Experiences with Massive PKI Deployment and Usage

Daniel Kouřil, Michal Procházka

{kouril,michalp}@ics.muni.cz

Masaryk University, Botanická 68a, 602 00 Brno, and
CESNET, Zikova 4, 160 00 Praha 6,
Czech Republic

Abstract

The Public Key Infrastructure (PKI) and X.509 certificates have been known for many years but their utilization in real-world deployment still presents many shortcomings. In this paper we present experiences gained during utilizing a large-scale distributed infrastructure serving a lot of users. We will start with a short presentation of the emerging European grid infrastructure and the EU Ithanet project, which both provide examples of real-world environments based on PKI. Then we will describe problems that arose during PKI deployment in the grid and solutions developed for their eliminations or mitigations, including our contributions. Two main views will be presented. First, we will focus on issues that can be addressed by operators of certification authorities and/or the infrastructure. In this part the problem of trustworthiness of CAs will be presented as well as accreditation procedures, which are employed to assess individual CAs and to give guidance to relaying parties as to accepting the CAs. Following that, a short survey of revocation checks mechanisms will be given and discussed.

The second part of the paper will deal with aspects related to users' usage of PKI. We will describe methods how a certificate can be obtained, especially focusing on utilization of the identity federations and their combination with standard PKI. We will also describe approaches to support the Single Single-On principle, based on short-lived or proxy X.509 certificates. Attention will be also paid to ways of protecting of private keys. We will describe issues related to usage of the smart card technology and employment of online credential repositories.

Keywords: PKI, grids, authentication, certification authorities, Single Sign-On principle.

1 Introduction

People can usually achieve a particular goal easier and faster if they collaborate with and share their knowledge between each other. The same approach can be applied in research, too. In order to be able to produce significant results, contemporary research projects usually to concentrate various experts. It is common for the activities to attract people from multiple different institutions that may even be located in different states. Such arrangement often makes it possible to concentrate a substantial knowledge and expertise of many people, which is necessary for current research to be performed not in isolated islands but as broad collaboration. Regardless how stimulating such an environment is, it also breeds brand new problems concerning the organization of collaborating people. One of the key problems is establishment of the virtual group of researchers and their mutual communication. Current research also often requires access to sophisticated devices and is resource-intensive in terms of computational, network or storage facilities. Providers of such resources do not want to grant access to anybody but want to keep control over users who are allowed to access. Also, activities sometimes require limiting access to their internal communication and document, since they may contain sensitive data that must not leak or be tempered with for any reason.

Achieving a required level of security may be easy in a closed environment that connects only several people who already know each other or even come from the same institution. However, moving one level higher, fulfilling security requirement gets more complicated in an environment linking hundreds or even thousands of users from many different countries or institutions want to collaborate.

One of the main problems to address is strong authentication in such an environment, which would provide such a level of confidence in users' identities that is acceptable for majority of the participants. An authentication system based on the *Public Key Infrastructure* (PKI) [1] uses 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 is 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). 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 to 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. Issued certificates are used to authenticate the users or services among each other, the set of entities that trust a particular CA (or multiple CAs) is called *relaying party*.

The principles described in previous paragraph apply to PKI based on the ISO X.500 schema and accommodate certificates following the X.509 standard [2]. There are also other mechanisms to provide public key infrastructure, with Pretty Good Privacy (PGP) being the most popular among them. In this paper we will only focus on X.509 certificates and set of CAs, since it provides a higher level of assurance for operations of distributed systems.

While the current PKI technology seems to be suitable and mature enough to cover large distributed user communities, there are still organizational and technical aspects that can negatively influence the overall level of security of the system. Over past years we have participated in several projects focused on establishment and routine operation of large infrastructures, where the PKI is used as the primary authentication mechanism. Despite the projects covered completely different user communities, interestingly enough the conclusions regarding user's view on the PKI is almost identical. In this paper we describe how PKI was implemented in such large environments and summarize precautions taken to make the PKI more usable for users and system operators.

2 PKI in real-world deployments

The PKI features fit 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 play an important role in the security of the larger systems. In this section we describe two environments, where PKI was selected as the authentication mechanism.

2.1 PKI in grids

Grid systems are an emerging concept supporting collaboration and resource sharing. A grid system allows its users to tie together various types of resources that often span multiple institutes or even countries. Most current systems provide facilities to perform high-performance computations or handle large amount of data. One of the most famous contemporary examples is the grid infrastructure built to process data generated by the Large Hydron Collider at CERN.

Grid systems not only provide the infrastructure to access the resource but also introduce other basic services provided as added value compared to ad-hoc systems that just couple the resources. One of such additional generic services is security provisioning. Grid users can easily establish secure communication between each other using the security function embedded in the basic grid middleware. Utilizing these services makes it easy for users to start their collaboration without having to bother with technical aspects of security mechanisms.

Unlike other distributed systems (e.g., peer-to-peer architectures), grid systems have always aimed at providing a high level of security. The PKI was natural choice, since the PKI features fit the requirements on building a robust environment with reliable authentication mechanism for connected users and services. Various approaches to building a grid system have appeared in the past, but majority of them based security on PKI. Several weaknesses have been spotted during the years of grid systems development, which also lead to several improvements and new approaches that increased the overall level of PKI-based systems.

2.2 PKI in the ITHANET project

We also participated in the EU ITHANET project, whose goal was to build an international network for thalassaemia research comprising Mediterranean and Black-Sea countries.

While preparing the collaborating infrastructure, we had to solve communication protection to prevent from leakage of sensitive data about patients. We designed and set up a solution based on virtual networks that made it possible for the participants to join the community. PKI was chosen as the authentication mechanism that is scalable enough to cover the community and also allows to be easily integrated with the VPN solution.

We established a dedicated CA for users who did not posses a digital certificate already and also provide them with client tools enabling to establish the virtual network tunnels. First results showed that users were not able to manipulate with the digital certificates and private keys properly. Further experiences revealed that the PKI technology and principles behind the digital certificates are too complex for user without any computing background. The conclusion was to make the security as transparent as possible for the end-users, since that is the only way how to retain a sufficient level of security.

3 Operating a PKI

Experiences from the grid systems suggest that it is possible to build a generalized PKI suitable for a wide range of applications. Being a provider of very basic middleware services, a grid infrastructure is not tied with any particular application. Since PKI is provided as part of the middleware layer, it also independent on applications being operated on the grid. Such an arrangement is different from the majority of current systems where the infrastructure provider and application provider are the same entities. For example, if contemporary electronic bank systems allow bank clients to authenticate using digital certificates, they require the people to obtain a certificate issued by the PKI managed by the bank. A similar situation can be seen in other areas, such as access to internal information systems at universities or corporations. Using a single PKI for every single application is clearly not a wise option, despite there can be several reasons for

a user to posses multiple digital identities. On the other hand though, establishing a trusted and properly operated PKI is a really difficult problem requiring a lot of resources. It is therefore useful if an established PKI is open for other applications that are willing to accept the rules of the PKI.

In this section we describe several issues concerning operating a trusted PKI. We also depict solutions and approaches that are used to address or mitigate the problems.

3.1 Trustworthiness of CAs

PKI in grid and other general infrastructures supposes that a user is in possession of a single key-pair and corresponding certificate, which is used to access many different applications. Such an arrangement makes the life of the user easier and more secure as well since there is only one private key to secure. In order to provide users with such general certificates, it is necessary to establish a CA that is willing and capable to issue them. The CA is a corner-stone in the generalized PKI, since trust in it determines the overall trustworthiness of the PKI-based system.

From the users' viewpoint the most important task of a CA is to ensure a sufficient level of authentication that each applicant must achieve during obtaining a certificate. Currently there are many tools enabling fast and easy establishment of technical backgrounds for a CA. However, establishment of a CA is principally an administrative problem since a lot various operations aspects must be covered to ensure the CA works in a well-defined manner. Every of aspects plays an important role and must not be neglected in the design of the CA. The CA operator must also guarantee that it is ready to provide long-term support on daily basis, to make sure that every certificate issued will be maintained properly throughout its lifetime. For example, if a need appears to revoke a particular certificate, the CA must process the request, revoke the certificate and update its revocation information. These steps must be done immediately whenever the need arises. Similarly, resource providers and infrastructure operators sometimes ask the CA to provide information about a particular certificate owner, e.g., during resolution a security incident in which the user is involved and the CA provider must be ready to hand out the information to them.

A large-scale distributed environment usually comprises a lot of CA, each one being operated by a different institution. With the high number of the CAs it is hard for the relaying parties to decide if a particular CA is trusted enough and if it fulfils the requirements of the end-users. Seriously established CA provides documentation that is available to the relaying parties, in which they describe the procedures taken to operations of the CA. Regardless how useful these documents are, they are hardly studied by the end-users who do not have the sufficient expertise or time to study them in detail.

A similar situation became evident in the early days of grid systems, since they engaged lot of CAs, making the orientation among them difficult for an ordinary user. To ease the life to relying parties the International Grid Trust Federation (IGTF) [3] has been established, which conducts accreditation of CAs based on policy documents the CAs submit for review. A list of IGTF-accredited CAs is made publicly available for all relying parties, along with other additional information, such as location of the CA policies, revocation information contact points, and so on. Having a single repository of trusted CAs is very convenient for the end-users, since they can easily install all CAs that were accredited without having to examine the individual CAs. Trusted anchors (i.e., the CA root certificates) make it possible for the users to establish mutual communication between each other without having to pass a difficult procedure of establishment a trusted relationship.

The IGTF is composed of representatives of CA managers, identity providers, and large relying parties. The IGTF maintains several authentication profiles, which specify the requirements on establishment and operation of a CA. Currently, there are profiles available for classic CAs, CAs issuing short-lived certificates, and CAs linked to existing systems for user management. The IGTF also defines procedures

for accreditation based on the profiles, which is performed by the IGTF members. Currently there are over 80 CAs accredited by the IGTF from the whole world.

3.2 Users' identification

In the standard PKI world, a certificate owner is identified by the name of the CA that issued the certificate (i.e. issuer name) and the subject name of the certificate. The IGTF makes the identification mechanism simpler by introducing a uniform name space for the subject names for all accredited CAs. As part of the accreditation procedure the applying CA must define one or multiple prefixes that will be used to build all the subject names of the certificates it is issuing. During the accreditation phase the IGTF verifies that the prefixes are not allocated to other CA and prefixes are reserved after a CA has been accredited. In such an arrangement subject names of different CAs cannot clash.

The specification of prefixes assigned is also available from the IGTF repository as a *signing policy* specification. The policy is checked by the relying parties whenever a certificate is being verified. The checks make it possible for the relying parties to verify that the CA really complies with its obligations. Using this name space policy it is possible to identify a certificate bearer only based on the subject name, since the issuing CA is determined by the prefix used. This naming schema makes life easier for e.g. administrators maintaining access control policies, since a single line with the subject name is sufficient to be used in configuration files.

3.3 Revocation checks

While verifying a digital certificate it is inevitable to also consult the CA to check that the certificate has not been revoked. Many environments based on PKI do not pay attention to perform proper checks of revocation information. Also several applications do not support revocation checks, which decrease the security level of the system. For example, popular browser Mozilla Firefox exposes a severe bug preventing from automatic CRL updates. Neglecting checks of revocation status may lead to severe violation of security since it is the only way of detecting that a certificate has been compromised.

There are two basic mechanisms available for revocation checks. The first one utilizes a list of revocation certificates that is periodically issued by each CA. The second way enables to perform on-line checks by direct contacting the issuing CA. As the latter approach the Online Certificate Status Protocol (OCSP) [4] is widely used.

The use of CRL is simple and supported by multiple contemporary applications. CRLs are fetched by the relying at regular interval (several times a day), which introduces a delay to the distribution of revocation information. Therefore, when a certificate is being verified, the relying party may not have the most current information available, even though the CA has already published an updated CRL.

If there are demands for having access to the most current information one has to employ on-line checks using OCSP. However, before selecting the particular mechanism, it is also important to take into consideration the time required by the CA to handle revocation. Usually a CA is expected to process a revocation request within a day. Especially operating an off-line CA where the cryptographic material is stored on a disconnected machine, the staff must visit the computer room, perform the revocation, store the CRL on an external device (USB stick) and bring it back to their desktop to publish it on the CA web page or other CRL distribution point. These actions are time-intensive and the staff need not perform them immediately. The decision concerning the revocation checks should therefore always consider this time, since if a private key is compromised, the attacker can manipulate with it all the time. Consideration must be made if it is really worth doing the on-line checks, which only shorten the overall time in which the private key is exposed to an attacker.

In large-scale PKI it has turned out that there is a non-negligible overhead concerning revocation checks. Periodic downloading of CRLs can easily presents a few millions requests each day against a server of a single CA, which can also require a large amount of data to be transported. Large data transfers can produce problems for mobile clients, for whom sufficient network parameters are not available. A similar problem faced operators of the Armenian academic CA, since Armenia had a very weak international connection, which was not sufficient for the repeated large downloads. The solution was to move the CRL to a server hosted in Europe, where much better connectivity is available.

4 Using a PKI

In the introduction section we described a PKI as a general service provided to the end-users without any link to existing application providers. In this section we describe approaches that make the PKI more user-friendly. Such adaptations may also lead to a more secure environment, since users do not seek for ways to bypass existing security procedures that have been set by the system designers.

4.1 Easy access to certificate

The basic mechanism of receiving a certificate is similar for most CAs. It usually consists of two phases, with the first one being generation of a key-pair and transferring the public key to the CA. In the second phase, the CA verifies the applicant, and possibly their possession of the private key. The order of the phases depends on the particular procedures employed by the CA. The main principle for key management is sufficient security of the private key, which often means that the private key cannot get out of the user's control, not even to the CA. In order to deliver their public to the CA the users create a standard certificate request, which is sent to the CA.

During the second phase the user must prove their identity as required by the CA policy. Highly trusted CAs require the applicants to visit personally a CA contact point and present their government id card. If a CA covers a large area, it usually operates a set of *registration authorities* located closely to the user communities. The RA takes care of the authentication step and communication the result to the CA, which is still the only subject that can access the signing private key.

Both the phases play an important role in the way how users perceive the PKI. Comfortable tools are necessary to generate key-pair, store the private key safely and create the certificate request. The interaction with the RA has also been identified as crucial since users often do not understand why they are sent to the RA, which may (along with non-intuitive clients tools) lead to negative attitude users towards the PKI.

From the point of view of a PKI operator it is therefore more suitable to introduce mechanisms that ease the process of obtaining certificates. A good choice is an *on-line CA*, which is available as an ordinary www page and is able to accept certificate request and issue certificates on demand without any intervention of the CA staff. Compared to classic CA, an on-line CA is more convenient to run since the operators do not have to work with disconnected machine and most operations are performed automatically.

Also users profit from the on-line service since all the cryptographic operations are done by the browser, based on proper HTML tags in the page. In order to generate a certificate a user only has to fill out an on-line form, which is acceptable for most users.

On the other hand, on-line CAs must take steps to secure the signing key properly, since the service is available on-line and therefore exposed to attacks from the network. The IGTF requires that an on-line CA stores the private key in a hardware device (HSM), which ensures that the key cannot be extracted even if the CA machine gets hacked. Currently, there are several commercial or open-source applications suitable to build an on-line CA.

In order to ease the second phase described, i.e., the identity vetting procedure, it is possible to make use a mechanism that links the local user management systems with the RA agenda. Such an arrangement can be achieved using the *identity federation* model.

An identity federation is an infrastructure connecting user management systems from different institutions to provide standardized access to information about users maintained by their systems. Federations provide a virtual bus layer to which systems for user management and end applications can connect and share authentication and authorization data. Every organization participating in a federation manages its users by a local user management system. An Identity Provider (IdP) service is built on the top of each local user management system, providing a standardized interface to access authentication information and other attributes about the users. Any party in the federation can get this information by calling the IdP service using 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 users are allowed to use a service, they have to present a set of attributes issued by their home IdP. These attributes are provided to users or to a service working on their behalf upon proper authentication of the user with the IdP.

An on-line can be operated as a standard SP in this model, leveraging the existing authentication methods and additional attributes. CESNET, the Czech NREN operator, is going to provide such a service on the top of Czech academic identity federation eduid.cz [5].

4.2 Single Sign-On

For each large system it is important to provide a single sign-on (SSO) mechanism, which makes the life of users' easier yet retaining a sufficient level of security.

The grid environment introduced a special type of public-key certificates - the proxy certificate [6]. 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 a 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.

The second possibility to provide a SSO mechanism in the PKI world is to use standard X.509 certificates with limited lifetime. Such certificates can be used in similar way how proxy certificates are. There are tools that enable to integrate retrieval of short-lived certificates from an on-line CA as part of the standard desktop logon process. In that way, the use of such certificates can be entirely transparent for the users, which is much more comfortable for ordinary users.

4.3 Private key protection

Deployment of the PKI based methods in large scale multi user Grid environments reveals drawbacks that are not easily visible in small closed installations. One of the most important factors with direct influence on the overall security of any PKI based environment is secure management of private keys. Too many users see their private key as just another file they can freely copy and distribute among machines. The files containing private keys are encrypted with a passphrase but the users often select too weak passwords that can be broken using a brute-force or directory attack. Also the file system access protection does not often provide sufficient level of security. Private keys stored is such files can be captured by a malicious

administrator or sometimes even ordinary users and can be further misused to perform an attack or unauthorized access to private data. The private key owner may not even notice the key compromise for very long time. On the other hand, offloading a private key to a disconnected computer makes the private key unusable. And even keeping it on a personal machine only usually leads to complications during the authentication process, making life with certificates difficult.

Current efforts to address these *private key hygiene* issues focus on removal of the long-term private keys from the user's desktop. One possibility is to use a specialized credential store service – *online credential repositories* – which maintains the long-term private keys and only provides access to short-term credentials (proxy certificates) derived from these long-lived ones. For instance, the MyProxy [7] service is very widely used for such a purpose. A MyProxy 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. MyProxy servers are used in multiple scenarios ranging from access to grid portals to support of long-running jobs.

The other option is to use a specialized hardware device which is able to maintain the private key and perform basic operations - hardware token (or smart card). Such a device ensures that the private key never leaves the token and prevents the key to be ever exposed to unauthorized users. The smart card technology allows to mount *two-factor authentication* where the user must prove to the end system something she has (i.e. the smart card) and also something she knows (i.e. the smart card password opening access to the private keys). These two factors must be presented at the same time. The biggest challenge tied with the use of smart cards lies in the user support, mainly if the PKI (and smart cards) are operated by providers different from the users‘ home institutes. In this scenario, the users‘ local support staff do not have either experience or mandate to solve problems caused by third-party devices, while the staff at the PKI provider do not know the users‘ local environment.

5 Acknowledgement

The work has been supported by the research intent “Optical Network of National Research and Its New Applications” (MSM 6383917201) of the Ministry of Education of the Czech Republic.

6 Conclusions

In this paper we presented several views and experiences gained concerning a design and operation of a large-scale PKI. We especially focused on real-world examples experienced by the grid community in Europe.

References

[1] Housley, R., and Polk, W., and Ford, W., and Solo, D. Internet X.509 Public Key Infrastructure - Certificate and Certificate Revocation List (CRL) Profile. IETF RFC 3280. April 2002.

[2] ITU-T Recommendation X.509. Information technology - Open Systems Interconnection - The Directory: Public-key and attribute certificate frameworks. http://www.itu.int/rec/T-REC-X.509/e. 2005

[3] Home page of IGTF. http://www.gridpma.org/

[4] Myers, M., and Ankney, R., and Malpani, A., and Galperin, S., and Adams, C. X.509 Internet Public Key Infrastructure Online Certificate Status Protocol – OCSP. IETF RFC 2560. 1999.

[5] Home page of eduid.cz. http://www.eduid.cz/

[6] Tuecke, S., and Welch, V., and Engert, D., and Pearlman, L., and Thompson, M..: Internet X.509 Public Key Infrastructure (PKI) proxy certificate profile. IETF RFC 3820. June 2004

[7] Basney, J., Humphrey, M., Welch, V. The MyProxy Online Credential Repository, Software: Practice and Experience. 2005.