Quantifying the Effects of More Timely Certificate Revocation on Lightweight Mobile Devices

Sufatrio and Roland H.C. Yap
National University of Singapore
Singapore
Outline

- Motivation and Goals
- Empirical Revocation Request
- Our Revocation Analysis Framework
- Application to Revocation Schemes
- Performance Evaluation
- Conclusions
Motivation

• Public Key Infrastructure (PKI):
  – Widely used in SSL, S/MIME, IKE, code authentication, signed documents, …

• Certificate revocation:
  – Overheads are important
  – Timeliness + scalability tradeoffs
  – Two standardized mechanisms:
    • CRL: high bandwidth, (generally) low timeliness
    • OCSP: high timeliness, high CA computation
New Concerns

Two emerging trends on the Internet:

1. Explosion of connected mobile devices:
   - Revocation must be lightweight on verifiers
2. Growth in both volume and value of transactions:
   - Higher timeliness revocation is needed

- A need to quantify (using a realistic framework) the effects of higher timeliness certificate revocation on lightweight mobile devices!
Goals and Approach

• Goal: to derive an analytical framework to quantify the overheads of certificate revocation with:
  - *Higher* timeliness guarantee (with revocation latency in hours or minutes, *not* days)
  - Based on a recent empirical study by Ma et al. that most revocations occur *early*, *not* (uniformly) at the certificates’ mid-life
  - Probability of certificates to be queried post-revocation *decreases* over time, *not* remaining constant

• We use the framework to compare different certificate revocation schemes:
  - Bandwidth, computation, storage costs
Certification System

• We are interested to measure the properties of a system in its *steady state*:
  – Commonly assumed
  – Fixed number of valid and not-yet-expired certs = \(N\)
  – Certificate issuance is balanced out by expiration

• (A single-CA) certification system:
  – Uniform cert lifetime of \(\beta\) days
  – Percentage of cert revoked \(b = 10\%\) (assumed from a report study by MITRE)
  – Time interval between two cert generations = \(\Delta X\)
  – Time interval between two CRL releases = \(\Delta t\)
Empirical Revocation Model by Ma et al.

- Ma et al. (Usenix Security ’06) found that:
  - Most revocations occur *early* in the certificates’ lifetimes
  - The certificate’s percentage *decreases* over time
- Revocation distribution can be modeled as a PDF: 
  \[ R(t) = ke^{-kt}, \quad k=0.26 \]
How to Derive the CRL Size?

- It is always assumed $\Delta X = \Delta t = 1$ day
- Function $f(v) = \text{the no. of new revocations between day } v \text{ and } v + \Delta t$ (with $\alpha = \frac{N}{\beta}$ as the no. of certs per generation or day):
  \[
  f(v) = \alpha b R(v) + \alpha b R(v - \Delta t) + \alpha b R(v - 2\Delta t) + \ldots + \alpha b R(v - (n - 1)\Delta t),
  \]
  (2)
- In the steady state condition:
  \[
  f(v) = f(\beta) = \alpha b R(1) + \alpha b R(2) + \ldots + \alpha b R(\beta)
  = \alpha b k e^{-k} \frac{1 - e^{-\beta k}}{1 - e^{-k}}.
  \]
  (3)
Visual Description

- The no of *new* revocations in the steady state:

How about $abR(\beta)$??
- Yes, they are revoked
- But, should *not* be added into CRL due to expiration! (We’ll fix this later)

Total no of revoked certs:
(to be listed in the CRL)

$abR(\beta) = 0$

+ $abR(\beta-1)$

+ $\ldots$

+ $abR(2)$

= $\mathcal{R}(\beta) \, f(\beta-1)$

with $\beta-1$ generations counted
Number of Entries in the CRL

- Defined as function $F(v)$: accumulation of $f(v)$
- In the steady state condition, the no of entries in the CRL (as in Ma et al.):

$$F(v) = F(\beta) = \sum_{t=1}^{\beta} f(t)$$

$$= \frac{\alpha bke^{-k}}{1-e^{-k}} \left[ \beta - \frac{e^{-k}}{1-e^{-k}} (1-e^{-\beta k}) \right].$$  \hfill (4)

- The formula corrected as: $F(\beta-1) = f(1)+f(2)+\ldots+f(\beta-1)$, \textbf{not} $f(\beta)$ as the last term
Visual Description of $F(v)$ (with $\beta$-1 Generations Counted)

Total cumulative revoked certs:

$$f(\beta-1) + f(\beta-2) + \ldots + f(1) = F(\beta-1)$$

$\text{Not } F(\beta)$

Notice that $f(v)$ in Ma et al. uses discretization on the PDF (leads to CRL entries of ~5.7M certs for $\Delta t = 1$ hour $\gg 100K$ certs)
Our Refinement on CRL Size:

(1) When $\Delta X = \Delta t = 1$ day

- We define $g(v)$ to replace $f(v)$: use integration

$$g(1) + g(\beta^{-2}) + \cdots + g(1) = F(\beta^{-1})$$

Total cumulative revoked certs:

$$\int_{1}^{\beta^{-1}} R(t) \, dt$$

$$\int_{\beta^{-2}}^{\beta^{-1}} R(t) \, dt$$

$$\int_{\beta^{-2}}^{2} R(t) \, dt$$

$$\int_{0}^{1} R(t) \, dt$$

Total $= \int_{0}^{\beta^{-1}} R(t) \, dt$
Our New Simpler Derivation

• Cert generation \(i\) day ago:

\[
h(i) = \alpha b \int_0^i R(t) \, dt. \tag{9}
\]

• There are \(\beta-1\) cert generations to be counted

\[
Rev\_Entries = \sum_{i=1}^{\beta-1} h(i) = \sum_{i=1}^{\beta-1} \alpha b \int_0^i R(t) \, dt
= \alpha b \left[ (\beta - 1) - e^{-k} \frac{1 - e^{-k(\beta-1)}}{1 - e^{-k}} \right], \tag{10}
\]
Our Refinement on CRL Size:

(2) When $\Delta X = \Delta t$ (in mins)

- We use two scaling factors:
  - $\gamma = \frac{\Delta X}{M}$ : $\Delta X$ in day unit
  - $\lambda = \frac{\Delta t}{M}$ : $\Delta t$ in day unit
- Number of certs per issuance:
  $$\alpha' = \frac{N\gamma}{\beta} = \alpha \gamma .$$
- The number of entries in the CRL (with $\eta=\beta'/\gamma−1$ as the no of cert generations counted):
  $$\text{Rev\_Entries}' = \sum_{i=1}^{\eta} h'(\gamma i) = \sum_{i=1}^{\eta} \alpha' b \int_0^{\gamma i} R(t) \, dt$$
  $$= \frac{N\gamma}{\beta} b \left[ \left( \frac{\beta}{\gamma} - 1 \right) - e^{-k\gamma} \frac{1 - e^{-k(\beta-\gamma)}}{1 - e^{-k\gamma}} \right],$$

Scaling Factors
Our Refinement on CRL Size:

(3) When $\Delta X = c.\Delta t$ (in mins)

Two cases:

– CRL release coincides with cert generation/issuance:

$$ Rev\_Ent\_at\_gen = \sum_{i=1}^{\frac{\beta}{\gamma}-1} h' (\gamma i) , \quad (16) $$

– CRL release takes place $j.\Delta t$ after the last cert generation:

$$ Rev\_Ent\_after\_gen(j) = \sum_{i=1}^{\frac{\beta}{\gamma}} h' ((i-1)\gamma + j\lambda) \quad (15) $$

$$ = \frac{N\gamma}{\beta} b \left[ \frac{\beta}{\gamma} - e^{-k\lambda j} \frac{1-e^{-k\beta}}{1-e^{-k\gamma}} \right] . $$

– Average:

$$ \frac{Rev\_Entries'}{c} = \frac{1}{c} \left( \sum_{j=1}^{c-1} Rev\_Ent\_after\_gen(j) + Rev\_Ent\_at\_gen \right) . \quad (17) $$
More Realistic Query Model on Revoked Certificates

• Once a certificate is revoked, does its probability to be queried remain the same as before (when valid)?
  • What assumptions on the probability over time?
  • Example:
    • A Web server with previously revoked cert now uses a new cert for its SSL-based transactions
    • A binary code signed with previously revoked cert is reissued with the new cert
  • Some verifiers may still have stale references
  • Overall reasonable to assume that the no of queries on revoked certs to **decrease over time**
Two Query Models on Revoked Certificates Used

- We model the probability with an exponential function (other functions are possible):
  \[ S(t) = a e^{-r \lambda t}. \]  
  (18)

- Two values of \( r \) are used:
  - \( r = 0.95 \) forming \( S_1(t) \) → accounting for 80% certs in the evaluation
  - \( r = 0.01 \) forming \( S_2(t) \) → accounting for the other 20% certs
Probabilities of Querying Revoked and Valid Certificates

- The no of queries on all revoked certs (of probability model $m$) throughout $\beta$ days:
  \[ Q_{rev\_\beta\_days}(m) = \sum_{i=1}^{\frac{\beta}{\lambda} - 1} \left( N_m \cdot b \cdot \int_{\lambda(i-1)}^{\lambda i} R(t) \, dt \right) \cdot \left( \sum_{j=1}^{\frac{\beta-\lambda}{\lambda}} S_{m}(j) \cdot q_{per\_interval} \right) \]  
  \[ \tag{19} \]
- Total no of queries on all revoked certs in $\beta$ days:
  \[ Q_{rev\_\beta\_days} = \sum_{m=1}^{n} Q_{rev\_\beta\_days}(m). \]  
  \[ \tag{20} \]
- Probabilities of queries on a revoked and valid cert:
  \[ Pr_{rev} = \frac{Q_{rev\_\beta\_days}}{Q_{all\_\beta\_days}}, \quad Pr_{valid} = \frac{Q_{valid\_\beta\_days}}{Q_{all\_\beta\_days}}. \]  
  \[ \tag{21} \]
Evaluation: Parameters

- Application of the analytical framework:
  - Measures the costs of various cert revocation schemes in the steady state under the same scenario
  - Revocation latency: 1 hour; 10 minutes
  - Single CA certification system with:
    - No of valid and not-yet-expired certs $N = 100,000$
    - Uniform cert lifetime $\beta = 365$ days
    - Percentage of cert revoked $b = 10\%$
    - Number of Verifiers = 30,000,000
    - Average daily queries needed by a Verifier = 30
    - A designated repository (also called CMAE) used in some schemes (e.g. CRL, CRS)
Performance Metrics

• We compare CRL, OCSP, CRS, CREV

• Performance cost metrics:
  – Daily certificate creation cost (only on CA)
  – Daily update costs: both computation and bandwidth costs
  – Daily query costs: both computation and bandwidth costs
  – Storage requirement (on CA and directory)

• CPU costs for PCs based on Crypto++ 5.6.0

• CPU cost for verifier based on RSA-1024 (CRT-RSA) on a HTC Touch Dual 2007 with Qualcomm 400 MHz processor
Sample Cost Derivations

- Data size of CRL ($L$ denotes length):
  \[ L_{CRL} = L_{CRL\_fields} + \left\lceil \frac{\text{Rev\_Entries}}{\ell} \right\rceil \cdot L_{CRL\_entry} \]
  with $L_{CRL\_fields} = 400$ bytes (CRL header and signature)
  $L_{CRL\_entry} = 39$ bytes (a revoked entry in CRL)

- Daily bandwidth cost CA-CMAE:
  \[ D_{BwCA-CMAE} = \frac{M}{\Delta t} \cdot L_{CRL} \]

- Average verifier’s computation cost in CRS:
  \[ Pr_{rev} \cdot OvhVer_{rev} + Pr_{valid} \cdot OvhVer_{valid\_avg} \]
  with:
  - Average valid cost: \[ OvhVer_{valid\_avg} = \frac{\ell+1}{2} \cdot C_{hash} \]
  - Average revoked cost: \[ OvhVer_{rev} = C_{hash} \]
Key Performance Factors

• Under latency of 1 day; 1 hour; 10 mins

<table>
<thead>
<tr>
<th>Performance Factors</th>
<th>Revocation Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 day</td>
</tr>
<tr>
<td>( \text{Rev_Entries} )</td>
<td>9880.33</td>
</tr>
<tr>
<td>( CRL\ Size ) (KB)</td>
<td>376.68</td>
</tr>
<tr>
<td>( Pr_{rev} )</td>
<td>0.005170</td>
</tr>
<tr>
<td>( Pr_{valid} )</td>
<td>0.994830</td>
</tr>
</tbody>
</table>

• No of CRL entry is \( \sim 98\% \) of all revoked certs:
  – Compared to \( 5,000 \) (mid-life cert revocation assumption)
  – With \( N = 100,000, \Delta t = 1 \) day, the difference in CRL data size = \( 190,720 \) bytes (\( 186 \) KB)

• Verifier is almost certain (\( Pr_{valid} \approx 0.995 \)) to perform repeated hash operations in CRS:
  – Compared to \( \sim 0.91 \) (constant post-revocation query probability)
  – Or \( \sim 0.95 \) (also assume mid-life revocation)
Evaluation: CRL, OCSP, CRS, CREV

- Revocation latency of 1 hour:

| Entity | (U=Update, Q=Query) | (sec/day) | CRL | OCSP | CRS | CREV-
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>D_OuthCA-CMAE (U)</td>
<td>sec</td>
<td>0.36</td>
<td>1.07×10^6</td>
<td>3.39×10^5</td>
<td>3876</td>
</tr>
<tr>
<td></td>
<td>储CA-CMAE (U)</td>
<td>sec</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-CMAE (Q)</td>
<td>MB</td>
<td>0.37</td>
<td>0</td>
<td>8357</td>
<td>1.53×10^5</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-CMAE (U)</td>
<td>MB</td>
<td>8.84</td>
<td>-</td>
<td>67.92</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-CMAE (Q)</td>
<td>MB</td>
<td>-</td>
<td>2.15×10^8</td>
<td>-</td>
<td>0.0077</td>
</tr>
<tr>
<td>CMAE</td>
<td>D_OuthCMAE (U+Q)</td>
<td>sec</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>StorCMAE</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-CMAE (U)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-CMAE (Q)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EVCP</td>
<td>D_OuthEVCP (U+Q)</td>
<td>sec</td>
<td>0.70</td>
<td>0.70</td>
<td>16.73</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>StorEVCP</td>
<td>MB</td>
<td>4.38×10^-4</td>
<td>6.0×10^-4</td>
<td>-</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-EVCP (U)</td>
<td>MB</td>
<td>-</td>
<td>0.01</td>
<td>-</td>
<td>3.15</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-EVCP (Q)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-EVCP (Q)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-EVCP (Q)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-EVCP (Q)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-EVCP (Q)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Verifier</td>
<td>D_OuthVer (Q)</td>
<td>sec</td>
<td>8.84</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-EVCP (Q)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-EVCP (Q)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>D_BwCA-EVCP (Q)</td>
<td>MB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Revocation Latency: mins

- OCSP: High CA’s computation
- CRL: High network bandwidth
- OCSP & CREV are lightweight (computation & bandwidth) on Verifier
- CREV: CREV-II improves CREV-I with smaller computation and bandwidth (in exchange for CA’s storage)
- CRS: High CA’s storage, Higher verifier’s computation
Comparison Results

- Under revocation latency of 10 minutes:

  - CREV: Computation + storage still manageable

<table>
<thead>
<tr>
<th>Entity</th>
<th>Daily Costs (U + Q)</th>
<th>Unit (/day)</th>
<th>CRL</th>
<th>OCSP</th>
<th>CRS</th>
<th>CREV-I</th>
<th>CREV-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td></td>
<td>sec</td>
<td>0.21</td>
<td>$1.33 \times 10^6$</td>
<td>2.03 $\times 10^{-4}$</td>
<td>22608</td>
<td>1296</td>
</tr>
<tr>
<td>CMAE</td>
<td></td>
<td>sec</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>EVCP</td>
<td></td>
<td>MB</td>
<td>0.37</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Verifier</td>
<td></td>
<td>MB</td>
<td>53.05</td>
<td>-</td>
<td>407.50</td>
<td>7506.56</td>
<td>1491.55</td>
</tr>
</tbody>
</table>

- CRL: network bandwidth ↑↑

- OCSP: CA’s computation ↑↑

- CRS: CA’s storage ↑↑, Verifier’s computation ↑↑

- CREV: Computation + storage still manageable

- OCSP & CREV: still lightweight on Verifier
Conclusions

• We have proposed a more realistic analytical framework for evaluating cert revocation schemes
• It enhances the revocation model in Ma et al., and incorporates more realistic features than existing work
• We show how to apply the framework to evaluate CRL, OCSP, CRS and CREV
• The tradeoffs in the schemes are clearly shown