Enabled XRootD SEs

➢ Yesterday’s talk by Pablo Saiz: “AliEn: ALICE Environment on the GRID”
Traditional Software Deployment

- Managed Central Software:
  - routine software builds
  - Catalogued & stored in AliEn

- Grid Site SW Operations
  - Jobs request SW from VO box service
  - VO box PackMan service pulls SW
  - SW deployed on shared area
  - WNs read SW from shared area

- Shared SW area
  - resource bottleneck
  - single point of failure
  - can require active repairs per site
Basic Torrent details

File chunks contain hashes of original file to provide data integrity

Tracker: map of seeders:files
Seeders: have & serve file
Leeches: pull & serve file chunks
AliTorrent Software Deployment

• Managed Central Software
  – Additional AliEn torrent store
  – Catalogue, seeder & tracker

• Grid site SW deployment
  – VO Box is not involved
  – Jobs pull SW from:
    • alitorrent.cern.ch seeder
    • local peers
    • other sites as available
      – though typically behind a FW

• Resolves:
  – Bottleneck & single point failures
  – Site level maintenance of shared area
AliTorrent Details

• Torrent Features
  – Distributed Hash Tables
    • Decentralized seeder lookup – seeders are trackers
  – Peer Exchange
    • Local peer information is propagated by seeders
  – Local Peer Discovery
    • Multicast to discover peers on same network

• ALICE/AliEn Features
  – Total software download is ~300-400MB
  – Enabled per site (VO box) with AliEn LDAP flag
AliTorrent as an opportunity

• AliTorrent use in AliEn
  – Reduces problems associated with SW deployment
  – Simplifies site operations by removing a VO box service

  ➢ Does not eliminate VO box model from ALICE Grid
  ➢ Does eliminate site-specific VO box requirement

• Elimination of site-specific VO box allows for remote use of other Grid resources ➔ OSG
AliEn-OSG Interface

• Developed in 2009 by ALICE & OSG teams
  – in production at ALICE-US sites since 2009
  – poster @ CHEP’09

• Features added to AliEn
  – VO box as a Condor-G submit host to local OSG CE
  – Condor module in AliEn
    • Builds submit files to launch JobAgents
    • monitors Condor queue for site utilization
  – Specifications put into AliEn LDAP
    • OSG-CE endpoint
    • Job occupancy targets
Extending OSG-AliEn Interface for AliTorrent

• Without torrent

```bash
#!/bin/bash

/project/projectdirs/alice/alicedev/alien.v2-19.168/bin/alien RunAgent
```

• With torrent

```bash
#!/bin/bash

DIR=`pwd`/alien_installation.$$ 
mkdir -p $DIR 
cd $DIR 
wget http://alien.cern.ch/alien-installer -O alien-auto-installer 
chmod +x alien-auto-installer 
./alien-auto-installer -type workernode -batch -torrent -install-dir $DIR/alien 
./alien/bin/alien RunAgent
```

• With torrent on OSG:
  • replace `pwd` above with $OSG_WN_TMP

```bash
DIR=$OSG_WN_TMP/ALICE/alien_installation.$$ 
```

R. Jeff Porter LBNL
First Results

- Initial goal to target NERSC Carver system
  - Carver: NERSC IBM IDATAPLEX HPC system with OSG CE
  - ALICE obtained a small allocation for development work
  - No VO Box would allowed on the system

- Conferred with NERSC security about torrent use
  - earlier requests for use of torrent were ‘tabled’ for review
  - Result: torrent not banned explicitly but NERSC relies on dynamic filtering “go ahead and try”

--- log file ----
It just worked ➔ WNHost = "c1554.nersc.gov"

Getting the torrent
It looks like the default IP address is 128.55.61.240
Seeding the torrent
AliEn workernode installation took approximately 133 seconds
Installation finished!
Performance

• Job latency from SW install: Traditional vs AliTorrent
  – Traditional install:
    • ‘copy’ from shared area
    • lone job latency: ~1 sec
    • steady increase vs #-jobs
  – AliTorrent install:
    • CERN to LBNL/NERSC
    • lone job latency: ~360 sec
    • sharp decrease vs #-jobs

• AliTorrent advantage
  – small numbers of long jobs where SW install time is insignificant
  – sites with large numbers of concurrent ALICE jobs
Expand to OSG Sites

- Single VO box submission to OSG → pilot job factory
  - OSG-CE Endpoints
    - Managed as a list
    - Endpoint:MaxJobs:RSL
  - Test on selected sites
    - OSG-ITB sites
    - ALICE-US sites LBNL & LLNL
  - Close SE → 700 TB SE @ NERSC
    - output destination

NERSC/PDSF → 10592 gt2 pdsfgrid.nersc.gov/jobmanager-sge https://pdsfgrid4.nersc.gov:25251/1337447241/
  10628 gt2 glcc88.ucllnl.org/jobmanager-pbs https://glcc88.ucllnl.org:40894/8704/1337454774/
  10629 gt2 glcc88.ucllnl.org/jobmanager-pbs https://glcc88.ucllnl.org:40902/9292/1337454810/
  10645 gt5 itbv-ce-pbs.uchicago.edu/jobmanager-pbs https://itbv-ce-pbs.uchicago.edu:36913/3026418950974269075/
  10647 gt5 itbv-ce-pbs.uchicago.edu/jobmanager-pbs https://itbv-ce-pbs.uchicago.edu:36913/3026418950974269075/
  10664 gt2 carvergrid.nersc.gov/jobmanager-pbs https://carvergrid.nersc.gov:25359/1337525731/
  10665 gt2 carvergrid.nersc.gov/jobmanager-pbs https://carvergrid.nersc.gov:60888/2359/1337525731/
  10674 gt5 osgitb1.nhn.ou.edu/jobmanager-condor https://osgitb1.nhn.ou.edu:64981/933117525531598993/
  10675 gt5 osgitb1.nhn.ou.edu/jobmanager-condor https://osgitb1.nhn.ou.edu:64981/933117525531598993/
Towards production operations

- Proof in principal is complete
  - AliEn Jobs run at sites submitted by remote VO box
  - successful running on multiple sites from single VO box
  - Plan to target OSG production sites

- Site policy caution:
  - limits can be broad
  - has hampered wider use

- ALICE Experience:
  - initial distrust turns into recognition of approach’s merit
  - simplifies site operations
Summary

- Software deployment on shared area
  - Bottleneck & site-level single point failure
  - site-level SW corruption requires admin intervention

- Torrent model → AliTorrent
  - Removes bottleneck & site-level single point of failure
  - Eliminates a site service & reduces site management
  - Performance capabilities meets typical ALICE workflow & site requirements
    - Eliminates requirement for site-specific VO box

- We have leveraged this capability to demonstrate AliEn workflow for opportunistic use of multiple OSG resources

- AliTorrent is a site-friendly tool for opportunistic (or general) use
  - don’t ask the site to “do” something → install or manage a service
  - ask the site to “not do” something → block torrent use