No domain left behind
Is Let's Encrypt democratizing encryption?

Aertsen, Maarten; Korczyński, Maciej; Moura, Giovane C.M.; Tajalizadehkhoob, Samaneh; Van Den Berg, Jan

DOI
10.1145/3106328.3106338

Publication date
2017

Document Version
Final published version

Published in
ANRW 2017 - Proceedings of the Applied Networking Research Workshop, Part of IETF-99 Meeting

No domain left behind: is Let's Encrypt democratizing encryption?

Maarten Aertsen
Delft University of Technology
maarten@rtsn.nl

Maciej Korczyński
Delft University of Technology
maciej.korczynski@tudelft.nl

Giovane C. M. Moura
SIDN Labs
giovane.moura@sidn.nl

Samaneh Tajalizadehkhoob
Delft University of Technology

Jan van den Berg
Delft University of Technology

ABSTRACT

The 2013 National Security Agency revelations of pervasive monitoring have led to an "encryption rush" across the computer and Internet industry. To push back against massive surveillance and protect users' privacy, vendors, hosting and cloud providers have widely deployed encryption on their hardware, communication links, and applications. As a consequence, most web connections nowadays are encrypted. However, there is still a significant part of Internet traffic that is not encrypted. It has been argued that both *costs* and *complexity* associated with obtaining and deploying X.509 certificates are major barriers for widespread encryption, since these certificates are required to establish encrypted connections. To address these issues, the Electronic Frontier Foundation, Mozilla Foundation, the University of Michigan and a number of partners have set up *Let's Encrypt* (*LE*), a certificate authority that provides both free X.509 certificates and software that automates the deployment of these certificates. In this paper, we investigate *if LE* has been successful in democratizing encryption: we analyze certificate issuance in the first year of *LE* and show from various perspectives that *LE* adoption has an upward trend and it is in fact being successful in covering the lower-cost end of the hosting market.

CCS CONCEPTS

•Networks →Public Internet; *Network measurement;* **•Security and privacy →Network security;** *Authentication; Key management;*

ACM Reference format:
Maarten Aertsen, Maciej Korczyński, Giovane C. M. Moura, Samaneh Tajalizadehkhoob, and Jan van den Berg. 2017. No domain left behind: is Let's Encrypt democratizing encryption?. In *Proceedings of Applied Networking Research Workshop 2017, Prague, Czech Republic, July 15 (ANRW'17),* 7 pages.
DOI: 10.1145/3106328.3106338

1 INTRODUCTION

The 2013 National Security Agency (NSA) revelations of pervasive monitoring and surveillance had a significant impact on the Internet industry. As a reaction, we have witnessed a surge on deployment of encryption technologies to curb these surveillance practices. For example, Google enabled encryption in the links between its datacenters [37] while Apple enabled encryption by default on its mobile devices [12]. The Internet Engineering Task Force (IETF) –a body that standardizes Internet-related protocols– issued RFC 7258 [13], making it clear that "pervasive monitoring is an attack".

We have also seen a surge on the encryption of web traffic in response to these revelations. For example, browser telemetry from both Mozilla Firefox and Google Chrome shows that more than 50% of page loads by their users is currently encrypted [3, 32]. However, a significant portion of web traffic is still unencrypted, and it has been argued that both the complexity and costs associated with obtaining and deploying the required X.509 certificates (issued by third-party paid certificate authorities – CAs) are major barriers for wide encryption of web traffic [21, p.86]. For example, some CAs charge up to $80 USD per certificate, per website, per year, and require manual setup.

To address these barriers against ubiquitous encryption, the Electronic Frontier Foundation (EFF), Mozilla Foundation, and University of Michigan set up *Let's Encrypt* [26] (*LE* hereafter), a CA that provides both *free* X.509 certificates and *automated* software to configure servers to use those certificates. By reducing both costs (to zero) and deployment complexity, *LE* aims to make encrypted traffic ubiquitous, democratizing certificate issuance and deployment. Slightly more than one year after launch, *LE* has issued 12 million certificates – making it one of the top three largest CAs [14, 15].

In this paper, we investigate *if LE* has been successful in democratizing encryption, and perform a comprehensive analysis on the issuance of *LE* certificates. Note that "democratization" here means that it makes encryption available to those for whom it was not previously available, ostensibly due to cost or administrative reason. We study the adoption of *LE* certificates within the different types of organizations with the focus on the lower-cost end of the hosting market, which employs shared hosting. We use as a starting point one year of data obtained from the Certificate Transparency (CT) logs [2] and make the following contributions: looking from various perspectives, we show that *LE* is indeed democratizing encryption – we show that 98% of the domains certified by *LE* fall outside Alexa 1M (§4.2), but that issuance is not restricted

ANRW'17, Prague, Czech Republic

978-1-4503-5108-9...$15.00
DOI: 10.1145/3106328.3106338

to the lower-cost end of the market. Moreover, we show that the success of *LE* is attributed by the adoption of major players (3 hosting providers are responsible for 47% of the *LE* certified domains, §4.3). We also show that issuance is predominantly for the lower-cost end of the market (shared hosting, §4.5), and that the majority of certificates are correctly renewed after their first expiration (90 days, §4.6). For the .nl top-level domain (TLD), we show that both old and new domains are benefiting from *LE* (§4.7). Last, we show that 63% of *LE* certified domains are correctly deploying their certificates (§4.8), which is a lower bound number that we determined by performing active https scans.

2 SSL/TLS CONNECTIONS AND CAS

To illustrate how encrypted traffic on the Web works, consider the following example. A user's browser connects to a web server to retrieve a webpage[1]. After establishing the TCP connection, the client (browser) and server start the SSL/TLS handshake, which we briefly summarize here and refer the reader to [7, 22] for more details. The browser first sends a client hello message, the server responds with a server hello message and a certificate message which includes its public key. Upon receiving the certificate message, the browser must validate the chain of certificates [6], and only after this step the SSL/TLS setup continues and the encrypted connection can be used.

However, there are two prior steps necessary to get the required certificate: an entity has to request a certificate from the CA for the particular fully qualified domain name (FQDN). The CA, in turn, issues a certificate, which is then deployed on the server.

Commercial CAs typically offer three types of certificates: domain validated, organization validated, and extended validation certificates. All of them employ the same encryption measures – they differ on how the CA verifies the user's identity (e.g. if the user is the legal owner of the domain and company for which a certificate is being issued).

Since *LE* automates issuance, it only provides domain validated certificates, where a user merely has to prove administrative control over the FQDN being certified. *LE* is the first CA to fully automate the process of validation and issuance, using the Automatic Certificate Management Environment (ACME) protocol [4].

3 METHODOLOGY

3.1 Certificate Transparency logs

The certificates issued by *LE* were obtained from Certificate Transparency (CT) logs [2], which provide a publicly available, append-only log of certificate issuance [25]. This is the complete set of all issued certs [34], resulting from *LE*'s commitment to full publication in CT. *LE* issued its first certificate on Sept 2015. Our data therefore contains one year of certificates based upon CT data (Sept 2015-2016).

For each certificate, we extract the respective validity period and FQDNs from the subjectAltName string. We then transform these FQDNs into a "normalized" *domain* form, which is defined as either the 2^{nd}–level or 3^{rd}–level if a given TLD registry provides such registrations (e.g.: example.co.uk or example.org) [23]. Therefore, we *do not* analyze the number of certificates issued by *LE* in this paper, we focus on their coverage of their "normalized" *domain* form. For instance, certificates for a.example.org and b.example.org would be mapped into one *domain* (example.org). In the rest of this paper, we use domains in the sense of "normalized" domains.

3.2 Passive DNS data

The CT logs only provide information about the domain names that have been issued *LE* certificates; it does not include information about *where* these domains are hosted. To determine that, we make use of passive DNS logs (Sept 2015-2016) obtained from DNSDB – a passive DNS database generously provided by Farsight Security [11]. These logs contain a mapping between A records[2] of domains queried on web within a specific time span, and their corresponding IP addresses. To our knowledge, DNSDB has the best coverage of the overall domain name space that is available to researchers. Using this data we determine IP addresses associated with *LE* domain names and found that DNSDB contains historical A records for 80% of *LE* domains in our data. An alternative to using passive DNS data would be to have performed DNS lookups for all domains covered by *LE*, but the required historical data for that purpose was not available.

3.3 Organization mapping and classification

Our next task is to identify organizations and ultimately shared hosting providers associated with *LE* domains and their IP addresses. We follow the procedure that is discussed in prior work [5, 29, 33]. We start by mapping the IP addresses associated with *LE* domains as observed by passive DNS data into the organizations to whom they are allocated. In short, for each IP address, we retrieve the organization using their respective registration information using MaxMind whois API [30]. The resulted list contains different type organizations including but not limited to hosting providers, Internet service providers (ISPs), university networks, broadband providers, mobile service providers, anti-DDoS services, content delivery networks (CDNs) and others. We extend the list of keywords and categories used in a prior study [8], and classify the organizations in our list into the above mentioned categories. Those organizations that cannot be classified, are put in an 'unknown' category. Each *LE* domain is then associated with an organization and organization type.

Finally, each *LE* domain associated with hosting providers is marked as *shared* hosting if it is associated with an IP address that hosts more than 10 domains, observed by the passive DNS data [33, 35]. This approach allows us to map *LE* domains not only into various types of organizations but also in different classes of hosting providers (dedicated hosting vs. shared hosting). This further enhances our knowledge about the segment of the market that is adopting more *LE* certificates.

[1] X.509 certificates can actually be used for other applications such as retrieving or delivering e-mail, but for the sake of simplicity, we only focus on web traffic here.

[2] A records are type of DNS records that map domains into IP addresses [31].

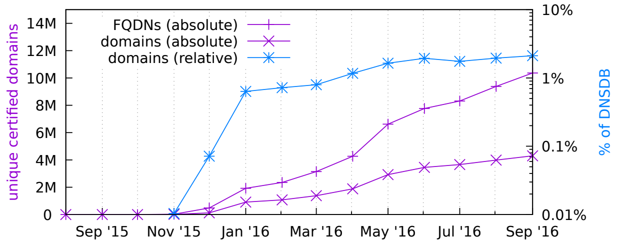


Figure 1: *LE* time series for FQDNs, domains, and DNSDB ratio

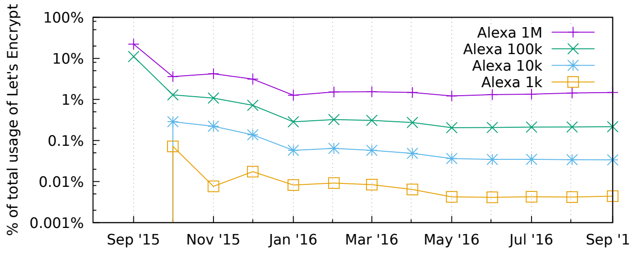


Figure 2: *LE* certificates issued to Alexa top-ranked domains

3.4 .nl domain registration information

To determine the domain age of certified *LE* domains, we employ registration information from the .nl TLD registry (SIDN). An alternative was to analyze whois records for all the other zones. However, given the fact that (i) most of TLDs do not offer historical domain whois service, (ii) it cannot be publicly accessed, (iii) and for those that do offer whois, the format is not standardized [28], we opt for singling out .nl as a case study.

4 ANALYSIS AND DISCUSSION

4.1 Absolute and relative growth

How big is *LE*? *LE* publishes statistics [18] showing a continuous growth in the number of daily issued certificates. *LE* is in fact the third biggest CA, according to other research [15].

Figure 1 shows a time series of the absolute number of unique *LE* certified domains, FQDNs, and domains relative to all domains observed in DNSDB (§3.2). First, we see a continuous growth in all metrics: by Sept 2016, there were ~10.4M FQDNs that had *LE* certificates, amount to ~4.3M domains (§3.1), on average 2.5 certificates per domain.

Moreover, to have an idea on how much of the domain name space uses *LE* certificates, we use DNSDB as a comparison and show that *LE* is used by 2% of all domains observed in Sept 2016. Given that *LE* has been only active for a year by the time of this analysis, 2% of a large sample of the domain space represents a significant growth.

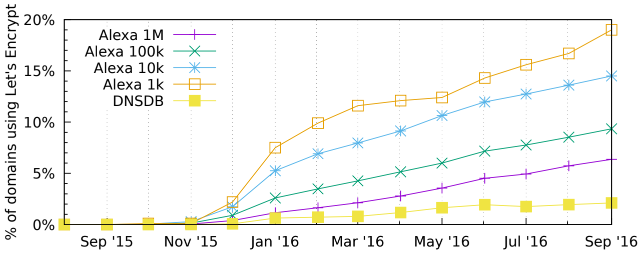


Figure 3: Relative usage of *LE* domains in Alexa rankings

4.2 Popular sites and *Let's Encrypt*

Although *LE* certificates are rarely issued and deployed for the most popular domains on their main websites [16], *LE* certificates may be issued for their subdomains. To examine this hypothesis, we first obtain a list of the most popular websites ranked by Alexa [1]. Then, we extract all FQDNs observed in DNSDB (e.g. example.org or subdomain.example.org) and match them against all *LE* certified FQDNs.

Figure 2 shows a time series of the relative contribution of Alexa ranked domains (1M, 100K, 10K, 1K) against the total number of domains with valid *LE* certificates. The contribution of Alexa 1M domains remains stable around 2% of total *LE* usage throughout 2016—a period of relatively rapid growth of *LE* issuance. By Sept 2016 about ~64K domains ranked in the Alexa 1M use *LE* in the sense that at least one *LE* certificate has been issued for a FQDN under its domain (e.g. subdomain.example.org)

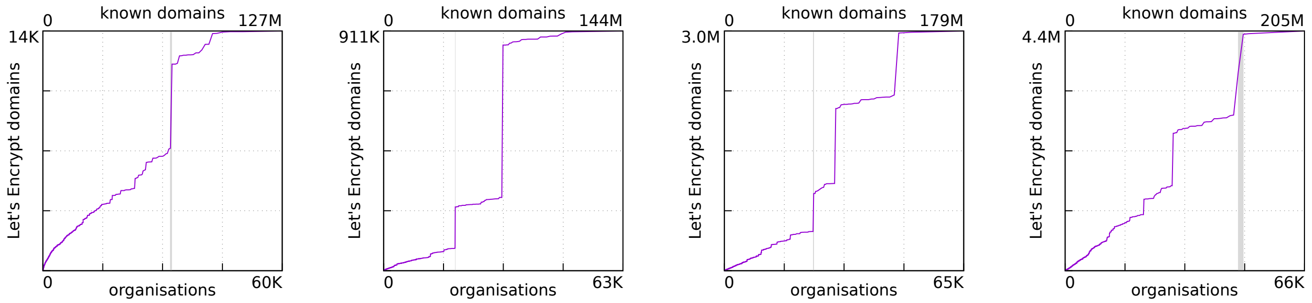
Figure 3 shows a time series of the relative growth of issuance within the Alexa rankings. By Sept 2016, as many as 19% of domains in the Alexa 1K have had issued at least one *LE* certificate. This suggests that 19% of domains associated with the most popular websites use and depend on *LE*'s service, but they do not necessarily issue and deploy certificates on their main websites (e.g., both wsj.com and welt.de are labeled as *LE* domains, yet do not use *LE* on their main websites).

Overall, we find that 98% of the domains certified by *LE* are less popular sites outside Alexa rankings – which is good for democratizing encryption – and 2% are popular websites, indicating that *LE* is not only constrained to the lower-cost share of the market.

4.3 Certificates distribution per organization

Which organizations are using more *LE* certified domains? Are there "big players" or are the *LE* domains distributed across small organizations? To answer this question, we use the methodology described in §3.3 and map *LE* domains to their respective IP address owners. We calculate size indicators, by aggregating these mappings per organization.

Figure 4 shows ECDF of *LE* certified domains per organization, for four selected months of issuance. We sort the organizations (x axis) by "domain density", the sum of the number of domains hosted (§3.3) in increasing order. Steps in these lines indicate bulk issuance of *LE* domains by an organization. For example, in Jan 2016, we see the large vertical line corresponding to deployment at Automattic/wordpress.com ($x = 0.5, \Delta y = 63.5\%$), which is especially noticeable when compared against Nov 2016. By Sept



(a) November 2015 **(b) January 2016** **(c) May 2016** **(d) September 2016**

Figure 4: ECDF of *LE* use versus organization size (domain density, the sum of the number of domains hosted). The bottom x-axis has domain density per organisation sorted in ascending order. The y-axis represents the total number of *LE* domains issued. The top x-axis represents the total of domains in DNSDB (the sum of all domain densities). The shaded area indicates domains that are not successfully attributed to an organization.

2016, we can observe three clear steps: Shopify ($x = 0.33, \Delta y = 6\%$), Automattic/wordpress.com ($x = 0.45, \Delta y = 22\%$) and OVH ($x = 0.7, \Delta y = 19\%$). All three companies have announced issuance for their customers and are jointly responsible for 47% of *LE* certified domains. It is exactly these companies, serving numerous, smaller customers that would otherwise not enable the use of encryption by their visitors.

We also find evidence which suggests that *LE* is popular among smaller organizations. Among all 66K identified organizations (§3.3), we find 14K that have domains certified with *LE* in Sept 2016. Notably, 9K have 5 or less *LE* certified domains. This corresponds to the lower left quadrant of Figure 4d, where smaller organizations are jointly responsible for 23% of all *LE* domains. We conclude that *LE* reaches both large hosting companies as well as smaller organizations with lower domain concentration.

4.4 Types of organizations

In the previous section, we analyzed the distribution of *LE* certified domains per organization. In this section, we classify these organizations according to their types (§3.3, [33]). We group organizations into education related, domain parking, hosting providers, ISPs, CDNs, DDoS-protection services and others (including government related).

The distribution of *LE* domains per organization category is shown in Figure 5. The majority of domains are associated with hosting organizations (68% in Sept 2016), while the share of DDoS protection services and CDNs remains low (2% and 0.1%, respectively). Popular websites are more likely to use CDN or DDoS protection services. The high adoption by web hosters versus the low adoption by CDN and DDoS services thus points to adoption by smaller and/or less popular websites. Note, however, that 29% of all domains were not attributed to any of the categories (‘unknown’).

4.5 Types of hosting: shared and non-shared

Hosting services typically are offered in multiple types at different prices. Resources such as CPU, memory, bandwidth and IP addresses could be dedicated to customers (“dedicated hosting”), or shared among them. Shared hosting is where prices are at their lowest level and profit margins are slimmer. Under these conditions, encryption deployment would be least expected.

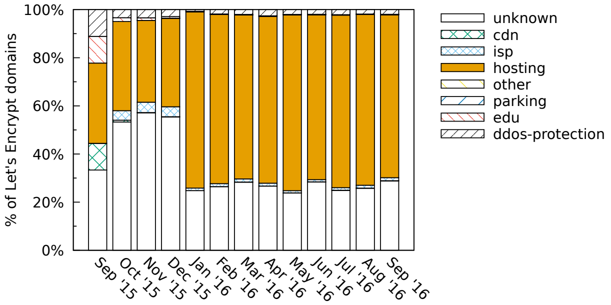


Figure 5: Distribution of *LE* domains per organization type (in % of domains)

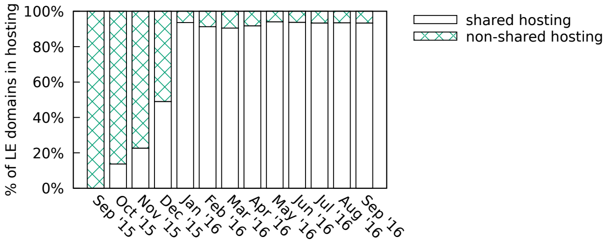


Figure 6: Distribution of *LE* accross shared and non-shared hosting (in % of domains)

We classify the IP addresses from the hosting organizations listed in §4.4 into shared and non-shared hosting via the methodology explained earlier in §3.3.

Figure 6 is a histogram of relative market share within the hosting segment, split between shared and non-shared hosting services. We find that from Jan 2016, *LE* use within hosting is predominantly

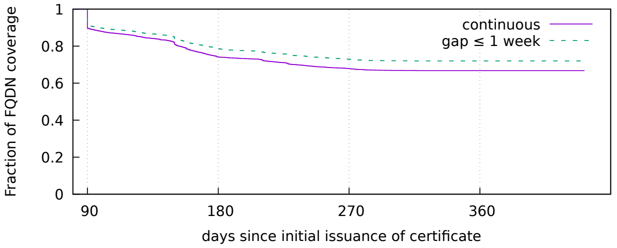


Figure 7: Survival analysis of *LE* certified FQDNs

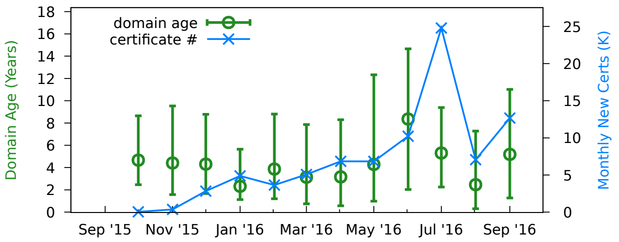


Figure 8: Median, Q25, Q75 and number of monthly new certificates for .nl domains

connected to shared hosting services, with a penetration above 90%. Recalling that by Sept 2016 the overall hosting segment is dominant over other types (67%), we find that *LE* has very high overall utilization in shared hosting, which has traditionally been the least likely candidate for adoption of encryption due to the associated costs. As in the previous section, we can see that *LE* covers the lower-cost end of the market.

4.6 Certification lifetime

Once domains issue an *LE* certificate, do they keep on renewing them every 90 days? Or do they let them expire? To answer this question, we carry out a survival analysis of *LE* certificates for each FQDN using a Kaplan-Meier Survival Estimate [20]. Survival for each FQDN is defined as a continuous chain of certificate renewals, whereas a "death" occurs when renewals stop. We do not take certificate revocation into account as it has very limited real world use [24]. The CT logs (§3.1) provide the complete information required for this metric. Alternatives, such as active scans, would only yield a sample. We identify three components that are likely to influence the outcome of this question: (i) renewal automation working correctly (not having automation set-up likely causes renewal failure); (ii) user satisfaction with the service and its certificates; (iii) the intended lifetime of the domains themselves.

Figure 7 shows the estimated survival function of *LE* certified FQDNs featuring two functions. The continuous function measures survival without any downtime: survival implies the issuance of certificates with perfectly overlapping validity periods. The second function measures survival with a maximum one week gap in between consecutive validity periods. This accounts for failure in automation, corrected after the previous certificate expires.

Since all certificates are valid for 90 days, we observe 100%, survivability for this period. After those 90 days we see drops: domains that either stop being re-certified, where automation was not successful or where the domain itself expired. The survival curve noticeably flattens after $x = 270$ days, indicating that the automation is effective.

The agreement between $gap = 0$ (continuous) and $gap \leq 1$ *week* indicates that beyond initial downtime, further survival is roughly similar. This may be explained by users that get continuous coverage after successful setup of automation. With more than 70% FQDN survival after a full year, we can conclude that the majority of *LE* users remain loyal to the service during our measurement period, which is not surprising given the size (§4.3) and type (§4.4) of *LE* users – predominantly (big) hosting providers.

4.7 New vs. old domains

What type of domains are more likely to employ *LE* certificates: newly registered domains or older domains? Since *LE* reduces the certificate cost to zero while providing full automation, one could hypothesize that registrars simply enable them by default on their registrations system, so every new domain could be automatically configured with an *LE* certificate.

To investigate this hypothesis, we focus on the .nl TLD as a case study (§3.4). There were 514,986 *LE* certificates issued for 191,176 unique FQDNs, during the monitoring period. In total, there were 85,223 unique .nl domains that had *LE* certificates.

Figure 8 shows the number of *LE* certificates issued for .nl domains for the first time (continuous line), and the median age, first quartile (Q25) and third quartile (Q75) (box plot). As we can see, for all months, the median age of the domains is above two years, with a large spread, suggesting that *LE* is being used both for older and newer domains. We can conclude that for this dataset, most of *LE* certificates are being used on already existing domains. In the absence of scan data for those domains, we cannot confirm if they had their first certificate issued by *LE*, or if they switched to *LE*.

4.8 Certificate issuing vs. deployment

So far we have covered the side of certificate issuance. However, another open question is to determine how many of these certificates are actually deployed. Answering this question is not straightforward: first, certificates can be used for other applications than the Web, such as e-mail or ftp. Certificates can also be deployed internally within networks or be used on non-standard ports.

To have a lower-bound estimate of *LE* certificate deployment, we randomly select 25,000 FQDNs for certificates that were issued (and therefore valid, not expired) between Nov 13 and 19, 2016 and scan them on https (TCP port 443) to determine if the certificates are actively deployed for use on the Web. We perform the scans on Nov 28, 2016.

Figure 9 shows the scan results. As can be seen, 15,803 (63%) of FQDNs have successfully deployed *LE* certificates. The remaining were divided into other errors, such as 2,465 (10%) having no DNS records – e.g. short-lived, possibly expired; 2,143 (9%) do not support TLS and 1,422 (6%) return an http error code and are likely not set-up for https in the first place. Interestingly, 2,846 (11%) deploy certificates not issued by *LE*. Here one could hypothesize that

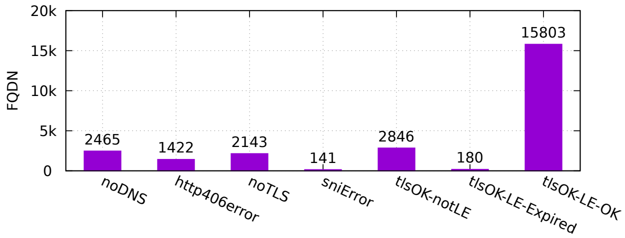


Figure 9: Scans of 25,000 *LE* covered FQDNs

either the hosting provider is waiting for paid certificates to expire or is just experimenting. In addition, 180 FQDNs had expired *LE* certificates.

Our results show that 63% of our sampled FQDNs (as a lower-bound value) have successfully deployed *LE* certificates. For a more comprehensive view on *LE* deployment, it is important to perform longitudinal active measurements on all popular services for all FQDNs covered by *LE*.

5 RELATED WORK

The ecosystem for certificates and their use has been analyzed by various studies, but none of them have singled out *LE* and analyzed its impact. For example, there are Internet-wide scan studies covering certificates [9, 10, 17, 27]. Several methods with the goal of mapping the CA ecosystem (e.g.: active scans, Certificate Transparency [2]) have also been compared [34]. Paid access reports have been previously issued (e.g. [36]).

Although *LE* is a new player in the CA market, there have been some preliminary efforts in measuring its adoption. For example, there are self-reported *LE* statistics pages (e.g. [18]) and a series of blog post by J.C. Jones [19]. They present a growth in the number of *LE* issued certificates, whereas we analyze its growth in terms of the absolute number of unique *LE* certified domains, FQDNs, and *LE* certified domains relative to all observed domains. We also show that the majority of certificates for FQDNs are correctly renewed after their first expiration.

Helme presented statistics on the *LE* coverage on the Alexa 1M most popular websites [16]. In our study, we show that although *LE* certificates are sporadically issued and deployed for the Alexa 1M domains on their main websites, *LE* certificates are issued for their subdomains.

Finally, to the best of our knowledge, this is the first work that singles out *LE* and shows what segments of the market are using and deploying their certificates. We demonstrate that *LE* is democratizing encryption, by being used mostly by the lower-cost share of the hosting market.

6 CONCLUSIONS AND FUTURE WORK

LE has been successful in disrupting the certificate industry, which has been slow in covering the lower-cost end of the market [3, 32]. By addressing the two major barriers inhibiting ubiquitous encryption (cost and complexity required in issuing X.509 certificates), *LE* has become one of the largest CAs within only one year after its first certificate was issued.

We have studied the certificate issuance in the first year of *LE* and showed that it has been playing a major role in democratizing encryption: it has been widely used, and mostly by the low-cost share of the market (shared hosting). We have also shown that once these barriers are eliminated, it enables big hosting providers to issue and deploy certificates in bulk, thus quickly and automatically enable encryption across a large number of domains. For example, we have shown that currently, 47% of *LE* certified domains are hosted at three large hosting companies.

The success of *LE* can also be measured by the fact that 70% of the *LE* certified domains remain active after the first issuance of the certificate (*LE* certificates expire after 90 days). Also, for one TLD zone (.nl), we show that *LE* certificates have been issued not only for newly registered domains, but also for several-year-old domains, likely benefiting from bulk issuing by their hosting companies.

Issuing a certificate is only one part of the story for encrypted communications: deploying it on the server side is also of essence. To measure the fraction of deployed *LE* certificates, we actively scanned a sample of 25K FQDNs. We showed that 63% of them are correctly deployed for https, which is a lower bound value given that these certificates can also be used for other applications.

As future work, it is of interest to continue observing how *LE* evolves and impacts the CA market. We will investigate whether growth of *LE* comes from sites that did not use certificates at all before or from sites for which hostmasters have switched from other CAs. Other open questions include deployment and configuration metrics, including for non-https services and *LE*'s growth relative to other CA's.

Acknowledgements:

The authors would like to thank J.C. Jones (Mozilla), Carlos Gañán (TU Delft) and the anonymous reviewers for their valuable comments on paper drafts, Maarten Wullink (SIDN) for sharing his https/DNS scanner (DNSEMAP) and the contributors of data to Farsight Security's DNSDB.

REFERENCES

[1] Alexa Top 1,000,000 Sites. http://s3.amazonaws.com/alexa-static/top-1m.csv.zip.
[2] Certificate Transparency - Known logs. https://www.certificate-transparency.org/known-logs.
[3] Google transparency report - https - https usage. https://www.google.com/transparencyreport/https/metrics/?hl=en.
[4] R. Barnes, J. Hoffman-Andrews, and J. Kasten. Automatic Certificate Management Environment (ACME). draft-ietf-acme-acme-03, July 2016.
[5] X. Cai, J. Heidemann, B. Krishnamurthy, and W. Willinger. Towards an AS-to-organization map. In *Proc. of IMC*, pages 199–205. ACM, 2010.
[6] 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 (Proposed Standard), May 2008. Updated by RFC 6818.
[7] T. Dierks and E. Rescorla. The Transport Layer Security (TLS) Protocol Version 1.2. RFC 5246 (Proposed Standard), Aug. 2008. Updated by RFCs 5746, 5878, 6176, 7465, 7507, 7568, 7627, 7685.
[8] X. Dimitropoulos, D. Krioukov, G. Riley, and k. claffy. Revealing the Autonomous System Taxonomy: The Machine Learning Approach. In *PAM*, 2006.
[9] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman. A Search Engine Backed by Internet-Wide Scanning. In *Proc. of ACM CCS*, 2015.
[10] Z. Durumeric, J. Kasten, M. Bailey, and J. A. Halderman. Analysis of the HTTPS certificate ecosystem. In *Proc. of IMC*, pages 291–304, 2013.
[11] Fairsight. DNSDB. https://www.dnsdb.info/.
[12] C. Farivar. Apple expands data encryption under iOS 8, making handover to cops moot. https://arstechnica.com?post_type=post&p=540095, 2014.
[13] S. Farrell and H. Tschofenig. Pervasive Monitoring Is an Attack. RFC 7258 (Best Current Practice), 2014.

[14] G. Gebhart and S. Schoen. Is Let's Encrypt the Largest Certificate Authority on the Web? https://www.eff.org/deeplinks/2016/10/lets-encrypt-largest-certificate-authority-web, Oct 2016.

[15] M. Gelbmann. The impact of Let's Encrypt on the SSL certificate market. https://w3techs.com/blog/entry/the_impact_of_lets_encrypt_on_the_ssl_certificate_market, Sep 2016.

[16] S. Helme. Security headers in the Alexa Top 1 Million - Let's Encrypt Usage. https://scotthelme.co.uk/security-headers-alexa-top-million/, Feb 2016.

[17] R. Holz, L. Braun, N. Kammenhuber, and G. Carle. The SSL landscape: a thorough analysis of the X.509 PKI using active and passive measurements. In *Proc. of IMC*, pages 427–444, Nov 2011.

[18] ISRG. Let's Encrypt Stats. https://letsencrypt.org/stats/, 2016.

[19] J. Jones. Blog series on growth of Let's Encrypt. https://tacticalsecret.com/tag/letsencrypt/, 2016.

[20] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[21] J. D. Kasten Jr. *Server Authentication on the Past, Present, and Future Internet.* PhD thesis, The University of Michigan, 2015.

[22] M. Korczyński and A. Duda. Markov chain fingerprinting to classify encrypted traffic. In *Proc. of IEEE INFOCOM*, pages 781–789, April 2014.

[23] M. Korczyński, S. Tajalizadehkhoob, A. Noroozian, M. Wullink, C. Hesselman, and M. van Eeten. Reputation metrics design to improve intermediary incentives for security of TLDs. In *2017 IEEE European Symposium on Security and Privacy (Euro SP)*, April 2017.

[24] A. Langley. No, don't enable revocation checkign. https://www.imperialviolet.org/2014/04/19/revchecking.html, Apr 2014.

[25] B. Laurie, A. Langley, E. Kasper, E. Messeri, and R. Stradling. Certificate Transparency. RFC 6962-bis-19 (Internet-Draft), Aug. 2016.

[26] Let's Encrypt. Free SSL/TLS Certificates. https://letsencrypt.org/.

[27] O. Levillain, A. Ébalard, B. Morin, and H. Debar. One year of SSL Internet measurement. In *Proc. of ACSAC*, pages 11–20. ACM, Dec 2012.

[28] S. Liu, I. Foster, S. Savage, G. M. Voelker, and L. K. Saul. Who is .Com?: Learning to Parse WHOIS Records. In *Proc. of IMC*, pages 369–380, 2015.

[29] Y. Liu, A. Sarabi, J. Zhang, P. Naghizadeh, M. Karir, M. Bailey, and M. Liu. Cloudy with a chance of breach: Forecasting cyber security incidents. In *USENIX Security*, pages 1009–1024, 2015.

[30] Maxmind. GeoIP2 ISP Database. https://www.maxmind.com/en/geoip2-isp-database, 2016.

[31] P. Mockapetris. Domain names - concepts and facilities. RFC 1034, Nov. 1987.

[32] Mozilla. Firefox telemetry. https://telemetry.mozilla.org/, Oct 2016.

[33] S. Tajalizadehkhoob, M. Korczyński, A. Noroozian, C. Gañán, and M. van Eeten. Apples, oranges and hosting providers: Heterogeneity and security in the hosting market. In *Proc. of NOMS*, Apr 2016.

[34] B. VanderSloot, J. Amann, M. Bernhard, Z. Durumeric, M. Bailey, and J. A. Halderman. Towards a Complete View of the Certificate Ecosystem. In *Proc. of IMC*, pages 543–549, 2016.

[35] M. Vasek, J. Wadleigh, and T. Moore. Hacking Is Not Random: A Case-Control Study of Webserver-Compromise Risk. *IEEE Transactions on Dependable and Secure Computing*, 13(2):206–219, 2016.

[36] W3Techs. Changes in the usage of IdenTrust. https://w3techs.com/technologies/changes/sc-identrust, Nov 2016.

[37] C. Welch. Google encrypts Gmail between data centers to keep the NSA out of your inbox. http://www.theverge.com/2014/3/20/5530072/google-encrypts-gmail-between-data-centers-to-keep-out-nsa, 2014.