PROJECT DASH: SECURING DIRECT MYSQL DATABASE ACCESS FOR THE GRID

A. Vaniachine, D. Malon, E. May, D. Ratnikov, ANL, Argonne, IL 60439, USA
M. Vranicar, J. Weicher, PIOCON Technologies, Naperville, IL 60563, USA

Abstract

High energy physics applications on computational grids require efficient access to terabytes of data managed in relational databases. The Database Access for Secure Hyperinfrastructure (DASH) project develops secure high-performance database access technology for distributed computing. In the DASH proof-of-concept prototype the Globus grid proxy certificate authorization is automatically integrated within MySQL database code using aspect-oriented software development approach. Pushing grid authorization into the database engine eliminates inefficient data transfer bottleneck and provides end-to-end data security for distributed applications. To provide on-demand database services capability for Open Science Grid, the Edge Services Framework activity builds the DASH mysql-gsi database server into the virtual machine image, which is dynamically deployed via Globus Virtual Workspaces.

DATABASES AND GRIDS

High energy physics collaborations are deploying grid technologies to address petabyte-scale data processing challenges. Besides the file-based event data, their applications require access to terabytes of non-event data (detector conditions, calibrations, etc.) managed in relational databases. In addition to serving data to applications databases play a critical role in the grid middleware: monitoring, catalogues, etc. Crosscutting distributed computing infrastructure, a hyperinfrastructure of databases emerged on the grid (Figure 1).

As grid computing technologies mature, more research is focusing on database and grid integration [1, 2]. Some of these rely on traditional approach of accessing database via an extraneous middleware layer separate from the database system core, which prevents efficient data transfers [3, 4]. Bridging the gap between data accessibility and the increasing power of grid computing requires new approach [5, 6].

The DASH project is building technologies for secure high-performance database access used for distributed event production and processing in High Energy and Nuclear Physics (HENP). To enable fast secure database-resident data transfer for computational grids the DASH technology leverages transport-level security efficiency similar to the https advantages introduced in the Globus Toolkit 4.0 [7].

END-TO-END DATA SECURITY

There are two different models in providing secure database access on the grid. In a traditional approach a separate middleware security layer (an extra server) does the grid authorization. In an alternative approach instead of surrounding database with external secure layers the safety features are embedded inside of server.

Middleware Approach

Encapsulating grid security in a separate middleware layer is a traditional technique used, e.g., in the OGSA-DAI project [2, 3]. It results in the traditionally weak database authorization techniques behind the secure layer,

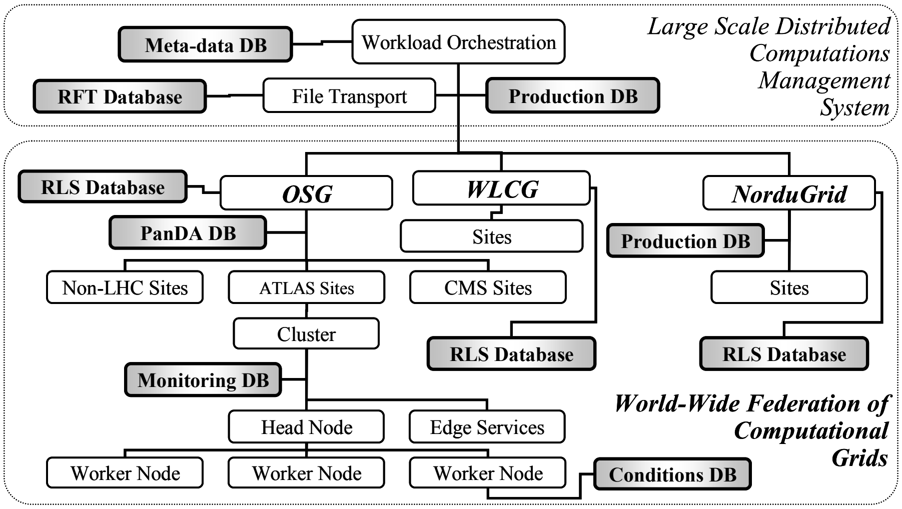

Figure 1: A hyperinfrastructure of distributed database services crosscutting through computational grids.

the clear-text passwords 'hidden' in the deployed configurations, limited control over the secure transport channel, cryptographic handshake for every XML message that prevents efficient data transfers.

Embedded Security

In an alternative approach the grid authorization is embedded in database server, which is possible with the open-source databases. The embedded security approach is listed among the top ten innovations in security by the panel of experts convened by Battelle [8]. In addition to the elimination of the clear-text passwords through the deployment of the same grid security model cross-cutting all data flow channels the inefficient data transfer bottlenecks are eliminated.

End-to-End Secure Transport

To overcome database access inefficiencies inherent in a traditional middleware approach the DASH project implements secure authorization on the transport level. Pushing the grid authorization into the database engine eliminates the middleware message-level security layer and delivers transport-level efficiency of SSL/TLS protocols for grid applications. The database architecture with embedded grid authorization provides a foundation for secure end-to-end data processing solutions for the experiments.

DASH TECHNOLOGIES

The DASH proof-of-concept prototype provides Globus grid proxy certificate authorization technologies for MySQL database [9] access control. Direct access to database servers unleashes a broad range of MySQL server functionalities for HENP data processing applications: binary data transport, XA transactions, etc.

Aspect-Oriented Software Development

To avoid a brittle, monolithic system DASH project uses an Aspect-Oriented Software Development (AOSD) approach [10, 11]. By localizing Globus security concerns in a software aspect, DASH achieves a clean separation of Globus Grid Security Infrastructure dependencies from the MySQL server code. During the database server build, the open-source AspectC++ tool [12] automatically generates the transport-level code to support a grid security infrastructure (Figure 2). Figure 3 shows that Globus security concerns crosscut the transport-level action flow in many places, which is a typical use case where AOSD technology adds considerable value by managing complexity in an automated way.

Enhanced Security

Use of the Aspect-Oriented Software Development approach enabled straightforward implementation of the broad range of security enhancements in the DASH server: strict checking for the expiration time of the proxy certificate, verification of the hostname presented in the server certificate checking (required to reject impersonation), etc. Future DASH server enhancements will include support for the grid job certificate in authorization for access to the database resources.

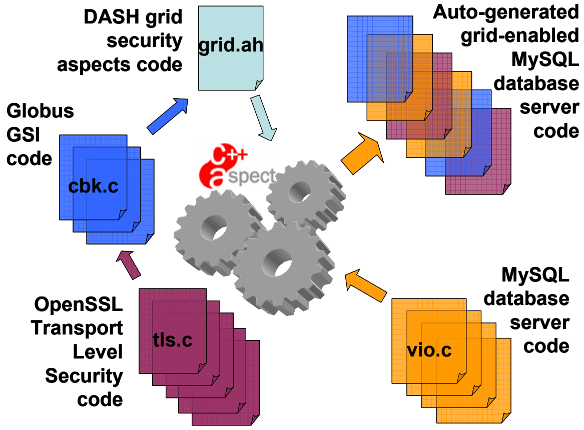

Figure 2: Automatic Code Generation.

Packaging Challenge

The grid-enabled mysql-gsi server executable requires relatively large Globus GSI libraries. To simplify server maintenance and installation it is advantageous to build the mysql-gsi server executable that contains those libraries. Initial feedback from the DASH beta-testers suggested that to avoid explicit Globus GSI libraries dependencies the preferred mysql-gsi server distribution would be the static build. However tests of pprototype servers built with DASH technology in ANL, BNL, CERN and University of Geneva showed that static builds only works best on the platforms (Linux distributions) very close to those that of the build machine. Due to sensitivities to the minor variations in the glibc library version we had to build the static distribution on exactly the same Linux distribution where the server was going to be tested by the beta-testers. We are now addressing that issue by developing the dynamic build that will have the static globus gsi and openssl libraries built-in.

Additional Benefits

Direct access to database servers unleashes a broad range of vendor-specific server capabilities for data processing applications. In addition, grid proxy certificate technology opens technical opportunities to enable fine-grained delegation of rights for access control (attribute certificates). Grid-enabled relational database server technology has the potential for application beyond the domain of high energy physics, and is of interest to bioinformatics and other data-intensive sciences.

SCALABILITY CHALLENGE

Large-scale world-wide distributed simulations performed by the ATLAS Collaboration in 2005 show steady progress in grid computing capabilities (see Figure 4). The high level of sharing of computational resources achieved on the grids result in increased

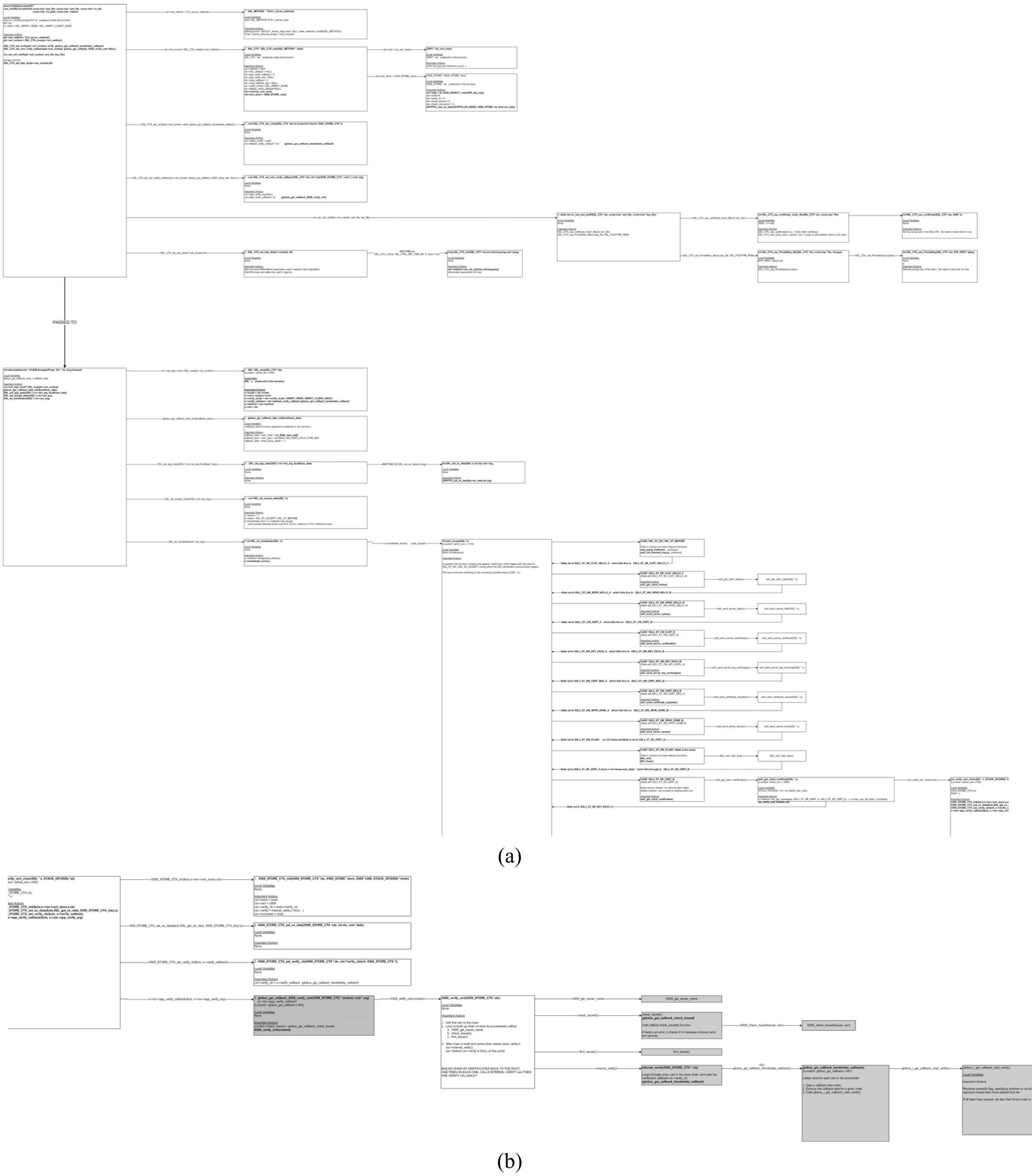

Figure 3: Left (a) and right (b) portions of the server-side grid-enabled SSL/TLS handshake action-flow diagram. (This diagram is best viewed by setting the highest level of magnification.)

variations in demand for database services, because of the chaotic nature of shared resource availability [5]. For efficient production on the grid the database services capacities deployed should be adequate for peak demand from distributed applications. Thus, the deployment of the database services capacities via the traditional ('static') technologies [13] requires over-provisioning of resources.

To provide on-demand database services capability for the Open Science Grid (OSG) [14], the Edge Services Framework activity [15] builds the DASH mysql-gsi database server into the Xen virtual machine image [16], which could be dynamically deployed via Globus Virtual Workspaces [17].

Through our OSG Edge Services Framework (ESF) activity collaboration to achieve the ESF proof-of-concept milestone the first ESF virtual machine was deployed by the CMS experiment at the US CMS Tier 2 facility with the first ESF service on that virtual machine was deployed

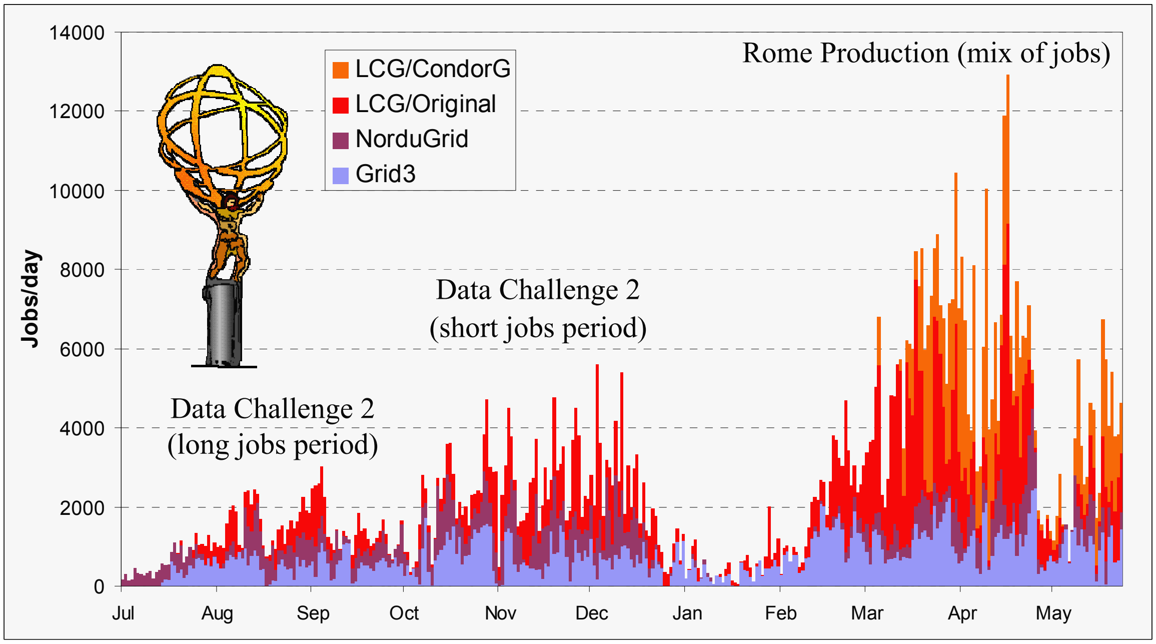

Figure 4: Growth in daily production rates of ATLAS computational tasks running on a world-wide grid federation.

by the ATLAS experiment: the grid-enabled MySQL database built by the DASH project. The proof-of-concept grid job submitted to the OSG (US CMS) production site from the submission host at the US ATLAS Tier 1 facility used it's US ATLAS production grid proxy certificate to get authorization for access to the DASH server data [18].

ACKNOWLEDGEMENTS

We thank all our collaborators for their interest and support, in particular, our Open Science Grid collaborators in the Edge Services Framework activity chaired by Kate Keahey and Abhishek Singh Rana and DASH beta-testers from the ATLAS experiment Frederik Orellana, Yuri Smirnov and Jason Smith. We also wish to thank David Axmark, Tzilla Elrad, Ian Foster and Ruth Pordes for discussions and interest in this work.

Argonne National Laboratory's work was supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under contract W-31-109-Eng-38. PIOCON Technologies, Inc work was supported by U.S. DOE Small Business Innovative Research Program under DOE Grant No. DE-FG02-05ER84369.

REFERENCES

[1] http://www.gridforum.org/6_DATA/dais.htm

[2] http://www.ogsadai.org.uk

[3] N. Hong. "OGSA-DAI Status Summary." 3rd OGSA-DAI UG Meeting, Edinburgh, U.K., 2005. http://www.ogsadai.org.uk/docs/UG3/Jun05_Users_Group.ppt

[4] M. Branco, D. Malon, A. Vaniachine. "Secure Grid Data Management Technologies in ATLAS." in Proc. of the 2004 Conference for Computing in High Energy and Nuclear Physics (CHEP04), Interlaken, Switzerland, *CERN Yellow Report 2005-002*, p.864 (2005).

[5] A. Vaniachine, D. Malon, M. Vranicar. "Advanced Technologies for Distributed Database Services Hyperinfrastructure." *International Journal of Modern Physics* **A 20** (16): 3877, 2005.

[6] K Bhatia, et al. "Engineering an End-to-End GSI-based Security Infrastructure." *Technical Report SDSC-TR-2005-1*.

[7] http://www.globus.org

[8] http://www.battelle.org/forecasts/defense.stm

[9] http://www.mysql.org

[10] T. Elrad, R. E. Filman, A. Bader. "Aspect-Oriented Programming." *Communications of the ACM*. **44** (10):29, 2001.

[11] C. Tull, P. Calafiura; "Aspect-Oriented Extensions to HEP Frameworks" in Proc. of the 2004 Conference for Computing in High Energy and Nuclear Physics (CHEP04), Interlaken, Switzerland, *CERN Yellow Report 2005-002*, p.621 (2005).

[12] http://www.aspectc.org

[13] http://lcg3d.cern.ch

[14] http://www.opensciencegrid.org

[15] A.S. Rana et al. "An Edge Services Framework (ESF) for EGEE, LCG, and OSG." CHEP06: http://indico.cern.ch/contributionDisplay.py?contribId=214&amp;sessionId=7&amp;confId=048

[16] http://www.cl.cam.ac.uk/Research/SRG/netos/xen

[17] http://workspace.globus.org

[18] A. Vaniachine. "DASH: Database Access for Secure Hyperinfrastructure." Globus OSG SC|05 flyer http://osg-docdb.opensciencegrid.org/cgi-bin/ShowDocument?docid=307