

A Research on Traceability in Pilot-based Workload Management Environment for the Grid

Sang Un Ahn, Ilyeon Yeo, and Sangwook Bae

Abstract— Pilot-based workload management system (WMS) is designed to ensure in advance that the Grid computing resources are suitable to run the payload submitted by users (Grid jobs). Also it is designed to run a series of user jobs in a local resource management system (LRMS) queue. However traditional LRMS is not capable to trace the series of user jobs pulled by the pilot. Therefore the accounting information in the LRMS is inconsistent with VO monitored information, the actual consumed resources by user jobs. In this paper, the traceability for Grid job in LRMS with pilot-based WMS is reviewed and a tag-based logging feature is proposed to address the problem.

Research Keywords— Grid, Pilot-based WMS, Traceability

1 INTRODUCTION

The inherent characteristics of computing resources in the Grid are distributed and heterogeneous which makes challenging to manipulating the Grid resources. The Grid for LHC experiments, which is called Worldwide LHC Computing Grid [1], has implemented a pilot-based workload management system (WMS) to facilitate Grid usage for their scientists, e.g. users. The pilot-based WMS is designed to ensure in advance that the available Grid computing resources are suitable to run the payload actually submitted by users, and to maximize the resource utilization by running a series of user jobs within a queue of local resource management system (LRMS), i.e. batch system secured by the pilot job in a way that it reduces overhead created from repeated submission of user jobs. LHC experiments have different implementations of pilot based WMS: AliEn [2] for ALICE, DIRAC [3] for LHCb, glideinWMS [4] for CMS and PanDA [5] for ATLAS.

However traditional LRMS cannot trace the actual user jobs pulled by the pilot job. This introduced the inconsistency at some extent of accounting infor-

mation, e.g. total amount of wall time (or CPU time) consumed by jobs and total number of jobs processed, between LRMS and WMS in the Grid. Also it may cause potential vulnerability of local site administration because sites cannot monitor the actual jobs. In order to address the untraceableness of user jobs in the pilot-based WMS, we reviewed the state of the art of pilot-based WMS and propose a tag based logging feature to improve the traceability on Grid jobs in the LRMS.

This paper is composed of a review on pilot-based WMS in Section 2, followed by the traceability of a Grid job in LRMS in Section 3 and the conclusions in Section 4.

2 PILOT-BASED WORKLOAD MANAGEMENT SYSTEM

In order to exploit computing resources distributed around the world, the resources should be seen as a federated one by the experiment (or virtual organization, VO, concept of the Grid). Grid implements Computing Element (CE) concepts to interface local resource management with global WMS managed by VO. A CE, can be called as “Grid Gate”, is composed of the following components: security (authentication and authorization of Grid jobs, i.e. “Gatekeeper”) and local resource manager [9]. For example, the CE accepts Grid jobs submitted by VO WMS via proper authentication and authorization processes (based on

-
- Sang Un Ahn is with Korea Institute of Science and Technology Information, 245 Daehak-ro Yuseong-gu 34141 Daejeon, South Korea. E-mail: sahn@kisti.re.kr
 - Ilyeon Yeo is with Korea Institute of Science and Technology Information. E-mail: ilyeon9@kisti.re.kr
 - Sangwook Bae (corresponding author) is with Korea Institute of Science and Technology Information. E-mail: wookie@kisti.re.kr

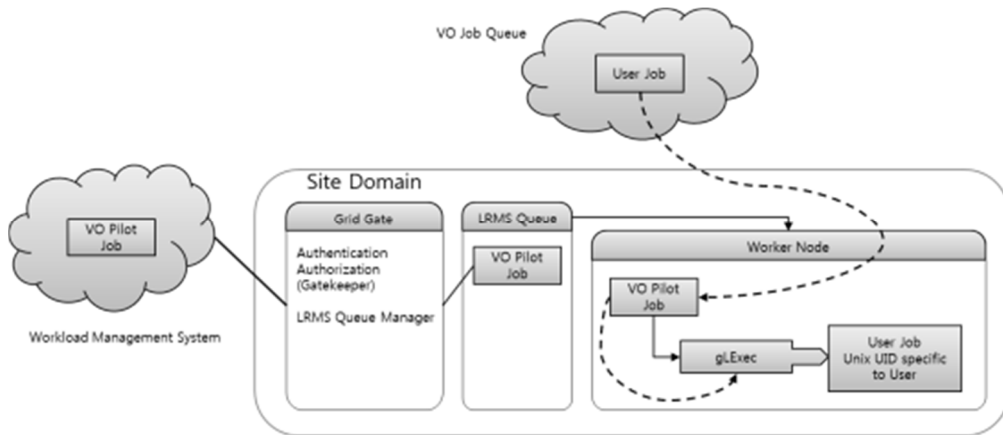


Fig. 1. Grid Job flow of Pilot-based WMS. gLExec maps UID of Grid job into Unix UID in LRMS

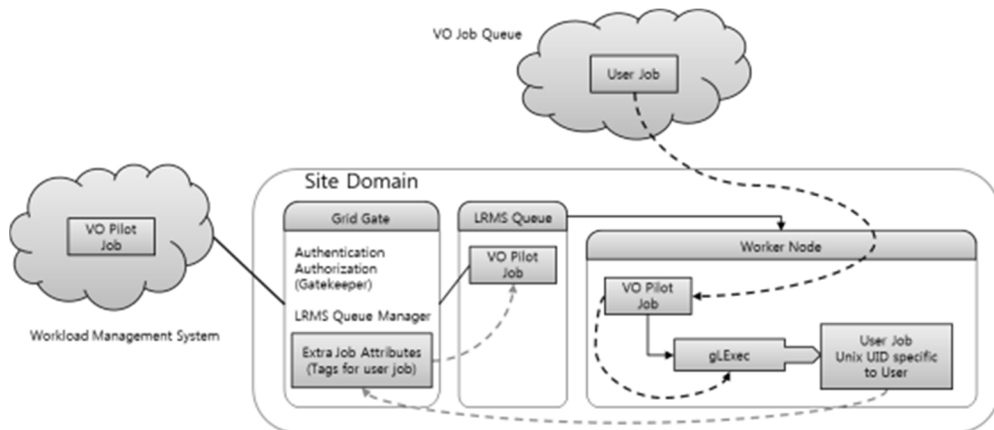


Fig. 2. Enabling additional attributes for user job in logging system of LRMS in order to trace user job information

X.509 certificate), and then inserts Grid jobs to LRMS queue.

The pilot-based WMS submits pilot jobs instead of the actual user job submission. Once the pilot job submitted to LRMS via CE and being running, it pulls the actual user job queued in VO job queue, generally managed by WMS of the VO, where the Grid jobs are submitted by Grid users (VO users). In this process, as depicted in Figure 1, gLExec [10] maps and modifies UID of Grid user job to be matched with Unix local UID.

The advantages of pilot-based WMS are: ensuring that the job environment should be properly configured in advance of actual user job submission which also can be used as service availability checks on local service, reducing the overhead created from repeated submission of user jobs to the remote sites in a way that a series of user jobs is set to be run where the pilot job secured in the LRMS within the time window allowed to the pilot job (usually set by site system ad-

ministrators).

3 TRACEABILITY OF A GRID JOB IN LRMS

From the site's point of view, the activities of the job submitted to its LRMS are invisible. While a series of user jobs spawned by the pilot job is running, the LRMS does not recognize it and it regards the user jobs as a single local user job. In this circumstances, the accounting information collected by LRMS can be inconsistent with the one monitored by VO. For example, the number of processed user jobs monitored by VO can be larger than the number of processed jobs counted by the LRMS by a factor of 3 or 4. Also the total amount of wall time (or CPU time) consumed by user jobs can be different between VO and local site. The LRMS only knows the total amount of wall time (or CPU time) consumed by the pilot job while VO can monitor the total amount of wall time (or CPU time) consumed by each user job (pulled by

the pilot job). In principle, the total wall time (or CPU time) of the pilot job should be longer than the sum of total wall time (or CPU time) of each user job because there should be at some extent overhead created from initialization of pilot job and idle time for pulling user jobs from VO queue and termination time after running multiple user jobs.

In order to address the inconsistency of accounting information, we propose a tag-based logging feature which adds extra attributes (or extension of job description) for the pilot job in the LRMS as shown in Figure 2 in which site administration can trace the details of pilot job activities, such as initialization, idle-to-pull, really-running, done, termination, etc. Also elapsed time for each state should be recorded accordingly. The implementation of this feature depends on the kind of LRMS, i.e. batch system. The advantage of the proposed feature is to secure the least traceability on the jobs running in the LRMS.

4 CONCLUSIONS

Pilot-based WMS is used to manipulate distributed and heterogeneous computing resources in the Grid for LHC experiments. Despite of the advantages of using it, lack of traceability in the LRMS can cause the inconsistency of accounting information with VO monitoring. In order to address the problem, we proposed a tag-based logging feature, which adds extra attributes, for pilot job in submission phase to enable the least traceability for the Grid job in the LRMS. Implementation and testing of the proposed feature on the currently available batch systems are remained for the future work.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) through contract N-16-NM-CR01-S01 and the Program of Construction and Operation for Large-scale Science Data Center (K-17-L01-C05-S01).

REFERENCES

- [1] J. Shiers, "The Worldwide LHC Computing Grid (worldwide LCG)," Proc. Conf. on Comp. Phys. 2006 (CCP 2006), pp. 219-213, 2007. (Conference proceedings)
- [2] S. Gagnasco, L. Betev, P. Buncic, F. Carminati, C. Cirstoiu, C. Grigoras, A. Hayrapetyan, A. Harutyunyan, A. J. Peters and P. Saiz, "AliEn: ALICE environment on the GRID," Proc. Conf. Comp. on Computing in High Energy and Nuclear Physics (CHEP 2007), pp. 062012, 2008. (Conference proceedings)
- [3] A. Casajus, R. Graciani, S. Paterson, A. Tsaregorodtsev, "DIRAC Pilot Framework and the DIRAC Workload Management System", Proc. Conf. on Computing in High Energy and Nuclear Physics (CHEP 2009), pp. 062049, 2010. (Conference proceedings)
- [4] I. Sfiligoi, "glideinWMS – A generic pilot-based Workload Management System", Proc. Conf. on Computing in High Energy and Nuclear Physics (CHEP 2007), pp. 062044, 2008. (Conference proceedings)
- [5] T. Maeno, "PanDA: Distributed Production and Distributed Analysis System for ATLAS", Proc. Conf. on Computing in High Energy and Nuclear Physics (CHEP 2007), pp. 062036, 2008. (Conference proceedings)
- [6] A. Paolo et al., "The gLite Workload Management System", Proc. Conf. on Computing in High Energy and Nuclear Physics (CHEP 2007), pp. 062007, 2008. (Conference proceedings)
- [7] D. Groep, O. Koeroo, G. Venekamp, "gLExec: gluing grid computing to the Unix world", Proc. Conf. on Computing in High Energy and Nuclear Physics (CHEP 2007), pp. 062032, 2008. (Conference proceedings)

Sang Un Ahn received a Ph.D. in Subatomic Physics from Blaise Pascal University in 2011 and in Nuclear and Particle Physics from Konkun University in 2012. He is currently a Senior Researcher in the department of Global Science experiment Data hub Center at KISTI (Korea Institute of Science and Technology Information). His research interests are workload scheduling and processing, data management and policy, and quantum computing.

Ilyeon Yeo received the B.S. and M.S. degrees from the School of Electronic Engineering at Kyungpook National University, in 2000 and 2002, respectively. He has been serving as a Senior Researcher of Global Science experimental Data hub Center at Korea Institute of Science and Technology Information (KISTI) since 2012. He served as a Senior Researcher of Knowledge Information Center and National Science and Technology Information Service at KISTI since 2002. His research interests include Parallel Computing, Information Retrieval, Database, Grid Computing, and Security

Sangwook Bae received the MS degree in Computer Information Communication Engineering from Konkuk university, Korea in 2009, and the PhD degree in Computer Information Communication Engineering from Konkuk university, Korea in 2014. Now, he is a senior researcher in Korea Institute of Science and Technology Information (KISTI) which is a national laboratory specialized to Information Technology, Korea. His research interests include batch system, IPv6, future network and security.