A Scalable PKI for a National Grid Service

Jens Jensen, David Spence, Matthew Viljoen
Rutherford Appleton Laboratory
The National Grid Service for the United Kingdom

February, 2007

Abstract

In this paper we describe work to expand the PKI for the UK National Grid Service (NGS), to integrate it with site authentication and improve usability. This work is complementary to the UK Shibboleth deployment. As the NGS grows to support wider and larger scientific communities, we investigate how we can improve usability by tying in Virtual Organisation management into the PKI framework.

1 Introduction

1.1 General Introduction

The UK National Grid Service (NGS) [13] runs Globus-based Grid middleware which depends on X.509 certificates for user authentication (Globus Security Infrastructure, GSI [31]). The UK e-Science Certification Authority [24] provides medium assurance [6] certificates for Grid users and e-Science projects in the UK, including the NGS. This Certification Authority (CA) must provide medium assurance certificates (section 2.4.1) because it is approved internationally (section 1.3) to identify users and hosts in the UK to international Grid collaborations. Conversely, although it primarily serves the UK, scientists using the NGS have collaborators across the world, and the NGS trusts certificates from other internationally approved CAs in order to facilitate these collaborations. In particular, interoperability between NGS and TeraGrid [22] is considered important.

Medium assurance, among other things, implies that users will have shown photo id to a Registration Authority (RA) operator, and that certificates have a maximal lifetime of 395 days (one year, plus 30 days within which users should request rekeying). Furthermore, many relying parties that use these credentials insist on having "meaningful" commonNames (i.e., bearing a reasonable resemblance to the person's authenticated identity). For many purposes, this is too strong an authentication and doesn't scale well to a large number of users ($\gg 10^4$ certificates), so we are deploying a scalable hierarchy primarily for NGS to expand the user base, with features to ease the account management. It is deployed alongside the UK Shibboleth federation [25], and is complementary to it, but is still entirely independent of it. We explain how they will interoperate.

Deploying a PKI for academic institutions is of course not a new idea. The innovation presented in this paper lies mainly in deploying it specifically for a national Grid, so we can tie in attribute and Grid account management, and in deploying the PKI alongside, and interoperating with, the Shibboleth federation. We will also briefly discuss other usability issues.

1.2 Related Work in this Area

From a high-level view, the architecture of the work presented in this paper is similar to that of SWITCHaai [21], partly because this work aims to solve some of the same problems. The principal difference is that SWITCHaai is entirely Shibboleth based.

From a more practical point of view, this work is similar to USHER, the US Higher Education Root[26]. We will look closer at this similarity in section 2.6.

The work presented here depends on technology developed in other similar projects, namely, MyProxy [14], SHEBANGS [17], and ShibGrid [18], as well as other related single sign-on work [7, 9].

1.3 International Accreditation

Although not fundamental to this paper, it will be helpful to briefly mention as explanatory background information that international Grid CAs are accredited by so-called Policy Management Authorities (PMAs). The Grid world is currently covered by three such, who together form the International

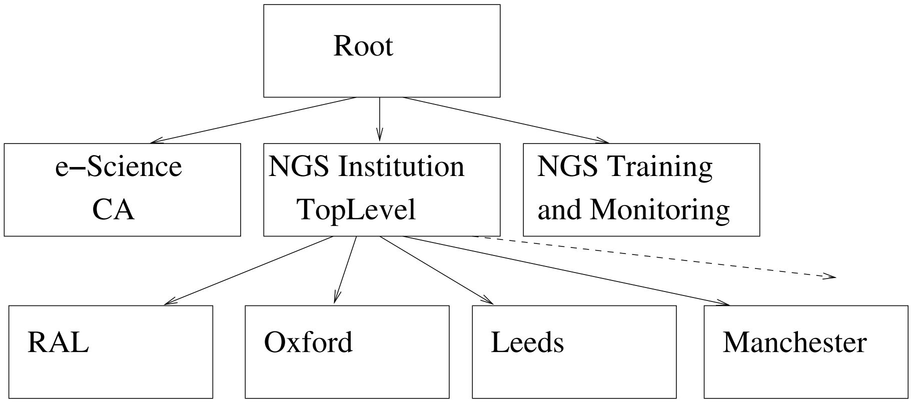

Figure 1: The UK e-Science Hierarchy

Grid Trust Federation, IGTF. Accredited CAs are then trusted by national and multinational Relying Parties (RPs), including the NGS, but the RPs are of course free to trust unaccredited CAs—indeed, this is often necessary to enable certain communities to access the Grids. Loosely speaking, it is the PMAs who impose upon their members that they operate to what we have called "medium assurance" in this paper, as a condition for accreditation.

2 Deployment

This paper discusses deployment of a hierarchy of *credential conversion* CAs, where each institution effectively runs its own CA which converts the institutions' site authentication to a short-lived X.509 credential (in the Grid CA context, such a CA is often referred to as a SLCS, a Short-Lived Credentials Service [8] (pronounced "slicks")). "Short-lived" usually means 12 hours, but is allowed to be "anything up to 10^6 seconds." [8]. One core difference is that although a SLCS acts as a CA, it does not need to issue CRLs.

This deployment brings us a wide user base, where essentially anyone from any such institution can get a certificate, but we lose the right to manage and control the user data that originally identified the user.

Indeed, our principal challenge was to widen the user base for the NGS, enabling also students and visiting scientists to obtain accounts via the authentication framework.

A secondary challenge was to improve the usability of the PKI. Usability is often seen as an obstacle to widespread Grid use; in some scientific communities users are unable or unwilling to learn basic PKI. The security required by a medium assurance CA will prevent these communities from engaging with Grid work.

2.1 The UK PKI Hierarchy

In July 2006 the UK e-Science CA deployed a new PKI based on a hierarchical model. This hierarchy was introduced at the same time as rolling over the certificate of the UK e-Science medium assurance CA. The new certificate for this CA is now subordinate to a root. Other CAs offering different services or providing other assurance levels have been deployed as part of this hierarchy. An example is a low-assurance training CA used for people new to the NGS.

Figure 1 on this page shows the UK hierarchy: a root CA ties together the internationally approved (medium assurance) CA, as well as the specialised hierarchy for NGS: the institutional credential conversion hierarchy and the training and monitoring CA.

The institutional credential conversion CAs, the SLCS, are anchored by a common SLCS trust anchor which itself is subordinate to the root. This anchor CA represents a single point of trust for any relying party (RP) that wishes to trust *all* credential conversion services in the UK, at least in principle. Its policy also ensures that a minimum set of requirements or level of assurance can be enforced between all credential conversion services, because it can impose specific policy and practices constraints upon its subordinates. These, of course, should not be set too high, or we would lose participant institutions.

Since a SLCS needs to be an online service in order to function, it is imperative that the private key be adequately protected to prevent it from being compromised. The SLCS anchor's policy states that subordinate certificates can only be issued when they are requested by a security device conforming to the FIPS 140-2 level 2 standard and the private key must not be exportable in any unencrypted form from that device (such devices can be obtained as USB tokens and are relatively inexpensive). Although no recovery is possible in case of hardware failure, the extra assurance won by so securing the key far outweighs the risks. Indeed, should the key be lost, a new certificate can quickly be issued by the parent CA.

Security concerns are also addressed in the SLCS anchor's policy by imposing the requirement upon the SLCSes that communication channels between them and their local authentication service, as well as the NGS authorisation services, are secure. In addition to this, each SLCS is required to log all credential conversion so that NGS traceability and accountability requirements of access mechanisms are satisfied.

2.2 Architecture

Figure 2 on the next page gives an overview of the architecture described so far and how this fits in to the wider NGS and site infrastructures.

As we are seeking to lower the barriers for users to user the NGS, we assume that the user will be using some easy-to-user *User Interface* software to manage their access to the Grid, to which we can make minor changes to support the SSO infrastructure (although we do not preclude the use of common command line tools).

The key component of the infrastructure is the *Credential Translation Service* or SLCS. There is one of these per site and its main function is to validate the user's site identity by calling out to the site's authentication infrastructure and then generates a Grid credential (a short-lived X.509 certificate) for the user. In this respect, it is similar to the Shibboleth IdP—see section 2.7.

The Credential Translation Service can also call out to a VOMS server to obtain a VOMS attribute certificate for the user. This would normally be to the NGS VOMS server, but others could be imagined, such as a site VOMS server.

2.2.1 Where Are You From?

Most of the work described in this paper applies to local clients, running within the site. However, we have also thought about central interfaces (shared between sites).

Such interfaces could also run on the user's machine (e.g. an applet), but would typically be on an NGS server (e.g. the NGS portal), or a third party server (e.g. a project portal). Unlike local clients which "know" which credential conversion service to contact, a discovery mechanism is needed for non-local services, and, worse yet, potentially a different mechanisms for each one. Any such service would of course play a role of the WAYF in a Shibboleth Federation where each client selects his or her home institution and is redirected, but for this project we wanted to simplify the selection, so the client does not even need to select a home institution.

The easiest way to accomplish this is to configure a lookup mechanism which notices which address the client is coming from, and uses it to contact the right credential conversion server. We do indeed lose the ability to support "roaming" clients with such a simple scheme, but gain the simplification of not asking the user for redirection. For central portals, the conversion service must also be reachable through the site firewall which complicates site deployment slightly.

On the server side, a trusted repository of the SLCS sub-CA certificates will have to be provided, along with the repository provided by the PMAs for traditional CAs.

2.3 Implementation

While there is no requirement to do so, the "obvious" way to implement a site SLCS is via MyProxy [14].

We discuss the MyProxy solution further in this section, but it is also worth mentioning Microsoft's Windows Server 2003 which also contains features to run a CA, but will not, as far as the authors are aware, be able to contact an external VO attribute server to add attributes to the certificates.

By using MyProxy we can bridge many of the site authentication infrastructures in use to the GSI/PKI world. MyProxy can be configured to provide certificates generated by an internal or external CA. We can also support advanced users who have long-term certificates by enabling them to upload proxies generated from their certificates (they are not permitted to upload the certificates and private keys themselves, by the CA's policy). This was the approach taken in the ShibGrid project [18]; see also Shibboleth discussion in section 2.7. For users who do not have long-term certificates, the service generates short term certificates and keys.

MyProxy can also support SASL [12] (allowing Kerberos [19] authentication), PAM (allowing LDAP [29], RADIUS[16] and many other authentication systems) and PubCookie.

Although it is simple to use the MyProxy command-line tools to leverage this functionality, this could potentially present users with quite a barrier to overcome. Many of the target audience traditionally balk at security in general, certificates in particular, and anything that cannot be clicked with a mouse. Therefore, we aim to simplify the process as much as possible, and hide the certificate/proxy process. For our own site's authentication infrastructure, Microsoft Active Directory (which happily, for the purposes of this work, is

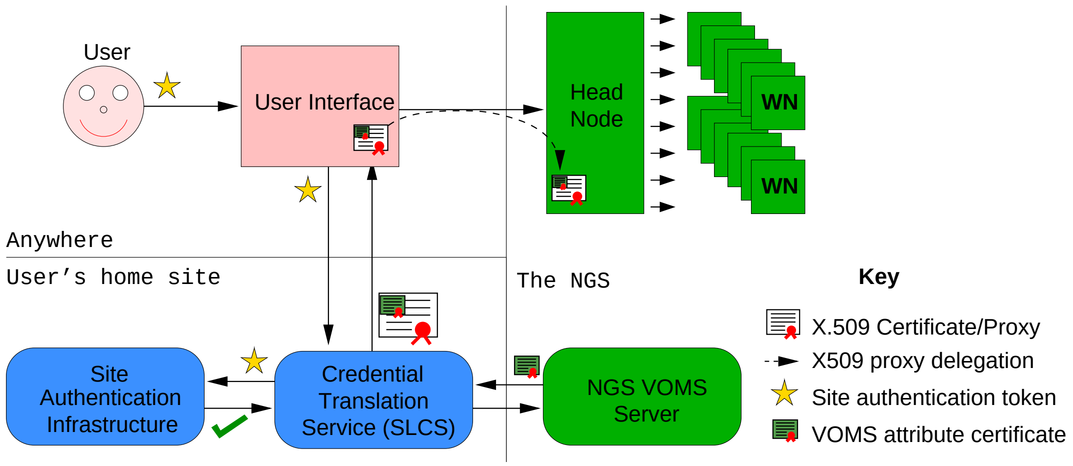

Figure 2: The architecture for credential conversion in the NGS

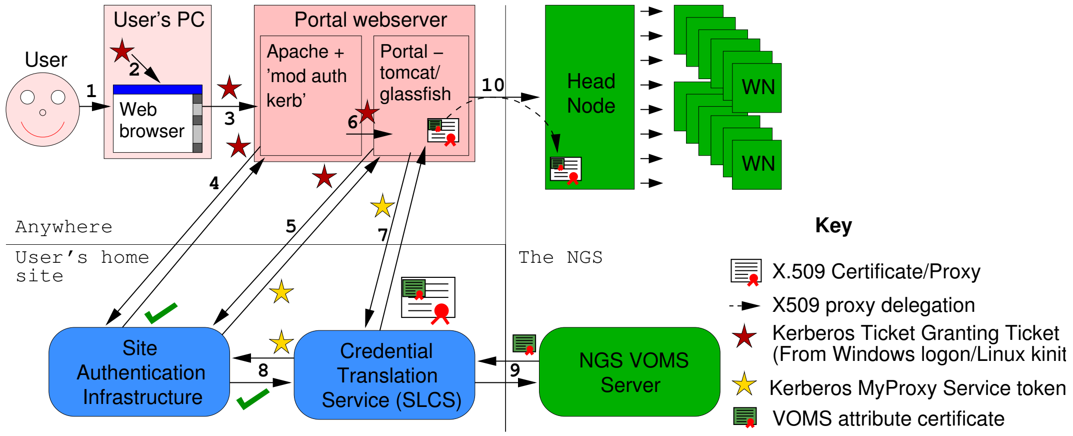
User's PC
User
1
2
Web browser
3
Portal webserver
Apache + 'mod auth kerb'
Portal – tomcat/ glassfish
6
10
Head Node
WN
WN
4
5
7
Anywhere
User's home site
The NGS
Key
X.509 Certificate/Proxy
X509 proxy delegation
Kerberos Ticket Granting Ticket (From Windows logon/Linux kinit
Kerberos MyProxy Service token
VOMS attribute certificate
Site Authentication Infrastructure
8
Credential Translation Service (SLCS)
9
NGS VOMS Server

Figure 3: Authentication process from a portal.

equivalent to Kerberos 5), we have integrated support for Kerberos authentication to MyProxy into two easy to use Grid access methods:

- **Portal access** (Figure 3.) Normally, users who wish to use a Grid portal would have to upload a proxy of their Grid certificate to a MyProxy server before logging on to the portal. At the portal they would then have to provide the hostname of the MyProxy server along with the username and passphrase they used to store their proxy. The portal would then contact that MyProxy server with the given detail to obtain a certificate. In our set-up they instead simply visit the portal which picks up their Kerberos Ticket Granting Ticket (TGT) and it uses this to contact the Kerberos-enabled MyProxy-with-CA, which generates a certificate for the user. The portal has to be trusted for delegation by the Kerberos Key Distribution Centre (KDC) for the user to delegate its TGT to the portal.

- **GSISSH Terminal** (Figure 4 on the next page.) We have also developed additions to a Java-based Grid Security Infrastructure enabled secure shell (GSI-SSH) terminal, GSI-SSHTerm [20], which runs in a user's web-browser or as a standalone Java application and provides terminal access to Grid resources. These additions automatically call out to the Kerberos-enabled MyProxy with CA to attempt a conversion of the user's local Kerberos credential to a Grid credential, when a user tries to log on to a remote resource.

Both of these methods rely on a specially patched version of the Java Commodity Grid (CoG) Kit [28] which allows SASL+Kerberos authentication to MyProxy servers. In choosing to start with Kerberos infrastructure support we hope to support

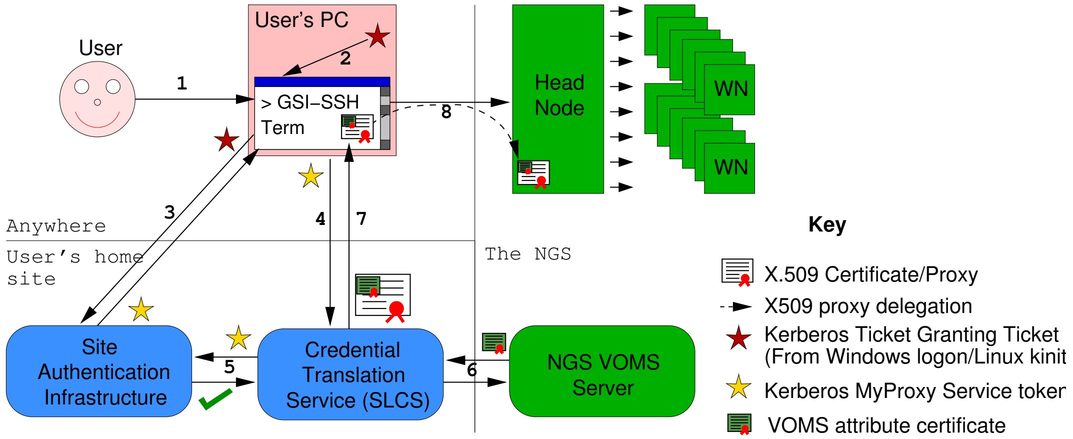

Figure 4: Authentication process from the GSISSH Terminal.

many institutions as this also encompasses systems built on Active Directory. Both these methods fall-back to username/password authentication (but using site passwords) to the MyProxy with CA if Kerberos tokens are absent.

In the case of Portal access we have had experience with both the Tomcat and Glassfish servlet engines. In both cases these would run behind an Apache instance running on the same machine and utilising mod_auth_kerb.

On the server side each institution must install their own dedicated MyProxy server which is configured to work correctly with their site authentication infrastructure. Although the configuration will be different depending on the site authentication infrastructure there should be enough commonality within the same technology to make the provision of example configurations useful.

The policy for the MyProxy's CA certificate means that it must be stored on a hardware token, for example a USB key-token. We have already undertaken the necessary changes to the MyProxy server code to allow it to connect to any hardware token supported through the openssl "engine" mechanism[1]. This set includes nCipher products, CryptoSwift and key-tokens supporting PKCS#11 and PKCS#15 (through components from the OpenSC project[2]). We selected to use the Aladdin eToken, through the PKCS#15/PKCS#11 interface. We have of course contributed these changes to MyProxy.

Another issue is which Distinguished Name (DN) to give to users. The policy for the CA certificate requires that the MyProxy CA will sign its certificate in a particular namespace. Further to this we only require that each user's DN is unique to that user, traceable to that user and consistent over time for that user. For some institutions it may simply be that they append `/CN=<userID>` to the CA's namespace, for others they will call out via LDAP to obtain more user information to generate a DN that is closer to those currently in use the UK e-Science CA. Within our own Active Directory domain, for example, we chose to use `/UID=<userID>/CN=<firstName> <lastName>`.

2.4 Scalability, Policy, and Assurance

2.4.1 Levels of Assurance

The Level of Assurance (LoA) of a CA or of a certificate issued by that CA is an indication of the extent to which an entity has been identified as the owner of a credential issued by that CA. Whereas the US government has proposed using four LoAs [15], Grid CAs have traditionally employed two LoAs. These map approximately to the US governments LoA levels 1 and 2 (*ibid*, sections 2.1, 2.3) with only level 2 being accepted internationally, level 1 is usually for internal use. This government-proposed mapping policy was expanded to recommended practice by the US National Institute of Standards and Technology in NIST-800-63 [5], which may well be indirectly behind the Grid requirements for "medium assurance" described below.

1. **Low LoA** Little or no identity verification has taken place during the issuance of the certificate. Such CAs are typically used to issue training or test certificates to access resources.
2. **Medium LoA** The certificate applicant is required to meet a representative of the CA, the Registration Authority (RA) in person during

[1] http://www.openssl.org
[2] http://www.opensc-project.org/

the certificate application process. Furthermore, the applicant is required to present an original photo id document as proof of identity, while the RA is required to retain a copy of the document for a specified period of time (at least as long as the lifetime of the certificate). The RA may stipulate the accepted forms of documents, typically an id card of the institution where the RA is situated, a driving license or passport. The Classic Authentication Profile of the IGTF requires all Grid CAs to be at least Medium LoA to be accredited to the IGTF.

3. **High LoA** If a Medium LoA CA is considered inadequate for particularly high risk relying parties then additional requirements may be stipulated covering identity verification or how certificates may be issued. For example, resources protecting particularly high-risk data may choose only to accept certificates from a CA which includes biometric verification of applicants and who only issues certificates protected by secure cryptographic devices which can prevent the private key of the certificate being exported.

It is worth noting that the Grid's medium assurance falls short of meeting NIST's Level 3. For example, most CAs make "best efforts" to revoke (issue a new CRL) within "one working day" (of the revocation request being approved) rather than 24 hours. Globus proxies and delegations permit credentials to be passed around and may live longer than 24 hours ("freshly issued", [5] 8.2.3.1.) "Picture id" is required for all, but many CAs accept photo id issued with a site account in addition to driver's licence or passport. Indeed, the emerging MICS profile (work in progress, mainly led by TAGPMA) enables a CA to skip the RA step and issue certificates directly for site accounts. On the other hand, no Grid CA except the project catch-all CAs accepts remote verification. A full comparison is interesting but beyond the scope of this paper.

HEBCA has started looking at the Grid assurance levels, to evaluate the feasibility of bridging the Grid PKI to the US Higher Education. Although in progress, we describe this work further in section 2.6.

2.4.2 Scalability

Consider a medium assurance CA issuing certificates to 10^6 users (the estimated number of higher and further education users in the UK is of this order of magnitude). Even assuming that no new users are added, the Registration Authorities (RAs) must process 10^6 renewal requests per year, which translates to over 2500 per day (if every day is a working day). With 250 operators, that is 10 per day per operator, assuming it is distributed evenly (even though the RA need not verify the user's identity, they must still perform a simple check that the user is still associated with the project or organisation).

Moreover, a new requirement is being added by the internal Grid CA community, that users must reauthenticate with their RA every N years, where N depends on how the private key is stored, as well as other factors; $N = 5$ is typical. That means on average 200000 reauthentications every year.

Even if people didn't have to reauthenticate, a medium level Grid CA still requires them to generate their own key pairs, and on relatively secure systems, e.g. their own desktop machine, not a shared service. Thus, people have to manage and convert their own keys and certificates, and the support load required to support 10^6 PKI-novice users managing their own keys would be beyond the capacity of NGS support.

One of the aims of this project is to *be scalable to* a million users distributed over 10^3 institutions, increasing the usability at the cost of the assurance that the CA can guarantee.

Another aspect of a medium assurance CA, as mentioned above, is that people have to show photo id during the identification process. People may argue that site authentication is typically at least as good as that: your employer probably saw your passport, birth certificate, etc. However, there are two problems with this approach: firstly, the CA has no access to this information, so cannot, for example, rely on it to ensure that the distinguished name (DN) in the certificate is never reallocated to another person. Secondly, the CA cannot *guarantee* that this process has taken place: many sites have visitors or contractors who are also in the site database, and may not be managed as strictly as those in the payroll database.

The other aspect of scalability is that of the number of CAs. In this model, we have one CA per institution. However, current Grid middleware must have *each* CA installed, whether it is an end entity issuing CA or not. This is fine for the "standard model" with one per country [4], but not with $N \gg 1$ per country for a global collaboration. It should be feasible for NGS to trust M countries and N institutes, though—the number of certificates is $M + N$, rather than MN. A similar situation is found in TeraGrid which is itself served by more than one CA. Furthermore, several related middleware problems have been found where the software has "mysteriously" failed once the number of CAs has reached a certain limit (usually around 60), but these are mostly fixed.

2.4.3 Account management

Another scalability problem is that of authorisation, which has traditionally been done by the Subject DN, in the gridmap files.

To get around this problem, NGS will, like many other Grids, be using VOMS to manage its user authorisation. VOMS [2] provides a central Virtual Organisations (VOs) service that manages VO membership and roles. Ordinarily the user, using either their certificates, or more commonly a Globus proxy [23], accesses the VOMS server and gets another proxy, this time with attributes describing their VO membership and optional roles.

One of the advantages of generating the certificates specifically for the NGS Grid, and the fact that they are short-lived, is that we can embed authorisation attributes in them, without requiring the user to perform the second step of contacting a VOMS server. This is particularly important because we aim to simplify or even hide the process from the user, so if the attributes can be embedded in the proxy generation step, that simplifies the process for us and improves usability for the user.

Such work was done in the Manchester "SHEBANGS" [17] ("Shibboleth Enabled Bridge to Access the NGS") project, where a MyProxy server is contacted for user attributes which are then embedded in a certificate (independently of VOMS, but compatible). We can leverage this work to provide attributes for NGS.

The site service needs to call out to, or cache information from, the NGS VOMS server, but should of course be able to fall back to "plain" (non-attribute) certificates in case the VOMS server is unreachable. This attribute management will have to be independent of the site (Shibboleth) Attribute Authorities, because they pertain to NGS work, not to site databases; nevertheless, there may be cases where NGS software will need to authorise based on both types of attributes. Combining these attributes is future work, but may be able to use the work from GridShib [30] which aims to make site attributes available to Grid resources.

2.5 Usability

X.509 digital certificates are the most widely used method of authenticating users to Grids. Prior to using the Grid, users are often required to request certificates using some form of a user agent, typically a web browser. Once issued, the certificate typically needs to be converted to a form that is usable by Grid middleware, stored in an adequately secured manner. New Grid users therefore need to learn the fundamentals of key management as well as the processes of revocation and renewal and under which circumstances these must be done.

It is thus clear that, unlike other applications of PKI such as smart cards where the mechanics of PKI are shielded from the end user, Grid users are required to have a basic grasp of using digital certificates. Whilst this may be reasonable for the majority of Grid users at present who come from a scientific computing background, there are increasing trends, not only in the UK but worldwide, for Grid computing to be used in multidisciplinary research, and particularly so for the NGS.

The authors of this paper, who not only manage and run the UK e-Science CA but also work with the helpdesk which deals with user queries related to the CA, frequently encounter frustration from end users who view digital certificates and their usability issues as a barrier to using the Grid. This may be partly alleviated with a CA that is easy to use and well documented; work has been done to address these issues [10]. However, the overhead of learning about digital certificates remains, and if the current authentication methods continue to be used, these usability problems will be encountered by the increasing number Grid users from non-computing domains.

In the work described in this paper, usability is improved because basic certificate management can be done by portals and other tools on behalf of the user, and locally at the user's site. It enables sites to provide the "single sign-on," i.e., single password, mechanism by integrating the site's credential conversion MyProxy with the site's authentication system; this is also one of the advantages of Shibboleth. Moreover, we can improve usability further by removing the need for the user to call out separately to a VOMS server.

For certain types of client tools, we can even hide the certificate/proxy generation from the user completely—every single step is performed transparently by the tool on the user's desktop, contacting the local conversion service using the user's cached *desktop* login credential, and the conversion service in turn contacts the global NGS VOMS server. Thus, the user need not even know that the client tool has generated a certificate on behalf of the user.

As discussed in section 2.3 we can do this with Microsoft Active Directory—and equally Kerberos V—but the account management step is currently missing. We have the account request step in ShibGrid, but of course it requires Shibboleth. The obvious solution is to build a desktop registration client which the user can use to request a personal account. A smarter solution would see the user registering with GridShib-exported site attributes, being joined to an NGS VO by a site-local administrator, and would access NGS resources on the basis of attributes alone. However, as discussed in section 3 this requires a greater trust in the site's operations.

2.6 The USHER Hierarchy and Levels of Assurance revisited

In this section, we briefly cover the USHER work since it is similar to the PKI-aspect of the work described in this paper.

The US Higher Education Root [26] is operated by Internet2 to establish a PKI for educational institutions. It consists of a root which issues certificates to institutional CAs, and it imposes requirements on the institutional CAs. USHER imposes the requirement that certificates are issued by subordinates

> "using a process that is at least as strong as its existing practice for managing accounts for central services such as electronic mail, calendaring, and access to central file storage[27]"

This is equivalent to the assurance provided in Shibboleth, and to what we have in our project—with the important difference that we do not have an explicit commitment from the site. In fact, we do not know the exact practices implemented by the site to meet these requirements, nor do we have the ability to audit the site's practices.

The principal difference between USHER and our PKI deployment is that our PKI deployment is much simpler, partly because we have no need for the institutions to trust each other's credentials. Only the resource providers need to trust the credentials of all the institutions, and the resource providers, although individually members of participant institutions, are all part of the NGS. There is a world of difference between asking the University of Oxford, say, to trust a CA, or to ask the NGS administrator at Oxford to trust the same CA. In a sense, we achieve the same result as USHER via the back door—via the project, not via the institution. The drawback is that our CAs are not widely trusted, but we can live with that because we need them only for the NGS.

Another simplifying fact is that we have no need for long-term credentials, so can rely on the institutional CAs to perform the conversion as and when it is needed.

Of course, our work is more than a PKI deployment: we are helping sites deploy MyProxy-based credential conversion servers, along with client tools and NGS-specific software such as portals that use it. Thus, our PKI deployment is tailored to fit the project rather than being a general purpose PKI.

HEBCA/USHER started work [3] to map Grid (more precisely, IGTF) *authentication profiles* to their own assurance levels (HEBCA is the EDUCAUSE US Higher Education Bridge CA)—note that Grid authentication profiles are all roughly "medium assurance." This sort of policy mapping exercise is commonly done for bridge CAs [11], although profiles are often not directly comparable even when they both claim to follow RFC 3647. In this case, the exercise (*ibid*, p. 9) showed the IGTF "classic" authentication profile (the first implementation of what we have referred to as "medium assurance" in this paper for Grid CAs), being equivalent or slightly worse to HEBCA level "Rudimentary" and the Federal PKI level C-4.

Although interesting from an academic point of view, this work should not yet be relied upon for mappings. Usually clarifications and policy adjustments bring partner policies closer together.

From our (NGS') perspective, the importantance lies in establishing the policy mappings, even just tentatively. NGS has in the past been required to "interoperate" with TeraGrid, and this has so far been accomplished between the IGTF-approved CAs, with lengthy separate reviews for those pending such accreditation. For example, the UK e-Science CA is trusted by TeraGrid for this reason. Experience has shown, however, that the US is a large country with many diverse PKIs, and usually the NGS has to trust one or more non-IGTF-accredited CAs. We have had requests from users in the US with no "obvious" CA for NGS access, and in the future it would be convenient if their institutions could get CAs via USHER, at least if they have a need for more than a few certificates (otherwise a project-related "catch-all" CA could do the job). Mapping the USHER policies with known Grid policies (namely, medium assurance) will greatly help the NGS evaluate the trust of those CAs.

As regards the UK and the PKI described in this paper with the institutions participating in NGS, there is in general not much we can do about getting the institution to commit to a certain level of assurance. As we described above and discuss in more detail in section 3 we are not exposed to the institution's user identity process, nor can we demand legally binding documents to this effect, since this was supposed to be a lightweight PKI to complement the Shibboleth deployment.

2.7 Shibboleth Interoperability

2.7.1 Credential Conversion with Shibboleth

One could create a central CA portal to which each user connects to obtain a certificate. To ensure that the home site database is queried, users must use Shibboleth to access this CA. Indeed, this is more or less the model SWITCH used for one of their CAs. In the UK, we chose the approach of distributing the subordinate credential conversion CAs to sites, for three primary reasons:

Firstly, the practical reason, because there isn't yet a widespread Shibboleth deployment in the UK.

Secondly, because of the personal data; we cannot rely on sites being able (or willing) to release sufficient personal data from their Attribute Authorities to uniquely identify the user and map them to the same DN every time: at the time of writing (Oct 06) it is not clear whether sites in the UK Federation are required to publish anything other than eduPersonScopedAffiliation (for an explanation of the eduPerson schema please see [1]). We would need at least eduPersonTargetedID but that in turn will not be sufficient to satisfy those Grid resources that require "meaningful" commonName (CN). Even if NGS itself chooses to accept pseudonymous identities, which could be imposed, if at all, only on the core sites, some affiliated resources have already that pseudonymous identities will not be accepted.

In fact, the base UK Shibboleth Federation aims to be pseudonymous, with explicit agreement between Identity Providers (IdPs) and Service Providers (SPs) whenever extra attributes are required. This means the institutions have to worry about data protection issues. For some, the provision of an institution's IdP may even be out-sourced with no link back to the institutions user database, which implies that these attributes cannot be provided. Finally, some sites with NGS users may just not join the Shibboleth federation.

Thirdly, using "smart clients" we can leverage the sites' internal authentication infrastructure without compromising site security. Unlike the Shibboleth portals where users need to select their home site and then log in again to their home site, our smart clients can pick up the user's site authentication token and transparently generate a VOMS-proxy with which the clients can access the NGS Grid on behalf of the user. Not all clients are "smart" enough to do that, but as mentioned earlier (sections 2.2 and 2.5) for parts of the userbase we aim to even hide that the proxy exists. As mentioned in section 2.2, compared to Shibboleth, we lose the ability to support roaming (off-site) users with the work described in this paper.

It is worth looking at other aspects of the distributed vs. central issues in more detail:

For the central model, running a central high-availability CA is a big commitment. The distributed model distributes the burden: a site's server can still go down, but at least only that site is affected, not the whole Grid. Another advantage of the model proposed in this paper is that not even the VOMS server needs to be high availability: if it is unavailable, sites can fall back to cached information. Information that is potentially slightly out of date is better than none at all.

Furthermore, if there is a central CA portal creating credentials for the users, then users need to use that credential either wholly within that portal, or export it from the portal. For NGS, though, there will be more than one portal, and not all NGS work will be done via portals. Nevertheless, using portals that call out to a central high-availability MyProxy is an option we may use in the future, when we are sure that all of the NGS is covered by the Shibboleth federation. This, too, would alleviate the lack of support for roaming users, since they could use the site credential conversion when on-site, and Shibboleth as a more complex alternative when off-site. In that case, a Shibboleth federation covering the NGS userbase adds value to the work described in this paper—or vice versa—although there is a danger, with more than one issuing authority, that the user will have more than one DN. This is discussed further in section 2.7.2 below.

In the distributed model the certificates are generated locally (within the user's home institution) and can be used locally, on the user's desktop, as we mentioned above, and the service does not need to be exposed to the outside world. The credential conversion service can pick up attributes that the institution's Shibboleth IdP may not publish, such as the commonName (CN). Whether the service is allowed to expose this to the NGS is a question of site data protection policy. The local credential conversion services also allow more robust traceability if pseudonymous DNs are used.

Moreover, an institution providing user account management with higher assurance can use that, internally or externally, without having to live with certificates created with a Shibboleth federation's lowest common denominator level of assurance.

Finally, it is relatively easy to add a new institution to the NGS framework whereas joining the UK Shibboleth federation is more work—the latter requires a legally binding commitment on behalf of the institution as well as setting up and running a high-availability IdP.

2.7.2 Accessing the Grid via Shibboleth Portals

Some version of the central Shibboleth-portal model, as discussed in the previous section, is likely to be implemented in the long term, via an NGS portal. Work is already being done to "Shib-enable" one NGS portal [18], and it will be feasible to integrate certificate generation into the portals. In this central model, the private key will be generated remotely (by the portal), and the certificate by a central service accessed by the portal. The challenge here is to ensure that both views are consistent: the portal views — there may be more than one portal — and that of the site's credential conversion service.

Within this world a user may have many DN-based identities: a DN from the UK e-Science medium assurance CA or one of its international peers, a DN from the Shibboleth credential conversion service (usually from a central Shibboleth portal), a DN from her institution's local credential conversion service and maybe even DNs from other institutions where she may have accounts. How does a Grid resource know that all these DNs are the same user?

Maybe it doesn't have to know. As long as authorisation *only* depends on the VO attributes (neither site attributes, nor the identity), the user will have the same access rights, assuming of course that the user gets the same VO attributes for each identity. However, current middleware deployed on NGS sites still depends on gridmap files, i.e., identity-based authentication.

Tying the same VOMS attributes to several identities, or more generally run a database that knows about the identity mappings, is a problem we cannot currently solve. It relies on the user's collaboration, but even honest users may be put off by the work required to register DNs centrally—as we explained (section 2.5), many target users don't want to know about certificates. We could go some way towards this goal by comparing commonNames (CNs); for the less common names (less common commonNames if you will) it would give us an indication of when two different DNs represent the same user, and this task could even be automated but would still need human review.

Meanwhile, our assumption is that this will not be a problem as most users will, in general, stick with the same identity. Users might make changes to which way they log on, but they probably will not keep changing between different methods, especially if their institution only supports one form of credential conversion service. This behaviour will be enhanced by allowing the same set of resources, services and access methods through all types of identity.

2.8 Status

We finally get to the status of the deployment. The hierarchy has been set up, but the access has only been tested locally within two of the core sites (namely, University of Oxford and Rutherford Appleton Laboratory). Independently, we have tested the attribute services with both VOMS and SHEBANGS, but this work has so far only been in the test phase because the NGS does not yet use VOMS for VO management in production.

Furthermore, we already provide both "normal" and single sign-on MyProxy services for the NGS. Ongoing work over the next months will see wider deployments to core sites and further integration of the independent components, primarily combining the SHEBANGS work with the site credential conversion service at University of Manchester.

Deployment will be further enhanced by deploying the required software—all but the keys—on a single CD image, similar to the NAREGI CA-on-a-CD or Scott Rea's (Dartmouth College) OpenCA-on-a-CD (but of course using MyProxy instead of OpenCA).

3 Security Issues

No CRLs are published. It is common practice that proxies [23] live about 12 hours, but as we mentioned, they could be valid for anything up to 10^6 seconds. This leaves a window in which a compromised credential could be misused, but the same problem is found in general with long lived X.509 certificate where the user must first notice that the credential has been compromised, must then request revocation, under some circumstances that request will have to be approved (usually when it hasn't been signed with the user's private key), then the CRL has to be issued, and finally, the RPs will have to download the CRL. We thus do not consider the lack of CRL a particular security risk, although the *architecture* should encourage users to protect their private keys, since, as mentioned above, we cannot expect users to be experts. Hiding the key from the user helps as long as the local client is careful to store it on local disk. Backups are not necessary because the client will be downloadable and the identity can be easily regenerated by the user.

The quality of site identification has been a contentious issue for many years, particularly for credential conversion CAs seeking to be accepted by international Grid projects. It is hard to persuade collaborators that your identification process is good enough when the CA manager is not responsible for the user data, and in general has no access to it. It is all the more difficult when the site database contains external users, contractors, or temporary staff. In the few cases where Grid projects have accepted such a CA, the credential conversion has been at the *same site* as the CA (FNAL, CERN), and the CAs have been able to enforce a security level necessary for medium assurance by using existing assurance level flagging in the site database (or very occasionally the Grid project has pragmatically decided to trust the CA until a problem occurs). Indeed, most sites would be reluctant to share their user data with a CA auditor, because it would violate data protection policies. However, although the sites provide no such assertion to the NGS in this framework, they do in general use their site databases also for access to more "precious" resources, e.g., internal financial information. We have thus decided that we can reasonably expect sites to operate the databases to a "satisfactory" level, even if we have neither documentation for what this level is, nor a binding commitment to operate to it. This trust is particularly important with pseudonymous credentials where we rely on the site to maintain the mapping to the user's real world identity, but also for sites where access is potentially granted to the site using site affiliation such as the Shibboleth site attribute (scopedAffiliation), i.e., access is granted to all users on site at the site's discretion. A site found to abuse this access privilege can be revoked from accessing the NGS, and the embarrassment factor will most likely fall upon the site rather than the NGS itself.

The scalability issue, as described in section 2.4.2 means that, for the purposes of this NGS deployment, medium assurance is too strong. However, there may well be cases where the NGS *will* require a higher level of assurance. People requiring such access will either have to get a certificate from the UK e-Science CA (for example, this is required to access TeraGrid), or will have to prove independently that their site authentication was sufficiently strong. This assurance level can subsequently be managed as a VO attribute. CERN has taken this approach with their internationally approved MICS CA: it issues certificates only to users in the site database with one of fourteen different status, and one type of external contractors. Each of those status flags ensures that the user has shown appropriate photo id to the CERN user office.

Finally, for access to, e.g., certain medical images (non-anonymised), or financial data, it may be that medium assurance is not strong enough. High assurance cannot be provided within this PKI, partly because we have no explicit commitment from the institutions. Rather, it would have to be provided by either a separate CA, or with special certificates issued by the e-Science CA with flags (e.g., policy OID) to mark it as high assurance.

4 Acknowledgments

This work is funded by the National Grid Service in line with its provision of a national Grid CA and Grid support; the closely related work in ShibGrid and SHEBANGS is funded by the UK Joint Information Systems Committee (JISC) under the current (Apr '04-Apr '07) core middleware development programme.

This document typeset with LaTeX.

5 Conclusion

In this paper, we have described an architecture for a Grid PKI for the UK National Grid Service, NGS, deployed alongside and interoperating with the national Shibboleth deployment, as well as the existing global Grid PKI. The work aims to provide "NGS access for the masses" and be scalable to a large number of users, and to improve usability for non-technical users by making—for certain types of client tools—the entire certificate and proxy issuance process hidden from the user. We have described how the deployment is lightweight, requiring no legally binding commitment on behalf of the institutions. We have described scenarios where we grant access to the NGS based on site or Virtual Organisation membership attributes. We have discussed how we weigh the associated security concerns, improving them where possible and accepting them where not.

References

[1] Internet 2. Eduperson specification. Identifier: Internet2-mace-dir-eduPerson-200312, December 2003. Version 200312. Available at `http://www.nmi-edit.org/eduPerson/internet2-mace-dir-eduperson-200312.p%df`.

[2] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell'Agnello, Á. Frohner, A. Gianoli, K. Lrentey, and F. Spataro. VOMS, an authorization system for virtual organizations. *Lecture notes in Computer Science*, (2970):33–40, 2004. Grid Computing, First European Across Grids Conference, Santiago de Compostela, Spain, February 13-14, 2003.

[3] P Alterman and S Rea. Policy Mapping Grid CAs. `http://indico.na-df.rnp.br/indico/materialDisplay.py?contribId=10\&amp;%sessionId=1\&amp;materialId=slides\&amp;confId=15`, Nov 2006. 3*rd* TAGPMA meeting, TACC, Austin Texas (URL checked Feb 07).

[4] J. Astalos, R. Cecchini, B.A. Coghlan, R.D. Cowles, U. Epting, T.J. Genovese, J. Gomes, D. Groep, M. Gug, A.B. Hanushevsky, M. Helm, J.G. Jensen, C. Kanellopoulos, D.P. Kelsey, R. Marco, I. Neilson, S. Nicoud, D. O'Callaghan, D. Quesnel, I. Schaeffner, L. Shamardin, D. Skow, M. Sova, A. Wäänänen, P. Wolniewicz, and W. Xing. International grid CA interworking, peer review and policy management through the European DataGrid certification authority coordination group. In P. Sloot, A. Hoekstra, T. Priol, A. Reinefeld, and M. Bubak, editors, *Advances in Grid Computing — EGC 2005*, LNCS3470, pages 285–295, Amsterdam, The Netherlands, February 2005. Springer.

[5] William Burr, Tim Polk, and Donna Dodson. Electronic authentication guideline. NIST Special Publication 800-63 Version 1.0.2, April 2006.

[6] R Butler and T Genovese. Global grid forum certificate policy model. http://www.ggf.org/documents/GFD.16.pdf, June 2003.

[7] D. Byard and J. Jensen. Single sign-on to the grid. In *Proceedings of the 2005 UK e-Science All Hands Meeting*, September 2005.

[8] TAGPMA. Editor T. Genovese. Profile for short lived credential services X.509 public key certification authorities with secured infrastructure. Reference: IGTF-AP-SLCS-20051115-1-1, November 2005. Version 1.1. Available at http://www.tagpma.org/files/IGTF-AP-SLCS-20051115-1-1.pdf.

[9] J. Jensen, D. Spence, and M. Viljoen. Grid single sign-on in CCLRC. In *to appear in the proceedings of the 2006 UK e-Science All Hands Meeting*, September 2006.

[10] J. Jensen and M. Viljoen. Usability of the UK e-Science Certification Authority. *UK e-Science All-Hands meeting*, 2005.

[11] Mark Luker. A bridge for trusted electronic communications in higher education and federal government. http://www.educause.edu/ir/library/pdf/ERM0203.pdf.

[12] J. Myers. Simple authentication and security layer (SASL). Request for Comments (RFC) 2222, October 1997.

[13] National Grid Service. http://www.ngs.ac.uk/, Oct 2006.

[14] J. Novotny, S. Tuecke, and V. Welch. An online credential repository for the Grid: MyProxy. In *10th IEEE International Symposium on High Performance Distributed Computing (HPDC-10), San Francisco, California, USA*, pages 104–114, August 2001.

[15] US office of management and budget. E-authentication guidance for federal agencies. http://www.whitehouse.gov/omb/memoranda/fy04/m04-04.pdf, December 2003.

[16] C. Rigney, S. Willens, A. Rubens, and W. Simpson. Remote authentication dial in user service (radius). Request for Comments (RFC) 2865, June 2000.

[17] SHEBANGS (Shibboleth Enabled Bridge to Access the National Grid Service). Details at http://www.sve.man.ac.uk/Research/AtoZ/SHEBANGS, June 2006.

[18] D. Spence, K. Tang, R. Allan, M. Dovey, N. Geddes, J. Jensen, A. Martin, D. Meredith, M. Norman, A. Richards, A. Trefethen, M. Viljoen, and D. Wallom. Shibgrid: Shibboleth Access for the National Grid Service. In *Proc. IEEE 2nd Int'l Conf. on e-Science and Grid computing*, 2006.

[19] J. Steiner, C. Neuman, and J. Schiller. Kerberos: An authentication service for open network systems. In *Proceedings of the USENIX Winter Conference, Dallas, Texas, USA*, pages 191–202, February 1988.

[20] NGS Grid Support. GSI-SSH Terminal. http://www.grid-support.ac.uk/content/view/81/61/.

[21] SWITCHaai. http://www.switch.ch/aai/.

[22] TeraGrid. http://www.teragrid.org/.

[23] S. Tuecke, D. Engert, I. Foster, M. Thompson, L. Pearlman, and C. Kesselman. Internet X.509 public key infrastructure proxy certificate profile. Request for Comments (RFC) 3820, June 2004.

[24] UK e-Science Certification Authority. http://www.grid-support.ac.uk/ca/, Oct 06.

[25] UK Shibboleth Federation. http://www.ukfederation.org.uk/, Oct 06.

[26] Us Higher Education Root pki. http://usher.internet2.edu/.

[27] Usher Subscriber Expected Practices. http://usher.internet2.edu/docs/USHER-Expected-Practices-final.htm, Nov 2006.

[28] G. von Laszewski, J. Gawor, P. Lane, N. Rehn, M. Russell, and K. Jackson. Features of the Java commodity Grid kit. *Concurrency and Computation: Practice and Experience*, 14:1045–1055, 2002.

[29] M. Wahl, T. Howes, and S. Kille. Lightweight directory access protocol (v3). Request for Comments (RFC) 2251, December 1997.

[30] V. Welch, T. Barton, K. Keahey, and F. Siebenlist. Attributes, anonymity, and access: Shibboleth and Globus integration to facilitate Grid collaboration. In *Proceedings of the 4th Annual PKI R&D Workshop*, April 2005.

[31] V. Welch, F. Siebenlist, I. Foster, J. Bresnahan, K. Czajkowski, J. Gawor, C. Kesselman, S. Meder, L. Pearlman, and S. Tuecke. Security for Grid services. In *12th International Symposium on High-Performance Distributed Computing (HPDC-12), Seattle, Washington, USA*, pages 48–57, 2003.