6-2006

On the Release of Crls in Public Key Infrastructure

Chengyu Ma
Singapore Management University

Nan Hu
Singapore Management University, hunan@smu.edu.sg

Yingjiu Li
Singapore Management University, yjli@smu.edu.sg

On the Release of CRLs in Public Key Infrastructure

Chengyu Ma[1]
Beijing University
machengyu@pku.edu.cn

Nan Hu, Yingjiu Li
Singapore Management University
hunan,yjli@smu.edu.sg

Abstract

Public key infrastructure provides a promising foundation for verifying the authenticity of communicating parties and transferring trust over the internet. The key issue in public key infrastructure is how to process certificate revocations. Previous research in this aspect has concentrated on the tradeoffs that can be made among different revocation options. No rigorous efforts have been made to understand the probability distribution of certificate revocation requests based on real empirical data.

In this study, we first collect real empirical data from VeriSign and derive the probability function for certificate revocation requests. We then prove that a revocation system will become stable after a period of time. Based on these, we show that different certificate authorities should take different strategies for releasing certificate revocation lists for different types of certificate services. We also provide the exact steps by which certificate authorities can derive optimal releasing strategies.

1 Introduction

The introduction of world wide web technology has resulted in a faster and easier exchange of information. It also exacerbate the problems of verifying the authenticity of communicating parties and transferring trust over the internet. The public key infrastructure (PKI) has been considered as a promising foundation for solving these problems, especially in the context of secure electronic commerce. Since the authenticity of PKI is achieved through the verification of digital certificates, it is crucial to understand the nature of digital certificates in practice.

Digital certificates have been supported by a wide range of entities. For example, Korean Government invested heavily on promoting digital certificates to the public. The digital certificates have been issued for various applications such as internet banking, government e-procurement and stock exchange.[2] In the year 2001, the Ministry of Information and Communication announced that the total number of users of PKI would reach to 10 million by the year 2002. Korea also made its own 128 bit encryption algorithm called SEED and encouraged all financial services to use it. It also developed a national certificate system based on public key infrastructure.[3]

Unfortunately the glory of the PKI can be so dimmed if there is no efficient way to verify the validity of digital certificates. Checking the authenticity and expiration date of a digital certificate is never sufficient enough as it is possible that a certificate has been revoked before its expiration for various reasons, such as 1) key compromise, 2)certificate authority (CA) compromise, 3) affiliation change, 4) superseded, or 5) cessation of operation [6]. To make PKI a useful platform, it is critical to manage the certificate revocations efficiently.

Previous research has concentrated on the trade-offs that can be made among different revocation options [6, 13, 15]. The purpose is to see which revocation mechanism is more efficient in which scenario. In order to compare the performance of different mechanisms, people ran simulations based on theoretical assumptions. For example, Naor and Nissim calculated the communication cost by assuming a fixed length of certificate revocation list (CRL). Cooper [2] and Arnes [1] modeled the distribution of revocation information by assuming an exponential inter-arrival probability for the requests for CRLs. To the best of our knowledge, no rigorous efforts have been made to understand the probability distribution of certificate revocation requests based on real empirical data.

Another key conclusion of previous research is that CA should release CRLs at a fixed time interval because consumers may not need the most current CRL. As long as a user has a CRL that is recent enough to meet its operational requirement, it is acceptable in practice [8]. Rivest [13] has proposed that the recency requirement should be set by customers, rather than CAs. Unfortunately, all the conclusions are based on some theoreti-

cal arguments, there is a lack of guidance for CA operations. Given a recency requirement to be set by the consumers, CA must understand the following aspects for setting an optimal CRL releasing policy: 1) Why CA needs to follow a given time interval? Will the interval be the same for a new type of certificates versus a type of certificates that has been provided by CA for a certain period of time? 2) Will the interval be the same for a mature CA versus a Start-up CA? 3) If CA does follow a given CRL releasing interval, how does it know whether that interval is optimal or not?

In this paper, we study how often should a CA release its CRLs. We concentrate our analysis on CRL because it is the most common and simplest method for certificate revocation [6]. We have several interesting findings: 1) Contrary to the common sense, the probability that a certificate being revoked is a decreasing function over the certificate's life cycle. People tend to think this probability either flat over time (as memory-less Poisson) or increasing over time. 2) CA should take different strategies for publishing certificate revocation lists when dealing with a new type of certificates versus a re-serving type of certificates. 3) A mature CA and a start-up CA should also take different strategies for releasing CRLs. 4) We give an optimal releasing interval prescription for CA to balance the trade-off between cost and risk. In a very general case, a mature CA who deals with just one type of certificates can save almost 400,000 dollars more over one-year operation if it follows our strategy by deceasing its CRL releasing interval from 34 days to 17 days.

The rest of the paper proceeds as follows: First, we briefly review the major concerns about the public key revocation. We then discuss empirically how to collect data and derive the revocation distribution. Next, based on the empirical distribution, we give CA an optimal CRL releasing strategy. Finally we conclude our paper with a discussion of contributions, limitations, and future research directions.

2 Literature Review

Since its introduction, the public key infrastructure [5] has provided a promising foundation for verifying the authenticity of public keys and for transferring trust among users or business partners. The major issue in PKI is how to revoke a certificate before its expiration. It has been argued that the running expenses of a PKI derive mainly from administering revocation [14]. Various mechanisms have been designed to achieve efficient, timely, and scalable revocation of certifications [15, 6].

The certificate revocation list (CRL) mechanism was introduced in 1988 and since then it remains the most common and simplest method for certificate revocation. A CRL is a time-stamped list of certificates which have been revoked before their expiration. A CA issues a signed CRL periodically so as to maintain a good synchronization between certificate users and revocation source. Some extensions of CRL include delta-CRL (which only carries changes from previous CRL), partitioned CRL (which is partitioned into a family of CRLs), and indirect-CRL (which can be issued by different CA than issuer of certificates). Rivest proposed to use short-lived certificates so as to eliminate CRLs [13]. The major drawbacks of this approach include a high burden placed on certificate servers which need to sign more certificates, as well as the problem of key compromise which cannot be addressed without using a separate mechanism [9].

Micali introduced the certificate revocation system (CRS) which is different from CRL. In CRS, a CA signs a fresh list of all not-yet-expired certificates together with selected hash chain values. A user sending a request regarding the validity of a single certificate will get a response including two hash chain values. The hash chain values can be used to verify whether the queried certificate is valid or not for a certain time interval. The major advantage of this method is that the verification process is very efficient, thus can be performed on-line. However, as pointed out by Naor and Nissim [11], the main disadvantage of this system is the increase of the CA's communication cost.

The certificate revocation tree (CRT) mechanism was suggested by Kocher [7] which can be used by a verifier of a certificate to obtain a short proof if the certificate has not been revoked. A CRT is a hash tree whose leaf nodes correspond to a set of statements about certificates status. The set of statements provides information about whether a certificate is revoked or not. A proof for a certificate status consists of an appropriate path in the hash tree (from the root to a leaf) specifying for each node the values of its children. With CRT, a user may hold a short proof for the validity of his certificate such that the entire CRL is not necessary for verifying the status of the certificate. The drawback of CRT is its maintenance cost. Any change to the set of revoked certificates may cause re-computation of the entire CRT.

An alternative to the CRL mechanism is to use on-line certificate status protocol (OCSP) to reduce the latency between a revocation report and the distribution of revocation information to users [10]. Once a CA accepts a revocation report, any query (OCSP request) to the status of one or more certificates will be correctly answered by an on-line validation server (OCSP responder) with relevant status values (good, revoked, or unknown) and valid intervals. Though OCSP provides more timely revocation services, it imposes new security requirements as the certificate validators shall trust the on-line valida-

tion service.

Besides the above mechanisms, researchers have studied various aspects of certificate revocations including the meaning of revocation [3, 4], the model of revocation [2], communication cost of revocation [11], tradeoffs in certificate revocation schemes [16], and risk management in certificate revocation [8]. Though various tradeoffs have been studied for different revocation options, no attempt has been made to understand the probability distribution of request for certificate revocation. In this paper, we conduct such research for CRL releasing mechanism based on real data, and give concrete guidance for the optimal operation of CA in various scenarios.

3 Problem Formulation

In this paper, we study how often should a CA release its CRLs. There are several key assumptions in our study: 1) This is a monopoly case, which means either there is just one CA in the system or different CAs provide different types of certificate services. So CAs do not need to consider the competition effect. 2) CA already decides the issued age of a given type of certificates, where issued age is defined as the time difference between the expired date and the issued date. 3) To get started, we assume CA issues one type of certificates with the same issued age. These certificates are independent and identical in terms of risk and cost. Later we will move on to more general cases.

Given all of the assumptions, the goal of CA is to find out about how often it should release a CRL to minimize its operational cost over a given period. Here we define "how often" as the optimal time interval between two successive CRLs being released, and the "operational cost" as the sum of *variable cost, fixed cost,* and *liability cost* as defined below.

Normally CA takes a batch process for CRL release. There is a trade-off between cost and risk. In the case that consumer files a revocation request to CA but CA does not release a CRL on time, we assume that CA will bear the liability cost if there is any damage occurred between request filing and CRL releasing. Each time CA releases a CRL, it incurs both fixed cost component and variable cost component. The fixed cost does not change with the length of CRL. It indicates a fixed dollar amount each time CA spends for releasing one CRL, regardless of the number of certificates in that CRL. Variable cost is the cost associated with processing each individual certificate revocation request.

If CA releases the CRL too often, its liability cost is low, but its fixed cost and variable cost will be high. On the other hand, the saving on fixed cost and variable cost might not be offset by the increasing liability cost if CA releases the CRL too rarely. So CA needs to find an optimal interval for CRL release. In order to find this solution, CA must know the CRL length on a given day if it decides to publish CRL on that day. The length of CRL at any time t is related to three components: 1) Length of CRL at any time $t-1$. 2) How many revoked certificates including in the CRL at time $t-1$ will be expired at time t. According to CRL policy, if the certificate is expired, it should be excluded from CRL. 3) How many new revocation requests it will receive from time $t-1$ to t.

Parameter	Meaning of Parameter
a	Max number of days between two successive CRL released dates that is accepted by customers.
b	The average percentage of certificates revoked among that type of certificates issued.
c	The number of days between two CRL releasing dates.
t	Time parameter in the function of R(t),which is R(t)=ke^{-kt}.
Δt	Time interval between two generations of CRLs.
X	Date on which certificates get issued.
k	Parameter in the function of R(t),which is R(t) = ke^{-kt}.
n	Numbers of generations.
v	Any time between 0 and ∞ in the f(v),F(v), and P(v).
d	The upper bound of the number of certificate revocations in one CRL which is allowed by CA, before it releases the CRL.
q	The number of CRLs that CA will published during period β .
i	The ith CRL published by CA.
Nd_i	CA releases the CRL on the Nd_i day.
α	Number of certificates issued at different times.
β	Issued Age of CRL, which is equal to Expired Date Minus Issued Date.
μ	Shape parameter in Poisson distribution which indicates the average number of certificate revocations in a given time interval.
λ	Number of certificate revocations in CRL on the day β for Poisson case.
θ	Stable number of certificates in CRL on a given day after β if CA decides to release CRL on that day.
FC	The fixed cost of CA for publishing one CRL.
VC	The constant unit cost of CA for including one certificate into the CRL.
Υ	The expected risk/liability per revocation cost for CA for delaying publish that revocation for one day
$f(v)$	The number of new certificate revocations between day v and day v+Δt.
$F(v)$	The valid cumulative number of certificate revocations from time 1 to v.
$P(v)$	The percentage of certificate revocations occurred from time v to time v+Δ t.

Table 1: Notation

For ease of reference, Table 1 lists all the notation that will be used in this paper.

4 Data Collection

How many new revocation requests a CA received from time $t-1$ to t is really driven by the probability distribution of certificate revocation requests. From September 7th to September 13th , we collected a series of CRLs from VeriSign.com. As one of the biggest Certificate Authority in the world, VeriSign provides variant types of certificates and publishes different CRLs periodically. We randomly choose five different CRL files from

VeriSign website, which belong to five different classes. Table 2 provides the descriptions for these 5 CRL files which have 39,243 total revocation records. When a certificate is issued, its validity is limited by an expiration date. Note that the definition of issued age is:

$$\text{Issued Age} = \text{Expired Date} - \text{Issued Date}$$

However, there are circumstances where a certificate must be revoked prior to its expiration date. Thus, the truly existence age of the certificate is the time between the issued date and the revoked date.

$$\text{Existence Age} = \text{Revoked Date} - \text{Issued Date}$$

A certificate is valid for its issued age unless it is revoked. Each revoked certificate in a CRL is identified by its certificate serial number and the revoked date. Based on the serial number of a given certificate, we searched the VeriSign online database to get both the issued date and the expired date correspondingly. We cleaned those error records whose revoked date was later than the expired date or whose issued date was later than the revoked date.

5 Data Analysis

We present the summary statistics for our data in Table 3. The average issued age of these CRLs is 493 days, while the average existence age is much shorter, only 31 days. To further demonstrate what is happening here, we plot the number of revocations against existence age in Figure 1, and the percentage of revocations against existence age in Figure 2 for classes RSASecureServer and SVRIntl.

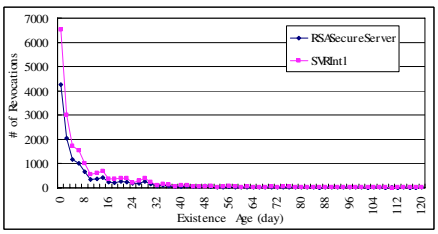

Figure 1: Number of revocations vs. existence age

The most interesting finding is that most of the certificate revocations occur at the first few days after issued, and the percentage of revocations decreases with elapsed time. More than 30% of revocations occur within the first two days after certificates get issued. This distribution pattern is very robust, and it is insensitive to which CRLs we investigated and which years we selected. It still holds when we pool five CRLs together.

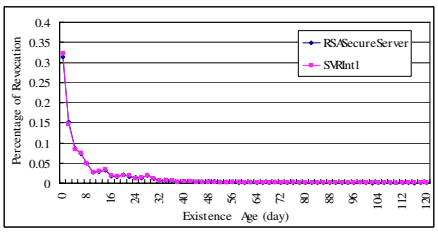

Figure 2: Probability of revocation vs. existence age

5.1 Empirical Model

In order to get the size of CRL at any time t, we must know the behavior of the probability density function (PDF) of certificate revocations over time. In the above, empirically we already show that the percentage of revocations decreases with elapsed time. Next statistically we derive the underline PDF.

5.1.1 Underline PDF for Certificates Issued at a Particular Time

We assume that there are α certificates issued at time X with issued age β. To get start, we assume that α is a constant number. Later we change α to a random number with a Poisson distribution to study the more general case. From time X to time $X+\beta$, on average $\alpha b\%$ of the certificates will be revoked. At time $X+\beta$, all the certificates issued at time X will be expired, no matter whether they have been revoked or not. Let $R(t)$ be the probability that any given certificate issued at time X will be revoked in the interval $[t, t+\Delta t]$, where t is between X and $X+\beta$. It also represents the revoked percentage, which is the number of revocations occurred in the interval $[t, t+\Delta t]$ divided by the total number of revocations occurred between X and $X+\beta$ (i.e., $\alpha b\%$). Following the empirical distribution observed in Figure 2, we use an exponential probability density function to model this distribution.

$$R(t) = ke^{-kt} \tag{1}$$

We use Maximum Absolute Deviation (MAD) to determine the parameter k. The MAD is proposed by Kolmogorov-Smirnov. It minimizes the largest gap between the cumulative relative frequency of a given data set and that of its fitted statistical distribution. In Figure 3, we present both the real empirical data and theoretically fitted PDF[4]. The PDF fits the empirical data very well when the parameter k is equal to 0.26, which is accepted at a 99% confidence interval.

File Name	Issuer	Publishing Time	Purpose	No.of Items	Max Existence Age (day)	Signature Algorithm
Class3CodeSigning2001.crl	VeriSign Class 3 Code Signing 2001 CA	September 04, 2005 6:00:08 PM	Code signing and object signing certificates used for Netscape browsers, Microsoft Internet Explorer browsers, Microsoft Office, Sun Java Signing, Macromedia, and Marimba.	1,993	380	md5RSA
CSC3-2004.crl	VeriSign Class 3 Code Signing 2004 CA	September 11, 2005 6:00:25 PM	Same as VeriSign Class 3 Code Signing 2001 except used for certificates with different expiration date.	228	364	md5RSA
Class3 NewOFX.cr	VeriSign Class 3 Open Financial Exchange CA	September 11, 2005 6:00:15 PM	Open financial exchange certificates, used for authenticating and securing commerce on the Internet.	515	302	md5RSA
RSASecure Server.crl	RSA Secure Server CA	September 08, 2005 6:00:25 PM	Secure server certificates used by a Root CA for managing PKI for SSL Customers and VTN Affiliates.	14,837	727	md5RSA
SVRIntl.crl	VeriSign International Server CA Class 3	September 08, 2005 6:00:16 PM	Global server certificates for managing PKI for SSL (Premium Edition) customers and VTN Affiliates.	21,839	720	md5RSA

Table 2: Descriptions of CRL files

Items	Issued Age(Unit:Day)	Existence Age(Unit:Day)
Mean	493	31
Median	366	5
Max	1173	727
Min	1	0
Q3	730	21
Q1	365	1

Table 3: Summary statistics

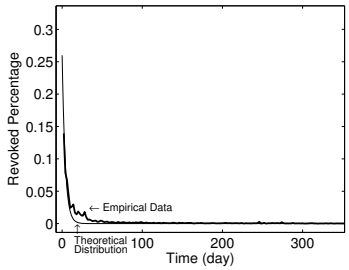

Figure 3: Empirical data vs. fitted exponential PDF

5.1.2 First Case: A Model for Pooling Certificates Issued at Different Time

In the previous section, we only consider the PDF of one population of certificates issued at time X. Now consider that CA issues certificates at different time. Each generation of certificates is composed of certificates issued at a particular time with the same issued age β, where the time interval between two successive generations is Δt. At any given time interval $[t, t + \Delta t]$, the revocation requests CA received originate from different generations.

In order for CA to decide when to release the CRLs, on daily basis CA must know: 1) The number of new revocation requests; and 2) The size of the CRL if it decides to release the CRL on that day. Based on the PDF derived from the previous section, we build a model to compute the total number of certificate revocations when pooling revocations from different generations.

The number of new revocation requests The PDFs of revocations from different generations of certificates follow the same exponential probability density $R(t) = ke^{-kt}$ as shown in Figure 4. Suppose that v is any time in $(0, \beta]$. Let $f(v)$ be the number of new certificate revocations between day v and day $v + \Delta t$, from all of the valid generations.

$$\begin{aligned} f(v) &= \alpha b\%R(v) + \alpha b\%R(v - \Delta t) + \\ &\quad \alpha b\%R(v - 2\Delta t) + \ldots + \\ &\quad \alpha b\%R[v - (n-1)\Delta t] \end{aligned} \tag{2}$$

where n is the number of generations in time period β; that is, $n = \lceil \frac{\beta}{\Delta t} \rceil$.

Assuming that Δt is one day, and that v is an integer, where v is in $(0, \beta]$, then we get the following equation.

$$\begin{aligned} f(v) &= \alpha b\%R(1) + \alpha b\%R(2) + \ldots + \alpha b\%R(v) \\ &= \alpha b\%ke^{-k} + \alpha b\%ke^{-2k} + \ldots + \alpha b\%ke^{-vk} \\ &= \alpha b\%ke^{-k}\frac{1 - e^{-vk}}{1 - e^{-k}} \end{aligned} \tag{3}$$

When v is in (β, ∞), we have

$$\begin{aligned} f(v) &= \alpha b\%R(1) + \alpha b\%R(2) + \ldots + \alpha b\%R(\beta) \\ &= \alpha b\%ke^{-k} + \alpha b\%ke^{-2k} + \ldots + \alpha b\%ke^{-\beta k} \\ &= \alpha b\%ke^{-k}\frac{1 - e^{-\beta k}}{1 - e^{-k}} \end{aligned} \tag{4}$$

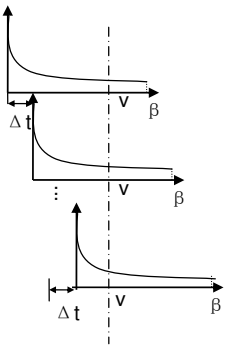

Figure 4: Model for pooling revocations from different generations

Equations 3 and 4 show that the number of new revocation requests CA received on daily basis increases with a decreasing rate as time elapses from day zero until day β. After that, the number of revocation requests on daily basis becomes a constant number. Figure 5 shows the number of new certificate revocations on daily basis in $(0, 2\beta]$ for the case where $\alpha = 1000$, $b\% = 10\%$, $k = 0.26$, and $\beta = 36$ days. From now on we omit the graph for $(m\beta, (m+1)\beta]$, where $m \geq 2$, because the shape in those regions are the same as that in $(\beta, 2\beta]$.

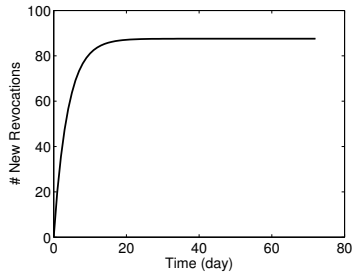

Figure 5: $f(v)$ behavior: the number of new certificate revocations on daily basis

The Size of CRL The CRL, if issued on day v, includes the new revocation requests on day v as well as the valid historical revocation requests (occurred before day v) whose expiration day is later than v. Let $F(v)$ be the valid cumulative number of certificate revocations from time 1 to v, where "valid" means not expired. This is also the size for the CRL if CA decides to publish it on that day. For any time $v \in (0, \beta]$, we have

$$\begin{aligned} F(v) &= \sum_{t=1}^{v} f(t) = \sum_{t=1}^{v} \alpha b\% k e^{-k} \frac{1-e^{-tk}}{1-e^{-k}} \\ &= \frac{\alpha b\% k e^{-k}}{1-e^{-k}} [v - \frac{e^{-k}}{1-e^{-k}}(1-e^{-vk})] \quad (5) \end{aligned}$$

Considering the $(\beta+1)$ day, the revoked certificates from the first generation are expired, and thus removed from the CRL. At this time, the number of valid generations is $(\beta+1)-1 = \beta$.

Considering the $(\beta+2)$ day, revocation from the first two generations are expired, and thus removed from the CRL. At that time the number of valid generations is also $(\beta+2)-2 = \beta$.

The rest may be deduced similarly. For any $v \in (\beta, +\infty)$, we have

$$\begin{aligned} F(v) &= F(\beta) = \sum_{t=1}^{\beta} f(t) \\ &= \frac{\alpha b\% k e^{-k}}{1-e^{-k}} [\beta - \frac{e^{-k}}{1-e^{-k}}(1-e^{-\beta k})] \quad (6) \end{aligned}$$

Equations 5 and 6 show that the size of CRL on daily basis with respect to time is a convex function. It increases with an increasing rate as time elapses from day 0 until the time reaches issued age β. After that, the size of CRL becomes a constant number. The reason it is not going to be infinite is that at any time some revoked certificates may be expired and removed from CRL. Figure 6 shows the daily CRL size in $(0, 2\beta]$ for the case of $\alpha = 1000$, $b\% = 10\%$, $k = 0.26$, and $\beta = 36$ days.

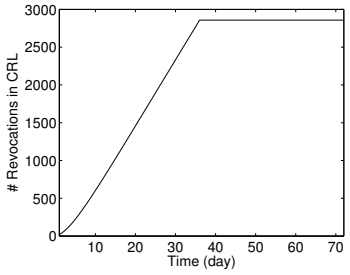

Figure 6: $F(v)$ behavior: daily size of CRL

Percentage of revocations Let $P(v)$ be the percentage of certificate revocations occurred from time v to time $v + \Delta t$, which is defined as the number of new certificate revocations at time v divided by the cumulative valid number of certificate revocations from time 1 to v. For any time $v \in (0, \beta]$, we have

$$\begin{aligned} P(v) &= f(v) / \sum_{t=1}^{v} f(t) \\ &= \frac{\alpha b\% k e^{-k} \frac{1-e^{-vk}}{1-e^{-k}}}{\frac{\alpha b\% k e^{-k}}{1-e^{-k}} [v - \frac{e^{-k}}{1-e^{-k}}(1-e^{-vk})]} \\ &= \frac{1}{\frac{v}{1-e^{-vk}} - \frac{e^{-k}}{1-e^{-k}}} \quad (7) \end{aligned}$$

For any time $v \in (\beta, +\infty)$, the number of new certificate revocations at time v is always equal to $f(\beta)$, and the cumulative valid number of certificate revocations from time 1 to v is always equal to $F(\beta)$. Thus,

$$\begin{aligned} P(v) &= f(\beta) / \sum_{t=1}^{\beta} f(t) \\ &= \frac{\alpha b\% k e^{-k} \frac{1-e^{-\beta k}}{1-e^{-k}}}{\frac{\alpha b\% k e^{-k}}{1-e^{-k}} [\beta - \frac{e^{-k}}{1-e^{-k}}(1-e^{-\beta k})]} \\ &= \frac{1}{\frac{\beta}{1-e^{-\beta k}} - \frac{e^{-k}}{1-e^{-k}}} \end{aligned} \quad (8)$$

Equations 7 and 8 show that the majority of the certificate revocations occur at early stage of the issued age of certificates, right after they were issued. Figure 7 shows the graph of percentage of certificate revocations in $(0, 2\beta]$ for the case of $\alpha = 1000$, $b\% = 10\%$, $k = 0.26$, and $\beta = 36$ days.

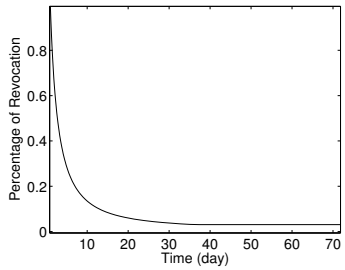

Figure 7: $P(v)$ behavior: percentage of revocations

5.1.3 Second Case: Overlap of Certificates with different Issued Ages

In the first case, we assume that CA issues a fixed number of certificates at different time, but each time it is the same type of certificates that issued with the same issued age. Now we relax these assumptions to a more general case by assuming that at any point of time CA can issue two types of certificates with different issued ages β_1 and β_2, where $\beta_2 > \beta_1$. We assume that these two types are independent of each other. Under these assumptions, we compute the new $F(v)$ and $P(v)$ in different intervals of $(0, \beta_1]$, $(\beta_1, \beta_2]$, and $(\beta_2, +\infty)$ correspondingly.

We overlap two types of certificates with the distribution functions $R_1(t) = k_1 e^{-k_1 t}$ and $R_2(t) = k_2 e^{-k_2 t}$. At the same time, each type of certificates is composed of generations of different certificates issued at different time.

The size of CRL on daily basis Now the daily CRL size $F(v)$ is a cumulative number of the revocations of two types of certificates. For any time $v \in (0, \beta_1]$, we have

$$F(v) = \sum_{t=1}^{v} f_1(t) + \sum_{t=1}^{v} f_2(t) \quad (9)$$

In the interval $(\beta_1, \beta_2]$, the size of CRL for the certificates whose issued age is β_1 has become stable and the value of $F_1(v)$ will be constant, while that for the certificates whose issued age is β_2 keeps increasing. For any $v \in (\beta_1, \beta_2]$, we have

$$F(v) = F_1(\beta_1) + \sum_{t=1}^{v} f_2(t) \quad (10)$$

In the interval $(\beta_2, +\infty)$, both CRLs become stable, and thus

$$F(v) = F_1(\beta_1) + F_2(\beta_2) \quad (11)$$

Figure 8 shows the graph of the cumulative numbers of valid certificate revocations in $(0, +\infty)$ for the case $k_1 = 0.26$, $\beta_1 = 36$, $k_2 = 1$, $\beta_2 = 72$, $\alpha_1 = 2000$, $\alpha_2 = 1000$, $b_1 = 10\%$, and $b_2 = 10\%$.

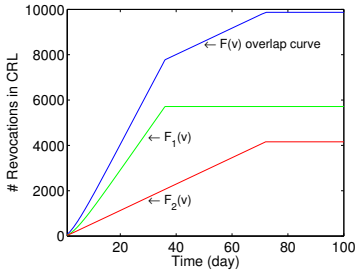

Figure 8: $F(v)$ behavior: overlap certificates with different issued ages

Percentage of Revocations Let $P(v)$ be the percentage of certificate revocations occurred between v and $v + \Delta t$ for the pooling case. For any $v \in (0, \beta_1]$, we have

$$P(v) = \frac{f_1(v) + f_2(v)}{\sum_{t=1}^{v} f_1(t) + \sum_{t=1}^{v} f_2(t)} \quad (12)$$

For any $v \in (\beta_1, \beta_2]$, we have

$$P(v) = \frac{f_1(\beta_1) + f_2(v)}{\sum_{t=1}^{\beta_1} f_1(t) + \sum_{t=1}^{v} f_2(t)} \quad (13)$$

For any $v \in (\beta_2, +\infty)$, we have

$$P(v) = \frac{f_1(\beta_1) + f_2(\beta_2)}{\sum_{t=1}^{\beta_1} f_1(t) + \sum_{t=1}^{\beta_2} f_2(t)} \quad (14)$$

Figure 9 shows the graph of percentage of revocations in interval $(0, +\infty)$ for the case $k_1 = 0.26$, $\beta_1 = 36$, $k_2 = 1$, $\beta_2 = 72$, $\alpha_1 = 2000$, $\alpha_2 = 1000$, $b_1 = 10\%$ and $b_2 = 10\%$. Overall, the behaviors of $F(v)$ and $P(v)$ are almost the same as those in case 1.

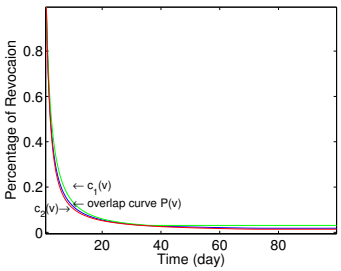

Figure 9: $P(v)$ behavior: overlap certificates with different issued ages

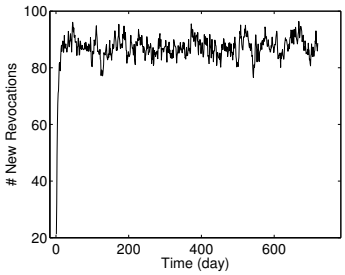

Figure 10: $f(v)$: daily number of new certificate revocations with Poisson distribution

5.1.4 Third Case: Simulation in the Case of Poisson Distribution

In our first case, we assume that CA issues a fixed number of certificates at different time. A more general case is that CA issues α certificates, where α is a random number following a Poisson distribution with parameter μ. For this case, the average number of revocation requests per interval is μ. The probability that there are x revocation requests occurred in each interval is $P_\mu(x) = \frac{\mu^x e^{-\mu}}{x!}$. Because the explicit forms of $f(v)$ and $F(v)$ are messy, we uses simulation to get some insights.

We conducted our simulation on a HP 1940 PC (with Pentium 4 CPU and 1.00GB RAM) using Visual C++. We follow the steps below in our simulation:

1. Firstly, generate a sequence of random number α_1, α_2, $\alpha_3, \ldots, \alpha_m$ based on the value of μ.
2. Secondly, compute the total number of new revocations on daily basis between 1 and 2β, which is $f(v)$.
3. Thirdly, compute the valid cumulative number of revoked certificates from day 1 to day 2β, which is $F(v)$.
4. At last, generate different groups of random numbers for α, repeat step 1 to step 4 for twenty times.

Figure 10 shows a typical case for daily numbers of new certificate revocations with Poisson distribution when $\beta = 360$ and $k = 0.26$. The number of new certificate revocations increases sharply to about 90 within a very short period of time after the certificates get issued. Instead of becoming stable as in case 1 and case 2, the curve fluctuates around 90 after a short period of time. The oscillation is driven by the randomness introduced by using Poisson distribution.

Figure 11 shows that $F(v)$ continues increasing from 1 to β, where $\beta = 360$. After β, it begins to fluctuate. In a actual business environment, a typical issued age β is 360 days, which is much bigger than Δt. In such a case, the stable value of λ in Figure 11 is so large that it dominates the fluctuation introduced by Poisson. Consequently, the curve is very smooth after β. This is consistent to the case 1 when we assume a fixed number of certificates issued at any point of time. When CA decides how often CRL should be released, it mostly cares about the distribution of $F(v)$. Because of the existence of the similarity between the fixed number case and the Poisson distribution case for the $F(v)$ distribution, later for our economic analysis, we will focus on the fixed number case.

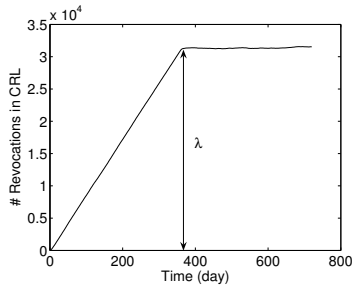

Figure 11: $F(v)$: daily size of CRL with Poisson distribution

5.2 Analytical Model: How Often Should CA Release CRLs

The key research question in our paper is to give a prescription to CA to decide how often it should release its CRLs. In order to answer that question, CA must know the distribution of new certificate revocation $f(v)$ and the distribution of certificate revocation list $F(v)$. CA needs to balance the liability cost of not releasing CRL on time and the fixed and variable costs of releasing CRL too often. So the goal for CA is to minimize the overall operational cost. Because the behaviors of $f(v)$ and $F(v)$ when time t is greater than the issued age β are different

from those when time t is smaller than β, we analyze the optimal strategies for CA for these two cases separately. For each case we assume a monopoly case so that there is no competition between CAs, and that the certificates are homogeneous in terms of risk, cost, and revocation probability. Also, different types of certificates are independent from each other. CA will get an optimal CRL releasing strategy for each type of certificates based on properties of the certificates.

5.2.1 Optimal Releasing Strategy When Time is Greater Than β

When time is greater than β, CA has run certificate services for at lease one issued age for that type of certificates. We will use the following variables in our analysis (the numbers in parentheses are the default values used in our computation).

- β: The issued age of one type of CRL. ($\beta = 360$)
- c: The estimated number of days between two CRL releasing dates. This is the decision variable that CA needs to optimize.
- θ: Estimated numbers of certificates in a CRL on a given day after issued age β if CA decides to release CRLs on that day. According to case 1, $\theta = F(\beta) = \frac{\alpha \cdot b\% \cdot k \cdot e^{-k}}{1-e^{-k}}(\beta - \frac{e^{-k}}{1-e^{-k}})$. ($\theta = 32,000$, $k = 0.26$, and $\beta = 360$)
- FC: The fixed cost for CA to publish one CRL .($FC = \$10,000$)
- VC: The variable cost for CA to include each individual certificate into a CRL. We assume the VC does not changed with the length of CRL. ($VC = \$1$)
- Υ: The expected liability cost per certificate revocation if CA delay publishing the revocation for one day. Therefore, the risk of delaying publishing a CRL of θ certificate revocations for n day is $\Upsilon \cdot \theta \cdot n$. If we assume that for the whole period of β, the expected liability cost that CA pays for the accident caused by the delay of publishing CRLs is Qm; (i.e.,$Qm = \$100,000$), then $\Upsilon = Qm/(\theta * \beta) = \$0.0087$ for $\theta = 32,000$ and $\beta = 360$.
- a: Recency requirement set up by the customers. It is the max number of days between two successive CRL releasing dates that is acceptable by customers. ($a = 50 days$)

Because $f(v)$ and $F(v)$ are stable after β, CA can take either a fixed interval strategy or a fixed CRL size strategy for releasing CRL. Fundamentally these two strategies are inter-exchangeable. For simplicity reason, we present the solution for the fixed interval strategy.

If CA releases one CRL every c days, the total cost for CA within period β is

$$cost(c) = [\Upsilon \cdot \theta \cdot \sum_{n=0}^{c-1} n + FC + \theta \cdot VC] \cdot \frac{\beta}{c} \quad (15)$$

Optimization model The problem converts to the following optimization problem:

$$\begin{cases} \min . \, cost(c) \\ s.t. FC \gg VC > 0, \Upsilon > 0, c \leq a \end{cases} \quad (16)$$

According to Karush-Kuhn-Tucker theorem[5][12], we get

$$\begin{cases} \frac{\partial[cost(c)+L(c-a)]}{\partial c} = 0 \\ c - a \leq 0 \\ L \geq 0 \\ L(c-a) = 0 \end{cases} \quad (17)$$

In order for $L \cdot (c-a) = 0$ to hold, it is required either (i) $L = 0$ or (ii) $c - a = 0$. We consider case (i) and case (ii) in the following.

Case (i) If $L = 0$, then

$$\begin{cases} \frac{\partial[cost(c)+L(c-a)]}{\partial c} = 0 \\ c - a \leq 0 \end{cases} \quad (18)$$

Compute the first derivation of the cost with respect to c, and get the optimal result c_0:

$$\begin{cases} c_0 = \sqrt{(\frac{2}{\Upsilon})(\frac{FC}{\theta} + VC)} \\ c_0 \leq a \end{cases} \quad (19)$$

Compute the second derivation of cost with respect to c, and the result is

$$\frac{\partial^2 cost(c)}{\partial c^2} = \frac{2(FC \cdot \beta + \theta \cdot VC \cdot \beta)}{c^3} \quad (20)$$

Because the second derivation of $cost(c)$ at point c_0 is

$$\frac{\partial^2 cost(c)}{\partial c^2}|_{c=c_0} = \frac{\Upsilon \cdot VC \cdot \beta}{\sqrt{(\frac{2}{\Upsilon})(\frac{FC}{\theta} + VC)}} > 0, \quad (21)$$

$cost(c)$ reaches its minimum value at c_0. The minimum operational cost of CA is

$$\begin{aligned} &[\Upsilon \cdot \theta \cdot \sum_{n=0}^{c_0-1} n + FC + \theta \cdot VC] \cdot \frac{\beta}{c_0} = \\ &\Upsilon \cdot \theta \cdot \beta(c_0 - \frac{1}{2}) = \\ &\Upsilon \cdot \theta \cdot \beta(\sqrt{(\frac{2}{\Upsilon})(\frac{FC}{\theta} + VC)} - \frac{1}{2}) \end{aligned} \quad (22)$$

Case (ii) If $c - a = 0$, the function is

$$\begin{cases} L = \frac{FC \cdot \beta + \theta \cdot VC \cdot \beta}{\Delta t^2} - \frac{\Upsilon \cdot \theta \cdot \beta}{2} \geq 0 \\ c = a \end{cases} \quad (23)$$

The minimal cost of CA is fixed and equal to

$$[\Upsilon \cdot \theta \cdot \sum_{n=0}^{a-1} n + FC + \theta \cdot VC] \cdot \frac{\beta}{a} = \frac{1}{2} \cdot \Upsilon \cdot \theta \cdot \beta \cdot (a-1) + \frac{\beta}{a}(FC + \theta \cdot VC) \quad (24)$$

when

$$a \leq \sqrt{(\frac{2}{\Upsilon})(\frac{FC}{\theta} + VC)} \quad (25)$$

We can see that the minimum releasing interval CA should follow is either a or $c_0 = \sqrt{(\frac{2}{\Upsilon})(\frac{FC}{\theta} + VC)}$ depending on whether $\sqrt{(\frac{2}{\Upsilon})(\frac{FC}{\theta} + VC)}$ is greater than a or not. It is clear that $\sqrt{(\frac{2}{\Upsilon})(\frac{FC}{\theta} + VC)}$ is an increasing function of FC and VC, but a decreasing function of Υ and β, where $\theta = F(\beta) = \frac{\alpha b\% k e^{-k}}{1-e^{-k}}[\beta - \frac{e^{-k}}{1-e^{-k}}(1 - e^{-\beta k})]$. That means if the fixed cost or the variable cost is higher, or the liability cost is lower, or the issued age of the certificates is shorter, CA should release CRLs less frequently.

Value study Now we use an empirical example to demonstrate how much money CA can save by following our strategy. The best waiting days to achieve the minimal cost is $c = \sqrt{(\frac{2}{\Upsilon})(\frac{FC}{\theta} + VC)} = 17.37 \leq a = 50$ days. Figure 12 shows the total cost for CA by using different releasing strategies. If CA deviates from the $c_0 = 17$days by using $2 \cdot c_0 = 34$days, CA ends up spending almost $\$400,000$ more for just one type of certificates within a period of β. This is not a trivial number given that there are multiple CAs providing numerous certificate services.

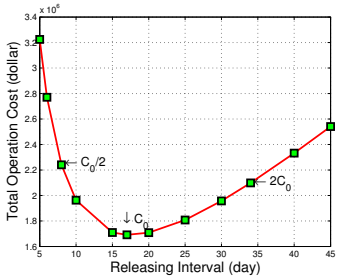

Figure 12: Total cost of CA with different releasing interval after β

5.2.2 Optimal Releasing Strategy When Time is Smaller Than β

There are two possible business scenarios for this case: 1) A "grown-up" CA that has been in CRL business for quite a while, but faces the situation of providing CRL services for a new type of certificates. 2) A "start-up" CA that just begin to provide CRL services. For these two cases, v is inside $(0, \beta]$, and $F(v)$ is a convex function with respective to v. CA can take either a fixed interval or a fixed size strategy for releasing CRLs. The fixed size means that a CA will release the CRLs whenever the number of certificates included in the CRL exceeds a fixed pre-specified number. Next we analyze both cases by using simulation. For each case, CA can get the parameter estimators either based on other types of certificates it provides before or from its industry peers.

Fixed interval strategy Similar to the analysis given for the case when time is greater than β, we can obtain the cost function when the time is smaller than β:

$$\begin{aligned} &cost(c) = \\ &\frac{\Upsilon ab\% k e^{-k}}{1-e^{-k}} \sum_{x=0}^{\frac{\beta}{c}-1} \sum_{n=0}^{c-1} (c-n)(1 - e^{-(xc+n+1)k}) + \\ &\frac{\beta}{c} FC + F(\beta) VC = \\ &\frac{\Upsilon ab\% k e^{-k}}{1-e^{-k}} (\beta c - \frac{\beta(c-1)}{2} + \\ &\frac{((c+1)e^{-2k} - ce^{-k} + e^{-(c+2)k})(1 - e^{-\beta k})}{(1-e^{-k})^2(1-e^{-ck})}) + \\ &\frac{\beta}{c} FC + \\ &\frac{\alpha b\% k e^{-k}}{1-e^{-k}} [\beta - \frac{e^{-k}}{1-e^{-k}}(1 - e^{-\beta k})] VC \end{aligned} \quad (26)$$

For each possible c, ranging from 1 to β, we compute the total cost for CA. Figure 13 shows[6] the total cost for CA by using different releasing intervals. We find that when $c = 28$ days, we get the minimal cost $\$2.08603 * 10^5$. This means that when time is smaller than β, CA should release CRLs once every 28 days. Recall that the optimal interval is 17 days when time is greater than β. It is easy to know that $cost(17) = \$2.64014 * 10^5 > cost(28) = \$2.08603 * 10^5$ in the period of $(0, \beta]$; therefore, CA should take different strategies for time periods $(0, \beta]$ and $(\beta, +\infty)$.

Fixed size strategy We will use the following variables for analyzing the fixed size strategy:

- d: CA will publish a new CRL if the number of certificate revocations exceeds d.
- q: Estimated numbers of CRLs that CA will publish during one issued age from time 0 to β.

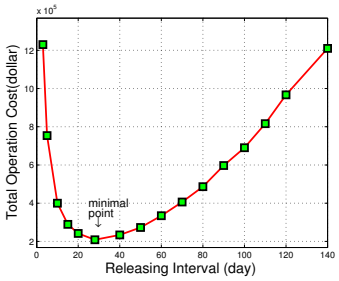

Figure 13: Total cost of CA with different releasing interval before β

- i: The i-th CRL published by CA, where $0 < i \leq q$.
- Nd_i: CA releases the CRL on the Nd_i-th day.
- $F(v)$: Size of CRL at time v.

Then we have

$$F(v) = \frac{\alpha b\% k e^{-k}}{1-e^{-k}}[v - \frac{e^{-k}}{1-e^{-k}}(1-e^{-vk})] \quad (27)$$

In order to estimate Nd_i, we need to compute the inverse function of $F(v)$, which is denoted as $G(d)$ (i.e., $G(d) = F^{-1}(v)$. After that, we can get the exact day Nd_i, on which CA needs to release its CRLs, by solving $Nd_1 = G(d)$, $Nd_2 = G(2*d), \ldots, Nd_i = G(i*d)$.

Given the LanbertW function defined as

$$LambertW(x) * exp(LambertW(x)) = x \quad (28)$$

we have

$$\begin{aligned} G(d) &= LambertW \cdot d(e^k - 1) \\ &\cdot \frac{e^{(-\frac{d(e^k)^2 - 2de^k + d + 100k}{100e^k})}}{100e^k k} \\ &+ \frac{de^{2k} - 2de^k + d + 100k}{100e^k k} \end{aligned} \quad (29)$$

We conduct our simulation step by step. Firstly, for each possible d chosen from 100 to $36,000$ ($a = 1000$, $b\% = 10\%$), compute $Nd_1, Nd_2, \ldots, Nd_i$. Secondly, calculate the time difference between Nd_i and $Nd_i - 1$, which we call c_i, to estimate the liability cost. Thirdly, compute $cost(d)$ for each individual d as following

$$\begin{aligned} cost(d) &= \frac{\Upsilon \cdot a \cdot b\% \cdot k \cdot e^{-k}}{1-e^{-k}} \\ &\times \sum_{j=0}^{q-1} \sum_{x=0}^{c_i - 1} (c_i - x)(1 - e^{-((i-1)c_i + x + 1)k}) \\ &+ q \cdot FC + F(\beta) \cdot VC \end{aligned} \quad (30)$$

We find that the minimal cost is $\$2.26790 * 10^5$ when $d = 2800$. The minimal cost is very similar to the result that we obtained using the optimal fixed interval. Therefore, when time is smaller than β, CA can take either a fixed size strategy or a fixed interval strategy. But CA can no longer follow the same optimal releasing interval as in the case when time is bigger than β.

Figure 14 shows the total cost for CA to use different size strategies. The cost for CA is minimal when $d = 2800$.

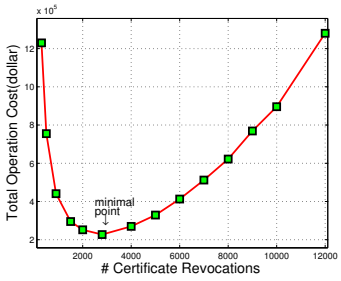

Figure 14: Total cost of CA with different size strategy before β

Figure 15 shows the relationship between releasing time and cumulative revocations when $q = 100$ and $k = 0.26$ for the fixed size strategy. Here we assume that $d = 2800$. That means whenever the size of CRL reaches 2800, CA will release it. That is the reason we see 2800, 5600, 8400, and so on along x-axils. As we can tell, as time moves away from time 0, the releasing interval between two successive CRL releasing dates remains almost unchanged. The fixed size strategy is almost equivalent to the fixed interval strategy at their respectively optimal points.

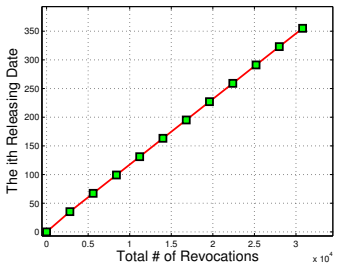

Figure 15: The i-th releasing day vs. cumulative revocations

To summarize, different types of CAs should take different CRL releasing strategies for the same type of certificate services, and the same CA should also use different mechanisms for different types of certificate services.

6 Discussions and Conclusions

In this paper, we analyze real empirical data collected from VeriSign to derive probability density function of certificate revocations. Unlike most previous research, our work is conducted based on real data. The contributions of this paper include: 1) We prove that a revocation system will become stable after a period of time; 2) CA should take different strategies when providing certificate services for a new type of certificates versus a re-serving type of certificates; 3) A start-up CA and a grown-up CA should take different strategies for CRL release; 4) We give the exact steps by which a CA can derive optimal CRL releasing strategies; and 5) We prove that a CA should release CRLs less frequently in the case that the fixed cost is higher, the variable cost is higher, the liability cost is lower, or the issued age of certificates is shorter.

There are several limitations for this study. First, this paper takes a static approach by assuming that there is no correlation between different types of certificates, and that customer behaviors do not affect CA's releasing strategy for deriving the optimal CRL releasing intervals. A more realistic approach is to use game theory to model the interactions between CAs and customers. Second, this paper assumes that CA offers certificates with a fixed issued age. To further minimize the total operational cost, CA may optimize not only the releasing time interval but also the issued age simultaneously.

References

[1] ARNES, A. Public key certificate revocation schemes. Master's thesis, Norwegian University of Science and Technology, 2000.

[2] COOPER, D. A. A model of certificate revocation. In *ACSAC '99: Proceedings of the 15th Annual Computer Security Applications Conference* (Washington, DC, USA, 1999), IEEE Computer Society, p. 256.

[3] FOX, AND LAMACCHIA. Certificate revocation: Mechanics and meaning. In *FC: International Conference on Financial Cryptography* (1998), LNCS, Springer-Verlag.

[4] GUNTER, C. A., AND JIM, T. Generalized certificate revocation. In *Symposium on Principles of Programming Languages* (2000), pp. 316–329.

[5] HOUSLEY, R., FORD, W., POLK, W., AND SOLO, D. RFC 2459: Internet X.509 public key infrastructure certificate and CRL profile, Jan. 1999. Status: PROPOSED STANDARD.

[6] JAIN, G. Certificate revocation: A survey. http://citeseer.ist.psu.edu/511985.html.

[7] KOCHER, P. C. On certificate revocation and validation. In *FC '98: Proceedings of the Second International Conference on Financial Cryptography* (London, UK, 1998), Springer-Verlag, pp. 172–177.

[8] LI, N., AND FEIGENBAUM, J. Nonmonotonicity, user interfaces, and risk assessment in certificate revocation (position paper). In *Proceedings of the 5th Internation Conference on Financial Cryptography (FC'01)* (2002), no. 2339 in LNCS, Springer, pp. 166–177.

[9] MCDANIEL, P., AND RUBIN, A. A response to "can we eliminate certificate revocation lists?". *Lecture Notes in Computer Science 1962* (2001), 245+.

[10] MYERS, M., ANKNEY, R., MALPANI, A., GALPERIN, S., AND ADAMS, C. X.509 internet public-key infrastructure — online certificate status protocol (OCSP). Internet proposed standard RFC 2560, June 1999.

[11] NAOR, M., AND NISSIM, K. Certificate revocation and certificate update. In *Proceedings 7th USENIX Security Symposium (San Antonio, Texas)* (Jan 1998).

[12] POLAK, E. Computational methods in optimization.

[13] RIVEST, R. L. Can we eliminate certificate revocations lists? In *Financial Cryptography* (1998), pp. 178–183.

[14] STUBBLEBINE, S. Recent-secure authentication: Enforcing revocation in distributed systems. In *Proceedings 1995 IEEE Symposium on Research in Security and Privacy* (May 1995), pp. 224–234.

[15] WOHLMACHER, P. Digital certificates: a survey of revocation methods. In *MULTIMEDIA '00: Proceedings of the 2000 ACM workshops on Multimedia* (New York, NY, USA, 2000), ACM Press, pp. 111–114.

[16] ZHENG, P. Tradeoffs in certificate revocation schemes. *Computer Communication Review 33*, 2 (2003), 103–112.

Notes

[1]This work was conducted when Chengyu Ma visited Singapore Management University

[2]http://sign.nca.or.kr/english/english.html

[3]http://www.mozilla.or.kr/zine/?cat=10

[4]We delete those records whose existence ages are zero.

[5]Karush-Kuhn-Tucker condition is a necessary and sufficient optimality condition for constrained optimization problems.

[6]For demonstration purpose, we assume that α is large enough so that CA can adopt a fixed interval determined by any optimal value of our model.