Temporal Analysis of X.509 Revocations and their Statuses

Adam Halim
Linköping University
Sweden

Max Danielsson
Linköping University
Sweden

Martin Arlitt
University of Calgary
Canada

Niklas Carlsson
Linköping University
Sweden

***Abstract*—Despite the X.509 public key infrastructure (PKI) being essential for ensuring the trust we place in our communication with web servers, the revocation of the trust placed in individual X.509 certificates is neither transparent nor well-studied, leaving many unanswered questions. In this paper, we present a temporal analysis of 36 million certificates, whose revocation statuses we followed for 120 days since first being issued. We characterize the revocation rates of different certificate authorities (CAs) and how the rates change over the lifetime of the certificates. We identify and discuss several instances where the status changes from "revoked" to "good", "unauthorized" or "unknown", respectively, before the certificate's expiry. This complements prior work that has observed such inconsistencies in some CAs' behavior after expiry but also highlight a potentially more severe problem. Our results highlight heterogeneous revocation practices among the CAs.**

1. Introduction

The security of almost every website visit hinges on the trust we place in the validity of X.509 certificates issued by Certificate Authorities (CAs). By issuing a certificate, a CA verifies the mapping between a domain (or organization) and one of the domain's public keys.

During normal circumstances, each certificate is valid for a pre-defined period (i.e., the *validity period*) specified in the certificate. However, there are several cases when the trust in a certificate must be *revoked* before its validity period expires [1]. For example, a CA may be asked to invalidate a certificate when the private key associated with the certificate is compromised, the domain is no longer used, or an organization has ceased to exist.

Today, a revocation is typically performed in one of two ways: using a Certificate Revocation List (CRL) or the Online Certificate Status Protocol (OCSP). CRLs work by having CAs periodically issue a timestamped list of revoked certificates (i.e., the CRL) that a browser can download and use to check whether a certificate has been revoked [1]. With OCSP, each CA maintains a server endpoint against which the browser can direct revocation status requests for individual certificates [2]. Due to performance and privacy related aspects associated with these protocols, some of the leading browsers do not perform these revocation checks [3], [4] but instead use proprietary alternatives to periodically push lists with prioritized revocations to their clients [5]–[8].

While the above-mentioned lists (e.g., Google's CRLSets [6] and Mozilla's OneCRL [7]) include some revoked certificates, these lists only include a very small and selective subset of all revocations [3], [4]. To study the revocations that take place in the wild, we instead study the revocation status of certificates by periodically making OCSP requests for a large number of certificates.

The temporal analysis approach is inspired by the recent work by Korzhitskii and Carlsson [9] who used a similar approach to study what happened to the revocation statuses of revoked certificates after a certificate expired. However, while their work raised several interesting problems with current practices (e.g., instances where some CAs changed the status from "revoked" to "good") their temporal analysis only included the time period after expiry, did not capture the timing of the revocations, and their data collection did not capture the revocation reasons provided via OCSP (only CRLs).

In this paper, we extend that work. First, we identify 36 million unique and newly issued certificates. We then monitor the status of these certificates for the first 120 days after issuance. Of particular interest here are the timing of the revocations, the reason for the revocations, and the degree that status changes happen also during the regular validity period of a certificate. For our data collection, we designed and implemented a data collection framework (Section 2) that allows us to quickly identify all newly issued certificates logged by the major Certificate Transparency (CT) [10] logs and that then periodically (every 24 hours) performs status checks of all these certificates. Tools and datasets will be shared with the paper [11].

Overall, our analysis highlights big differences in the revocation patterns observed for different CAs (e.g., in terms of revocation rates, timings, and revocation reasons). These differences highlight heterogeneous revocation practices but may also in part be an effect of individual CAs adapting their practices based on their individual customer bases. We also identify and discuss three types of status changes in which the observed status changes from "revoked" to "good", "unauthorized" or "unknown", respectively, before the certificate's expiry. The occurrence of such instances raises more questions regarding why they may have taken place, to what degree caching in CDNs may impact client security, and further emphasizes the need for revocation transparency protocols.

Outline: Section 2 presents our data collection. Sections 3 and 4 present a characterization of the revocations of different certificate types issued by different CAs. Next, we present the timing-based analysis (Section 5) and analysis of status changes that differentiate from current expectations (Section 6). Finally, Section 7 discusses related works before Section 8 concludes the paper.

2. Data collection

The collection was split into two partially overlapping phases. During the first phase, which lasted for seven days, we continually identified all newly issued certificates submitted to the most popular Certificate Transparency (CT) logs.[1] The use of CT logs is ideal for this purpose since both Google and Apple require a certificate to be logged in a CT log before their browsers trust the certificate [13], [14]. During the second phase, we monitored the status of each of these certificates every 24 hours for 120+ days. We next describe each phase in more detail.

Identification of newly issued certificates (Phase 1): We developed a multithreaded tool in Go based on the LogClient struct in Google's CT log repository as well as several functions in the `crypto/x509` and `encoding/pem` packages. At a high level a LogClient is created for each CT log of interest, all LogClients are stored in an array, and for each CT log, we then use `get-sth` and `get-entries` requests to (1) determine if the tree size has been updated and (2) download all certificates and their respective certificate chains, respectively. The retrieved certificates are then processed and uploaded to a database running MongoDB [15].

Daily status checking (Phase 2): We split the certificates from phase 1 into 24 collection groups (numbered $0-23$) based on the hour they were issued. We then performed periodic status checks (following the format specified in RFC6960 [16]) for each of these groups every 24 hours using `cron` [17]. This part was implemented using the `crypto/ocsp` package [18].

To perform an OCSP query, the certificate along with its issuer's certificate is sent to the OCSP URL. In response, the server sends the certificate's status. In the case of status "revoked" a revocation time and an optional revocation reason are also sent. For requests that throw an error, the response is logged together with a timestamp. For CRL checks, the program downloads the CRL and checks to see if the certificate is included in the list.

Duplicates removal: Certificates can be logged by multiple CT logs. While our tool includes several optimizations (see Appendix), we did not check if a certificate already has been logged. Instead, we identify and remove duplicates before the characterization.

Ethical statement: Data were collected from public infrastructures using public protocols. While our measurements add some load to the OCSP servers (see Appendix), this load is small compared to the overall request load they typically would see. Furthermore, we report and discuss odd certificate behaviors that we observe with the CAs.

3. High-level Characterization

Status report method: We first identified each certificate's OCSP servers and CRLs, when available. In total, 99.98% of the certificates provided an OCSP server, 16.54% a CRL, and only 0.0015% (539 in total) did not provide any of the two methods. The big differences in adoption rate is mostly due to some of the most popular CAs not using CRLs (e.g., Let's Encrypt) or only partially (e.g., DigiCert 42.3%, Sectigo 1.1%). In contrast, CAs are required (by the CA Browser Forum [19]) to provide OCSP servers that can answer status queries for every certificate they issue from the time the certificate is issued to the time they expire.

In the following, we focus on the certificates that provided OCSP responses. We tracked these certificates daily for 120+ days since first being issued (and logged).

Dataset and average revocation rate: The dataset contains 35,958,651 unique certificates issued and logged on May 2-8, 2021. Out of these, 222,540 (0.62%) were revoked during our observation period (phase 2). This revocation rate is comparable to what has been reported by other recent studies (during times when there is no mass-revocation event [20] or the revocations of the mass-revocation event are discounted from the rate [9]).

We next characterize the revoked certificates and discuss biases in the revoked set, who issued the certificates, and the revocation reasons provided by the CAs.

3.1. Certificate and key characteristics

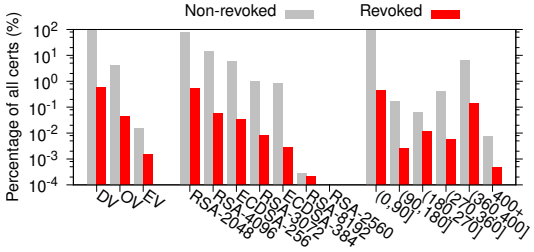
Consider first the revocation rates of different certificate subsets. Figure 1 provides a summary. Here, certificate categories are divided along three dimensions: (1) validation type, (2) public key, and (3) validity period.

Validation type: The validity type of each certificate was determined using their Object ID (OID) [19]. As expected, we observe a high skew in usage, with most certificates being Domain Validation (DV) certificates (95.97%). In contrast, Organization Validation (OV) certificates (4.01%) and Extended Validation (EV) certificates (only 0.02%) see much smaller usage. Interestingly, the revocation rates are the opposite, with EV certificates having a revocation rate of 8.6%, OV certificates 1.1%, and DV 0.60%. (In the figure we also include a line for the overall average of 0.62%.)

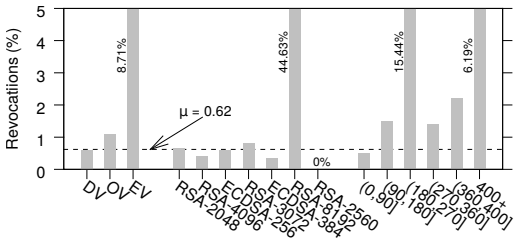
Public key type and size: Among the three most popular key types, certificates including RSA 2048 keys (78.7%) had a higher revocation rate than the two others (RSA 4096 responsible for 13.7% and ECDSA-256 for 5.8%). While this at first may suggest that websites using somewhat stronger keys are somewhat less likely to revoke the certificates, the differences are too small to draw any conclusions (e.g., 0.66% vs. 0.42% vs. 0.60%) and we note that the revocation rate for certificates with RSA 8192 is very high (44.6%). None of the 77 revoked RSA 8192 certificates came with a revocation reason and all but 4 were issued for keytalk.com.

Validation period: Ignoring a big spike in revocation rates for certificates with validation periods in the range (180,270], the revocation rates tend to be increasing with increased validation periods. While many of the observed differences are due to differences in the revocation rates of individual CAs, we expect that part of the observed trend is due to the shorter certificate lifetimes allowing websites to easily and naturally phase out certificates (hence reducing the need for some revocations due to "cessation of operation" or "affiliation change", for example).

1. To get a representative collection, we collected all certificates logged to the major, non-test logs of each log provider listed on Merkle Town [12]: Argon 21/22 (Google), Xenon 21/22 (Google), Oak 21/22 (Let's Encrypt), Nessie 21/22 (DigiCert), Yeti 21/22 (DigiCert), Nimbus 21/22 (Cloudflare), Mammoth/Sabre (COMODO), 360.cn 21/22. Only concerned with newly issued certificates, we did not use any 201X or 2020 logs.



(a) Percentage of certificates associated with each category



(b) Revocation rate for different certificate categories

Figure 1. Summary of the frequency each certificate category was observed and each categories respective revocation rates. Here, we break down categories along three different dimensions: (1) validation type, (2) public key type, and (3) validation period.

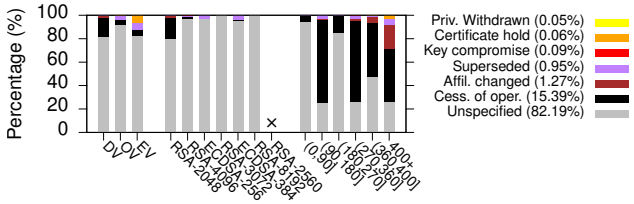


Figure 2. Certification reason breakdown for each category.

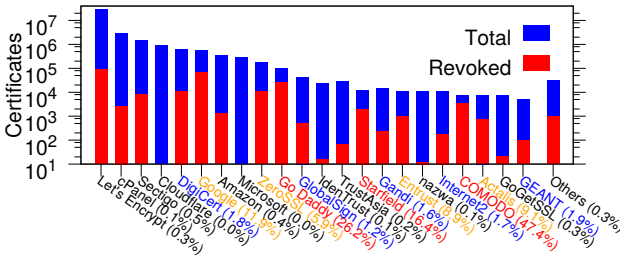


Figure 3. Certificates logged (blue+red) and revoked (red) per CA. We include all CAs that had logged at least 5,000 unique certificates.

Here, it is also important to note that the revocation rates of the certificates with lifetimes longer than 120 days (all belonging to the other categories) only can be seen as a lower bound, since some certificates that were not yet revoked at the end of our 120+ day collection period still may have been revoked by the time they expired. If looking at the full lifetime of the certificates the differences are therefore expected to increase.

The higher revocation rates for certificates with validation periods in the ranges (180,270] days (15.4%) can be explained by a large number of revoked COMODO certificates. For example, out of the 4,285 revoked certificates with a duration in the interval (180,270] days, 3,398 were issued by COMODO.

Similarly, Let's Encrypt and their 90-day certificates play a big role in keeping the revocation rates low of the certificates with a validity period in the range (0,90], as they are responsible for most of these certificates and had a below average revocation rate. While certificates with duration of 400+ days are much rarer (due to recent policy changes [19], [21]–[23]), also here, the larger revocation rate observed for these certificates (6.2%) most of the revoked certificates are issued by one CA; in this case, GoDaddy who is responsible for 106 out of 173 revoked certificates. In Section 4 we look closer at the revocation patterns of individual CAs.

3.2. Revocation reasons

Figure 2 summarizes the revocation reasons specified in the status responses [24]. We make several observations.

First, most revocations (82.2%) are for an "unspecified" reason. While prior work has seen similar numbers, this large fraction is important to highlight as it raises the question whether the community wants to push for increased revocation transparency? Without transparency (similar to what CT provides for issuance) it is difficult to identify and evaluate risks associated with malicious campaigns that manage to compromise several organizations' keys, for example. On the flip side, many organizations may not want to reveal that a key has been breached.

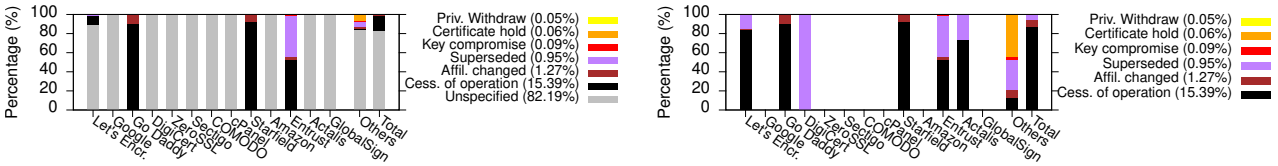
Second, most of the revocations associated with longer-lived certificates are due to "cessation of operation" and "affiliation changes". Furthermore, most of these certificates are DV certificates. Third, EV certificates stand out with the by far largest fraction of "certificate holds" (6.4%). Out of the 123 certificates that had revocation status "hold", 34 were EV certificates, 89 were OV certificates, and none were DV certificates.

Finally, we looked closer at the 194 certificates that had revocation reason "key compromised" (0.09%). In general, almost all these certificates used RSA keys (136 with length 2048 and 56 with length 4096) and only two used ECDSA-256. While these numbers may suggest a bigger (relative) fraction of RSA 4096 keys being compromised, the bias appears to be due to indirect biases in which keys certain CAs promote and how the keys may be handled. It should also not be seen as a reflection of the security of the keys themselves. We have also seen somewhat larger fraction of OV certificates (0.20%). For example, 32 of the 194 certificates were OV. All the other 162 certificates were DV certificates; none were EV. We did not observe any significant biases in the validity periods. Here, 144 certificates had validity periods of no more than 90 days, 3 fell within the (270,360] range, and 47 within the (360,400] range.

4. CA-based comparison

Big differences in revocation rates per CA: Figure 3 shows a histogram with the number of revocations (red) and issued (red+blue) certificates for all CAs with at least 5,000 issued certificates (23 CAs in total). Here, we group all revocations by other CAs in an "other" category and show the revocation rate of each CA as a percentage together with the label of each CA. To ease comparison, we also color-code each label: below 0.5% (black), 0.5-2% (blue), 2-15% (yellow), and above 15% (red).

We observe large differences in the percentage of certificates that each CA revokes, ranging from 0.0%



(a) All revoked certificates

(b) Only certificates with specified revocation reasons

Figure 4. Per-CA breakdown of revocation reasons. We include all 13 CAs with at least 500 unique revocations, ordered from most to least revocations.

(Cloudflare and Microsoft) to 47.4% (COMODO). The three CAs with the highest revocation rates are COMODO (47.4%), GoDaddy (26.2%), and Starfield (16.4%), The high revocation rate of COMODO is likely due to many customers moving to other CAs (when replacing their certificates [25]) and in the case of the latter two CAs the high rates appears to be due to their customer base frequently ceasing operation or changing domain (see analysis later in this section). These CAs appear preferred by transient domain owners. Among the top-10 CAs (with most issued certificates) six CAs had revocation rates below 0.5% (from left to right): Let's Encrypt (0.3%), cPanel (0.1%), Sectigo (0.5%), Cloudflare (0.0%), Amazon (0.4%), and Microsoft ($3.58 \cdot 10^{-6}$). These results are consistent with the results presented by [9].

In addition to differences in customer base, the CAs also target different use cases. For example, most of the certificates issued by Microsoft had the common name "Microsoft Azure TLS Issuing CA" with domain names within the Microsoft Azure domain, suggesting that they were issued to cloud applications running on Microsoft Azure. The lower revocation rates of such certificates may suggest that these certificates are less likely to be deemed to need revocation than regular web domains. However, given that several of the CAs target domain certificates, the big differences also raise questions regarding the revocation policies used by different CAs.

Big differences in use of revocation reasons: While most CAs do not give a revocation reason, some do. Figure 4(a) provides a breakdown of the revocation reasons used for each CA with at least 500 observed revocations (13 CAs in total) and an "other" category (with 1.2% revocations). (Figure 4(b) shows the corresponding statistics only including certificates with a specified revocation reason.) We first note that two of the three CAs that provide revocation reasons for almost all their certificates (GoDaddy and Starfield) also are among the three CAs we previously noted had the highest revocation rates. In the cases of GoDaddy and Starfield, most of the revocations are due to "cessation of operation" (and a noticeable fraction where there has been an "affiliation change"). The large number of "cessation of operation" observed with GoDaddy may be due to them being a popular web hosting company for individuals that may be more likely to cancel their web hosting plan compared to companies using many of the other CAs. This is also seen when looking at all the certificates for which "cessation of operation" was given as reason. Out of the 34,258 certificates with this revocation reason, 23,839 were issued by GoDaddy, 8,041 by Let's Encrypt, 1,817 by Starfield, 521 by Entrust, 8 by Actalis, and 32 by CAs outside the top-13. Given these numbers, it is perhaps not surprising that GoDaddy also dominates the use of "Affiliation change" as revocation reason. Out of the 2,827 certificates with this revocation reason, 2,601 were issued by GoDaddy, 143 by Starfield, 34 by Let's Encrypt, 26 by Entrust, and 23 by CAs outside the top-13 (with at least 500 revocations).

With Starfield being a spin-off from GoDaddy that uses the same revocation policies, it is also not surprising that their revocation percentages and reason breakdowns are similar. In the case of Entrust, the split is closer to 50-50 between "cessation of operation" and the certificate being "superseded". Most of the 2,114 cases of "superseded" were by Let's Encrypt (1,410 cases), Entrust (428), and GoDaddy (142). Also, Digicert (2), Starfield (17) and Actalis (3) had such cases. The remaining 88 cases were from CAs outside the top-13.

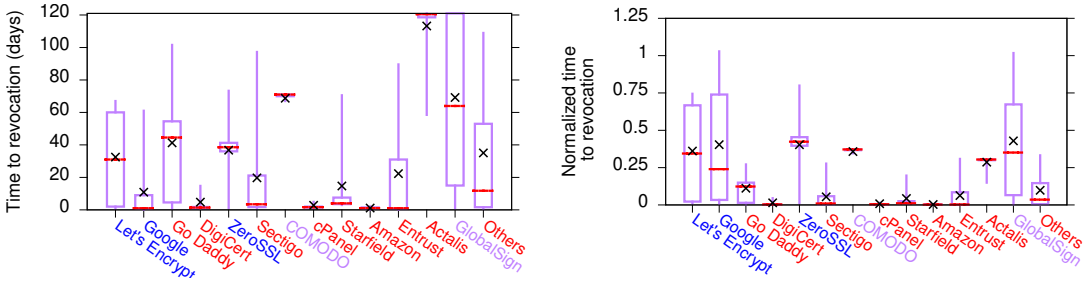
Among the less common reasons we have also observed 194 cases of "key compromise", 118 cases of "privilege withdrawn", and 123 cases of "certificate hold". Out of the "compromised key" cases, we observed 144 cases for Let's Encrypt, 27 for Entrust, 10 for GoDaddy, 6 for Starfield, and 7 for CAs outside the top-13 list. "Privilege withdrawn" were only observed by GoDaddy (108 cases) and Starfield (10 cases), and "holds" (123 cases) were only observed by CAs outside the top-13.

In summary, the above cases are interesting since they highlight that the CAs take highly diverse approaches. For example, we never observed a revocation reason for seven out of the 13 CAs with at least 500 observed revocations: Google, ZeroSSL, Sectigo, COMODO, cPanel, Amazon, GlobalSign. Furthermore, among the CAs for which we observed revocation reasons, the degree to which they provided a reason differed substantially (i.e., Figure 4(a)) and when they provided a reason the reasons the used differed significantly (as best exemplified by the numeric per-reason breakdowns above but also seen among the more popular reasons visible in Figure 4(b)).

5. Timing-based analysis

We have observed significant CA-based differences when revocations take place. Figure 5 shows whisker plots for the revocation timings broken down per CA using both (a) the absolute time since issuance and (b) the relative time normalized with regards to the validity period of each certificate. Here, we show the 5-percentile (bottom marker), 25-percentile (bottom of box), median (middle/red line), 75-percentile (top of box), 95-percentile (top marker), and the average (black $\times$ marker).

Before interpreting these results, we note that different CAs have different validity periods. To capture this, in the above figures, we use blue x-axis labels for the CAs for which most observed certificates have a validity period no longer than 91 days, red labels for those with mostly validity periods in the range 360-400 days, and purple labels for the CAs with mostly intermediate validity periods.



(a) Absolute time (b) Normalized time

Figure 5. Revocation timing observed for different CAs. We include all 13 CAs with 500+ unique revocations, ordered from most to least revocations.

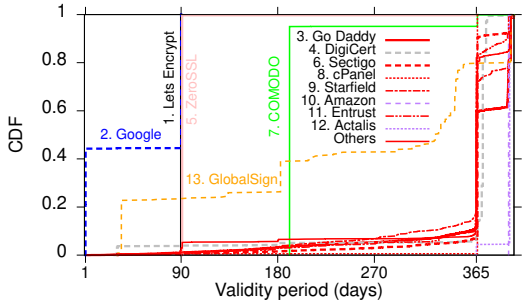


Figure 6. Validity period of revoked certificates broken down per CA.

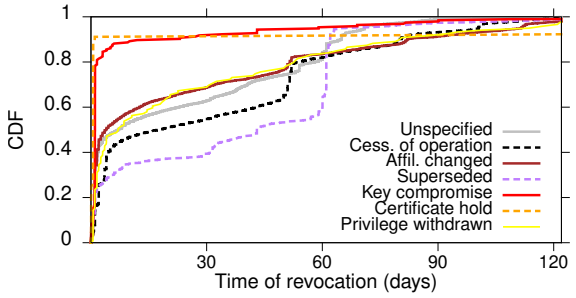


Figure 7. Time until revocation broken down based on revocation reason.

Furthermore, Figure 6 shows the cumulative distribution functions (CDFs) of the observed validity periods of each CA. Here, a very interesting observation is that 44.3% of the revoked Google certificates have a validity period of only one day. These certificates appear to be test certificates (mapping to one of many generic subdomains belonging to one of two Google-owned domains: haplorrhini.com, gkemanagedcerts.certsbridge.com) or to tarsier-monitoring.appspot.com (also Google-owned). This observation also explains why we see very early revocations (e.g., median of 1.04 days in Figure 5(a)) but the normalized revocation timings look somewhat higher (e.g., median of 0.24 in Figure 5(b)).

Early revocations: For seven of the CAs (Google, DigiCert, Sectigo, cPanel, Starfield, Amazon, Entrust), most revoked certificates (indicated by the medians) were revoked within 4.0 days of issuance and (with exception of Google) within 0.011 of their respective lifetimes. For these CAs it is clear that many of the revocations are requested almost immediately after initial issuance.

Spikes in revocations: In contrast, we see sharp intermediate spikes in the revocation rates (e.g., see small inter-quartile distances) for three of the CAs: (1) ZeroSSL typically issues 91-day certificates and has clear spikes around 36-41 days. (2) COMODO typically issues 191-day certificates and has a clear spike around the 70-day age. (3) Actalis mostly issues 396-day certificates and we observed a clear spike around age 118-120 days. While the first two cases provide clear well-defined spikes, we are cautions to say too much about the Actalis case (as this spike took place close to the end of our analysis period).

Steadier revocation rates: Let's Encrypt and GlobalSign had steadier revocation rates than most other CAs. This is captured by the bigger relative variations in revocation times using both absolute and normalized metrics.

Early revocations: While early revocations dominated across reason codes, revocations of "key compromise" and "holds" were in general earlier. This is shown in Figure 7.

6. Questionable status changes

Finally, we report on the cases were the reported status changed from "revoked" to "good", "unauthorized", or "unknown" (error response). Table 1 in the Appendix provides a detailed summary of these results and in the following we present a self-contained analysis of these results. In contrast to the work by [9], who only considered status changes that took place after a certificate has expired, we consider only status changes that took place before expiry. Furthermore, here we only report statistics for certificates that otherwise validated against the three major root stores operated by NSS, Apple, and Microsoft.

"Revoked" to "good": Without revocation transparency, revocation history can go missing. During the studied 120-day period, we observed 2 GoDaddy certificates that changed status to "good" for two days just to be changed back to "revoked" after that. The two GoDaddy certificates (amanprintersdelhi.in and thegoni.co.uk) are DV certificates with revocation reason "Cessation of Operation". In both cases RSA-2048 were used and in both cases the only snapshots that the waybackmachine (https://web.archive.org/) finds are from after the status changes. Since when these pages were first recorded by the engine (Nov. 2021 and Dec. 2021, respectively), there have been few or no changes.

When writing up the paper we also observed 28 certificates issued by QuoVadis and 5 certificates issued by HydrantID that have changed their status to "good" again (after the 120-day period but before expiry). These certificates typically cycled through statuses "revoked" and "unauthorized" and/or "unknown", before finally switching to "good" (or flipping between "good" and "unauthorized/unknown"). In all these cases the revocation reason was "hold". While it technically is okay to temporary invalidity a certificate (using reason "hold"), the information that a certificate has been on "hold" probably should be preserved as it suggests that the certificate's integrity at some point may have been in question.

"Revoked" to "unauthorized": For the second class, we observed big diversity in the type (50 DV, 15 OV, 7 EV), issuer (34 GoDaddy, 12 QuoVadis, 9 Google, 7 DigiCert, 4 Fiducia & GAD IT AG, 3 Amazon, 2 HydrantID, 1 Let's Encrypt), and reason code (27 "Cessation of Operation", 20 "Unspecified", 18 "Certificate Hold", 6 "Affiliation changed", 1 "Superseded"). Given the small numbers and high diversity, we first suspected that this behavior may be driven by CAs trying to satisfy special requests made by their customers rather than common practices by the CAs. However, looking closer at the time sequences, we observed that the status often repeatedly changed between "revoked" and "unauthorized" (sometimes via "unknown"). After talking to some involved CAs, we now believe that the issue may be cache related and associated with us periodically hitting out-of-date responses stored on CDN nodes. It is concerning that this issue often persists for long time periods, suggesting the CDN caches can have out-of-date information for a long time period. For example, in 50% of the cases we observed "unauthorized" status (typically repeatedly and going back and forth) more than 110 days after the certificate was revoked. This also suggests that most certificates were revoked early (120 day period).

We also found it interesting that this group consists of a significant fraction of EV and OV certificates. Compared to DV certificates, these certificates make up a smaller fraction of the total number of certificates (e.g., Figure 1) and are more expensive. The seven EV certificates are also interesting since EV certificates are considered to provide a higher level of validation. While this does not always translate into higher security, some clients may still place higher trust in EV certificates than DV certificates. All seven EV certificates observed associated with this class were issued by QuoVadis, had reason code "Certificate Hold", and included an RSA-2048 key. These certificates were also among the set of certificates that eventually were reported to have "good" status.

"Revoked" to "unknown": This case was by far the most observed case of the three (changing from "revoked" status to something else). While the first two cases occurred 2 and 72 times, respectively, this case was observed 3,490 times among the certificates that chained back to a valid root in all of the three considered root stores (NSS, Apple, Microsoft). This corresponds to $9.0 \cdot 10^{-6}$, $3.2 \cdot 10^{-4}$, and 0.016 (1.6%) of all revoked certificates. Also here, the top-5 CAs differed significantly compared to both the ranking of the CAs with most revoked certificates (Figure 3). For example, the five CAs with the most such instances (i.e., 1,286 GoDaddy, 1,292 DigiCert, 403 Sectigo, 136 cPanel, 136 Amazon) had rankings 3, 4, 6, 8, 10 with regards to number of revoked certificates (e.g., order in Figures 4 and 5). Furthermore, these CAs had rankings 6, 5, 3, 2, 7 with regards to number of issued certificates (Figure 3).

As shown in the appendix, we have observed a relatively larger fraction of the revoked OV/EV certificates and certificates that specified a revocation reason (other than "key compromise"). While discussions with CAs suggests that some of these instances may be due to servers periodically reporting error codes, we have not observed anything suggesting that the CAs have risked the safety of their customers.

7. Related work

Certificate Transparency (CT) [10] has been studied from many perspectives [26]–[33], including as a whole [26], [27], compliance [29], and the logged certificates [32], [33]. Here, we use CT logs as a data source for newly issued certificates.

Both CRLs [1] and OCSP [16] have several drawbacks [3]–[5], leading to them often being ignored [3] or replaced by proprietary revocation lists [5], [6], [8]. Several other revocation solutions [20], [34]–[38], novel PKIs [39]–[43], and a revocation transparency protocol [44] have been proposed but are not currently deployed. Chuat et al. [4] presents an evaluation framework and comparison of different revocation protocols.

The revocation rates have been studied under both normal circumstances and during mass-revocation events [3], [9], [45], [46]. The work closest to ours is the work by Korzhitskii and Carlsson [9]. In their work they used periodic measurements against OCSP servers to study what happened to the revocation statuses of revoked certificates after a certificate has expired. In contrast, we follow certificates from the day of issuance, focus on the time period before expiry, and provide an analysis of both the timing and reason for the revocations.

The revocation rates observed in this paper are similar in magnitude to those recently observed by Korzhitskii and Carlsson [9] and Smith et al. [20]. Other researchers have studied certificate replacements associated with the mass-revocations [45]–[47], regular certificates [25], [48], and invalid certificates [49].

Other related measurements have shown that most (94%) OCSP responses are served using CDNs [50] but that OCSP responders still were not sufficiently reliable to support the *OCSP Must-staple* extension [51]. Liu et al. [3] also observed that only 0.35% of the revocations were covered by Google's CRLSets [6].

8. Conclusions

We have presented a temporal analysis of the revocation status responses provided by OCSP responders. The analysis identifies and compares revocation patterns observed (e.g., with regards to rates, timing, reasons, etc.) for certificates issued by different CAs and associated with different validation type, key usage, and validation periods. We also observe a non-negligible number of certificates for which the CA changes the status from *revoked* to *good* (2 cases), *unauthorized* (79 cases), and *unknown* (3,490 cases). While such changes are easier to explain after expiry [9], these cases raise new questions regarding why these instances occur. We are currently contacting several of the CAs to ask for the reasons behind these instances. Overall, our analysis highlights big differences in the revocation patterns observed for different CAs which may stem from differences in their customer base and how much they adapt their practices based on the customers they serve. In future work, we plan to extend the analysis to a 400+ day period (capture the lifetime of most modern certificates) and look closer at some of the temporal behaviors of the observed revocations.

References

[1] D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley, and W. Polk, "Internet x.509 public key infrastructure certificate and certificate revocation list (crl) profile," RFC 5280, 2008.

[2] A. Retana and D. Cheng, "Ospfv3 instance id registry update," Internet Requests for Comments, RFC Editor, RFC 6969, 7 2013.

[3] Y. Liu, W. Tome, L. Zhang, D. Choffnes, D. Levin, B. Maggs, A. Mislove, A. Schulman, and C. Wilson, "An end-to-end measurement of certificate revocation in the Web's PKI," in *IMC*, 2015.

[4] L. Chuat, A. Abdou, R. Sasse, C. Sprenger, D. Basin, and A. Perrig, "SoK: Delegation and Revocation, the Missing links in the Web's Chain of Trust," in *Proc. IEEE EuroSP*, 2020.

[5] A. Langley, "ImperialViolet," https://www.imperialviolet.org/2014/04/19/revchecking.html, accessed: 2021-3-18.

[6] Google. CRLSets. Last accessed: September 2021. [Online]. Available: https://dev.chromium.org/Home/chromium-security/crlsets

[7] Mozilla, "OneCRL (CA/Revocation Checking in Firefox)," https://wiki.mozilla.org/CA:RevocationPlan#OneCRL, 2020.

[8] J. Larisch, D. Choffnes, D. Levin, B. M. Maggs, A. Mislove, and C. Wilson, "CRLite: A scalable system for pushing all TLS revocations to all browsers," in *Proc. IEEE S&P*, 2017.

[9] N. Korzhitskii and N. Carlsson, "Revocation statuses on the Internet." in *PAM*, 2021.

[10] B. Laurie, A. Langley, and E. Kasper, "Certificate transparency," Internet Requests for Comments, RFC Editor, RFC 6962, 6 2013.

[11] A. Halim, M. Danielsson, M. Arlitt, and N. Carlsson, "Temporal analysis of x.509 revocations and their statuses (code + datsets)," https://www.ida.liu.se/~nikca89/papers/wtmc22.html, 2022.

[12] Cloudflare, "Merkle town - explore the certificate transparency ecosystem," https://ct.cloudflare.com/logs.

[13] D. O'Brien, "Certificate Transparency Enforcement in Chrome and CT Day in London," https://groups.google.com/a/chromium.org/d/msg/ct-policy/Qqr59r6yn1A/2t0bWblZBgAJ, 2018.

[14] Apple. (2021) Apple's Certificate Transparency policy. [Online]. Available: https://support.apple.com/en-us/HT205280

[15] MongoDB, https://www.mongodb.com/.

[16] S. Santesson, M. Myers, R. Ankney, A. Malpani, S. Galperin, and C. Adams, "X.509 internet public key infrastructure online certificate status protocol - ocsp," RFC 6960, Tech. Rep., 2013.

[17] die.net, "crontab(5) - linux man page," https://linux.die.net/man/5/crontab.

[18] Golang, "golang/crypto," https://github.com/golang/crypto/blob/master/ocsp/ocsp.go.

[19] CA/Browser Forum, "Baseline Requirements for the issuance and management of publicly-trusted certificates, v1.8.0," https://cabforum.org/baseline-requirements-documents/, 2021.

[20] T. Smith, L. Dickinson, and K. Seamons, "Let's Revoke: Scalable Global Certificate Revocation," in *Proc. NDSS*, 2020.

[21] Apple, "About upcoming limits on trusted certificates," 2020. [Online]. Available: https://support.apple.com/en-us/HT211025

[22] Google, "Certificate lifetimes," https://chromium.googlesource.com/chromium/src/+/master/net/docs/certificate_lifetimes.md, 2020.

[23] Mozilla, 2020. [Online]. Available: https://blog.mozilla.org/security/2020/07/09/reducing-tls-certificate-lifespans-to-398-days/

[24] D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley, and W. Polk, "Internet X.509 Public Key Infrastructure certificate and Certificate Revocation List (CRL) profile," RFC 5280, 2008.

[25] C. M. Bruhner, O. Linnarsson, M. Nemec, M. Arlitt, and N. Carlsson, "Changing of the guards: Certificate and public key management on the internet," in *Proc. PAM*, 2022.

[26] J. Gustafsson, G. Overier, M. Arlitt, and N. Carlsson, "A first look at the CT landscape: Certificate Transparency logs in practice," in *Proc. PAM*, 2017.

[27] Q. Scheitle, O. Gasser, T. Nolte, J. Amann, L. Brent, G. Carle, R. Holz, T. C. Schmidt, and M. Wählisch, "The rise of Certificate Transparency and its implications on the Internet ecosystem," in *Proc. IMC*, 2018.

[28] B. Li, J. Lin, F. Li, Q. Wang, Q. Li, J. Jing, and C. Wang, "Certificate Transparency in the wild: exploring the reliability of monitors," in *Proc. ACM CCS*, 2019.

[29] E. Stark, R. Sleevi, R. Muminovic, D. O'Brien, E. Messeri, A. P. Felt, B. McMillion, and P. Tabriz, "Does Certificate Transparency Break the Web? Measuring Adoption and Error Rate," in *Proc. IEEE S&P*, 2019.

[30] C. Nykvist, L. Sjöström, J. Gustafsson, and N. Carlsson, "Server-side adoption of Certificate Transparency," in *Proc. PAM*, 2018.

[31] J. Amann, O. Gasser, Q. Scheitle, L. Brent, G. Carle, and R. Holz, "Mission Accomplished?: HTTPS Security After DigiNotar," in *Proc. IMC*, 2017.

[32] O. Gasser, B. Hof, M. Helm, M. Korczynski, R. Holz, and G. Carle, "In Log We Trust: Revealing poor security practices with Certificate Transparency logs and Internet measurements," in *Proc. PAM*, 2018.

[33] D. Kumar, Z. Wang, M. Hyder, J. Dickinson, G. Beck, D. Adrian, J. Mason, Z. Durumeric, J. A. Halderman, and M. Bailey, "Tracking certificate misissuance in the wild," in *Proc. IEEE S&P*, 2018.

[34] S. Micali, "Enhanced certificate revocation system," *Massachusetts Institute of Technology, Cambridge, MA*, pp. 1–10, 1995.

[35] M. Naor and K. Nissim, "Certificate revocation and certificate update," *IEEE JSAC.*, vol. 18, no. 4, pp. 561–570, 2000.

[36] A. A. Chariton, E. Degkleri, P. Papadopoulos, P. Ilia, and E. P. Markatos, "DCSP: Performant Certificate Revocation a DNS-based approach," in *Proc. European Workshop on System Security*, 2016.

[37] ——, "CCSP: A compressed certificate status protocol," in *Proc. IEEE INFOCOM*, 2017.

[38] M. Pachilakis *et al.*, "Design and Implementation of a Compressed Certificate Status Protocol," *ACM Trans. Int. Tech.*, vol. 20, 2020.

[39] D. Basin, C. Cremers, T. H.-J. Kim, A. Perrig, R. Sasse, and P. Szalachowski, "ARPKI: Attack resilient public-key infrastructure," in *Proc. ACM CCS*, 2014.

[40] T. H.-J. Kim, L.-S. Huang, A. Perrig, C. Jackson, and V. Gligor, "Accountable key infrastructure (AKI): A proposal for a public-key validation infrastructure," in *Proc. WWW*, 2013.

[41] P. Szalachowski, L. Chuat, T. Lee, and A. Perrig, "RITM: Revocation in the middle," in *Proc. IEEE ICDCS*, 2016.

[42] P. Szalachowski, L. Chuat, and A. Perrig, "PKI safety net (PKISN): Addressing the too-big-to-be-revoked problem of the TLS ecosystem," in *Proc. IEEE EuroS&P*, 2016.

[43] J. Yu, V. Cheval, and M. Ryan, "DTKI: A new formalized PKI with verifiable trusted parties," *The Computer Journal*, vol. 59, 2016.

[44] B. Laurie and E. Kasper, "Revocation transparency," *Google Research*, 2012.

[45] L. Zhang, D. Choffnes, D. Levin, T. Dumitracs, A. Mislove, A. Schulman, and C. Wilson, "Analysis of SSL Certificate Reissues and Revocations in the Wake of Heartbleed," in *Proc. IMC*, 2014.

[46] Z. Durumeric, F. Li, J. Kasten, J. Amann, J. Beekman, M. Payer, N. Weaver, D. Adrian, V. Paxson, M. Bailey, and J. A. Halderman, "The Matter of Heartbleed," in *Proc. IMC*, 2014.

[47] O. Omolola, R. Roberts, I. Ashiq, T. Chung, D. Levin, and A. Mislove, "Measurement and analysis of automated certificate reissuance," in *Proc. PAM*, 2021.

[48] A. Mirian, C. Thompson, S. Savage, G. M. Voelker, and A. P. Felt, "HTTPS Adoption in the Longtail," Google and UC San Diego, Tech. Rep., 2018.

[49] T. Chung, Y. Liu, D. Choffnes, D. Levin, B. M. Maggs, A. Mislove, and C. Wilson, "Measuring and Applying Invalid SSL Certificates: The Silent Majority," in *Proc. IMC*, 2016.

[50] L. Zhu, J. Amann, and J. Heidemann, "Measuring the Latency and Pervasiveness of TLS Certificate Revocation," in *Proc. PAM*, 2016.

[51] T. Chung, J. Lok, B. Chandrasekaran, D. Choffnes, D. Levin, B. M. Maggs, A. Mislove, J. Rula, N. Sullivan, and C. Wilson, "Is the Web ready for OCSP Must-Staple?" in *Proc. IMC*, 2018.

[52] MongoDB, "cursor.batchsize()," https://docs.mongodb.com/manual/reference/method/cursor.batchSize/.

TABLE 1. SUMMARY OF QUESTIONABLE STATUS CHANGES CHANGES AFTER INITIAL REVOCATION (AND BEFORE THE EXPIRY DATE OF THE CERTIFICATES (I.E., "NOT AFTER" TIME). ALL REPORTED HERE VALID USING ALL THREE ROOT STORES (NSS, APPLE, MICROSOFT).

Status change	Total	Example stats
Revoked → Good	2	Both are DV certs issued by GoDaddy, had revocation reason "Cessation of Operation" (amanprintersdelhi.in and thegoni.co.uk). The public keys were RSA-2048.
Revoked → Unauthorized	72	Certificate type (50 DV, 15 OV, 7 EV), issuer (34 GoDaddy, 12 QuoVadis, 9 Google, 7 DigiCert, 4 Fiducia & GAD IT AG, 3 Amazon, 2 HydrantID, 1 Let's Encrypt), keys (71 RSA-2048, 1 ECDSA-256), reason (27 "Cessation of Operation", 20 "Unspecified", 18 "Certificate Hold", 6 "Affiliation changed", 1 "Superseded").
Revoked → Unknown	3,490*	Top-5 CAs (1286 GoDaddy, 1292 DigiCert, 403 Sectigo, 136 cPanel, 136 Amazon, 237 other), Certificate type (2,253 DV, 1,207 OV, 30 EV), Key type (3,386 RSA-2048, 69 RSA-4096, 32 ECDSA-256, 3 ECDSA-384), reason (2,078 "Unspecified", 1,217 "Cessation of Operation", 123 "Affiliation changed", 52 "Superseded", 13 "Certificate Hold", 6 "Privilege withdrawn", 1 "Key compromise")

*We also observed one such case were the certificates was not valid for any of the root stores. (OV cert, RSA-2048, by Chunghwa Telecom.)

Appendix

1. More details about the data collection

Phase 2 optimizations: We did several optimizations for phase 2 of the data collection. For example, we again leverage goroutines to allow concurrent certificates checks, use a small `batchSize` [52] to speed things up and only update the DB when there are status changes or new errors occur. To save resources when the servers do not respond, requests are timed out after ten seconds. Another optimization was to tune the semaphore size (controlling the number of checks performed simultaneously). After testing with different sizes, we found that a semaphore size between $500 - 1,000$ gave a good balance of performance while not risking overwhelming the machine with work. In our experiments we used a semaphore size of 700. Finally, during testing, we found that most certificates had chain certificates in common. To avoid duplicates and save space, the tool only stores unique issuer certificates in the database.

Hardware and resource usage: The tool was deployed on a machine with an Intel Core i5-2500k and 8 Gb of RAM. When revocation checks are running, around 50 % of the CPU and less than 2 Gb of RAM was utilized. Running revocation checks in parallel with gathering newly issued certificates resulted in RAM usage peaks of 2.5 Gb.

2. Summary details of discussed status changes

Table 1 summarizes the key statistics for the cases were the reported status changed (1) from "revoked" to "good", (2) from "revoked" to "unauthorized", and (3) from "revoked" to "unknown". Again, we only report statistics for certificates that otherwise validated against the three major root stores operated by NSS, Apple, and Microsoft.

3. Detailed stats for "revoked" to "unknown"

Similar to for the second category, this category included a relatively larger fraction of the revoked OV (1,207 / 15,737 = 7.7%) and EV (30 / 532 = 5.6%) certificates compared to DV (2,253 / 206,271 = 1.1%) certificates. This suggests that this miscellaneous behavior is more likely to occur for more expensive certificates. We argue that these cases should not occur regardless of price.

There is a noticeable over representation of RSA-2048 (3,386/185,508 = 1.8%) as all other key-types had well-below average rate of such cases (at most 0.33%, with RSA-4096). For the revocation reasons, we observe only a single instance when the revocation reason is "key compromise" (1/194 = 0.52%). Otherwise, the certificates with a specified revocation reason all had above average observation rates: "cessation of operation" (3.6%), "affiliation changed" (4.4%), "superseded" (2.5%), "certificate hold" (10.6%), and "privilege withdrawn" (5.1%).

Finally, let us look closer at the 30 EV certificates in this category. Here, we observed three reasons (23 "Unspecified", 4 "Certificate Hold", 3 "Affiliation changed"), five issuers (22 DigiCert, 4 QuoVadis, 2 Entrust, 1 WCA Global, 1 D-TRUST SSL), and two key types (28 RSA-2048, 2 RSA-4096). Perhaps most interesting is that we again see several QuoVadis certificates and that there were so many DigiCert certificates in this class. Having talked to DigiCert, it appears that most of these issues may be related to some of their servers periodically having reported 304 errors for some of these certificates.