Stephen Chan, NERSC/LBNL
Matt Crawford, FNAL
9 October 2003

Certificates for Automated Clients
<draft-ggf-caops-auto-client-certs-00.pdf>

Status of this Memo

This document provides information to the grid community. It does not define any standards or technical recommendations. Distribution of this memo is unlimited.

Abstract

In evaluating Grid technology for production use, several members of the DOE Science Grid community feel there is a need for another class of certificates to be used for authenticating automated clients that connect to Grid servers. The usage of these new certificates is different enough so that current CA certificate policies will need to be updated.

1. Introduction

Current DOE Science Grid (DOESG) and European Data Grid (EDG) certificate policies cover issuing x509 certificates for individual identities and for authenticating hosts and services. In evaluating Grid technology for production use, several members of the DOESG community feel there is a need for another class of certificates to be used for authenticating automated clients that connect to Grid servers. The usage of these new certificates is different enough so that current ceritificate policies will need to be updated.

We propose that a new namespace branch, differentiated by an "OU=robots" Relative Distinguished Name (RDN), be created that can be used to distinguished certificates used for automated processes. In addition, we propose that specific extensions be added to the certificate to identify responsible parties and legitimate source domains. In terms of policy, we recommend that sites that wish to use automated certificates publish a policy statement describing how they intend to use and manage these certificates so that RA’s and resource providers can make informed trust decisions.

2. Motivation

One of the major differences between a production environment and a research environment is in procedure: production environments typically have a large collection of procedures that are used to maintain "production quality", examples are service monitoring systems, backup systems and various housecleaning utilities. Production environments also typically support repetitive, ongoing processes - either internal system processes or processes relating to the applications being run at the site, for example, processes that move datasets from one site to another.

These procedures and repetitive processes are typically automated, run across multiple machines, and generally run using an identity with the necessary privileges to perform their tasks, and little else. As Grid technology is put into production, we are finding that we need to perform the same kind of repetitive, automated tasks, but using Grid credentials over Grid services. However, there does not seem to be any allowance within the DOESG and EDG certificate policies for issuing certificates to support these kinds of tasks.

3. Discussion

Currently, 2 classes of certificates are issued:

Personal identity certificates

> These are used to identify individuals, and have the "OU=people" attribute in their subject distinguished names. Typically, these certificates are used to authenticate the client side of a client/server request, and the certificate must be on any machine from which a client request could originate. The private key is supposed to be while stored encrypted, and user action is required to generate a proxy certificate based on the key. The accountable party is the individual identified on the certificate.

Host/Service certificates

> These are used to authenticate a host, or a server running on the host, and generally have the "OU=services" attribute in their subject distinguished names. These certs are generally used to authenticate servers to clients, and are uniquely located on that machine. There is usually no reason for a client connecting to a server to present an "OU=services" certificate (for example, you would not expect to see such a cert in a grid-mapfile). The private key is stored unencrypted, and is used "as-is," without generating a proxy certificate. The accountable party is nominally the individual that submitted the certificate signing request, but in fact it is the IT group that maintains the machine or service identified on the certificate.

A certificate for an automated client has attributes that span both these types of certificates:

1) The private key needs to be stored unencrypted so that automated tools can use it. This is similar to a service’s credential.

2) The certificate may be used to generate proxy certificates, and is used to authenticate the initiator of a connection. One could reasonably expect to see the subject names of these certificates in a grid-mapfile, as you would a personal identity certificate.

3) The accountable party may not be clearly identifiable in the Sub ject Distinguished Name in the certificate, but might nominally be the person submitting the certificate signing request. However, the true accountable party would be the group that operates the automated client. This is similar to a service.

4) Many of these automated clients do the same task from multiple machines. For example, automated backup clients all do essentially the same task, and usually run as the same restricted identity across multiple machines. For data replication clients, the replication service may operate from multiple machines. Getting and managing multiple certificates for a small number of machines is annoying, getting and managing certificates for hundreds of machines (such as the typical HENP cluster) becomes problematic. IT staff may be inclined to reuse the same certificate on multiple hosts. If this is not acceptable practice, there needs to be some mechanism to prevent or detect it.

Because these certificates are neither fish nor fowl in the current model, there needs to be clarification on how they can acceptably be handled. Another set of use cases that hasn’t been discussed, but are related, is authentication of peer to peer services.

The discussion on the DOESG mailing list also seemed to identify these main issues:

Identity/accountability -- who exactly would be accountable for these certs? The correct answer from the perspective of an organization is that an IT group would be accountable for the certificate. Individuals within an IT group may share or change responsibilities, and turnover is not unheard of - at which point, the certificate will no longer accurately reflect the accountable party (arguably, it may never have in the first place). In a managed, production environment, an organizational abstraction is the actual accountable party, and not an individual. However it might be argued that assigning accountability to an abstraction dilutes accountability.

Shared certificates -- the notion of sharing the certificate across multiple machines increases the potential damage from a compromised private key. However, this same form of risk is managed on a daily basis by IT groups that have standard passwords for system accounts (such as root). The exposure is contained by the fact that there is a de facto trust boundary at the site perimeter - no machine managed by another group shares the same "root" password. If shared certificates were issued, it would be wise to ensure that all machines that trust that certificate are under the same administrative domain.

4. Analog from existing practice

Kerberos is already used by automated clients at many sites. To explore the problem space, we examine the practice at one such site, Fermilab.

Service principals are used directly as clients when it is the service itself that is acting as the client of another service. For example, when a mass storage systems initiates a transfer to or from another mass storage system, it identifies itself as the same name by which it’s known to clients.

When the automated process initiating a connection isn’t a known service itself, it uses what’s known at this site as a "project principal." The name of such a principal has three components (besides the realm) which identify its function, the responsible organizational unit, and the single host from which it is expected to authenticate itself. For example, "level3/cdf/b0dax42.fnal.gov@FNAL.GOV" acts as part of the Level-3 trigger for the CDF experiment, and authenticates using a keytab file (analogous to the private key corresponding to a certificate) stored on the host b0dax42.fnal.gov. Since a Kerberos initiator must first communicate with the KDC to obtain credentials, the KDC has a record of where the authentication requests for this principal come from and, to the extent that DNS can be relied upon, may reject requests from elsewhere than that one host or raise an audit flag for them.

A large class of automated client requests are performing actions for a specific user, generally through the unix "cron" scheduled execution facility. Fermilab users have the ability to create and destroy per-user, per-host principals for themselves with names of the form username/cron/hostname@FNAL.GOV and to create a keytab file on the corresponding host. Tying each such principal to a host and a user facilitates recovery from any security incident involving the user or the host.

5. Proposals

One of the most basic, and probably least controversial proposals is to create a separate namespace for this new class of certificates. Reusing service certificates sets up a precedent of having "server side" certificates suddenly become client side proxy certificates, and makes it harder to clearly identify misuse of certificates. Using personal certificates for these services requires that individuals expose their personal certificates to theft because the private key is stored in the clear (or stored together with its decryption key). This is especially dangerous for administrators, whose personal accounts may contain sensitive information and or privileges. In addition, it is generally desired that these automated client identities accounts map onto severely restricted local accounts, instead of a normal user account.

We propose that a new namespace branch be created, alongside the customary "OU=People" and "OU=Services," differentiated by an "OU=robots" RDN. The commonName RDN(s) in such names shall include a descriptive name for the automated function. This function should be specific, similar to "CN=NERSC backuprobot" for a NERSC client used to perform backups, or "CN=ATLAS filereplicator" for a process that performs file replication for the ATLAS collaboration. When possible, the fully-qualified domain name of the host that holds the private key shall appear in another commonName RDN.

The descriptive name RDNs are purely informational, although Registration Agents or CAs should deny certificate requests that they deem to be too generic or insufficiently descriptive.

In addition, all such certificates must include a SubjectAltName extension. If non-critical, it simply provides contact information for the party responsible for the actions of the automated function. The automated process identified by the certificate is considered to be an agent of the party identified in the SubjectAltname.

The contact information must lead to members of formal organizations that operate datacenters with a clearly identified systems administration staff, or members of a multi-institutional collaboration, vetted by a principal investigator or spokesperson. A minimum of one rfc822Name within the SubjectAltName extension is strongly recommended.

Including a critical SubjectAltName extension is experimental. Such a critical extension should contain, in addition to the contact information, one or more iPAddress and/or dNSName elements and relying parties are requested to accept the credential only when presented from a peer on one of those addresses or hosts. (But see Security Considerations below.)

Parties that request these certificates must publish a policy statement that describes how the certificates will be used, who will have access to the private key and what measures will be used to ensure the security of these certificates at their site. This document must be reviewed and approved by the appropriate certificate authority bodies before certificates will be signed.

5.1. Alternatives

The more controversial proposals would include somehow altering the policies to make allowance for the notion that "permanent" IT groups (as opposed to ad hoc virtual organizations) can be accountable parties. Possibly the most controversial is how to handle the notion certificates with shared access to the private key. This is a convenience for IT organizations and is a fact of life for almost all service credentials. But when considering automated processes as initiators of Grid activities, the security of their credentials needs to be approached with caution.

None of the proposed changes can be effected unilaterally, and we also believe that the issues are of concern to European Data Grid users as well as the DOE Science Grid users. In fact, it will likely be a concern for all large Certificate Authorities, and the situation provides an opportunity for the DOESG and EDG to exercise leadership in setting standards for how these certificates can be handled.

6. Security Considerations

Three main security considerations have come out of discussions:

6.1. Accountability

How do we identify a responsible party for the robot certificate? This is addressed by the required SubjectAltName component - verification of the SubjectAltName is handled by the RA using standard methods. In addition, the policy statement is another aspect of accountability. By publicly stating the manner in which the certificates can be used, as well as providing contact information for a responsible individual, we believe that an acceptable standard of accountability can be maintained.

6.2. Misuse

This is an important consideration. In fact, the ability to partition off automated client certificates into a separate namespace allows resource providers finer grained access control, which has the potential to curb misuse. In addition, the ValidDomains attribute and the descriptive Common Name attribute serve to identify legitimate uses for the client certificates. If normal user certificates or host certificates were used, it would be harder to distinguish either legitimate uses (for host certificates) or legitimate client machines (for user certificates).

6.3. Proper Management of Client Certificates

This is a legitimate concern, and is the motivation behind the policy statement required of all parties that request automated client certificates. Ultimately, proper handling depends on the professionalism of the staff managing the certificates - there is no clear legalistic method to verify this, so this depends initially on the registration authority’s confidence in the policy statements of the certificate requestor, and subsequently on the authorization decisions of the resource provider.

Arguably, the security of any certificate is only as good as the people who handle them. Production operations staff have been operating system accounts for decades (for example: "root" in Unix), and a strong case can be made that such groups have legitimate uses for automated client certificates, as well as the professional experience to management them securely. With the arrival of finer grained authorization for Globus services, the security of single purpose client certificates can be further enforced by technical means. Partitioning off a new namespace only simplifies the application of these authorization tools, thus increasing security.

6.4. Other security considerations

Attempting to restrict the source hostname or address from which a credential may be used by means of a critical SubjectAltName extension is only a weak security measure, since addresses can be spoofed and DNS is generally insecure.

Storing a single private key on multiple systems is to be avoided if at all possible, and multiple automated clients performing similar or connected functions should have separate Subject Distinguished Names. Preferably those names will identify the host on which the unique instance of the associated private key resides.

Finally, as with service private keys, care must be taken to prevent the private key from being accessible to unauthorized parties through system backup media.

7. Author Information

Stephen Chan
NERSC/LBNL
One Cyclotron Road
Berkley, CA USA 94720

Matt Crawford
Fermilab MS-369
PO Box 500
Batavia IL 60510
USA

Intellectual Property Statement

The GGF takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the GGF Secretariat.

The GGF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this recommendation. Please address the information to the GGF Executive Director.

Full Copyright Notice

This document and the information contained herein is provided on an "AS IS" basis and THE GLOBAL GRID FORUM DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

References

RFC3280 Housley, R., W. Polk, W, Ford, and D. Solo, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile," April, 2002.