Building a foundation of trust for an expanding PKI ecosystem architecture

Critical considerations when protecting the new corporate network, mobility, cloud applications, and the Internet of Things

ENTRUST
SECURING A WORLD IN MOTION

Contents

Introduction

Business applications today are increasingly dependent on the use of trusted digital credentials. Credentials are the certificates and keys that controls how users, entities, and a growing number of devices connect to systems and access critical resources and data. The Public key infrastructure (PKI) is the set of hardware, software, policies, processes, and procedures required to create, manage, distribute, use, store, and revoke digital certificates and public-keys. As the foundation that enables the use of technologies such as digital signatures and encryption across large user populations, PKIs deliver the essential elements necessary for a secure business environment and the trusted ecosystem essential for e-commerce and the growing Internet of Things (IoT). According to a recent study, IoT is the most important trend driving the deployment of applications using PKI, increasing significantly from 21% of respondents in 2015 to 47% in 2020.[1] With a growing demand for trusted digital certificates, PKIs must meet that challenge.

PKIs help establish the identity of people, devices, and services – enabling controlled access to systems and resources, protection of data, and accountability in transactions. With evolving business models becoming more and more dependent on electronic interaction requiring online authentication and compliance with stricter data security regulations, next generation business applications are becoming more reliant on PKI technology to guarantee high assurance.

As the core component of a PKI responsible for establishing a hierarchical chain of trust, certificate authorities (CAs) issue the digital credentials used to certify the identity of users. CAs underpin the security of a PKI and the services they support, and therefore can be the focus of sophisticated targeted attacks. Casualties of these attacks have included CAs such as DigiNotar which were put out of business after compromised and attacks where unauthorized certificate issued by an intermediate CA, was used to create bogus end-entity certificates that ultimately affected numerous Internet websites.[2] In order to mitigate the risk of attacks against CAs, physical and logical controls as well as hardening mechanisms have become necessary to ensure the integrity of a PKI.

This paper examines the security risks of typical enterprise and government PKIs. The paper describes how, as more high-value business applications and growing number of devices increasingly depend on trusted digital credentials, higher assurance

[1] 2020 Global PKI and IoT Trends Study, Ponemon Institute.
[2] Google Chrome will banish Chinese certificate authority for breach of trust, Arstechnica, 2015. http://arstechnica.com/security/2015/04/google-chrome-will-banish -chinese-certificate-authority-for-breach-of-trust/

solutions are now necessary to reinforce security and mitigate growing risks. Analyzing such aspects as the number of certificates being used by individuals and devices, the importance and value of the applications they support, and whether these applications are subject to higher levels of scrutiny due to government or industry regulatory compliance, are some of the critical factors to consider in assessing whether a PKI can still meet the demands of an evolving ecosystem. With the backdrop of well-known attacks on sensitive data, it has become increasingly critical for organizations architecting PKIs to implement strong encryption and digital signatures. Options that should be considered include using robust algorithms and longer key lengths, or newer approved technologies such as elliptic curve cryptography (ECC) for mobile devices with computational limitations. With these, organizations should step back and look at their entire infrastructure to determine the appropriate assurance level for their PKI based on the critical systems they support today and those that they will support in the future.

Why is your PKI more important than ever?

PKIs provide a framework that enables cryptographic data security technologies such as digital certificates and signatures to be effectively deployed on a mass scale. As a foundational element of many trusted systems, PKIs are already present in more places than one would generally think. PKIs support identity management services within and across networks, and underpin online authentication inherent in secure socket layer (SSL) and transport layer security (TLS) for protecting internet traffic, as well as document and transaction signing, application code signing, and time stamping. PKIs support solutions for desktop login, citizen identification, mass transit, mobile banking, and are critically important for device credentialing in the IoT. Device credentialing is becoming increasingly important to impart identities to growing numbers of cloud-based and internet-connected devices that run the gamut from smart phones to medical equipment. In the next two years, an average of 41% of IoT devices in use will rely primarily on digital certificates for identification and authentication.[3]

[3] Ibid.

Software and firmware that runs on IoT devices also need digital certificates to affirm its integrity and protect from malware. With an estimate of 20 billion IoT devices now deployed, the number of digital certificates is expected to explode in the coming years, and demand for PKIs to grow rapidly.

Using the principles of asymmetric and symmetric cryptography, PKIs facilitate the establishment of a secure exchange of data between users and devices – ensuring authenticity, confidentiality, and integrity of transactions. Users (also known as "Subscribers" in PKI parlance) can be individual end users, web servers, embedded systems, connected devices, or programs/applications that are executing business processes – for simplicity in this paper we refer to these generically as "users". Asymmetric cryptography provides the users, devices or services within an ecosystem with a key pair composed of a public and a private key component. A public key is available to anyone in the group for encryption or for verification of a digital signature. The private key on the other hand, must be kept secret and is only used by the entity to which it belongs, typically for tasks such as decryption or for the creation of digital signatures.

In order to bind public keys with their associated user (owner of the private key), PKIs use digital certificates. Digital certificates are the credentials that facilitate the verification of identities between users in a transaction. Much like a passport certifies one's identity as a citizen of a country, the digital certificate establishes the identity of users within the ecosystem. Because digital certificates are used to identify the users to whom encrypted data is sent, or to verify the identity of the signer of information, protecting the authenticity and integrity of the certificate is imperative in order to maintain the trustworthiness of the system.

With evolving business models becoming more and more dependent on electronic transactions and digital documents, and with more Internet-aware devices connected to corporate networks, the role of a PKI is no longer limited to isolated systems such as secure email, smart cards for physical access or encrypted web traffic. PKIs today are expected to support larger number of applications, users and devices across complex ecosystems. And with stricter government and industry data security regulations, mainstream operating systems and business applications are becoming more reliant than ever on an organizational PKI to guarantee trust.

PKIs today are expected to support larger number of applications, users and devices across complex ecosystems – a task that they were not originally designed to do.

The CA and the changing security ecosystem

CAs manage the lifecycle of all digital credentials within a PKI, including their issuance, renewal, and revocation. The digital credential, often referred to as an X.509 certificate[4], validates the ownership of a public key by the named subject of the certificate. When receiving digitally signed information, the certificate enables users (signers ("Subscribers") and verifiers ("Relying Parties")) to validate that the private key used to create the signature indeed belongs to the person or entity that created the signature. The CA is the third party which both the owner of the certificate and the party using the certificate trusts. Because of this critical dependency, CAs underpin the security of not only the PKI, but of all transactions and exchanges that are protected by the certificates that they issue.

Medium sized and large organizations and government agencies often deploy their own CAs and issue certificates for their own use. Others may use managed/hosted CA services provided by a "service provider" or may use one of the many "Commercial CAs" that provide certificates for use on the Internet. Organizations providing managed/hosted CA services or "Commercial CAs" typically charge a fee for the issuance of certificates. Managed/hosted services can be accessed by multiple private organizations with "Commercial CAs" also being accessible to the general public. Both managed/hosted CA services and "Commercial CAs" therefore serve the purpose of establishing trust between all the parties in a transaction making use of certificates, effectively acting as a trusted third party.

Applications most often using PKI credentials include SSL/TLS for public facing websites and VPN, enterprise user authentication and device authentication, cloud-based applications and e-mail security, and mobile authentication. With a growing dependence on the PKI, it is imperative for organizations to ensure that the certificates being issued to support these applications are authentic. Underpinning the security of certificates are the private keys of the CAs that are used to sign them.

With increasing focus on cloud-based applications, device credentialing and authentication, and code signing, the role of the CA is becoming even more critical. Inability of PKIs to support new applications is a significant issue faced by many organizations. On average,

[4] X.509 is an International Telecommunications Union (ITU) standard that defines the format of digital certificates used by PKIs.

companies today are using their PKI to support seven different applications. And increasing effort to secure those PKIs is needed as part of creating a “foundation of trust.”In fact, the IoT continues to be the strongest and fastest growing force affecting PKI planning and evolution.[5]

Whether originating from private deployments, closed hosted services, or publicly available ones, certificates issued by the CA must be trusted by the Relying Parties and users who are proving their identity. Compromise can lead to fraudulent transactions, counterfeit applications, codes, or devices with identities that may be difficult or impossible to distinguish from legitimate ones.

Externally vs. internally hosted CAs

The CA manages the population of digital certificates within the community of users that it serves and also uses certificates to perform its own certificate issuance operations. The CA issues user certificates by signing them with its own private key and presenting its own public key and certificate to enable those user certificates to be validated. To protect certificates from forgery it is imperative that the CA signing key be secured and that the CA signing certificate itself be authentic.

In order to properly issue digital certificates in a scalable and trustworthy manner, organizations generally rely on a hierarchical chain of trust that includes a “Root” CA and “Subordinate” CAs. The chain of trust up and down the hierarchy has its foundation in the Root CA that provides the anchor or highest level of trust in the system. This approach enables the Root CA to distribute its certificate issuance load across the subordinate CAs to improve capacity, manageability, and resiliency across the system. CAs issuing certificates to Subscribers are known as “Issuing CAs”.

CAs manage the lifecycle of all digital credentials within a PKI, including their issuance, renewal, and revocation.

[5] Global PKI and IoT Trend Study, Ponemon Institute, 2020.

The use of subordinate CAs also allows for application segmentation, regional separation and the support of specialist functionality such as the establishment of policy authorities, known as “Policy CAs”. Using Policy CAs it is possible to segment different organizational areas or geographies or those operating in a different legal jurisdiction. Equally, it is possible to control the purposes of certificates that CAs under the Policy CA may issue providing segregation of PKI functionality. Under each “Policy CA” a number of “Issuing CAs” may operate.

Further, Registration Authorities may also be used to support different phases in the issuance process such as performing extensive checks on the identity of the users requesting certificates. “Registration Authority” functions may also be performed by the CA in simpler PKI deployments.

Unless subject to complex legal or jurisdictions conditions which necessitates a need to segregate operations of the PKI, organizations should opt for the simpler two tier PKI hierarchy. (Root CA and Issuing CAs.) A schematic representation illustrating how the CA function fits within the organizational PKI is shown in Figure 1.

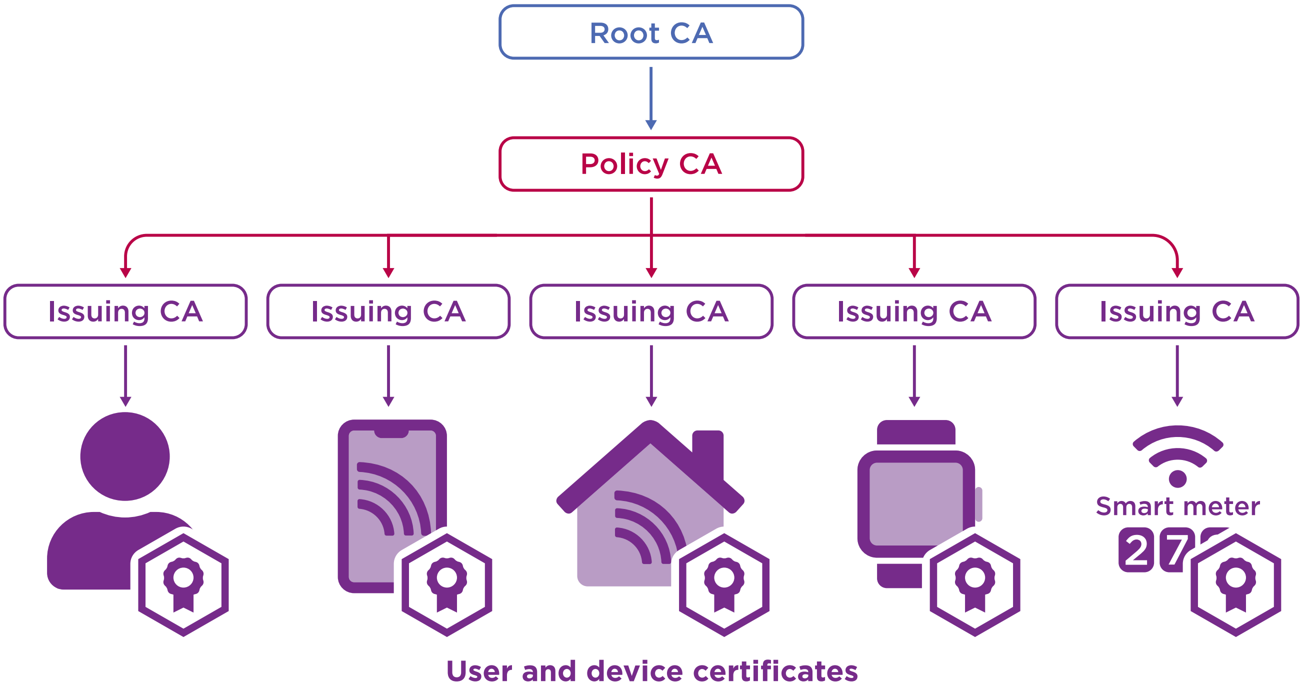

Figure 1. The role of the CA within the expanding PKI ecosystem

Once users are approved for a certificate, the issued credential binds their real world identity to a public/private key pair. The certificate itself typically includes the name of the user it was issued to, the public key component, a validity date range, and the name and signature of the issuing CA. The name and signature of the issuing CA are critically important as they are used to determine the authenticity of the certificate and therefore the trustworthiness of the identity. The user that the certificate is issued to then possesses both the certificate (which includes the public key) and the private key.

In many cases the private key is generated locally by the user, and the corresponding public key is sent to the CA with the certificate request for signing by the CA. The certificate containing the public key is then typically published in an open directory while the private key is kept secret by the user.

In addition to publishing the certificates, PKI directory services can also make certificate status information available to users – the most important aspect of which is whether any given certificate is still valid or whether it has been revoked or cancelled prior to its natural expiry date. Failure to spot that a certificate has been revoked may result in a fraudulent transaction, for example by a terminated employee, being accepted as if it were legitimate.

PKIs generally employ one of two methods to communicate the status of certificates. The first is through Certificate Revocation Lists (CRLs) which are issued periodically to online repositories associated with the certificate directories. These CRL Distribution Points (CDPs) provide a snapshot of revoked credentials at a certain point in time. The second method uses the Online Certificate Status Protocol (OCSP) to provide a dynamic capability that can deliver real-time verification of a certificate's validity. Both methods enable users to validate a certificate with the expectation that the service providing this attestation is trustworthy (the trust is inherited from the CA that either issued/digitally signed the CRL or issued the certificate that the OCSP service is using.) With increased reliance on the PKI, certificate revocation capabilities also become critically important. Recent studies have revealed troubling statistics, finding that significant number of organizations in the study (32%) are not using OCSP and/or CRLs.[6]

[6] Ibid.

Security considerations

The security of CAs is paramount in order to ensure the trustworthiness of the entire PKI. Because of the hierarchical structure of CAs, the Root CA is established as the "trust anchor". For this reason, a compromise of the Root CA private key can have severe ramifications. Replacement of the Root CA in its entirety or reissuance of the Root CA certificate containing a new public key and associated private key would require the re-issuance of certificates to all Subordinate CAs and subsequently, reissuance of all certificates to Subscribers. Since Root CA certificates are often widely published and embedded into devices and applications, the process itself would be very difficult to undertake. Robust protection of the Root CA private key is therefore essential.

To ensure the security of the Root CA, these are generally deployed off-line, where they are detached from the network, and use a hardware security module (HSM) to generate and protect the signing key. HSMs are devices designed specifically to isolate keys and signing operations from the CA software, host platform, and operating system – all of which are vulnerable to tampering and other forms of attack. HSMs also help to automate otherwise manual key control processes and procedures, and provide powerful controls to ensure correct authorization for the use of the protected root key material as well as the secure backup of the key material for recovery if necessary. HSMs can therefore help stretch the life of keys as well as allow the use of larger keys without performance compromise relative to software.

Besides the Root CAs, Subordinate CAs and other components within the PKI are typically deployed in physical or virtual servers across the organization, and should be afforded protection against potential attacks, including network based attacks, malware, malicious insiders, and any other threat that might compromise or disrupt their operation. As with Root CAs, the use of HSMs should always be considered. And since CAs are generally in operation over long periods of time, consideration should also be paid to the consequences that hardware and operating system updates and modernization can have on the system to ensure that changes made to improve operations do not undermine the security of the original design. When the CA certificate is due to expire, a carefully planned update procedure is essential as precipitous CA key and certificate renewals are one of the major causes of PKI failures.

A careful planned update procedure is essential as precipitous CA key and certificate renewals are one of the major causes of PKI failures.

Within the certificate revocation mechanism of the organizational PKI, there is also a high value key that signs the OCSP responses, usually issued from the Subordinate CA that the OCSP Responder is providing services for. OCSP is also an area where HSMs should be considered for use. As well as the security aspect, HSMs can also provide important performance benefits, and OCSP is one of the few areas in PKI where performance is vital.

Other areas of the PKI that hold high value private keys include the Registration Authorities (RAs). Examples of Registration Authorities include Network Device Enrollment Services (NDES) used to issue credentials to devices authorized to become part of a network. Increasingly, bring your own device (BYOD) schemes in the enterprise and the IoT are also expanding these services and the need for high assurance.

Ownership within your organization

As organizations increasingly depend on their PKIs to support more critical applications, clear ownership of the organizational PKI is indispensable to guarantee a foundation of trust across all services that depend on it. A coherent approach that applies standard of due care and best practices across the organization should be taken. Appropriate resources and skills should be employed to assess the demands place on the PKI and the threats it may be exposed to in a changing environment. A focus on “start with security” will pay dividends here.

A typical threat assessment of an organizational PKI should start with the issues that impact the security posture of any critical IT system. This should underscore the level of exposure that servers used to host CA software and associated repositories of certificate status information might have to internal and external unauthorized entities. The assessment should focus on:

- Threats to PKI and impacts of potential security breaches
- Access controls and authentication mechanisms
- Open ports, connections, and default syntax policies
- Firewalling and compartmentalization mechanisms
- Security maintenance and patching practices
- New code reviews and implementation procedures
- Virus and malware prevention and detection
- Audit and compliance requirements processes
- Forensic analysis to prove the integrity of the PKI

An analysis of existing certificate policies, processes, and procedures should also be part of a comprehensive review. Policies, processes, and procedures need to be periodically examined and adjusted to ensure their relevance and effectiveness. If HSMs are being used, the exercise should also include an assessment of whether they are being managed and administered appropriately.

Maintenance is a particularly important issue since CA systems may become so isolated from core IT systems that they are excluded from regular upkeep. There is also the risk that the CA might not be kept up to date with the latest security patches for fear that such activity might introduce possible problems. Although Subordinate CAs and particularly Root CAs might be isolated from direct network connections, security patches should still be kept up-to-date, always ensuring that the source of these updates is fully authenticated and the code closely scrutinized so that they do not act as conduits for possible malware.

A careful planned update procedure is essential as precipitous CA key and certificate renewals are one of the major causes of PKI failures.

Although these attack vectors apply to any IT system, their impact is amplified by the critical role that CAs and a PKI play in an organizations' trust infrastructure and the costly repercussions of failure. In the next section we turn our attention to vulnerabilities and consequences that are specific to a CA.

Exploits and wider consequences

Depending on which area of a PKI is potentially exploited, there can be different levels of repercussions across the system. For instance, an attack on the root signing keys will impact the entire system as it will compromise the trustworthiness of any and all certificates issued by the Subordinate CAs. Therefore, the security of the Root CA private key is always the most important aspect to consider. Organizational PKIs not using HSMs to protect their private keys and not employing mechanisms to effectively revoke certificates leave themselves vulnerable to disruption with potential severe consequences.

A compromise of a subordinate CA's signing keys may have more limited impact, but that depends on the size and nature of the community to which it issues certificates; the more pressing problem from an infrastructure perspective is to ensure that the system as a whole can be trusted.

Potential exploits to CAs can come through network-based attacks and can generally manifest themselves in three scenarios. First, malware can compromise CA software and generate fraudulent requests or approvals for what would appear to be legitimate certificates. Second, malicious code or insiders can attempt to steal private signing keys that would enable the certificate approval process and allow bogus certificates appearing to be legitimate to be issued on demand. Third, signing keys can be substituted with rogue keys that are known to the attacker rather than stolen. Any such attack scenarios clearly have far reaching impact on the organization, shattering the trust of the entire system.

Attacks where malware takes control of the CA software may take advantage of inherent weaknesses in software – vulnerabilities that are often associated with the way the software is configured. A threat and vulnerability assessment performed by a qualified independent assessor can identify weak points in the infrastructure and the configuration aspects of the CA to strengthen the security posture of the organizational PKI.

To protect against these threats requires more than just a focus on protecting the CA signing keys while they are in use. It requires an appraisal of the entire key and certificate management process and the various operational tasks that impact that lifecycle. Over the last decade a number of “standards of due care” have become widely established for key management. These are covered later in this paper, and should be followed to safeguard the generation, use, and exchange of keys between systems for backup and recovery purposes – subject to administrative mechanisms that enforce separation of duties.

While protecting the signing keys used by the CA is an important security aspect, it is only one part of the security spectrum that should be considered when evaluating your PKI. The National Institute of Standards and Technology (NIST) has in the past highlighted how CAs have increasingly become targets for sophisticated cyberattacks due to the high reward potential for an attacker. Network-based attack scenarios such as those where an attacker seizes control of the servers running the CA software have led to serious consequences for credential providers. For this reason, it is critical that subordinate CAs implement appropriate levels of protection including robust controls over the certificate issuance, status reporting, and revocation processes.

Assessing your PKI dependence

As organizations grow and more high-value business applications become dependent on the PKI, it is important to step back and assess if the PKI can adequately support growing security demands, just as a building's foundations have to be inspected if additional floors are to be built. Regardless of whether an organization owns its own CA, pays for an outsourced service, or uses a publicly available one, the value of their PKI depends on the level of trust that it delivers. PKIs originally deployed to support low value operations and low volume certificate issuance/management might no longer be capable of supporting more sensitive applications which may be subject to higher level of scrutiny from a security compliance perspective. Similarly certificate volumes may have increased over time and now far exceed original planning assumptions, or certificate policies, including items such as algorithm choice and key length, may no longer be appropriate.

Why should you be concerned over the strength of your PKI?

Software-only systems (i.e., systems that do not employ dedicated hardware such as HSMs for cryptographic operations) can be inherently vulnerable to many of the threats outlined earlier and for that reason, best practices should be followed when deploying PKIs and IoT to strengthen them against attacks. From the "2020 PKI Global Trend Study", in the next two years, an average of 41% of IoT devices in use will rely primarily on digital certificates for identification and authentication. In addition, only 39% of organizations in the study indicated they use HSMs to secure their PKIs, and 32% have no certificate revocation mechanism in place.[7]

Consider a scenario where a high technology medical equipment company injects digital certificates into their equipment at manufacturing to enable them to be authenticated as a legitimate product once they are delivered to hospitals and put into use. Such equipment contains encrypted private keys that enable them to certify their authenticity to gain access to an ecosystem as a legitimate component. If a CA issuing the keys and certificates

[7] Global PKI and IoT Trend Study, Ponemon Institute, 2020.

for these is compromised and bogus equipment is manufactured, the authenticity of legitimate equipment can no longer be trusted. Illegitimate equipment posing as legitimate could gain access to critical systems and be used as a vector to introduce malware and to stage cyber-attacks against the systems it connects into. The repercussions of such attack scenarios could be massive considering the operational cost of recalling and reissuing certificates to equipment, not to mention the reputational impact on the company.

Software publishers who depend on digital signatures to attest to the authenticity of their product would see their business plummet if their customers could not trust that the software did indeed come from a legitimate source. While software verification processes are usually carried out in the background, consumers have an expectation that they can trust what they load into their systems, particularly when these are updates they are paying for from reputable vendors.

A compromised CA would be able to issue fraudulent certificates which would allow unsuspecting customers to install what appears to be legitimately signed software from a bogus source. Such a scenario would potentially infect the customer's platform, affecting the developer's reputation, and could even put them out of business.

There are several "real-world" examples of code signing certificates and private keys being acquired and used by unauthorized 3rd parties.[8] Even when the code signing private key is stored in a HSM, it is still critical that organizations implement "defense-in-depth" security practices to ensure that access to the code signing key is carefully controlled. If this is not done, the consequences can be embarrassing for those concerned.[9]

Private CAs that make it their business to sell trust, so that applications using the PKI services can be considered trustworthy, should be particularly sensitive to the security of their CAs. Such services supporting online invoicing and document notarization can and have lost their entire business due to compromises of the CA. Scenarios such as the ones presented above highlight why a system-wide approach to security must be deployed and help put in perspective the cost of an HSM in light of the potential remediation costs. Code signing is the last step in the development process, creating the actual files that will be delivered to the user or devices. How that process is structured is central to its security and effectiveness.

[8] The Scary and Terrible Code Signing Problem You Don't Know You Have, SANS Institute, 2020, https://www.sans.org/reading-room/whitepapers/critical/scary-terrible-code-signing-problem-you-36382

[9] Malicious Code-Signing Becomes Dark-Web Cottage Industry, InfoSecurity, 2015, http://www.infosecurity-magazine.com/news/malicious-signing-dark-web-cottage/

Factors to consider when determining your requirements

The volume of certificates issued by a CA, the number of applications they support, their value, and whether these applications are subject to higher levels of scrutiny due to government or industry regulatory compliance issues, are some of the critical factors to consider in assessing whether a PKI can still meet the demands of an evolving organization. Other aspects that can also drive security requirements include:

- Geography and topology including partners and external parties
- Approval processes including supervision and accountability
- Auditing and compliance procedures
- Speed of issuance and validation, and associated latencies
- Existing cryptographic policies and legacy systems
- Available budget

For existing PKIs, technological changes and policies since the system was originally deployed must also be considered. Many organizations are migrating to longer key lengths for popular algorithms such as RSA due to computational advances, and others are considering alternatives that have matured such as ECC for uses in mobile devices where computational power is limited. Equally, organizations should have already moved away from the SHA-1 to the SHA-2 hashing algorithm due to security concerns and NIST recommendation and best practices.[10]

While the risk of an attack can never be eliminated, awareness of evolving capabilities of attackers, their motivations based on the value of your applications and the data they process, and potential for exploits will enable you to assess and prepare for possible consequences.

[10] https://csrc.nist.gov/publications/detail/sp/800-131a/rev-2/final

A stronger PKI for the next generation

Factors that organizations should consider when strengthening their general security posture, and which also help strengthen the security of the PKI to better support higher value business applications, include procedural as well as technical aspects. Procedural aspects involve policies on access controls, separation of duties, key lengths, and auditing mechanisms that help mitigate risks. Technical aspects include the manner by which the CAs are architected and configured, and how CAs and their signing keys are protected. When building a stronger PKI, enhancements must be proportional to the value of the systems and the data that it processes. To determine that balance, a comprehensive risk assessment should be the first step.

From a PKI policy and procedural perspective, the most important aspect to keep in mind is the significance of the Certificate Policy and Certificate Practice Statement (CP/CPS). The preparation of these documents, which are often legally binding, describes how certificates are handled within the organization and thus establishes how technical solutions are put in place to support them.

The CP focuses on the certificates themselves and the CPS on the CAs that issue the certificates. As part of the certificate lifecycle management process the CP requires that the key generation, key storage, backup, recovery, and distribution processes be well documented and in compliance with regulatory requirements.

Because the CPS translates certificate policies into operational procedures, it has to properly align with not only the security aspects they must address, but also the operational and legal requirements of the business. Other factors may include how the organization coordinates and brings together the individuals required for a keying ceremony of a Root or Subordinate CA and how they are selected so they represent the right parts of the organization. Key ceremonies should be fully planned, rehearsed, and enacted in line with the CP. Proper documentation of these ceremonies is vital if any part of the PKI is to be audited by a third party accreditor such as may be necessary for compliance with standards such as WebTrust[11] or tScheme.[12]

Crafting a balanced CP/CPS is therefore essential when building a strong PKI since it describes how certificates are issued and managed throughout their lifecycle. Together the CP and CPS define the level of trust that the organization can place in the certificates they issue.

[11] http://www.webtrust.org/
[12] http://www.tscheme.org/

Architecturally, an organization having to issue certificates to increasing numbers of high value applications should consider using a distributed certificate revocation capability, so that there is no single point of failure should there be an incident where the trust of a CA is compromised. When the volume of certificates increases subordinate CAs can be deployed to balance the certificate issuance load for greater reliability. If a certain number of certificates are being used for higher value operations, these might be placed under a separate PKI/CA structure that is hardened with specialized procedures and technical means that can provide greater degree of protection and backup capabilities to ensure resiliency. Segmentation of the CAs in this manner will also facilitate auditing requirements that often come with higher levels of scrutiny of high-value applications. When separate PKIs are maintained with different assurance levels, these should be afforded with matching levels of protection for their signing keys.

Together the CP and CPS define the level of trust that the organization can place in the certificates they issue. Key ceremonies should be fully planned, rehearsed, and enacted in line with the CP. Proper documentation of these ceremonies is vital if any part of the PKI is to be audited by a third party accreditor for assurance purposes.

System level CA requirements for a higher-assurance PKI

PKIs should always be designed to be trustworthy and resilient. Root CAs that anchor system trust should never be connected to the network, not even for regular maintenance purposes such as during software updates. Critical elements should always be air-gapped to protect against possible network-based attacks, and the security patches installed on these systems should be fully vetted to ensure they are authentic and not a conduit for malware.

Business continuity and disaster recovery plans should be developed and put in place; and accurate, up-to-date documentation should be kept to ensure that the system configuration can be replicated and that critical keys can be securely backed up and recovered if needed. The CA hierarchy should be designed with built-in redundancy so that there are no single points of failure and so that cryptographic processes do not represent operational bottlenecks that impact performance.

The CA should ensure that private keys are kept secret and only used by their owners or authorized CA software. Public keys should always be bound to an identity through certificates signed by the CA, and certificate status information should be protected at all times during storage and distribution. Certificate revocation information should be signed by the CA that issued the certificate or certificates that have been revoked or alternatively, by another CA if designated as authoritative for such a purpose. Organizations must keep in mind that unless configured otherwise, all systems relying on certificates will stop working if the certificate revocation sources were to become unavailable. Designing a system with redundant revocation resources is therefore critically important to ensure the resiliency and high availability of a PKI.

The procedures undertaken for the revocation of any certificate should be documented in the CP with the operational mechanisms for doing so documented in the CPS. The CP should detail who may submit revocation requests and how they may be submitted. For static CRLs, every entry should state the time at which the next scheduled CRL will be issued in order to ensure better control. And if using OCSP, responses should always be signed to prevent bogus responses from revoking valid certificates.

As a number of high profile breaches have shown, it is not sufficient to mitigate the risk of a Certificate or Registration Authority breach by relying upon the ability to revoke a certificate. A major benefit of a PKI over other technologies is the scale and inherent reliability that comes from the ability to validate credentials locally and offline. Deployment of a secure revocation infrastructure is complex, and it can be easier and more cost effective to secure the certificate issuance process than to deploy a revocation infrastructure that does not impact performance or reliability. In order to correctly revoke a certificate following a breach it is necessary to:

- Determine that a breach has occurred. This may take days, weeks or longer.
- Determine which certificates must be revoked; potentially a time-consuming reconciliation process and it may not always be possible to determine the identity of all unauthorized certificates.
- Understand and assess the impact of revoking a certificate. The impact may be low if a certificate is a single rogue credential or potentially very large if a CA certificate must be revoked, leading to the indiscriminate revocation of legitimate and unauthorized certificates. This may mean that it is necessary to delay certificate revocation until contingency plans or replacement credentials can be issued: potentially an expensive and lengthy process.

Organizations deploying a PKI should assess the risk and impact of a breach in the credential-issuing infrastructure and determine their approach to certificate revocation at different levels of the certificate hierarchy. Where an organization deploys a robust revocation solution, this must be supported by audited processes that will detect any breach and trigger the required re-issuance and revocation of credentials.

When building a strong PKI, organizations should consider the number of users, their mobility, and the manner by which they are authenticated. A high-assurance PKI will typically depend much more on sound policies and procedures than on specific technical mechanisms. Best practices should therefore focus on ensuring that the right policies and procedures are in place before jumping into technology options. Specific mechanisms and tools can certainly strengthen the security of the PKI, but these must be deployed once a solid policy foundation is established.

Cryptographic level best practices for a higher-assurance PKI

A strong PKI must have high assurance cryptographic technology at its core. This section describes ten cryptographic best practices which will specifically strengthen the security of the PKI to better support higher-value business applications. When it comes to deploying secure PKI based applications these should be regarded as effectively representing standards of due care:

- Know the origin and quality of your keys. Critical signing keys should come from a high quality entropy source using true random numbers. HSMs offer an environment where keys are generated using a certified key generation process and mechanisms tested to deliver the appropriate security.
- Know exactly where your keys are and who/what systems can access them at all times. CA private keys used to sign certificates should be maintained within a FIPS 140-2 validated[13] and protected environment and any time they leave the device, they must be encrypted only with approved algorithms and key lengths. When using HSMs you know your keys are in one location and not scattered across the software environment in potentially multiple locations with varying access restrictions. Imported keys not generated by an HSM should not be used given the serious concerns over the quality of the keys and the level of protection that they might have received during their original use prior to movement to the HSM and in transit.

[13] The Federal Information Processing Standard (FIPS) 140-2 standard defines internationally recognized design practices for cryptographic products.

- Ensure each key is only used for one purpose. Critical applications should be governed by distinct and specific key management activities. Strong control of keys ensures strong control of certificate issuance, and the capability to be able to prove this is important for auditing purposes. HSMs are designed precisely to deliver these important services that facilitate the enforcement of the organizational security policy.
- Formalize a plan to rotate, refresh, retain and destroy keys. Overworked keys are a liability and obsolete keys are an unnecessary risk. HSMs are purpose built to safeguard and securely manage sensitive keys throughout their lifecycle.
- Only use globally accepted and proven algorithms and key lengths. The use of high performance HSMs with built-in cryptographic accelerators will allow larger key sizes (e.g., 4096 bit RSA) and stronger hash functions (e.g., SHA-256) to be used; this will provide effective security long after smaller keys have become vulnerable to attack by faster processors and sophisticated cryptanalysis. With new NIST recommendations for use of stronger algorithms and longer RSA key lengths beyond 1024 bit[14], organizations should also consider the benefits of ECC supported by HSMs. As a next generation approach to public key cryptography, elliptic curves deliver robust security at shorter key lengths – improving processing efficiencies.
- Adopt independently certified products wherever possible. If you're tempted to write cryptographic software yourself, you may be jeopardizing your security. Commercially available HSMs are generally designed to FIPS 140-2 and Common Criteria[15] standards. Only FIPS 140-2 Level 3 approved HSMs using NIST validated algorithms should be employed. FIPS 140-2 Level 3 HSMs are certified tamper resistant – providing a secure environment where keys can be protected from extraction.
- Implement dual control with strong separation for all sensitive operations. With unbreakable cryptography, the attacker will go after the keys and the people that manage them. Avoid super users and single points of attack. HSMs enable organizations to enforce these policies by implementing multiparty supervision of administrative activities so no one single individual can have access to sensitive keys.

[14] Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths, NIST Special Publication 800-131A, Rev 2, March 2019.

[15] Common Criteria is an international standard for computer security certification.

- Ensure your keys are securely backed up and available to your redundant systems. CA private signing keys should be backed up, stored, and retrieved only by trusted and authenticated entities using prescribed mechanisms to enforce separation of duties. Use of a HSM can help to enforce this level of separation. Backup copies of CA private signing keys should be subject to the same or greater level of security controls as active signing keys. Recovery processes should also be in place for cryptographic keys used by any high-value application. Protecting these application keys should be considered just as important as safeguarding the CA/PKI keys.
- Control access to cryptographic functions and systems using strong authentication. Security relies on consistency; strong keys should not be able to be accessed by weak means. The use of certified and tamper resistant HSMs will improve confidence that keys are protected throughout their lifetime.
- Never allow anyone or any "open" system to come into possession of the full plaintext of a private or secret key. Theft of these keys can enable attackers' access to past and future data without detection. HSMs provide separation from the IT environment by moving them away from standard servers and into a hardened device.

A hardened, high assurance PKI provides an environment that protects security critical cryptographic keys from theft and misuse. Binding certificate issuance to identity checks and approvals using an HSM, controlling the rate of issuance of certificates, and maintaining key counters have been important lessons learned from CA security compromises. Breach identification, recovery, and contingency planning are important steps that can be taken to strengthen the security of a PKI.

Certification levels used by the FIPS standard define increasing qualitative degrees of security given to products based on algorithm testing performed, authentication methods used, and physical tamper protection mechanisms employed. In the case of Common Criteria, evaluation assurance levels are based on security and functional requirements established for the specific class of the product.

Conclusion

The security robustness of PKIs must be continually reassessed based on an organization's increasing dependence on its services, evolving security mandates, and external and internal threats. Considering that PKIs came into the mainstream in the early 2000s, many deployments are now getting their first significant lifecycle overhaul – usually as part of a comprehensive IT modernization program. There are many reasons to consider when reassessing, upgrading, or migrating your PKI. In addition to the increasing demands highlighted in this paper, the use of more robust longer keys, SHA transition, and new and more efficient algorithms should be taken into account.

Now is the time to reassess the robustness of the PKI foundations to ensure they can support the additional security demands that have evolved over the years. As user populations have grown and more sensitive applications become dependent on the digital certificates issued by CAs, the PKI can become a target for sophisticated attacks.

Just like a family's insurance needs must be re-examined regularly to ensure a sufficient and reliable safety net, organizations must stand back and look at the growing set of applications that depend on the security of the PKI to determine if it is up to the job. The continuous assessment of an organization's PKI should not just be concerned about the mitigation of the impact of a possible compromise, but rather in finding ways to reduce the risk of compromise by instituting best practice policies, procedures, and mechanisms.

When swapping or overhauling your PKI, consider the security challenges that mobile, cloud computing, and Internet connected devices will certainly bring in the near future. Think of the ways in which the CA can be hardened to meet the evolving security needs of your organization. Certificate registration, issuance, revocation, and the associated signing functions that establish the trust in these services all rely on effective long-term protection of keys. If you are also considering migrating to longer key lengths to comply with NIST recommendations, now is the best time to overhaul your PKI and future-proof your security needs with the protection that an HSM can deliver. Whether you migrate to longer RSA keys now or choose to implement new/more efficient ECC, HSMs can provide the security you need to protect high-assurance applications for years to come. Best practices such as the ones outlined in this paper provide a framework for the protection of the CAs and the hardening of the organizational PKI to meet increasing security demands.

Cloud applications and the IoT are the newest disruptors to future PKI planning. Organizations must not only tend to the digital certificate needs of today, but must also prepare for the future – a future with never-before-seen diversity and scale.

To find out more about Entrust nShield HSMs

HSMinfo@entrust.com

entrust.com/HSM

ABOUT ENTRUST CORPORATION

Entrust keeps the world moving safely by enabling trusted identities, payments and data protection. Today more than ever, people demand seamless, secure experiences, whether they're crossing borders, making a purchase, accessing e-government services or logging into corporate networks. Entrust offers an unmatched breadth of digital security and credential issuance solutions at the very heart of all these interactions. With more than 2,500 colleagues, a network of global partners, and customers in over 150 countries, it's no wonder the world's most entrusted organizations trust us.

Learn more at
entrust.com/HSM

Contact us:
HSMinfo@entrust.com