TORRENT BASE OF SOFTWARE DISTRIBUTION BY ALICE AT RDIG

V. Kotlyar[2], E. Ryabinkin[3], G. Shabratova[1], I. Tkachenko[3], A. Zarochentsev[4]
[1] *Laboratory of High Energy Physics, Joint Institute for Nuclear Research, Dubna*
galina@mail.cern.ch
[2] *State Research Center of Russian Federation Institute for High Energy Institute for High Physics, Protvino, Russia*
Victor.Kotlyar@ihep.ru
[3] *National Research Center "Kurchatov Institute", Moscow, Russia*
gridops@grid.kiae.ru
[4] *Saint-Petersburg State University, Saint-Petersburg, Russia*
andrey.zar@gmail.com

The experience of few RDIG sites in the implementation of such service for processing ALICE jobs will be presented in this report.

The GRID framework of LHC experiment ALICE – AliEn [1] is an open source framework built on Web Services and a Distributed Agent Model. In this model Job Agents are submitted onto a grid site to prepare the environment and pull work from a central task queue located at CERN. The communication between each ALICE site and central ALICE services is realized by ALICE-specific VO box. This is a single point contact. The deployment of job-specific software was performing from early AliEn days via PackMan [2]. This service at VO box simplifies deployment of job software, done onto a shared file system at site, and adds redundancy to the overall GRID system. Last year there was developing, testing and implementing a peer-to-peer method [3] based on BitTorrent for downloading job software directly onto each worker node at several ALICE sites. Today the main part of sites supporting ALICE migrates to the pear-to-pear download of application software.

PackMan usage for deployment of job-specific software

PackMan is a Transport Package building tool for packing up Templates, TVs, Snippets, Chunks and other Packages into a Transport Package. This software packages enables users to easily install and remove software on Linux. In case of PackMan application for ALICE GRID sites operate in such way:

- Jobs request Soft Ware from VO box service;
- VO box PackMan service pulls Soft Ware;
- Soft Ware deployed on shared area;
- Working Nodes read Soft Ware from shared area.

Figure 1 gives a scheme of traditional PackMan usage for software transfer to sites. This scheme has some advantages and disadvantages.

Advantages

There is necessary only one service/site managing for installation of require packages. The routine software builds with catalog & stored in AliEn is managed by Central Software.

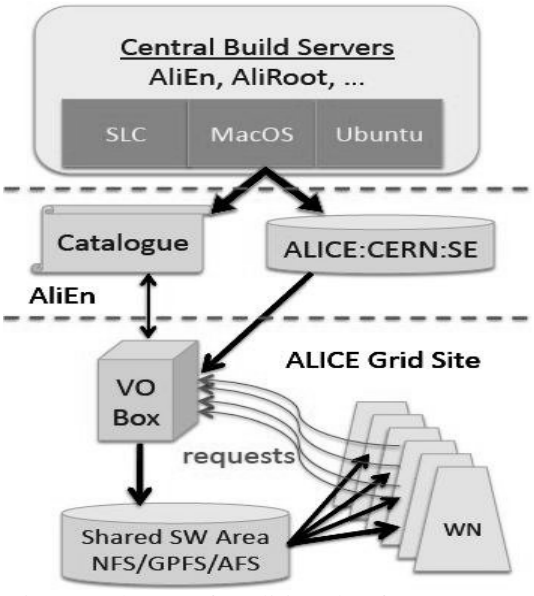

Fig. 1: Scheme of traditional software transport to working nodes with PackMan help

Disadvantages:

- Shared software area is a single point, which is a source of failure / bottleneck
- It is not simple to redeploy rebuilds of the same version. This can require active repairs per site
- Need to keep a short list of active software packages.

Peer-to-Peer method

Of the many p2p file-sharing prototypes in existence, Bit-Torrent is one of the few that has managed to attract millions of users. BitTorrent relies on other (global) components for file search, employs a moderator system to ensure the integrity of file data, and uses a bartering technique for downloading in order to prevent users from free riding. This method has been proposed by ALICE for deployment of job-specific software.

Basic Torrent details

The basic architecture of p2p data and principal scheme of operation with these data presented on Fig 2 and Fig 3 correspondingly. There are using a such definitions used in torrent method:

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

In order to provide data integrity, file chunks contain hashes of original file.

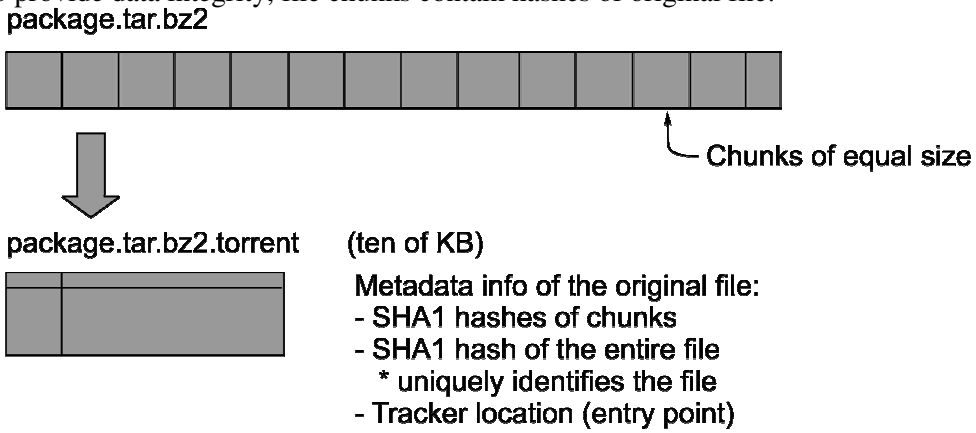

Fig. 2: Structure of p2p data

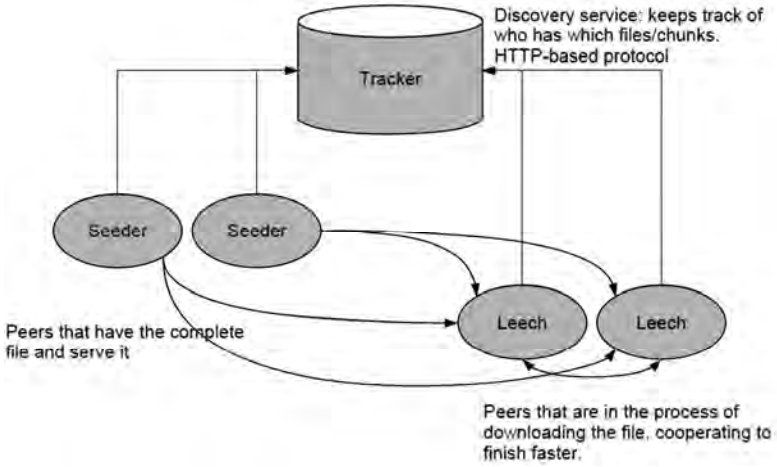

Fig. 3: Scheme of p2p operation.

Implementation pear-to-pear in GRID infrastructure of ALICE – AliEn

Fig 4 presents a principal scheme of pear-to-pear operation for uploading application software to working nodes of site.

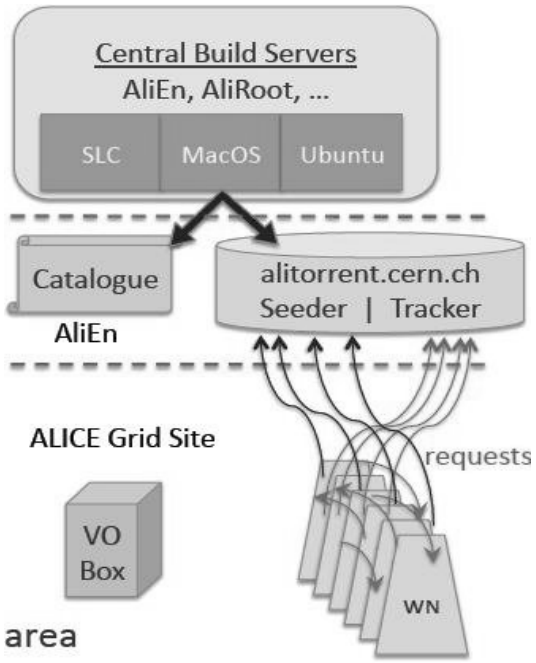

Fig. 4: Pear-to-pear upload of job-specific software

There are new duties applying to manage Central Software as according to p2p scheme VO box from the site side is not involved in the software deployment:

- New AliEn torrent storage has been added for management of p2p scheme.
- In addition to Catalogue there has been stored also seeder & tracker information.

Jobs now pull software from several sources:

a) central alitorrent.cern.ch seeder,
b) other worker nodes will fetch the content mostly from local nodes, if available,
c) worker nodes from site A are usually firewalled from site B, so no inter-site traffic,
d) if initial download is not possible via torrent, fall back to wget and then seed the fetched files.

In addition there is using special features of area2c by ALICE:

i) DHT (Distributed Hash Table) which let decentralize distributed lookup system,
j) Peer exchange. In this case information about local peers will be quickly propagated between peers,
k) LPD (Local Peer Discovery) with application of multicast mechanism to find out other peers in the local network.

So according to all these features, the system can work even without access to the central seeder and tracker.

For successful operation of p2p scheme, the following requirements have been applied to Fire Wall:

- Outgoing access from the WNs to alitorrent.cern.ch:{8088, 8092},
- Please don't allow incoming connection requests from the world to the WNs. But don't be surprised if they do talk to other outside nodes (users that have the package...),
- Allow WN-to-WN connections on at least by TCP,UDP/6881:6999 – aria2c listening ports and UDP,IGMP → 224.0.0.0/4 – local peer discovery.

These "tools" have been integrated in the *alien-installer* and http://alimonitor.cern.ch/packages. The usage of p2p upload (download) is activated by a flag in LDAP. This flag switches modes:

name=<CE_NAME>,ou=CE,ou=Services,ou=<SITE>,ou=Sites,o=alice,dc=cern,dc=ch
installMethod=Torrent

Some practical remarks

The volume of transported software does not exceed 400Mb/job (AliEn itself is packaged in a small (35MB) archive, AliRoot, Root & deps. : max. 300MB/job).

Network load is not so large. CERN seeder limited to 50MB/s. In practice the machine has an average of 8MB/s outgoing. So the fraction that goes to any particular site is negligible.

AliTorrent Software Deployment Advantage:

- Reduces problems associated with SW deployment
- Simplifies site operations by removing the PackMan VO box service. This action does not eliminate VO box model from ALICE Grid. It does eliminate site-specific VO box requirement
- Elimination of site-specific VO box allows for remote use of other Grid resources (for example OSG)
- Eliminate Bofleneck & single point failures

Applications:

Torrents@ALICE: technical details (Experience of RRC-KI and IHEP sysadmins):

Preface

Outline some practical points of using Torrent-based software distribution for ALICE VO as they are seen from the prospective of our Tier-2 site (RRC-KI).

We have been running Torrent-based software distribution since October 2011, so we have around a year of experience with this scheme and so far we had seen no major troubles connected with Torrents@ALICE.

If you have any questions, corrections, suggestions or other stuff; do not hesitate to ask, either during the presentation or by e-mail: gridops@grid.kiae.ru

Software

1. ALICE uses aria2c client, http://aria2.sourceforge.net/.
2. Job downloads the package with Torrent client from ALICE HTTP server (http://alitorrent.cern.ch/), unpacks and starts it.
3. Torrent description files (.torrent) are downloaded from the same server.
4. Torrent client ends when the job ends and we have a set of torrent downloads per each job: no shared cache.
5. Downloaded data lives inside the working directory of the job; in the case of CREAM CE it is CREAMxxxx directory that is removed automatically by the wrapper script.
6. The first downloaded item is the slim AliEn package for the LCG worker nodes.
7. After this, the pilot code is started and the usual sequences of operations are performed.
8. Local PackMan uses Torrent for downloading the needed software packages.

Firewall rules

Just as per aria2c manual, http://aria2.sourceforge.net/aria2c.1.html

```
# Actual downloads.
$IPTABLES -p tcp -m multiport --source <WNs> --dports 6881:6999 -j ACCEPT
# Distributed hash table.
$IPTABLES -p udp -m multiport --source <WNs> --dports 6881:6999 -j ACCEPT
# Peer discovery via multicast.
```

$IPTABLES -p udp --source <WNs> --destination 224.0.0.0/4 -j ACCEPT
$IPTABLES -p igmp --source <WNs>--destination 224.0.0.0/4 -j ACCEPT

Moving a site to the Torrent-based scheme

1. Tune the firewall on the WNs (and only them).
2. Announce to the ALICE mailing list. alice-lcg-task-force@cern.ch, that you're up to using Torrents.
3. Install the up-to-date AliEn on the VO-BOX locally (not on the shared file system). Of course, you can still use the shared file system, but there is no point in doing so if you have a reliable local disk.
4. LDAP entry for your site will be changed to use Torrents at PackMan.

Moving a site to the Torrent-based scheme

1. After LDAP modification it is wise to check that
2. you have no host entry
3. you have no forbidWnInstall entry in the active PackMan leaf for your site:
4. AliEn LDAP lives at ldap://aliendb06a.cern.ch:8389,
5. you're site's leaf is ou=<SITE>,ou=Sites,o=alice,dc=cern,dc=ch
6. PackMan sub-leaf is name=<NAME>,ou=PackMan,ou=Services
7. NAME is the value of attribute packman from the sub-leaf host=<VO-BOX FQDN>,ou=Config.
8. Now watch for new ALICE jobs and aria2c processes at your worker nodes.
9. When things are settled, you can get rid of the VO shared area for ALICE completely.
10. You're done.

Strong and weak points

1. Software distribution was slimmed down (fits in 1 GB per job) –**good**
2. Local disks are used for the software – **good**
3. Once you have some ALICE jobs at your cluster, torrent downloads are blazingly fast – **good**
4. You have multiple copies of software at a single node – **bad**
5. There is a limit for Torrent download rates (1 MB/s), so it will not fill up the network pipe – **good**
6. No monitoring what was downloaded and how fast – **bad**

References

[1] P. Saiz, et al., AliEn –*ALICE environment on the GRID*, Nucl. Instrum. Meth., A502 (2003) 437.

[2] http://packman.links2linux.org/

[3] R.J Porter, I.Saketer, C. Grigoras, et al, *Employing peer-to-peer software distribution in ALICE Grid Services to enable opportunistic use of OSG resources,* contribution 499 at CHEP2012, New-York, 2012; P. Saiz, et al, *AliEn: ALICE Environment on the GRID***,** contribution 516 at CHEP2012, New-York, 2012.