



Distributed Certificate Validation

Abstract

This paper investigates the performance and availability issues associated with current validation architectures and presents a distributed validation approach that scales to 100s millions while reducing deployment costs by 60-80%.



Scalable Validation for Large PKI Deployments

Introduction

The US Government is preparing to roll out large-scale, public key infrastructures. These include the Federal e-Authentication initiative and the various DoD Services' and Agencies' PKI initiatives, such as the KMI program. In each of these programs the number of end users will ultimately be in the 10s to 100s of millions. Once the infrastructure is in place, the number of public key enabled (PKE) applications will grow rapidly to meet the continuous quest to conduct business faster, cheaper and more securely.

A critical factor in the success of these programs will be the performance of the infrastructure as seen by the end user. These end users will quickly become disillusioned with the system if their PKE applications are slow, if they are forced to wait while the system performs security checks, or worst of all, if the system becomes unavailable due to a sudden surge of users, an information warfare attack (e.g., denial of service attack) or a physical disaster that destroys a critical part of the infrastructure. Scenarios such as these could result in end users refusing to use their PKE applications and resorting to conducting business transactions using alternate, unsecured means.

This paper investigates the performance and availability issues associated with the current validation architectures and presents a distributed validation approach that scales to 100s millions while reducing deployment costs by 60-80%.

Validation Problem Statement

One area that has already been identified as having a significant impact on overall PKI performance is that of certificate validation. Since the security of any PKE transaction is based on the current status of the participants' certificates, this status (i.e., valid, revoked or suspended) must be checked for every user transaction. This results in certificate status checking being a high-volume operation. Further complicating the situation is that the source of status for each certificate must be secured and, to a large extent, centrally managed.

There are currently two approaches under consideration for use in the Federal and DoD PKI programs: Certificate Revocation Lists (CRLs) and Online Certificate Status Protocol (OCSP). It is universally recognized that a CRL approach will not scale to user populations in the millions since these lists are anticipated to grow in size to several megabytes. This makes CRLs an unacceptable choice for downloading to each relying party/end user in the system. Consequently, program planners are looking at the implications of deploying an OCSP-based validation system.

OCSP solves the payload size problem present in the CRL approach since it only sends the status information of the certificate in question (not the entire list), is inter-operable with CRL issuing Certification Authorities (CAs), and provides centralized certificate status management. However this approach dramatically increases the network traffic for data that is critical to each transaction. The result is a degradation of overall system performance and validation responder availability.

Concerns over performance and availability issues have raised several fundamental OCSP deployment related questions:

- 1. How many OCSP responders to deploy?**
- 2. Where to put the OCSP responders?**
- 3. How to deploy OCSP responders in a tactical environment?**

An additional concern is:

- 4. How does a relying party know it can trust the response it receives from an OCSP responder?**

Since OCSP responses are signed, one must be concerned about the security status of the keys used in this signing process. Validating an OCSP responder's signature is not a simple operation (e.g., who validates the responder's certificate?).

Obviously the answers to these questions should be based on an architecture that will provide the best possible operational performance. The appropriate choice for this situation is a distributed architecture which puts certificate validation responders close to the users, regardless of their environment. Unfortunately, cost and complexity considerations are limiting the possible answers to these questions.

The reason cost and complexity play such a large role in the OCSP responder decisions is the following:

- Each OCSP server contains both the raw certificate status information and the secret key that is used to digitally sign each certificate response. The responses are signed to protect their integrity as they travel from responder to the relying party application. Therefore each deployed OCSP server needs to be housed in a secure facility and operated in a secure manner by trusted employees.

- Managing and coordinating such a deployment is both costly and complex. The cost of housing and operating a secure facility is significant, comparable to that required for a Certification Authority.

As a result, the answers being given to the first three questions above are:

1. Limit the number of responders deployed
2. Restrict where the responders are deployed
3. No good answer for the tactical environment since it will be difficult to secure the responders even if a lot of money is spent on protecting them.

The above discussion can be summarized as follows: Concern over security results in a centralized OCSP architecture which significantly limits the system’s ability to scale. Consequently both performance and availability needs will suffer.

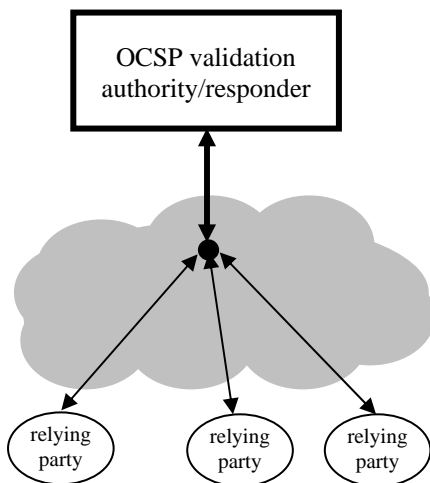


Figure 1: Centralized OCSP Architecture

In traditional OCSP validation there is one “authority” which communicates directly with all relying applications to provide them with validation proofs. This centralized approach can result in poor performance, weakened security, and a single point of failure.

Performance Improvements using a Distributed Architecture

Achieving high performance and high availability in an environment where there are millions of users, all getting data from a single or limited source, has already been studied and solved in the commercial world. The answer is to deploy a distributed architecture. Email is a good example of a distributed architecture. Another example is that used by Akamai Technologies, Inc. to deliver web content worldwide.

Akamai hosts web content for over 1,400 of the world’s leading e-businesses on a globally distributed network of more than 15,000 servers in 68 countries. They found that

a major performance factor is “the single point-of-failure of the first mile where the Web applications interface with the Internet, and the limitations of the routing and switching equipment that make up the backbone of the Internet”. Akamai addressed this problem by moving the content closer to the end users. Performance improvements using this approach have been measured by Akamai’s customers at over 400%.¹ While Akamai does not have the same security concerns that a validation system must address, there can be little argument that the distributed approach will provide the best possible performance and availability. How this can be achieved and still meet the security requirements is discussed below. First let us substantiate the performance claim.

A recent independent analysis of Internet performance issues² concluded that performance, as perceived by the end user, drops as the network becomes congested or as the distance to the data source increases. In fact, even with recent improvements in processor speeds, optimization of TCP usage, and migration to 1 Mbps or higher access lines, the actual measured performance has remained constant due to increases in network delays. These delays are caused by heavier network usage and increased distances to servers. Figure 2 below shows that the percentage of the delay in delivering content from a central server to an end user due to network delay increased from 33% in 1995 to 69% in 1999 and is forecasted to reach 84% in 2003.³ The conclusion of this study is that:

“There is no substitute for getting the content closer to the user.”

While this study is focused on delivering web content to browsers over the Internet rather than certificate validation to PKE applications, the conclusion is still relevant.

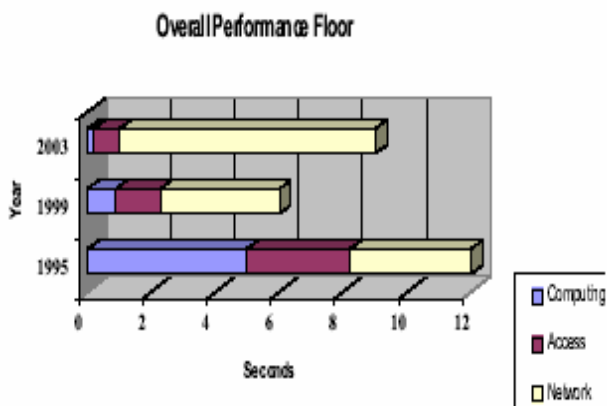


Figure 2: Increase of Network Delay on the Internet

*Network delay is the largest and fastest growing component of network performance degradation.
Data © 2000 Sevcik*

¹ Technical paper, “A Distributed Infrastructure for e-Business – Real Benefits, Measurable Returns”, Akamai Technologies, Inc., ©2000.

² Sevcik, Peter, “Performance Issues facing the World-Wide Web”, *Business Communications Review*, Volume 29, Number 9

³ *ibid.*

In a separate white paper⁴ Akamai has measured a response improvement of 200% during sudden surges in traffic on a distributed server architecture versus a centralized server architecture (see Figure 3).

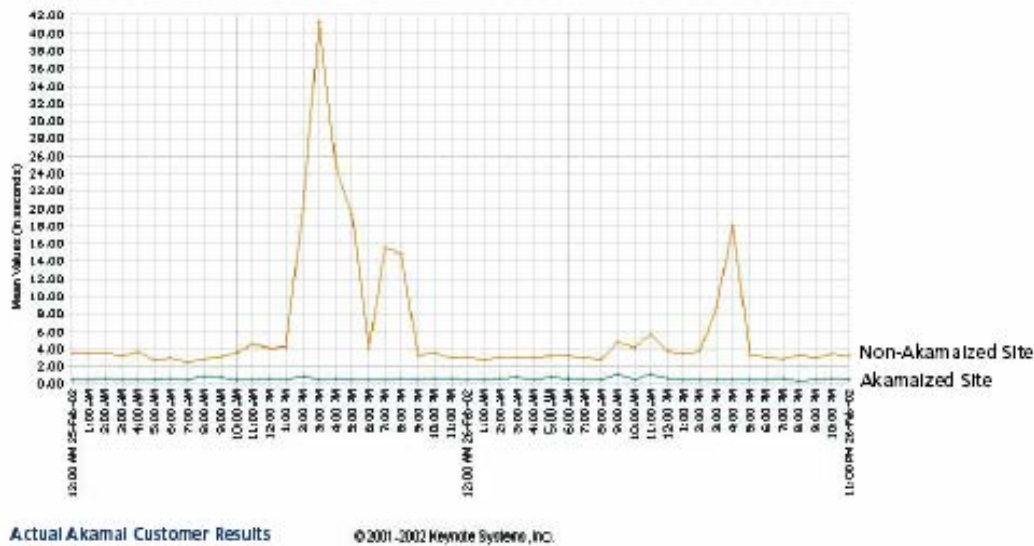


Figure 3: Smoothing of Network Usage Surges by Distributed Architecture
(Data reproduced by permission of Akamai Technologies, Inc.)

The Solution – Secure Distributed Validation

In a distributed validation approach there are no constraints limiting the number of certificate status responders and no restrictions on the environments in which they can be placed. This is achieved by:

- pre-computing status “proofs” for each individual certificate, and
- protecting the integrity of these proofs so that they can be freely distributed

Figure 4 shows a distributed validation architecture where multiple validation responders have been located close to the end user relying party applications without any constraints

⁴ Technical paper, “Why Performance Matters”, Akamai Technologies, Inc., ©2000.

as to the environment in which they can be placed. To accomplish this, the servers must contain no secret information.

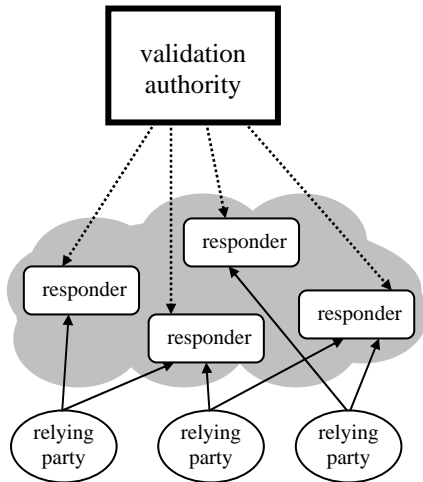


Figure 4: Distributed Validation Architecture

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

CoreStreet’s Secure Distributed Validation Design Principle

*The design principle of secure distributed validation is the **separation of security sensitive data and trusted operations** from the **delivery process** of providing certificate status to relying party applications.*

In this approach the validation authority contains all the sensitive data and performs all trusted operations. This can be done using a single validation authority which simplifies the securing of its operations and centralizes its management. Periodically the validation authority pre-computes individual, time-bounded status proofs,⁵ the publishing periodicity being determined by local policy (e.g., hourly, daily ...). The integrity of these proofs can be protected either by digitally signing them, as is done in traditional OCSP or by using validation tokens (or V.Tokens) generated through a hashing technique that renders their integrity self-validating.⁶ These proofs can therefore be freely distributed since they:

- **Do not require secure channels for transmission,**

⁵ The validity period for these validation proofs can be any length and is determined by policy.

⁶ For a description of V.Token technology see “Real Time Credential Validation: Secure, Efficient Permissions Management”, a white paper available at www.CoreStreet.com.

- **Do not require secure storage.**

The architecture depicted in Figure 4 reflects the fact that responders in a distributed validation approach are no longer security sensitive and can be placed close to end user relying party applications in unsecured, office type environments.

Benefits of Secure Distributed Validation

The benefits of the distributed validation approach are numerous. They include:

1. **Truly scalable:** true scalability is achieved by separating the delivery process from the security sensitive operations associated with certificate validation. The barriers to true scalability – performance, availability, security and cost have been eliminated.
2. **High availability:** high availability is now achieved because end user applications have access to a local responder. This is analogous to placing email servers on local area networks to be close to end users for improved availability.
3. **High performance:** the distributed validation architecture takes advantage of the lessons learned in the commercial world by decreasing the distance between a relying party application and a responder, eliminating a choke point at the responder, the largest cause of poor performance.
4. **Improved survivability:** the single point of failure threat has been significantly reduced. Distributed denial of service attacks are virtually eliminated by the deployment of multiple, geographically dispersed responders.⁷ Physical attacks on the validation authority itself are also ineffective since responders will continue to operate for some period of time which will allow for a “recovery” period during which a backup validation authority can be brought on-line.⁸
5. **Cost effective:** since responders do not require secure communication, housing or operation there is little cost associated with deploying them in a widespread

⁷ On October 21, 2002 a “distributed denial of service” attack was launched against the 13 “root 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.

⁸ For example, if the validity period of the validation proofs is 24 hours and new validation proofs are released every 12 hours, each responder can continue to operate for at least 12 hours after a validation authority has been disabled or destroyed.

fashion. In addition, industry standard web server platforms can be used, dramatically reducing the cost of deployment.⁹

- 6. Flexibility and Adaptability:** each responder can support more than one validation authority. This allows independent authorities to retain complete control over their domain (i.e., without relinquishing any trusted operations or data to another authority) while sharing a common delivery infrastructure.
- 7. Improved global reach:** responders can now be located in the far reaches of the globe without introducing poor performance at the end user due to distance dependent network delays.
- 8. Tactical environment solution:** since responders do not hold any security sensitive data they can be located in tactical environments where the threat of being overrun is real. The architecture is also ideal for rapid deployment scenarios as adding and deleting responders is quick and easy.
- 9. More secure:** two elements of security have been significantly improved over the traditional OCSP validation model:
 - a.** Certificate status requests go only to responders, not to the validation authority. Since the validation authority does not allow any inbound communication from the outside world the threat of an outside attack is virtually eliminated.
 - b.** Scaling the validation system to serve increasingly larger user communities does not require distributing security sensitive data or trusted operations to multiple locations. Therefore the ability to securely manage this operation is greatly enhanced.

⁹ See Appendix B for a quantitative deployment cost comparison.

The CoreStreet Approach

CoreStreet's Real Time Credentials (RTC) solution offers a distributed certificate validation product that can support user populations in the 100s of millions with high availability, high performance, improved security and lower deployment costs. RTC employs the distributed validation architecture described above and supports both digitally signed validation proofs and self-validating V.Token proofs.

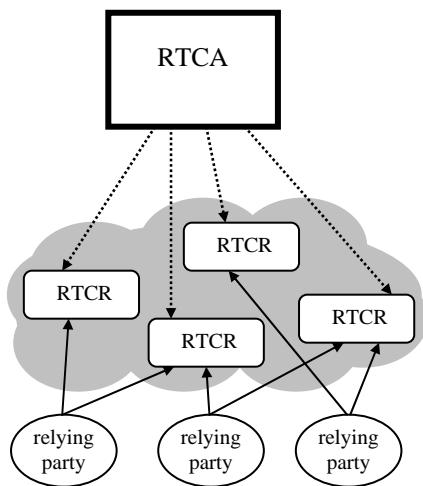


Figure 5: CoreStreet Real Time Credentials Distributed Validation Architecture

CoreStreet's Real Time Credentials validation solution provides guaranteed scalability without sacrificing performance, availability, security or cost.

Validation proofs are periodically generated at the RTC Authority (RTCA) and distributed as digitally signed files (via an intermediate server) to the RTC Responders (RTCR). A single RTCA server running on an Intel or Sparc platform can comfortably support a population of 10 million users with daily proofs. Larger populations can be supported by simply adding larger and/or additional RTCA platforms. Each RTC Responder is capable of receiving either an OCSP request or a V.Token request for the status of a certificate and returning the appropriate response. Since this operation is a simple table lookup the response is returned in ~2 millisecond. This is a significant improvement over traditional OCSP where each response must be signed before it is delivered to the relying part application.¹⁰

Additional CoreStreet RTC Benefits

CoreStreet's RTC distributed validation implementation offers the following additional benefits over traditional OCSP validation:

¹⁰ See Appendix A for relevant deployment parameters.

- 1. Off-line validation:** because the CoreStreet validation proofs are unforgeable and unalterable, they can be presented to the relying party from any source – including by the user himself. This provides great flexibility for applications that are difficult or impossible to connect to a network. For example, a user could retrieve his validation proof for the day using his Common Access Card and then present it, along with his certificate, to a disconnected application. The application can then locally authenticate and validate the user.
- 2. Minimum bandwidth solution:** CoreStreet has two validation solutions that support low bandwidth environments: V.Tokens and MiniCRLs. Each V.Token is only 20 bytes long making it ideal where there are severe bandwidth limitations from the responder to the client. MiniCRLs offer a factor of 30x size reduction over a traditional CRL making it ideal where there are severe bandwidth limitations from the RTCA to the RTC responders.
- 3. Dynamic privilege management:** CoreStreet’s Distributed OCSP and V.Token technologies enable, for the first time, the ability to dynamically manage multiple privileges associated with a single certificate without having to reissue or modify that certificate in any way. In addition, CoreStreet’s technology allows these privileges to be managed by independent, autonomous authorities.
- 4. Self-validating authority:** One of the more difficult issues associated with deploying OCSP responders is answering the question “How does a relying party know it can trust the response?” CoreStreet’s V.Token technology makes this a moot issue as the integrity of a V.Token is self-evident. There are no signing keys involved in this approach and therefore no certificates associated with a signing key that must be checked to validate the integrity of the response.

Alternately, with CoreStreet’s Distributed OCSP approach there is only one responder (versus 10s or 100s) with a single signing key. In this case a single certificate must be validated. This makes the issue much easier to solve as approaches such as a “short-lived certificate” become practical options.

Format Selection

CoreStreet’s RTC proofs using either digital signature or V.Token formats offer secure credential validation for a wide variety of applications.

Distributed OCSP – these digitally signed proofs offer the simplest integration based on their compatibility with existing protocols and standards. CoreStreet’s

Distributed OCSP proofs are completely compatible with OCSP, are syntactically constructed as standard OCSP responses and can be used by any OCSP compatible relying party application. Depending on signature algorithm, key length, and the inclusion of privilege management, digital signature proofs range in size from 150 and 350 bytes, and can be processed and verified in fewer than ten milliseconds on a typical computer. A digital signature solution offers excellent scalability to about ten million independent credentials and hundreds or thousands of responders.

V.Token proofs - provide great scalability and performance for environments where the bandwidth to the client application is limited and when there are no requirements for legacy compatibility (V.Tokens require adding 40 bytes of data to the Subject Directory Attribute extension of an X.509 certificate and use client software capable of interpreting the proofs). Depending on the one-way hashing algorithm and protocol selections, V.Token proofs range in size from 16 and 100 bytes and can be generated and evaluated in less than one millisecond. A V.Token solution can easily support hundreds millions of credentials along with tens of thousands of responders. V.Tokens are ideal for use in handheld wireless devices and broadcast environments.

MiniCRLs - provide great scalability and performance for environments where the bandwidth from the RTCA to the RTC responders is limited, such as in the case of ship-to-shore communications channels. This approach is also ideal for a large number of users in low bandwidth broadcast environments. MiniCRLs represent the absolute minimum size (uses one bit per issued certificate) for conveying certificate status information. The effective size is reduced to approximately one-half bit per certificate using standard compression techniques. A segmentation technique is used to keep the size of the data sent to client applications small. The MiniCRL approach uses a client side plug-in to interpret the certification status information.

Using either format, CoreStreet's Real Time Credentials solution provides for unequaled availability, scalability, and security for mission-critical credential management.

Summary

Secure distributed validation is not just a better way to provide certificate validation for a PKI deployment; it is the only solution that guarantees scalability without sacrificing performance, availability, security and cost. In addition it provides solutions to tactical and bandwidth limited operations that could not otherwise be achieved.

CoreStreet is the leading provider of real time credential validation. CoreStreet's RTC architecture, which separates security sensitive validation operations from certificate status delivery, is unique among industry validation providers. CoreStreet offers system planners and implementers the first, truly scalable solution for certificate validation. Our unique technology also provides the ability to expand the use of credential validation to off-line, disconnected applications, low band-width scenarios and dynamically managed privileges – all built on a secure, cost effective foundation that is flexible, adaptable and scalable.

Contact us at:

CoreStreet Ltd.
One Alewife Center
Suite 200
Cambridge, MA 02140

Email: info@corestreet.com
Tel: 617-661-3554

Appendix A: CoreStreet RTC Deployment Parameters

The following table provides relevant deployment parameters for the three approaches provided by CoreStreet’s Real Time Credential distributed validation solution. These parameters were developed under the following assumptions:

1. 1 million end users
2. Period of validity for validation proofs is 1 day
3. File compression of 50% for downloaded files
4. T1 transmission speeds from RTC Authority to RTC Responders
5. Signing key for Distributed OCSP is 1024 bit RSA

Times were measured using a single midrange Intel server with a hardware accelerator.

Table A: CoreStreet Distributed Validation Deployment Parameters

Parameter	Distributed OCSP	Distributed V.Token	MiniCRL
Storage rqmts at RTCA	1 Gbyte	120 Mbytes	120 Mbytes
Processing time at RTCA (using an HSM)	~10 minutes	17 cpu minutes	5 cpu minutes
File size sent from RTCA to RTC responders	14 Mbytes	17 Mbytes	90 kB
Download time to RTC responders	1.3 minutes	3 minutes	1 second
Storage rqmts at RTC responders	50 Mbytes	50 Mbytes	30 Mbytes
Size of proof sent to client	2.5 kB	400 bytes	3-4 kB
Relying application processing time	10 milliseconds	1 millisecond	10 milliseconds

A typical Sparc RTCA server with a hardware security module can generate Distributed OCSP proofs for 1 million certificates in approximately 10 minutes, and a million V.Token proofs in less than 20 minutes. A single RTC responder can provide more than 1000 responses per second, allowing it to service millions of relying party requests per day. RTC responders can also simultaneously support Distributed OCSP, V.Token and MiniCRL requests. This provides the capability of using a single system to mix and match validation approaches to meet multiple and distinct operational needs.

Appendix B: Validation Approach Cost Comparison

Note: Prices shown are for illustration purposes only and do not represent the current product pricing. For current information, please contact your CoreStreet representative. (See www.CoreStreet.com/contact)

One of the major differentiators of CoreStreet’s Secure Distributed Validation is its cost effectiveness. This appendix quantifies these cost savings by comparing the deployment and recurring infrastructure costs for providing the same level of validation service. Only those costs directly related to architectural choices are examined. For the purposes of this discussion “same level of service” is defined to mean deploying the same number of geographically dispersed responders to provide the same performance and availability to the relying party applications. Two different approaches are compared, traditional OCSP (T-OCSP) and distributed OCSP (D-OCSP). The infrastructure costs for deploying validation tokens (V.Tokens) and MiniCRLs are essentially the same as for D-OCSP.

Assumptions:

1. Number of certificates being managed (millions)	10
2. Number of responders deployed	10 & 100
3. Minimum freshness time (update period in hours)	2
4. Key type	RSA
5. Key length (i.e., OCSP responder key)	1024

Estimated traditional OCSP component costs:

1. T-OCSP responder hardware/site (includes HSM)	\$25,000
2. T-OCSP responder site security setup (one time/site) ¹¹	\$50,000
3. T-OCSP responder security operations (yearly/site)	\$50,000

Estimated distributed validation component costs:

1. D-OCSP RTCA hardware (includes HSM)	\$113,000
2. RTCA site security setup (one time, single site)	\$50,000
3. RTCA site security operations (yearly, single site)	\$50,000
4. RTC responder hardware/site (no HSM needed)	\$3,000
5. T1 monthly rates	\$1,000

Traditional OCSP deployment requires:

- Network connection to each deployed responder to receive periodic CRL updates
- Physical & electronic security protection for each deployed responder (estimated)
- Trusted operators using dual access control (cost is estimated)

¹¹ While this cost will vary by site, \$50,000 is very conservative and could easily be 4 to 10 times more.

Distributed OCSP deployment requires:

- One RTCA (the “back-end”) that needs physical and electronic security protection (cost is estimated, could be collocated with CA at little or no extra cost)
- Single set of trusted RTCA operators using dual access control (cost is estimated)
- Network connection from the RTCA to each deployed RTC responder
 - D-OCSP requires T1 speeds for up to ~ 10 million certs
- No secure storage, comms or operators for any of the deployed RTC responders
- Standard COTS servers for each RTC responder

Cost Calculations:

- T-OCSP setup costs = responder hardware + firewall + network setup + site security preparation
 - = (#Resp) x (\$25k + \$10k + \$2k + \$50k)
 - = (#Resp) x (\$87k)
- D-OCSP setup costs = RTCA hardware + firewall + file servers + network setup + site security prep + responder(hardware + T1 setup)
 - = (1) x (\$113k + \$10k + \$10k + \$2k + \$50k) + (#Resp) x (\$3k + \$2k)
 - = \$185,000 + (#Resp) x (\$5k)
- T-OCSP recurring costs = leased lines (to download CRLs) + security personnel
 - = (#Resp) x {[#Mbytes/(T1 transfer rate)] x (\$T1) + \$50k}
 - = (#Resp) x {(140 Mbytes/187,500 bytes/sec) x \$12k + \$50k}
 - = (#Resp) x {(1) x \$12k + \$50k}
 - = (#Resp) x (\$62k)
- D-OCSP recurring costs = leased lines + security personnel costs
 - = (#Resp) x {[#Mbytes/(T1 transfer rate)] x (\$T1)} + (1) x (\$50k)
 - = (#Resp) x {(140 Mbytes/187,500 bytes/sec) x \$12k} + \$50k
 - = (#Resp) x {(1) x \$12k} + \$50k
 - = (#Resp) x \$12k + \$50k

Table B-1 provides a cost comparison for a deployment of 10 responders.

Table B-1: CoreStreet Distributed Validation Deployment Parameters

Cost Element	# Rspdrs	T-OCSP	D-OCSP
1 st year setup costs	10	\$870,000	\$235,000
Yearly recurring costs		\$620,000	\$170,000
Savings in first year (setup + ops)		\$1,085,000	
Recurring cost savings		\$450,000	
<hr/>			
1 st year setup costs	100	\$8,700,000	\$685,000
Yearly recurring costs		\$6,200,000	\$1,250,000
Savings in first year (setup + ops)		\$12,965,000	
Recurring cost savings		\$4,950,000	

Note 1: This comparison points out explicitly the cost difference in scaling up from 10 to 100 responders. While T-OCSP costs grow linearly, D-OCSP costs grow much more slowly.

Note 2: Because the D-OCSP proofs have been presigned, the responder to relying party response time is 20x faster for D-OCSP than for T-OCSP. This has not been factored into this calculation. Thus the savings are even greater than shown.

It is clear from this comparison that CoreStreet's Secure Distributed Validation provides the flexibility needed to achieve high availability without incurring the significant cost penalty inherent in the traditional OCSP approach.

Note: Prices shown are for illustration purposes only and do not represent the current product pricing. For current information, please contact your CoreStreet representative. (See www.CoreStreet.com/contact)