

GridPP

Grid Security in a production environment: four years of running www.gridpp.ac.uk

GridSite has been developed by the GridPP collaboration to bridge the gap between the Web and Grid. GridSite adds support for several Grid security protocols to the Apache webserver platform, and so allows users to use Grid security credentials to access or modify Web sites. It also lets software authors reuse familiar Web tools when building Grid services.

The architecture of GridSite has evolved during the past 3 years, influenced by operational experience with production systems, and the project has led to new developments, such as the Grid Access Control Language (GACL). Finally, GridSite has been made to interoperate with other deployed security systems, both as producers and consumers of GridSite's authorization information.

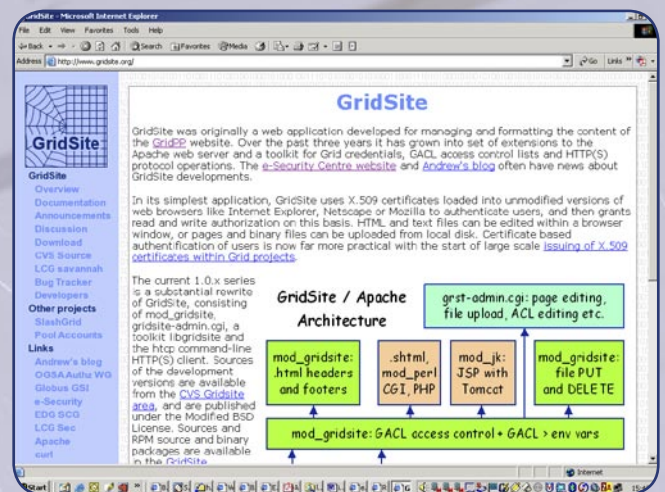
• Supporting a complex collaboration

The GridPP collaboration involves a community of about 100 hundred particle physicists, computer scientists and site administrators. Most of the members are located at one of twenty particle physics groups at UK universities and CCLRC, or at international laboratories, and are mostly participants in running or planned experiments. GridPP has developed Grid middleware as part of the EU DataGrid project and is involved in the EGEE and LHC Computing Grid deployment projects.

These various affiliations create a need for overlapping domains of authorization, which correspond to software development subgroups, home institute or experiment associations. Furthermore, different areas of the website have to be accessible to members in one or more of these domains, with write access held by fewer members than read access. Given our involvement in developing Grid middleware itself, it was natural to fulfil these Web site requirements using Grid technologies.

• Authentication and Virtual Organization (VO) structure

Most collaboratively maintained websites use some form of username/password authentication to identify users. However, for the GridPP website, all of the user community were being issued with X.509 user certificates as part of the project's Grid deployment, which offered an alternative authentication method.



As described above, GridPP involves many working groups of several different classes - for example, middleware development groups, applications communities and deployment teams. An individual may be a member of several groups, and access control to areas of the website needs to reflect this structure.

However, every member has a unique X.509 certificate Distinguished Name ("DN"), which can be used both for Web authentication and accessing our deployed Grid infrastructure. Consequently, we have used the simplest way to represent a group, by simply listing its members' DNs. These "DN Lists" are both stored internally and published as plain text files via HTTPS. They are also uniquely identified by their URI (which is an HTTPS URL.)

This allows the VO group information to be used by other services, as a basis for their own access control decisions; and DN List groups have also been made available via other protocols.

• GACL access policies

The groups defined by the DN Lists provide only half of the access control system. Some way of associating groups or individuals with rights to use or modify objects in the system is also required. The access policies are defined in terms of DNs of individuals or the URLs of DN Lists. We also define a set of permissions which matching users are then given - currently, Read, List, Write, Execute and Admin.

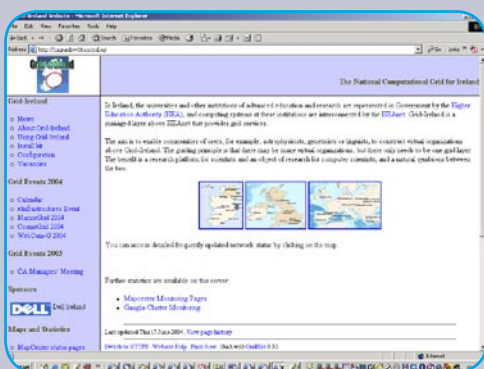
These policies are represented in an XML-based form, which we call the Grid Access Control Language (GACL.) For convenience, the policy file governing access to a directory on the web server is stored in a file named .gac1 in that directory. If none is present, the .gac1 file in the parent (or higher directories) is used.



The GridPP website - built with GridSite

• Web management interface

One of the requirements of the original GridSite design was to give users of any platform an intuitive interface to manage the web pages and files stored on the server. This has been provided by a management interface implemented as HTML forms, which allows authenticated users to create pages, edit text or HTML files, and to upload individual or collections of files - all without leaving their standard web browser. All of these operations are controlled by the relevant GACL policies, which can themselves be created or edited. An editor is also provided for the DN Lists that enforces their simple format.



The Grid Ireland website - built with GridSite

• Operational experience

During 2001 to 2003, the GridPP web server was operated using GridSite. This proved to be very successful, and many members took responsibility for managing their own areas of the site. We were thus able to remove most of the content management responsibilities from the "webmaster" team, who were able to concentrate on keeping the system operational, supporting users' queries, and developing extensions to the software.

However, during this period we identified several limitations of the architecture. In particular, as the service became more popular, the performance implications of using the Common Gateway Interface (CGI) to communicate with the webserver became a problem. The CGI interface requires that a new Unix process is spawned for each page or file request. In addition, GridSite did not take advantage of HTTP connection reuse, causing particular performance problems with HTTPS. Our CGI approach also prevented us from easily using other web technologies, such as PHP or other CGI programs.



Andrew McNab, who developed GridSite, receives a certificate from the CERN Director General for his contribution to Grid computing

• Redesigning the system

During 2003 we redesigned GridSite to operate as a shared object module within the Apache 2.0 web server, rather than as a standalone CGI program. This effectively made GridSite a fully fledged part of the Apache environment, and gives us several operational advantages, such as: reducing process creation overhead; using Apache's configuration and logging infrastructure; support for dynamic content systems such as PHP, mod_perl, JSP etc.; and connection reuse which improved HTTPS performance. This modularised system is now in production use on several sites, including the GridPP and LCG Grid Operations Centre sites.

Other modifications during the project have included:

- Redesigning our support for Globus GSI proxies to the Apache 1.3 SSL handler, to remove strict dependency on Apache and OpenSSL versions. So, new rebuilds of GridSite are not necessary each time an Apache or OpenSSL security vulnerability is found and corrected.
- Adding support for EU DataGrid project Virtual Organisation information, either using the LDAP system or the more recent VOMS system.
- Providing a Virtual Organisation service for the BaBarGrid project, for the BaBar experiment at SLAC.

• Further work

Although the GridSite system's use of X.509 user certificates for authentication is practical for its current user community, we recognise the need to support usernames and passwords as an alternative. In particular, we are investigating ways to add support for Shibboleth as a way of providing this. We are also adding support for the XACML policy language as a standards-based alternative to GACL.

Andrew.McNab@man.ac.uk

www.gridpp.ac.uk/gridsite