



The GridPP website, 2000-2007

1. The GridPP Project

The GridPP Project began in 2000 as "The UK HEP Grid", with a collaboration of particle physicists and computer scientists at a dozen sites around the UK, and the intention to build a computing and data grid connecting resources and scientists. The researchers and system managers involved had worked together before as part of experimental particle physics collaborations, but new collaborative structures and tools were required to link staff at sites together as part of the grid infrastructure, rather than via experiment-based collaborations. As part of the Testbed Support activity, the University of Manchester group provided a website for the UK HEP Grid project, which became www.gridpp.ac.uk

2. Static HTML pages

The first version of the website was produced with static HTML pages and using an unmodified Apache httpd server at Manchester. Once the site reached several pages, a very rudimentary system of repeated headers and footers was used by running the page sources through the C pre-processor. As the UK HEP Grid project grew, it became clear that members at other sites would need to be able to modify pages too. One option was simply to create more Unix user accounts on the webserver machine and hand out passwords or manage ssh keys. However, this was expected to scale poorly as the project grew to tens of active members, and instead GSI-SSH was used, with users mapped to a single web user account via a locally managed grid-mapfile. This removed the need to manage "secrets" (ie passwords or keys) on the web server machines. However, it in turn raised problems, in that the identity of the user making each change was lost and users had write access to the whole website. For these reasons, the GridSite system was developed to provide finer grained access, and also remove the need to work at the command line.

3. GridSite 0.1: Certificates and ACLs

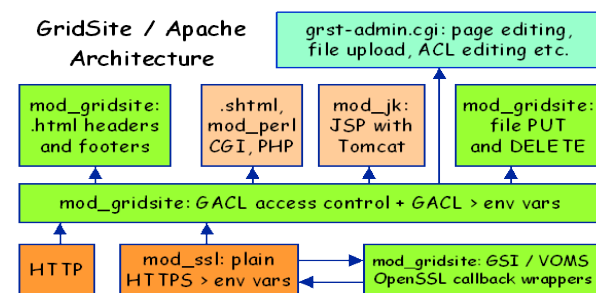
By 2001, the UK HEP Certification Authority had been set up and was issuing X.509 user certificates to all members of the project, which could be used for grid job submission. Since these certificates were also compatible with standard web browsers across Unix, Windows and Mac platforms, we were able to design a web-based management system using the same credentials. "GridSite" was implemented as a CGI filter, which was run by the Apache web server for each request for an HTML web page or graphics file. Access control was by hidden files each directory, listing the X.509 DNs of users with read or write access. Users with write access could edit pages via web forms generated by the GridSite CGI executable.

4. GridSite 1.0 and GACL

Although the CGI form of GridSite worked well, it imposed an overhead in that the CGI process had to be restarted for each page or image request, and it could not be used to control access to other dynamic content, eg PHP dynamic web pages. The Access Control List format was also limited and did not accommodate references to groups of users, which led to large and difficult to manage ACLs. The system was again redesigned and reimplemented in 2003, using the XML-based Grid Access Control Language (GACL) which Manchester had developed and which was already being used by components of the European Data Grid middleware. GACL allowed references to X.509 DNs, groups defined as lists of DNs and VOMS attribute names.

5. mod_gridsite

The other half of 2003's redesign saw the monolithic GridSite CGI replaced with an Apache module, `mod_gridsite`, and a reusable toolkit of grid security utilities and parsers, `libgridsite`. Since `mod_gridsite` functions are executed within the Apache process, no CGI overhead is added and GridSite can be configured with additional `httpd.conf` directives. `mod_gridsite` also dynamically modifies the OpenSSL callbacks used to check X.509 certificate chains, to accept GSI Proxy Certificates and allow GridSite-secured web services to be accessed by running grid jobs and grid services.



6. Site redesign

During 2004, the layout of the GridPP website was redesigned to make it easier to navigate and to improve the quality of the graphics, logos etc. Most of this work was done by Richard Hirst, Steve Lloyd and Sarah Pearce of Queen Mary, London, and it was recognised at the UK e-Science All Hands Meeting with the Gold Prize for the Best Project Website.



6. Wiki, Subversion & Web Services

During 2005 and 2006, additional dynamic content features were added to the site using PHP, including RSS aggregators, project news listings and a modified version of the MediaWiki software used by Wikipedia. The modular architecture introduced with `mod_gridsite` allows this free mixing of static and dynamic content, whilst retaining the same X.509-based security structure, even for third-party, non-grid systems like MediaWiki. The project wiki has since become one of the most active areas of the site, largely due to the informal style of wiki articles which can be gradually improved as new components of the grid are better understood. We have also integrated an unmodified version of the Subversion software version control system, which is being used to store LCG/EGEE site configuration information, with access controlled by X.509 user certificates and managed by `mod_gridsite`. This last development is an example of the bridging of the web world with the grid, where GridSite is also used as a basis for some LCG/EGEE web services.