



WHITEPAPER

Vulnerability Analysis of Certificate Validation Systems

The US Department of Defense (DoD) has deployed one of the largest Public Key Infrastructure (PKI) in the world. It serves the Public Key Enabled (PKE) applications used by more than 4 million DoD employees and contractors. A critical component of this PKI is the certificate validation approach. To be successful, the chosen validation approach must scale to tens of millions of users while providing high performance, high availability in a cost effective and secure manner.

This paper addresses the security issues inherent in the three most common certificate validation approaches: Certificate Revocation Lists (CRLs), Traditional Online Certificate Status Protocol (T-OCSP), and Distributed Online Certificate Status Protocol (D-OCSP).

Traditional Validation

Each of these approaches has a different architecture. In the CRL approach the Certificate Authority (CA) periodically publishes the CRL to one or more directories from which it can be retrieved by relying parties. The basic T-OCSP architecture is shown in Figure 1. Both the CRL and T-OCSP approaches are highly centralized.

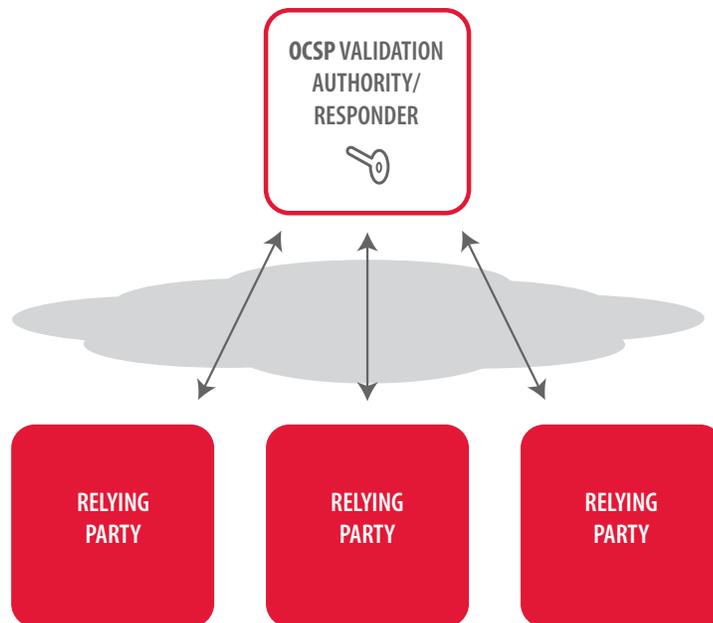


Figure 1: Traditional OCSP Architecture

In traditional OCSP validation there is one, or at most a small number of “authorities” which communicate directly with all relying applications to provide them with validation proofs.

Distributed Validation

Figure 2 shows the basic D-OCSP architecture. This approach is based on the design principle of separating the security sensitive data and trusted operations from the delivery process of providing certificate status to relying party applications. This design approach allows for the distributed architecture shown below.

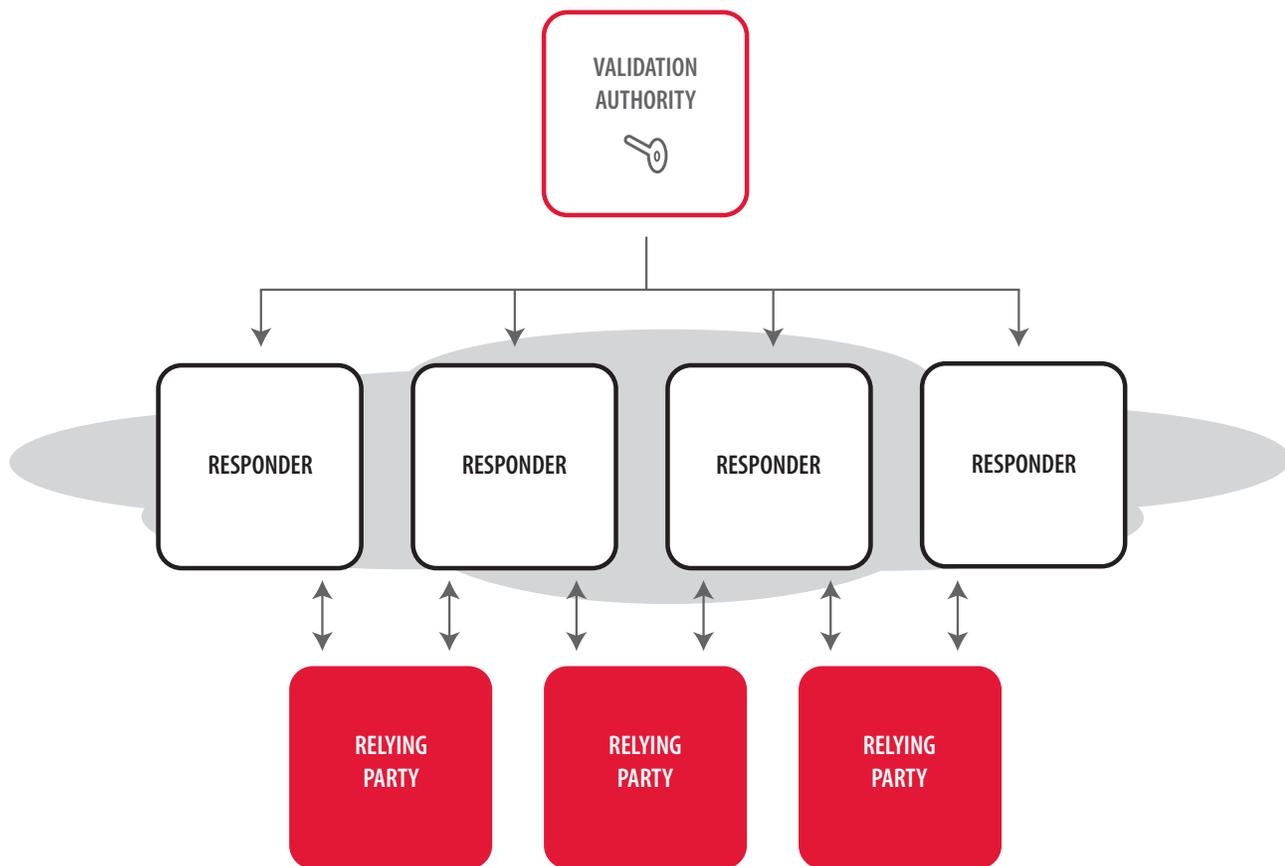


Figure 2: Distributed Validation Architecture

In distributed validation there is one authority which controls the release of validation proofs and multiple (unlimited), “non-vaulted” responders that provide these proofs to relying party applications.

The focus of this analysis is on external threats to these three certificate validation approaches.

Methodology

The methodology used here to evaluate the security of the three certificate validation systems focuses on four areas: vulnerabilities, threats, potential for exploitation and countermeasures. These are defined as follows.

- **Vulnerability:** is a weakness in the system or a point where a system is susceptible to attack. This weakness could be exploited to violate system security.
- **Threat:** is a possible danger to the system such as a person, system component or event that might result in a compromise of the secure operation of the system.
- **Likelihood of exploitation:** is result of an analysis of how likely a given vulnerability will be exploited. It considers both the environment in which the system will run and the countermeasures used to mitigate the threats.
- **Countermeasure:** is an action, device, procedure or technique for protecting the system against threats to its secure operation.

Every computer based system has security related vulnerabilities. Some of these may be easy to exploit (e.g., systems with passwords that are easy to guess) while others may require the expenditure of so much time and so many resources that it's highly unlikely that they will be exploited in the near future (e.g., brute force attack to crack a 2048 bit RSA key).

Vulnerabilities are characteristics of the system under analysis. They can exist for a variety of reasons such as poor design, poor coding, insecure operation or the difficulty and/or cost of building and implementing a secure system.

Threats are properties or elements of the environment in which the system under analysis will operate. Different environments pose different threats. A threat is an actual scenario under which a vulnerability is exploited and can arise from any number of sources: external attacks, internal attacks, accidental user error, system process error, etc. Threats may grow as vulnerabilities are discovered. However, the vulnerabilities of the system remain constant. A system may have a vulnerability which is not exploitable in the specific environment in which it runs.

In this paper we discuss four vulnerabilities common to the three types of certificate revocation systems: CRLs, T-OCSP and D-OCSP. We identify threats associated with these vulnerabilities and assess the likelihood of exploitation by an attacker. Where appropriate we also discuss potential countermeasures.

A distributed architecture has been shown to be the best defense against Denial of Service attacks.

Denial-of-service

Denial-of-service (DoS) is an action or series of actions or circumstances that prevents a system or any of its resources from functioning efficiently and reliably. Denial-of-service means that system users are unable to get the resources they need when they need them. For networked certificate validation systems, availability is particularly important since even a minor slowdown in service can have a reverberating effect for the entire network.

In recent years DoS attacks have become very fashionable.¹ Systems that are centralized and do time consuming work to complete a transaction are highly susceptible to DoS attacks. Note that a DoS attack affects everyone using that service.

Vulnerability

All three validation systems are vulnerable to DoS attacks. The original US DoD CRL approach was particularly vulnerable to DoS because of the considerable amount of time it took to download the large (5-6 Mbytes) CRL files. This is clearly demonstrated by the fact that DoD users were experiencing unacceptable denial-of-service just from ordinary use. The DoD is now using distributed OCSP for certificate validation.

T-OCSP is also highly susceptible to DoS attacks since in this approach each response is digitally signed in real time, a time consuming process. In addition, T-OCSP is a highly centralized approach because of the cost associated with standing up multiple T-OCSP authorities which must be housed and operated securely, equivalent to what is required to operate a CA.

The D-OCSP approach is the least vulnerable to DoS attacks. D-OCSP has a highly distributed architecture in which responders are cheap to deploy and do not require secure housing and operation. A distributed architecture has been shown to be the best defense against DoS attacks (Ref. footnote #1).

Threat

There is a threat of denial-of-service due to an overloading of the system with certificate status requests as a result of an unanticipated demand for service.

A second threat is that loss of connectivity to the server or servers providing the certificate validation service will result in a denial-of-service to the entire community.

An additional threat is that an attacker (or group of attackers) can swamp the certificate validation service with bogus requests thereby denying availability of the service to legitimate users.

Likelihood of Exploitation

DoS attackers would have the easiest time exploiting a CRL based validation approach simply because it takes so long to download large CRLs. In fact, the original DoD system for providing CRLs was exhibiting a DoS state just from

¹ On October 21, 2002 a “distributed denial-of-service” attack was launched against the 13 root DNS servers that provide the primary roadmap for almost all Internet traffic. The attack, the largest such attack to date, failed because of the distributed nature of the Internet root server architecture. Five of the root servers withstood the attack and remained available for legitimate Internet traffic throughout the strike.

It is virtually impossible for an external intruder to successfully attack either the CRL or D-OCSP systems since they do not allow any incoming traffic to their trusted elements.

requests by legitimate users. Similarly T-OCSP systems are also susceptible since it takes approximately 20 times longer to sign a response than it does to generate and send a request. The D-OCSP approach is the hardest to exploit for two reasons: 1) the architecture provides many more sources from which to retrieve the certificate status data, and 2) the time to generate and send a response is about the same as the time to generate and send a request. DoS attacks are extremely serious since they impact the entire user community.

Countermeasures

The most effective countermeasure for mitigating denial-of-service threats from either overload or a DoS attack is through the use of multiple redundant sources of certificate status information. The threat from loss of connectivity can be mitigated through use of redundant networks and/or the use of local sources of certificate status information.

The D-OCSP architecture is ideally suited for this countermeasure as adding additional unprotected servers is cheap and can be placed locally without the need to be housed securely. In the T-OCSP case, adding additional servers is expensive to deploy and operate, creates key management issues and complicates failover procedures. Finally, adding additional servers in the CRL case is not very effective since the time it takes to download a large (e.g., 5 MByte) file is significantly longer than the time required to request a download. Therefore a distributed denial-of-service attack will probably always be successful against a CRL system.

Intrusion

Intrusion refers to the deliberate attempt to break into a computer system. Often the intruders are more interested in the challenge of breaking into a high profile web site than in exploiting the ability to manipulate the system once in. However, in the case of certificate validation systems the intruder's motivation is probably more aligned with exploitation, making this a high risk for attack with intent to compromise. In either cast, the publicity associated with a successful intrusion can be very damaging.

Vulnerability

Any system that is connected to a network is subject to intrusion. The vulnerability of the CRL and D-OCSP validation systems to these attacks is virtually zero because neither allows incoming traffic to the server which performs trusted operations and stores security sensitive data. In contrast, the T-OCSP system does allow incoming traffic to its trusted authority, thereby making it vulnerable to intrusion attacks.

Threat

Whenever a networked computer system allows incoming traffic from the outside world there is the threat that the outside user will uncover a security flaw in the system to gain access to the sensitive data or trusted operations.² In the case of certificate validation systems the threat is that the intruder could manipulate the status of certificates, making revoked ones valid or valid ones revoked.

² On October 25, 2002 a CERT Advisory (CA-2002-29) was issued alerting users of Kerberos of a "remotely exploitable buffer overflow" that would allow an attacker to gain root privileges enabling him to run any code or make any changes to the Kerberos key distribution center. This example illustrates that even mature security software is vulnerable to intrusion attacks if connected to a network. The only way to eliminate this vulnerability is to not allow network access.

Likelihood of Exploitation

It is virtually impossible for an external intruder to successfully attack either the CRL or D-OCSP systems since they do not allow any incoming traffic to their trusted elements. The degree to which an external intruder can attack a T-OCSP system will depend upon the robustness as well as the correct deployment and operation of the intrusion prevention elements of the system.

Countermeasures

Standard intrusion countermeasures, such as the introduction of firewalls, can be used to mitigate this threat. These do not eliminate the threat in the T-OCSP case however since this approach is designed to allow external, unknown users to connect to the trusted element of the system.

Response Freshness

An important concept for any certificate validation discussion is that of response freshness. Response freshness measures how current the certificate validation status is when it reaches the relying party application. In general, the freshness clock begins when new revocation data is received by the Certification Authority (CA), which then must be added to the next CRL. In an ideal world a new CRL would be issued whenever a certificate is revoked. This approach however, is impractical for most deployments. Consider, for example, the US DoD scenario. The DoD has approximately 4.4 million users and a certificate revocation rate of about 17%. This results in a certificate being revoked, on average, every 2 minutes.³ Clearly it would be impractical to issue a new CRL every 2 minutes. Current DoD policy is to issue CRLs every 24 hours that are valid for 96 hours. The window of vulnerability therefore ranges from 24 to 96 hours, depending upon when the relying party receives its next CRL update.

Vulnerability

All three validation systems are vulnerable to a lack of freshness of the validation responses. The degree to which stale responses may be used is determined by policy where the risk of using stale data is weighed against the practicality of issuing CRLs for a large deployment.

Both the T-OCSP and D-OCSP approaches depend upon receiving revocation updates from the CA in the form of a CRL. All three systems therefore have essentially the same freshness.

It is important to note that response freshness, in either the T-OCSP or D-OCSP approach, is not determined by when the OCSP response is signed. The freshness is determined by when the CRL was signed.

Since the D-OCSP approach requires a few minutes to pre-sign all the OCSP responses there is a slight delay in making the new revocation data available to the responders. In the DoD example, with the status of 9 million certificates to be pre-generated, this difference is equivalent to delaying the issuance of the CRL by approximately 10–15 minutes. This vulnerability can be reduced by adding either faster or additional processors to speed up the pre-signing process.

³DoD certificates are valid for 3 years so the revocation rate is approximately 250,000 per year per certificate type. In the DoD case each user actually has 3 different certificates which means 3 different CRLs would need to be issued every 2 minutes.

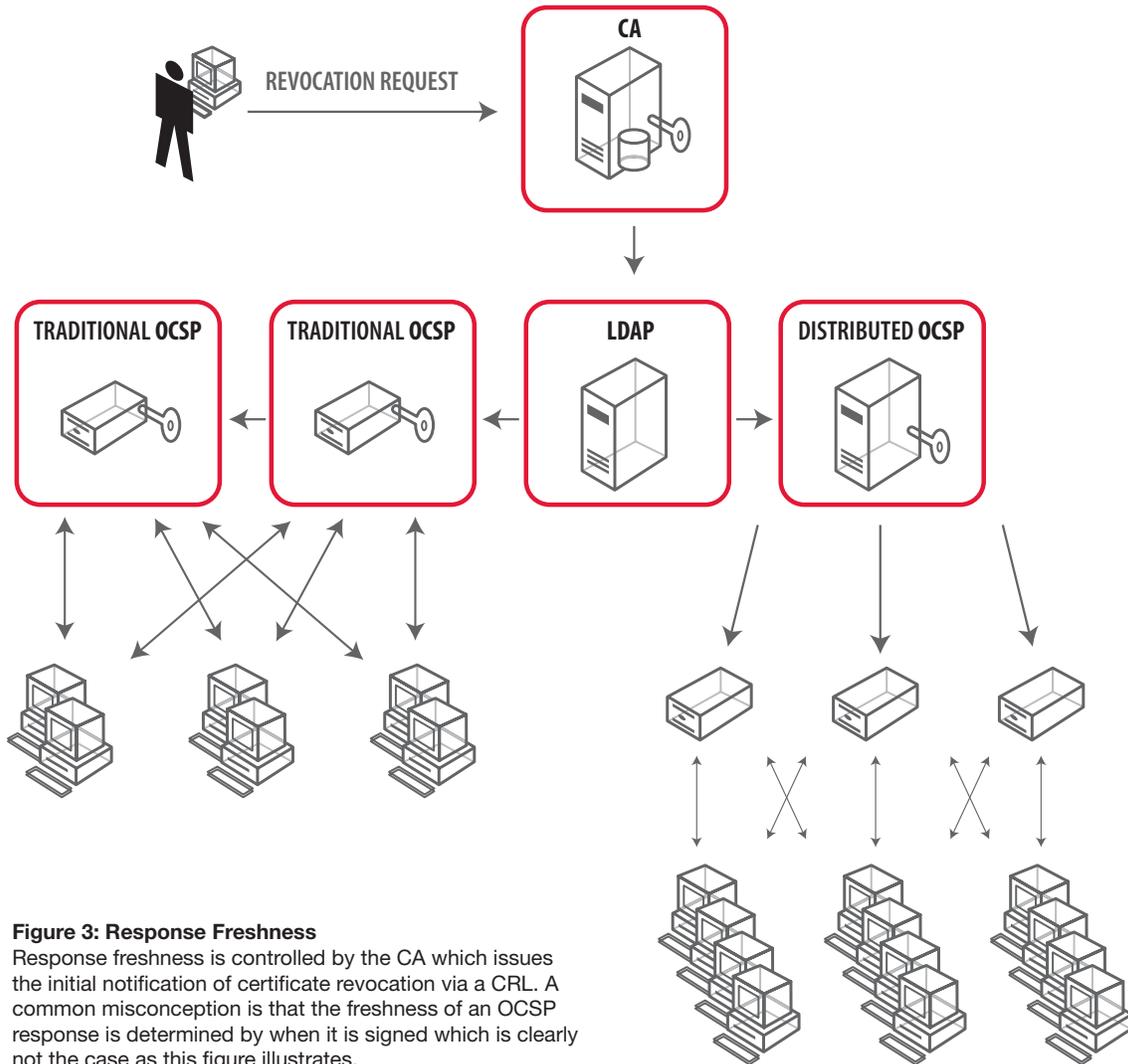


Figure 3: Response Freshness

Response freshness is controlled by the CA which issues the initial notification of certificate revocation via a CRL. A common misconception is that the freshness of an OCSP response is determined by when it is signed which is clearly not the case as this figure illustrates.

Intrusion is a vulnerability only for T-OCSP since this approach is designed to allow external, unknown users to connect to the trusted authority.

Threat

The security threat is that someone whose certificate has been revoked can still use that certificate until the revocation data is received by the relying party application.

Likelihood of Exploitation

The most significant threat is associated with certificates that are revoked because of the resignation or termination of a legitimate certificate holder. In this case the holder of the certificate can still authenticate herself to the system and therefore continue to access or use resources that rely on the validity of that certificate. Certificates that are revoked because of a potential compromise (e.g., loss of an identity token such as a national ID smart card or DoD Common Access Card) are not easily exploited since using the certificate requires access to the private key to authenticate the user.

Countermeasures

Since the threat of exploiting stale data is limited in practice to legitimate certificate holders that are terminated or resign, a procedural process of physically repossessing the identity token or other private key storage device as part of the termination process would mitigate this vulnerability.

Replay

Replay is defined as the recording of a legitimate message for the later, unauthorized resending of the message. One example is that of a funds transfer message — which would be a very profitable replay attack.

Vulnerability

In the DoD example, the certificate revocation information is updated every 24 hours and is valid for 96 hours. In such a system, all three validation approaches are susceptible to replay attacks. The vulnerability arises as follows. CRL_1 is released with the certificate in question still valid. After the release of CRL_1 , but before CRL_2 is released, the certificate in question is revoked. It is now possible for an attacker to replay the data from CRL_1 while preventing the relying party from receiving the updated information (i.e., from CRL_2 , CRL_3 , and CRL_4). In the DoD example, the period of vulnerability is 72 hours. Note that a “nonce” can be used to eliminate this vulnerability as discussed in Countermeasure section below.

Threat

The threat is that someone will record a certificate status response (CRL or OCSP) when the certificate is valid and then replays the response after the certificate has been revoked. In the DoD example the threat is limited to the period for which CRL_1 is valid.

Likelihood of Exploitation

The exploitation of this vulnerability is technically difficult and requires some explanation. The figure below shows the setup required.

To exploit this vulnerability an attacker would have to first know which certificate (or certificates) are about to be revoked. This would be difficult for a random attacker to know so the most significant replay threat is from legitimate certificate holders that are about to resign or be terminated. The setup then begins (1) when the last CRL for which the certificate in question (e.g., #123) is still valid (CRL₁ in this case). The attacker must somehow ensure a request (2) is made to the CRL or OCSP responder for the “good” certificate in question.

This will result in the relying party application submitting a status request (3) for certificate #123 to the responder from which it normally receives either a CRL or OCSP response. The attacker then “sniffs” and copies the response (4 & 5) as it is returned to the relying party application.

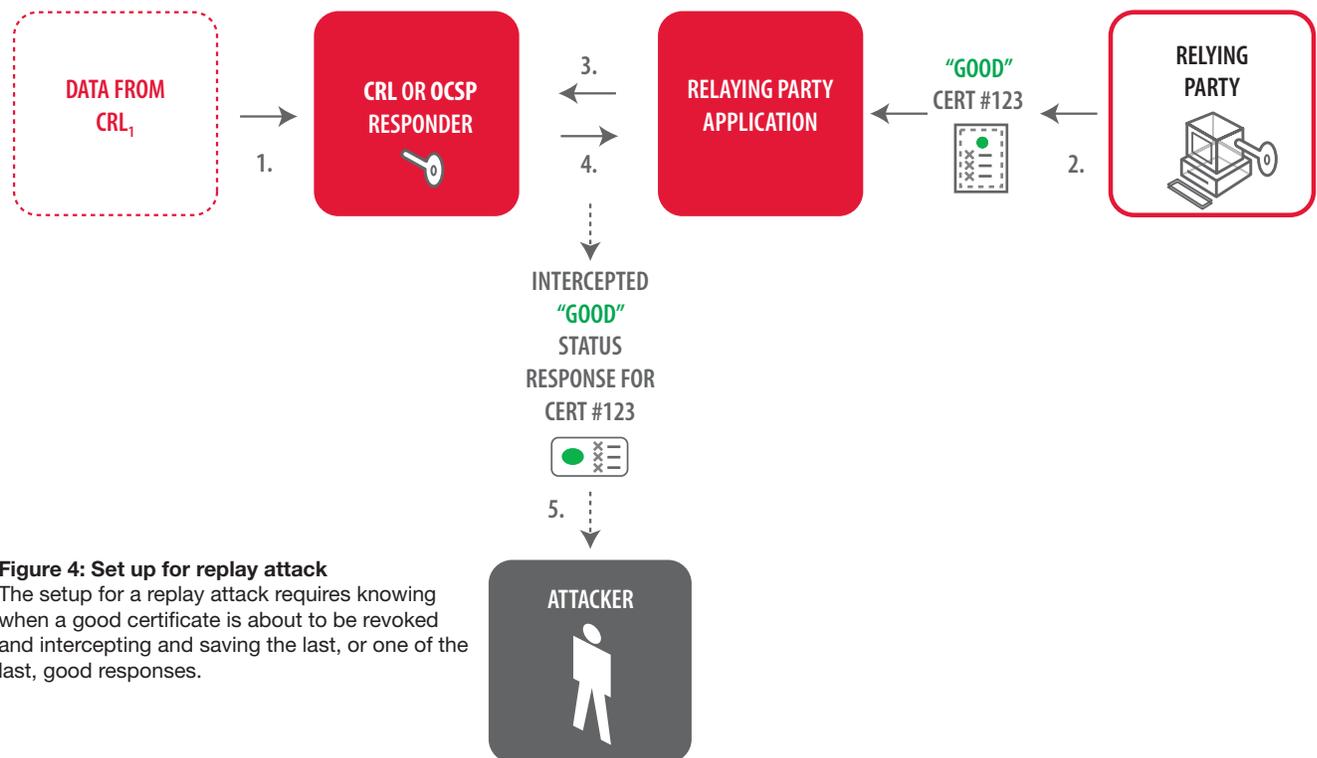


Figure 4: Set up for replay attack

The setup for a replay attack requires knowing when a good certificate is about to be revoked and intercepting and saving the last, or one of the last, good responses.

In the current scenario, execution of the replay attack can commence once CRL_2 is issued and can continue until CRL_1 expires (i.e., in the DoD example, this would be 96 hours after it was issued). Now the attacker submits a request (2) to the relying party application using the revoked or “bad” certificate. The attacker must have gained control of the associated private key (i.e., both possess the private key and be able to properly authenticate to the system it is stored on to gain access to its use) since the relying party application will require proof of possession of this key before acting upon the request. The relying party application then submits a request for the status of certificate #123 as usual. This time the attacker must intercept the revoked status response (4), prevent it from reaching the application and substitute (5) the old, saved response stating that certificate #123 is still valid. All this must be done before the saved response expires.

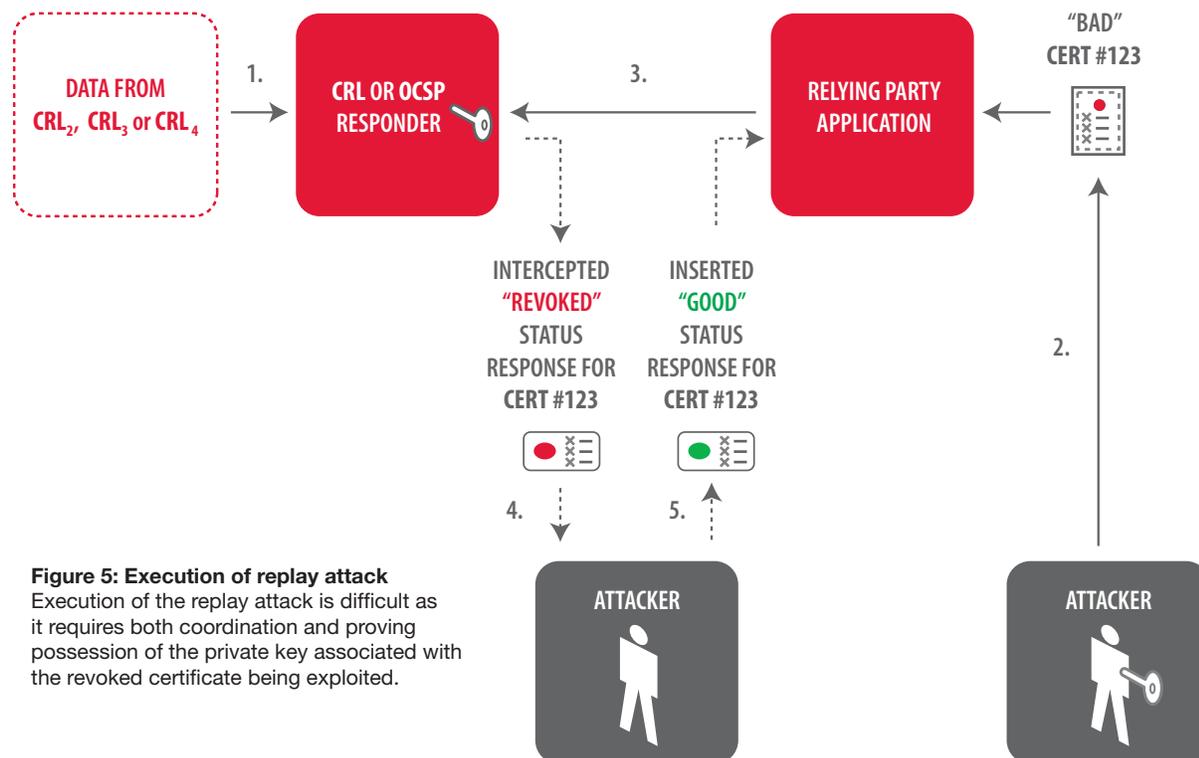


Figure 5: Execution of replay attack
Execution of the replay attack is difficult as it requires both coordination and proving possession of the private key associated with the revoked certificate being exploited.

Denial-of-service is the most critical vulnerability of the four since it is the easiest to exploit and affects the entire user population.

It is important to note that not all certificates are vulnerable to this attack, only those that have been revoked during the period of validity of a given CRL. Given how difficult it would be for an outsider attacker to gain access to a user's private key and to properly authenticate himself, the most significant replay threat is associated with certificates that are revoked because of the resignation or termination of a legitimate certificate holder who already possesses their own private key and can readily authenticate themselves.

Countermeasures

Knowing that the primary replay threat is limited to legitimate certificate holders that are terminated or resign, a procedural process of physically repossessing the identity token or other private key storage device as part of the termination process would mitigate this vulnerability.

Reducing the validity period of the CRL would also reduce the window of vulnerability. This may however result in DoS insecurity if the end user or OCSP authority is unable to get up-to-date CRLs during this shorter time period.

In the T-OCSP approach the replay threat can be virtually eliminated through the use of a “nonce”. Using this approach the relying party inserts a random number (nonce) into the request it sends to the responder. The responder then includes this random number in its signed response thus ensuring that the response has not been replayed.⁴ This approach however increases the denial-of-service vulnerability and eliminates the use of cached responses which are critical for systems that need to scale to large numbers of users.

⁴ For further discussion on nonces and their use see CoreStreet white paper *Nonce Sense, Freshness and Security in OCSP Responses*, CoreStreet Ltd., 2003 available at www.corestreet.com/library.

Conclusions

Denial-of-service is the most critical vulnerability of the four since it is the easiest to exploit and affects the entire user population. The countermeasure of adding additional servers from which relying applications can retrieve certificate status is only effective in the D-OCSP approach.

Intrusion is a vulnerability only for T-OCSP since this approach is designed to allow external, unknown users to connect to the trusted authority. How well the T-OCSP intrusion countermeasures mitigate this threat is hard to measure since it depends upon the absence of unknown design flaws as well as the proper implementation and operation of the countermeasures.

Freshness of the certificate status data is determined by policy. All three approaches have virtually the same vulnerability. The threat of exploitation is small since it is mainly limited to disgruntled legitimate holders of certificates that have been revoked because their owners have been terminated. Physically repossessing the private key from these individuals is an effective countermeasure.

Replay attacks are technically hard to successfully execute. In addition, the threat of exploitation is small since the threat is similar to that associated with the freshness vulnerability. Use of the nonce countermeasure in the T-OCSP approach eliminates the replay threat but increases the DoS vulnerability. Physically repossessing the private key from terminated legitimate certificate holders is an effective countermeasure for the CRL and D-OCSP cases.

From an overall security perspective the D-OCSP approach provides the most security from the external threats discussed here. It provides the best protection against denial-of-service attacks, is not vulnerable to intrusion attacks and has an effective countermeasure against a small, but real and difficult to exploit replay threat.

The T-OCSP approach is susceptible to denial-of-service attacks which are not easily mitigated by redundant server countermeasures. It is also vulnerable to intrusion attacks. The replay attack can be effectively eliminated through the use of a nonce countermeasure.

The CRL approach is the least acceptable from a security perspective because it is highly vulnerable to denial-of-service attacks. It is not vulnerable to intrusion attacks and has an effective countermeasure against the small, difficult to exploit replay threat.

Bibliography

Russell, Deborah and Gangemi, G. T. Sr., *Computer Security Basics*, O'Reilly & Associates, Inc., Sebastopol, CA, 1992.



About CoreStreet

Governments and enterprises rely on CoreStreet products to improve security, speed, and efficiency in critical day-to-day applications, including PKI-based IT security, physical access control, and secure ID checking.

For more information, including detailed product and solution information and technical briefs, see www.corestreet.com