

EXPERIENCE WITH PKI IN A LARGE-SCALE DISTRIBUTED ENVIRONMENT

Daniel Kouřil, Michal Procházka, Luděk Matyska
CESNET z. s. p. o., Zikova 4, 160 00 Praha 6, Czech Republic, and
Masaryk University, Botanická 68a, 602 00 Brno, Czech Republic
e-mail: {kouril,michalp,ludek}@ics.muni.cz

Abstract

PKI shows some weaknesses when used in real deployment, such as problems with securing the private key, combination of PKI with other authentication systems, availability of revocation information, user-friendliness of PKI system and relationship with federation systems. In this paper we present solutions that were used during PKI deployment in a grid distributed environment.

1 Introduction

Compared with local infrastructure operated by a single institutions the distributed systems—*grids*—have some specifics that make the environment far more complex. Users and resources are spread geographically, sometimes even span multiple countries, resources usually do not belong to a single administrative domain and usage of resources and behaviour of users is regulated by local policies and legislation. Therefore providing users with a uniform view on the environment is very difficult in the distributed systems. The situation gets even more complicated in the area of authentication and authorization. An authentication schema of distributed systems has to be flexible and scalable in order to fulfill requirements of the whole system. A combination of a user name and password is still the most popular authentication type but in its simplest form it does not comply with requirements on distributed system mainly because it does not provide the Single Sign-On functionality. Authentication based entirely on user name and password is not secure enough and it requires a centralised administration, which is not maintainable in large-scale systems due to large administrative overhead.

An authentication system based on the *Public Key Infrastructure* (PKI) [1] has a decentralized management of users information and therefore it is suitable as the authentication system for distributed environments. Authentication data of user in PKI world are represented by a personal public-key certificate providing a digital identification in the digital world. The relationship between the user and her digital certificate is approved by a *Certification Authority* (CA).

Each distributed environment is composed of various components. Usually the components are represented by particular institutions participating in the system, which provide their resources and users. Nowadays every institution already implements some type of an authentication system and it is clear that they cannot be convinced to use PKI. That is why we have to keep in mind that there should be solutions that provide a mapping mechanism between existing authentication systems and PKI.

In this paper we will describe our experience with deployment of PKI in a distributed environment in nation-wide grid METACentrum. From its beginning METACentrum have used Kerberos as an authentication system, therefore we will present mapping services between PKI and Kerberos implemented in the infrastructure. We will also focus on problems that we encountered during PKI deployment.

At the end of the paper we will talk about cooperation of PKI and existing federations, especially with an arising federation in Czech Republic.

2 PKI Technology

The whole architecture of PKI is based on three pillars. The first one is represented by a key-pair (consisting of a public and private key) that is bound with its owner using a *public key certificate*, most often encoded in the X.509 format [2]. The second component is a *Certification Authority* that signs the certificates with its signing key. The last pillar is formed by the *relaying parties* who trust the CA and accepts certificates it signed. All the three parts are tightly interconnected.

In PKI every entity holds its key pair that is used for asymmetric cryptography. That means the data encrypted by a public key can be decrypted only with the corresponding private key and vice versa. A personal digital certificate binds the key pair and its owner and provides information about the owner identity. Each certificate contains a public key and information about the person such as her name, institution and location. The certificate along with all necessary information is signed with the private key of a CA, whose identification is also included in the certificate. In order to make the CA operation scalable, the model of PKI introduced the concept of *Registration Authorities* (RA) that are responsible for proper authentication of applicants who ask for certificates. In this model the CA signs certificates requests that are validated by authorized RAs.

The basic principle of PKI-based authentication is built on encrypting random data string sent from the service to the client as a challenge. The client (user) then encrypts the string with her private key and sends the string back together with her certificate. The target service verifies the string was encrypted by the user who sent the certificate. It is done by decrypting the encrypted string with the public key from the certificate. If the decrypted string matches that one sent by the service, the user authentication succeeds. Authentication is successful only in case if the service trusts the CA that issued the user certificate. The authentication schema based on PKI assumes that a private key is held securely. Therefore special attention must be paid to *private key hygiene issues* to make users store and handle their private keys in secure manner.

When a private key is compromised it is necessary to revoke the corresponding public-key certificate. Every CA should publish a current list of revoked certificates that can be used by all applications and services to check if their clients' certificates are still valid. Each certificate has its own lifetime and a certificate is not considered valid anymore once its lifetime exceeded. Common lifetime of personal certificates is one year. Having lifetime assigned to certificates also mean that a certificate has to be renewed by its owner before expiration.

Every CA should have own policies describing the operational procedures, where all the processes of the CA management are specified. This policy should be published and available to all relaying parties. The International Grid Trust Federation (IGTF)¹ is a body to certify particular grid CAs according to their policies. Each CA must pass a

¹<http://www.gridpma.org>

review process that checks the CA policy fulfills the minimal criteria specified in the selected CA authentication profile. The profiles define e.g., that every user who wants to obtain a certificate has to personally contact an RA providing her national identification card. The relaying parties have assurance that they can track down a user through her CA in case she made a problem. The IGTF covers regional bodies handling their local CAs (currently the European EUGridPMA, Asia Pacific Grid PMA, and the Americas Grid PMA form the federation). So far the IGTF has accredited almost seventy CAs from all the world.

The grid environment introduced a special type of public-key certificates—the *proxy certificate* [3]. A proxy certificate is made by the user herself, with the user's private key acting as a CA signing key. The proxy certificate model is primarily used for delegation of user's identity into the grid world, to support batch job submissions and other operations that the user cannot directly assist with. Grid services use clients' proxy certificates to be able to contact other services on behalf of the clients. Grid credentials formed by the proxy certificates and associated private key are usually stored on a filesystem secured by proper filesystem permissions but without any additional protection by a passphrase. To reduce the potential damage caused by a stolen proxy credential they are usually short-lived with the lifetime set to couple of hours. Proxy certificates also make it possible to build an Single Sign-On system, where user creates a proxy certificate only once a day and using the proxy certificate she can then access grid services for the whole day without providing any other authentication data or creating new proxy certificate.

2.1 Problems in deployment of PKI

The PKI features fits the requirements on building a robust environment with reliable authentication mechanism for connected users and services. Because PKI provides a very good level of scalability it is suitable as an authentication mechanism for large-scale distributed environments with hundreds or thousands users. But PKI also has some limitations that can be encountered when one tries to deploy it in that scale. These limitations are not visible in small-scale solutions but they can play an important role in the security of the whole system. According to our experience the limitations are:

- Reliable protection of the users' private key
The keys used for asymmetric cryptography in PKI are very long strings of bytes that cannot be remembered by users and have to be stored in digital form in a computer. Most often they are kept in files encrypted by passwords. A corresponding password has to be provided by the user every time she needs to use the private key. Protection of the file is entirely under control of the user, which turns out to be a crucial problem as user too often do not secure the file sufficiently. Since the files are located on the desktops, it is impossible to make the users responsibly protect the file or even check if the file is really protected by a strong password or the file has right file permissions. Another problem is the possibility of unauthorized access to the password, either using a dictionary or brute-force attacks or a malicious software such Trojans catching the passwords. It is also very difficult to detect if the private key has been compromised.
- Distribution of revocation information on the time
Current well known practice of the CA is to publish a new revocation list after every change. All relaying parties of a particular CA have to have access to the fresh

revocation information. Only in that way we can ensure the highest trustworthiness of PKI. If the information is not actual the service cannot recognize certificates that have been revoked recently. Nowadays the most common mechanism of distributing revocation information uses the standardized *Certificate Revocation List* (CRL), which contains all revoked certificates of a CA. A CA publishes its CRL on the web page or makes it available from a directory service (LDAP) so the relying parties can retrieve it to local machines. The main disadvantage of this approach is a missing mechanism of notifying partners when a new CRL has been published. Currently each relying party has to periodically poll the CRL distribution points to check if a new CRL appeared. The second solution of learning if a certificate has not been revoked is the *Online Certificate Status Protocol* (OCSP) protocol [4]. Using the OCSP protocol a relying party can contact the CA online to check current status of the certificate that is being verified. This approach requires additional communication with the CA and pose additional overhead for authentication process.

- Cooperation with existing authentication systems
Distributed systems and grids are not usually build from scratch but they are often based on existing solutions at participating institutions. When designing a grid infrastructure it is therefore important to keep in mind that PKI should collaborate with existing authentication mechanism. PKI should not add additional overhead or obstacles for users who use the established authentication mechanisms in their institution. In ideal case PKI is deployed consecutively or as an addition to the existing authentication mechanism without bothering users.
- Easy access to obtain user certificates
Policy of each CA defines procedures and conditions that have to be followed to obtain a certificate. Most trustworthy CAs requires personal contact between the RA and the certificate applicant and the applicant has usually to provide a national identification document. In a distributed environment with users spread over a large geographical area it is very important to have an adequate number of RAs so that all users have easy access to a RA. Users also do not have any experience with PKI, which requires to have educated user support staff to be able to help the users.

In the next section we will describe a distributed environment that was used as a pilot infrastructure for PKI deployment. Also problems with deployment will also be mentioned.

3 Deploying PKI in METACentrum

The METACentrum project is a key activity of the CESNET association that operates the Czech NREN and conducts research in advanced network technologies and applications. Using the NREN as the base infrastructure, METACentrum develops and maintains a grid environment that is available to all researches in the Czech republic. METACentrum also participates in most national and international projects focusing on research and development in the grid area that are being performed in the country nowadays.

The METACentrum environment comprises computing and storage resources from several computing centers from all around the country to provide a grid infrastructure containing over 450 CPUs, 25 TB of distributed storage and 400 TB of backup and archive capacity. The environment is routinely used by over 200 users solving problems from a wide range of scientific and engineering areas.

The security framework of METACentrum is built upon the Kerberos protocol [5], which is a well-known system for authentication and key exchange using a trusted third-party model. The system makes use of Kerberos *tickets* that serve as identity credentials for the users accessing the end services. The tickets are issued by the Kerberos authentication server after users' authentication based on entering a username and a secret password. One of the major features that Kerberos provides is support of the Single Sign-On principle that allows the users to use the infrastructure in an easy yet secure manner. After the user initially logs in and retrieves her tickets the user's applications can authenticate to the services transparently using the tickets without users' direct intervention. An explicit login procedure is required just once a day producing a ticket valid for the rest of day.

Kerberos fits very well requirements for domain authentication inside an organizationally closed environment such as e.g., a university. It is not suitable for a dynamic distributed infrastructure, where various domains emerge and leave often. Using Kerberos in such an environment reveals some drawbacks of the protocol, mainly its low scalability. Therefore Kerberos is not a good choice to form an authentication framework in a general grid environment and rather PKI is used there usually. Having been used Kerberos internally, METACentrum started an activity to address the need to provide their users with an easy access to the international grid projects based on PKI. As we have always seen Kerberos as the ideal solution for our local infrastructure we did not want to get rid of it and replace Kerberos with PKI. Instead we decided to adapt our infrastructure to support user authentication using both these mechanisms with Kerberos remaining as the core middleware framework.

3.1 PKI in METACentrum

Since the beginning of this activity we have closely collaborated with the CESNET Certification Authority, which provides a production-level service for the whole research community in the Czech republic. Being approved by the IGTF, the CESNET CA establishes a very solid and trusted basis for a PKI. Immediately after starting our activity to incorporate PKI into METACentrum we formally established an Registration Authority of the CESNET CA that is operated by the METACentrum employees. Having such an RA allows to provide a bridge between our users and the CESNET CA and to focus on specific needs of the grid users and thus provide better user support. A member of the RA staff usually attends all events that METACentrum organizes for its users presenting the differences between Kerberos and PKI and describing the advantages of having public-key certificates issued by the CESNET CA (especially the possibility to join other grid projects and collaborate with researches abroad). Users attending the events can also pass the registration procedure on the spot and receive codes to generate their certificates. The RA members also offer to visit the users directly at their home institutions to facilitate their pass through the whole procedure of getting certificates.

METACentrum also actively develops tools to ease PKI usage. The software is available from the METACentrum portal in the form of both source code and binary packages for major platforms and operating systems. The tools comprise graphical SSH and SCP clients with PKI support and a command-line application to generate proxy certificates from the users' long-term credentials. We also offer a pilot implementation of a GUI for the MS Windows system to maintain proxy certificates in an easy-to-use way. Our goal is to provide users with a single place containing the whole equipment necessary to access any grid environment using PKI.

Since the very beginning we have focused on the problem of private key hygiene and proper protection of private keys possessed by our users. We got a grant from the CESNET Development fund to evaluate the possibilities offered by the smart card technology. During the project we tested several pieces of smart card devices (*tokens*) and chose the most appropriate one suitable for the METACentrum users. We eventually selected USB tokens iKey3000 that integrates in a single piece of hardware the full functionality of a smart card and reader and allows for better mobility compared with the latter solution. More information on the grant can be found in [6]. The hardware tokens provide a secure storage to place the user's private key so that it can never be extracted outside the device. Instead of using the private key directly, the applications requiring private key operations must contact the cryptographic chip on token, supply input data and ask for the result computed by the chip using the private key. Before using the token, the application must first authenticate itself to the chip using the owner password that the user must pass on to the application. Having private keys stored in the tokens where they cannot be accessed directly prevents from common attacks on PKI that focus on stealing or misuse the private keys stored in local filesystem or memory.

We prepared a user guide describing usage of the tokens and started distributing the tokens among our users. We found out soon that usage of tokens in the open METACentrum environment would be quite difficult issue mainly due to the lack of coordinated user support aiming at management of the users' desktops. Being a virtual environment the METACentrum does not provide any end user machines. Instead, in order to access the METACentrum resources the users use their standard desktops provided to them by their home institutions. These machines are maintained by the local institution administrators who do not have any relationship with METACentrum and cannot provide support for problems concerning the tokens. Instead the users encountering problems with the tokens have to contact the METACentrum user support and ask for an assistance. Even though our administrators are ready to provide support they cannot access the user machines directly and also do not have any influence on the local environment at the institution (e.g., permission to change the firewall rules etc.). Nevertheless we managed to address many support requests raised by user from various institutions but their resolutions was very difficult as we often had to simulate the users' environment first on our machines and only then started looking into the problems. While solving these kind of issues we found out very useful the concept of virtual machines containing images of main operating systems. The situation was complicated by the fact that support of smart cards on Linux is still not on a production level and there is couple of problems that sometimes make the use of tokens bothersome (e.g., simultaneous access of multiple applications to the token).

In order to distribute tokens among the users we chose an open mode that provides the users with full access to the token filesystem. The user is responsible for entire management of the tokens and can arbitrarily adapt the token data (i.e. format the token filesystem, load and remove credentials etc.). The advantage of such an approach is that the token owner is not limited in its use, on the other hand if the token PIN is lost the users cannot ask the administrators to unlock the token using the administrative PIN and the token must be re-formatted removing all the stored data. Also in this case we felt the lack of a local administrator sitting close to the end users and helping them with their problems.

During the activity we also amended current middleware tools to support smart cards so the users can use commands with the same interface they are used to. We also designed and implemented several changes to the METACentrum architecture, which added support of PKI. In particular we introduced the PK-INIT extension [7] to the Kerberos

authentication server, which allows our users to use PKI authentication to obtain initial Kerberos ticket needed to further access to METACentrum.

We have also put into operation an experimental installation of an on-line CA server that signs certificates to clients who authenticate with a Kerberos ticket. This CA can be seen as a transform service translating one type of credentials (Kerberos ticket) to another (PKI credential). Using the CA the users can gain better experience with PKI without having to pass the quite complex registration procedure to get a certificate issued by an production CA. Of course, this our experimental CA is only accepted by a small fraction of our services and cannot compete with the possibilities offered by possession of CESNET CA certificates.

In the grid world the concept of *on-line credential repositories* (OCR) is very popular. An OCR server provides a secure storage where the users can load their credentials assigning them a password that can be used later to download a proxy certificate derived from the credential stored in the repository. OCR servers are used in multiple scenarios ranging from access to grid portals to support of long-running jobs. We are currently evaluating an installation of the MyProxy OCR server [8] that supports authentication using one-time passwords. We are also testing software OTP generators (soft-tokens) that can be loaded into a cellular phone or PDA. After installing such an application the user can use her mobile device to generate the OTP necessary to obtain a grid credential from the MyProxy server. That approach can be used to access grid facilities from locations that cannot be entirely trusted (internet kiosks or cafes).

While deploying PKI we were also evaluating the mechanisms to access the revocation information and concluded that both the major methods (periodic retrieval of CRLs and OCSP) are not ideal in the dynamic grid environment. Based on our previous experience in the grid monitoring domain, we designed and implemented an alternative mechanism to CRLs distribution [9]. The mechanism allows the relying parties to subscribe with a CA to receive notification of new CRLs that are sent by the CA upon each change of the CRL. Using a grid messaging infrastructure ensures the transport fast delivery of the notification messages to all subscribers. Unlike OCSP this approach does not add any additional communication overhead to the authentication stage between the client and server.

4 Federations and PKI

As we mentioned earlier PKI is suitable for large scale environments with many users. But these environments have to support PKI internally or have to implement mapping services between PKI and local authentication system. These requirements cannot be always easily fulfilled due to various limitations. PKI targets solely on authentication while authorization in the outside environment is not covered and must be still solved by other mechanism. The concept of federation can address some issues caused by the mentioned drawbacks of PKI.

A federation is a infrastructure connecting user management systems from different institutions to provide a standardized access to user information maintained by their systems. Federations provide a bus layer to which the system for user management and end application can connect and share authentication and authorization data. Every organization participating in the federation manages its own users by a local user management system and other parties in the federation can access information from the system using a *Identity Provider* (IdP) service with a standardized protocol. End services (*Service*

Providers—SP) are able to process the data returned by the user's home IdP and use them to make access control decisions. Before a user is allowed to access a SP, she has to present a set of *attributes* issued by her home IdP, which are provided to the user or a service working on her behalf upon proper authentication of the user with the IdP.

If we get both systems PKI and federations and connect them together we can provide a powerful authentication and authorization system to the user. Users will be able to access large spectrum of services using one authentication mechanism. Federations also enhanced PKI with full-featured authorization mechanism. Federation also brings advantages to the SP which does not need to support variety of authentication and authorization systems, they need only support middleware used in federation.

Federations can also help users to obtain the certificates using a federated on-line CA, which will work as a common SP with authenticated access. Since it will be able to get all the user's attributes from her home institution thus no personal contact between the user and RA will be needed. Based on the user's attributes, the federated on-line CA could issue various types of the certificates, e.g., short lived, signing-only, etc. This concept will spread PKI also into the institutions that do not have available capacities to build and maintain RA. Also mobile users will appreciate the easy and secure access to their PKI credentials.

In METACentrum we deployed an experimental installation of a IdP based on middleware Shibboleth [10] using PKI as the authentication mechanism. Information needed for authorization are obtained from the METACentrum LDAP server, which contains all users of METACentrum. Using the METACentrum IdP, the users of METACentrum will be able to access all services which are available in growing Czech national federation. At this time METACentrum has one SP offering for new users to apply for a METACentrum membership.

METACentrum is now part of existing federation—Eduroam [11]. Eduroam is a federation that is mainly focused on federated access to the network in all participating organization. Also in this federation we provide PKI authentication for our users, which should works everywhere in Eduroam. But experiences from our users shows that support of PKI authentication in Eduroam correctly works only at institutions in Czech Republic. We often encounter problems or complete disfunction of PKI authentication in foreign countries.

Experiences from the deployment and operation of Eduroam in METACentrum were used during designing and deploying Eduroam at Masaryk University. Experts from METACentrum also participate in meetings and provide experience during planning federation at Masaryk University.

5 Future work

The main goal is to build a uniform environment from the users point of view but not only at layer of computing and storage resources but even in authentication and authorization. Therefore we plan to spread PKI to be only one authentication system used by user and all other authentication systems will be hid from the user and conversion among that systems will be made automatically and transparently.

Usage of PKI requires that users have to have the certificate and corresponding private key at the place from which they want to use the resources. That is why we are know focusing on the mobile users who do not have certificate by themselves.

6 Conclusion

In this paper we introduced a distributed environment—Czech national grid METACentrum where we deployed PKI infrastructure to build a more secure and scalable authentication system. We described our experience and results achieved during PKI deployment, and mentioned issues caused by the virtual character of the environment. We also briefly described how the emerging area of federations could help in smooth PKI deployment in the future.

Acknowledgments

The work was supported by the CESNET Research Intent MSM6383917201 and also by the CESNET Development fund project no 065/2003.

References

- [1] R. Housley, W. Polk, W. Ford, D. Solo. “Internet X.509 Public Key Infrastructure—Certificate and Certificate Revocation List (CRL) Profile”. IETF RFC 3280. 2002.
- [2] ITU-T Recommendation X.509: Information technology—Open Systems Interconnection—The Directory: Public-key and attribute certificate frameworks, 2005. <http://www.itu.int/rec/T-REC-X.509/e>
- [3] S. Tuecke, V. Welch, D. Engert, L. Pearlman, M. Thompson. “Internet X.509 Public Key Infrastructure (PKI) proxy certificate profile”. IETF RFC 3820. June 2004.
- [4] M. Myers, R. Ankney, A. Malpani, S. Galperin and C. Adams. “X.509 Internet Public Key Infrastructure Online Certificate Status Protocol—OCSP”. IETF RFC 2560. June 1999.
- [5] B. C. Neuman, T. Ts'o. “Kerberos: An Authentication Service for Computer Networks”. *IEEE Communications*, Volume 32, Issue 9, September 1994.
- [6] D. Kouřil, L. Matyska, M. Procházka. “Improving Security in Grids Using the Smart Card Technology”. In *Proceedings of the IEEE/ACM International Conference on Grid Computing (Grid 2006)*. IEEE Computer Society, 2006.
- [7] L. Zhu, B. Tung. “Public Key Cryptography for Initial Authentication in Kerberos (PKINIT)”. IETF RFC 4556. June 2006.
- [8] J. Basney, M. Humphrey, V. Welch. The MyProxy Online Credential Repository. *Software: Practice and Experience*, Volume 35, Issue 9, July 2005.
- [9] D. Kouřil, L. Matyska, M. Procházka. “A Robust and Efficient Mechanism to Distribute Certificate Revocation Information Using the Grid Monitoring Architecture” Accepted for the 3rd IEEE Int'l Symp. on Security in Networks and Distrib. Systems.
- [10] S. Cantor. “Shibboleth Architecture—Protocols and Profiles”. 10 September 2005. <http://shibboleth.internet2.edu/shibboleth-documents.html>
- [11] L. Florio, K. Wierenga. “Eduroam, providing mobility for roaming users”. In *Proceedings of the EUNIS 2005 Conference*, Manchester, 2005.