The Potential of an Individualized Set of trusted CAs: Defending against CA Failures in the Web PKI (Extended Version) *

Johannes Braun and Gregor Rynkowski

Technische Universität Darmstadt
Hochschulstraße 10, 64283 Darmstadt, Germany
{jbraun,grynkowski}@cdc.informatik.tu-darmstadt.de

Abstract. The security of most Internet applications relies on underlying public key infrastructures (PKIs) and thus on an ecosystem of certification authorities (CAs). The pool of PKIs responsible for the issuance and the maintenance of SSL certificates, called the Web PKI, has grown extremely large and complex. Herein, each CA is a single point of failure, leading to an attack surface, the size of which is hardly assessable.
This paper approaches the issue if and how the attack surface can be reduced in order to minimize the risk of relying on a malicious certificate. In particular, we consider the individualization of the set of trusted CAs. We present a tool called Rootopia, which allows to individually assess the respective part of the Web PKI relevant for a user.
Our analysis of browser histories of 22 Internet users reveals, that the major part of the PKI is completely irrelevant to a single user. On a per user level, the attack surface can be reduced by more than 90%, which shows the potential of the individualization of the set of trusted CAs. Furthermore, all the relevant CAs reside within a small set of countries. Our findings confirm that we unnecessarily trust in a huge number of CAs, thus exposing ourselves to unnecessary risks. Subsequently, we present an overview on our approach to realize the possible security gains.

1 Introduction

Nowadays, the extensive use of e-business, e-banking and e-government services over the Internet makes entity authentication, confidentiality and integrity indispensable in many cases. Entity authentication and thus secure connection establishment builds on the underlying Web public key infrastructure (Web PKI). The core of the Web PKI is the ecosystem of certification authorities (CAs) that are responsible for the issuance and the maintenance of SSL certificates. These certificates are issued to web service providers and are used in the SSL/TLS protocols.

However, the Web PKI fails in many points to provide the desired security [1–4]. One serious problem is that the security of the Web PKI suffers from the enormous size of the Internet. For the sake of interoperability (i.e., as much legitimate web service certificates as possible should be verifiable) the number of CAs, which are fully trusted by default in current browsers and operating systems, has continuously been growing over the past. Currently, there are approximately 1,500 directly or transitively trusted

* A short version of this paper appears in 2013 ASE/IEEE International Conference on Information Privacy, Security, Risk and Trust

CAs [5, 6]. As each of these trusted CAs can sign certificates for any web service or domain, trusting a single malicious CA, i.e., one that is in fact not trustworthy, can break the whole Web PKI's security. An adversary, who is in possession of a fake certificate issued by any one of the trusted CAs, can potentially intercept and manipulate the complete communication between any Internet user and the certified web server without the user even noticing the attack. With each additional CA, the risk of trusting a malicious or defective CA increases. Several security incidents in the past clearly show that this is more than just a hypothetical threat [2, 7–12]. These incidents reach from the erroneous issuance of CA certificates [12] to the complete compromise of CAs as in the case of DigiNotar [9].

As the risk of relying on a malicious CA grows proportional to the total number of trusted CAs, it is desirable to reduce this number to minimize the attack surface. However, a global limitation of the trusted CAs is problematic. It would lead to interoperability problems and browser warnings whenever a certificate issued by an unknown CA is presented. The problem with warnings is, that users get used to and tend to ignore them (see e.g., [13, 14]), even leading to a weakening effect.

The work at hand deals with the issue of the *individual* limitation of trusted CAs. We present a tool that allows to identify the set of CAs relevant to a specific user based on his browser history. We conduct a user study and evaluate how the currently deployed Web PKI is observed from a user's point of view. We show, that the set of CAs relevant to a user is indeed highly dependent on his individual browsing behavior. A thorough analysis and characterization of the individual view on the Web PKI can help to minimize the attack surface in the future. We show that there is an immensely high potential to improve the security by locally maintaining an individualized set of trusted CAs. Furthermore, we present our ongoing work to realize such a trust management system.

The paper is organized as follows. We provide background on the Web PKI and related work in Section 2. In Section 3, we present our tool *Rootopia*. We describe the setup of the user study in Section 4. In Section 5, we present the findings and evaluate the data collection method in Section 6. Then, we present an overview on a system that realizes the identified security improvements and discuss future work in Section 7.2. Finally, we conclude this paper in Section 8.

2 Background & Related Work

2.1 The Web PKI

The Web PKI is based on the X.509 standard [15], and the CAs accordingly issue X.509 certificates. Among others, X.509 certificates have an issuer and a subject field and contain a public key. The issuer field contains the Distinguished Name (DN) of the certifying CA, while the subject field contains the DN of the entity, whose key is certified. To ensure this binding, certificates are digitally signed by the issuing CA. The certificates are used during the SSL/TLS protocols to establish secure connections and in this context they are mostly used to authenticate web servers.

The Web PKI uses a hierarchical but tightly interwoven structure of CAs that digitally sign certificates. The Web PKI has a set of *Root CAs*. The Root CAs act as basis for the whole PKI. Root CAs sign certificates for *subordinate CAs* (Sub CAs which themselves sign certificates for other Sub CAs and web servers. This way, a

hierarchical structure is created. Besides that, CAs may mutually issue certificates (called cross-certification) to each other which makes the system even more complex.

A sequence of certificates starting with a Root CA's certificate and ending with a web server's certificate is called *certification path*. The process of checking the certification path for correctness and validity is called *path validation* [15]. The intention is, that the subject of a certificate in the path is either the issuer of the subordinate certificate or the web server. Furthermore, the signature on a certificate in the path must be verifiable with the public key in the preceding certificate.

During the TLS handshake, a web server presents its certificate along with the certification path to the client. The client validates the certification path and checks if it starts with a trusted Root CA and if the data contained in the last certificate identifies the communication partner. If so the client trusts in the authenticity of the server. The public key is extracted from the certificate and used to establish session keys to secure the communication.

The public keys of Root CAs are distributed within trusted lists called *root stores*, along with operating systems and browsers. Thus, browser and operating system vendors globally define – according to their specific policies [16, 17] which comprise certain security and audit requirements – which CAs are trusted.

Over the past, the number of CAs included in those root stores has been constantly growing. For example, the root store of the Mozilla browser comes together with the NSS crypto library and contains about 160 CAs [18, 19]. Another example is Microsoft's root store which contains about 264 CAs[1] [4], which are directly trusted. Together with the respective Sub CAs which are transitively trusted, there are approximately 1,500 CAs located in 52 different countries [5,6]. And each one can arbitrarily issue certificates for any domain.[2]

That such a global system, where trust decisions are made based on the uniform acceptance of 1,500 CA certificates is error prone can be seen from the various security incidents presented in the introduction. Compromising or compelling a single one of the trusted CAs allows a potential attacker to impersonate as any web server, or to mount a man-in-the-middle attack on any SSL/TLS secured connection, thus opening doors for Internet fraud and surveillance.

A simplified example of the resulting Web PKI is depicted in Fig. 1. Here, an exemplary certification path exists from the Root CA R-CA_1 to the end entity EE_1, where the arrows represent certificates. The circular arrows stand for self-signed certificates, which are often issued by Root CAs to themselves in order to publish their keys. To validate EE_1's certificate, one only needs to know the key of R-CA_1. All other keys are shipped within the intermediary certificates. In this small example, it can also be seen, that it can be difficult to determine all the trusted CAs. For example if R-CA_3 were removed from the root store, its direct Sub CA S-CA_4 would still be trusted due to the additional chain from R-CA_2. However, as there exists no public repository of all the certificates, the existing chains are in general unknown to users until they are presented during connection establishment.

We present related work dealing with the problems of the Web PKI in the following section.

[1] due to a silent update mechanism

[2] The name constraints extension [15] can be used to limit the power of a CA, however it is almost never used [6].

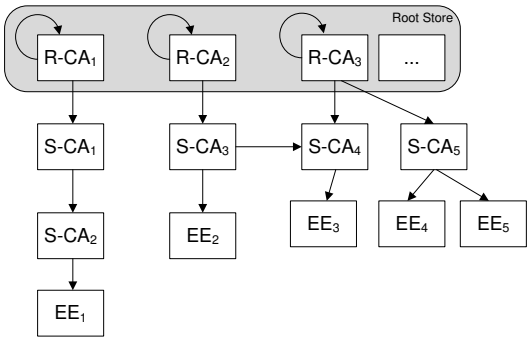

Fig. 1. Example Web PKI

2.2 Related Work

Several works exist, that aim at understanding the current deployment of the Web PKI. The approaches reach from active scanning the IPv4 space [5, 11, 20, 21] or the most popular web pages for SSL/TLS connections and passive monitoring of Internet traffic [18, 22]. The certification paths obtained during the SSL/TLS handshakes are stored in databases and used to analyze the quality of certification practices, determining the landscape of existing CAs and for the detection of malpractices and malicious certificates. These works provide valuable input to understanding the problems of the Web PKI to finally feature security improvements.

Notarial solutions like the ICSI SSL Notary [22], Convergence [23] or Perspectives [24] provide the possibility to a client to check the certificates it obtained from a server against the certificates in such an above mentioned database of formerly observed certificates or against the certificates the notary obtains when it accesses the server in question. In some cases also consensus decisions of several independent notary servers are involved. Such services help to detect targeted attacks and are for example used to detect and track the location of the adversary which mounts a man-in-the-middle attack [25]. While providing valuable input for a reconfirmation of certificates in doubt, notaries also come with disadvantages. For example there are communication overhead, privacy issues and the information notaries provide may be fallible especially when new certificates are in question.

Certificate or public key pinning [26] is a mechanism that lets users locally store certificates of websites when they are accessed for the first time and then reuse the stored public keys afterward. This implies the trust on first use approach and requires an adversary to be present during the first connection establishment. Browser add-ons realizing this mechanism [27, 28] often provide additional information to the users to support trust decisions, which in fact only helps users that have the expertise to understand this additional information. Public key pinning may also be realized in an application specific manner as in Google's Chrome browser, which is shipped with the public keys for Google services. This in fact led to the detection of the DigiNotar compromise [9].

Certificate transparency [29] is an experimental proposal for publicly logging the existence of SSL certificates to allow public auditing of CAs and the detection of erroneously issued certificates.

A recently often discussed alternative to the X.509 PKI is the binding of certificates – above all self-signed certificates – to Domain Name System (DNS) names using DNSSEC [30]. This mechanism removes many drawbacks of the current PKI system yet security then relies on the security of the DNS infrastructure. Furthermore, DNSSEC is still not completely accepted.

So far, detailed studies on the issue of individual user requirements concerning the Web PKI to our knowledge are not available. We are only aware of experimental self-studies [31, 32].

3 The Tool – Rootopia

Our tool called Rootopia is implemented in Java 1.7 to be as platform independent as possible. It currently runs under Windows and Mac OS X, but will soon be available also for Linux. It can process history files from Mozilla Firefox or Google Chrome which are stored in SQLite Databases. Microsoft Internet Explorer (IE), is supported in combination with the external tool IEHistoryView [33]. IEHistoryView extracts the IE history and stores it in a text file, which can then be processed by Rootopia. With the possibility of supporting these three browsers we are currently able to serve over 90% of users [34].

3.1 Functionality

Rootopia is run locally on the user's machine. It analyzes browser histories to enumerate the CAs the user relied on in the past. It determines when and how often the CAs were observed by the user. First, the hosts to which https connections were established in the past are extracted from the history. The https hosts are filtered for multiple occurrences and sorted by the date when they were first accessed. For each host, the date of the first and the last visit is stored to draw conclusions on the dates when the related CAs were observed. After processing the history, Rootopia establishes a TLS connection to each of the hosts and retrieves the certificates provided by the host server. The certificates are analyzed to identify the involved CAs and the corresponding certification paths. We use the path validation provided by the Java Cryptography Architecture with the default TrustManager, which implements the standard X.509 path validation. We only include valid certification paths in our analysis to ensure, that only valid and globally visible CAs are counted. In case path validation fails, we store the certificates to evaluate the failures. We refer the reader to Section 6.2 for a discussion.

The collected data is stored into different files to be available for further investigation and comparison. We decided to store the data within CSV files, as these are on the one hand easily machine readable and on the other hand can be conveniently viewed and processed by major spreadsheet programs. This enables users to see and decide which data they provide for analysis counteracting privacy concerns. Additionally, Rootopia provides a visualization of the connections and dependencies between observed CAs. Also a time course with the observed CAs is shown, so the user can see which CA has been seen for the first time at a specific date.

3.2 Collected Data

Now we describe the data groups that are collected and explain why this data is interesting. Table 1 shows the data sets we collect for each observed CA. We identify

Data ID	Meaning
DN	identifies the CA
CA kind	specifies the role of the CA (Root, Root/Sub, Sub) in which the CA was observed
certificate	the certificate(s) certifying the CA
first seen	specifies the date when a CA was first seen
last seen	specifies the date when a CA was seen for the last time
Sub CAs	DNs of the CAs certified by the CA
Super CAs	DNs of the CAs certifying the CA
# hosts	number of hosts that have the CA in their certification path
# visits	total number TLS connections involving the CA
EE CA	boolean, specifies if a CA certifies end-entities

Table 1. Collected Data for each CA

CAs with their distinguished names (DN) extracted from the issuer and subject fields of the obtained certificates. "first seen" is the date, when – according to the user's history – for the first time a connection to any host was established where the CA was involved into the certification path. "last seen" analogously specifies the date when for the last time a connection was established to any host involving the CA. With that data it is possible to examine, how the view on the Web PKI changes over time according to the user's browsing behavior.

The number of hosts and visits shows the relative importance of the respective CA for the user. The number of visits is the sum of the number of visits of the hosts related to the respective CA. Sub CAs and Super CAs represent the relationship between CAs. "CA kind" describes how the respective CA appears in the different certification paths. All Root CAs and Root/Sub CAs together define the absolutely minimal set of CAs that must be contained in a user's root store in order to be able to validate the certification paths for all his previously accessed hosts. Root/Sub CAs are such CAs, that are seen both, sometimes as Root CA and sometimes as Sub CA. "EE CA" is set to true if the respective CA has issued at least one end entity certificate. All those CAs together are the absolutely minimal set of CAs that have to be trusted. If one knows those CAs, all host certificates could be validated with a chain of length one, meaning that certification of Sub CAs could be completely ignored.

4 Web PKI User Study - Setup

For the pilot study whose results are presented in Section 5, we analyzed the histories of 22 volunteers from our university. After executing Rootopia on their browser histories, we collected the output data for further analysis. Besides that, we collected metadata using a questionnaire to be able to group the people into different categories. Within the questionnaire, we ask for different aspects which might have influence on the browsing behavior and thereby on the observed part of the Web PKI.

4.1 Ethics of Data Collection

Collecting user data involving the analysis of browser histories is a privacy sensitive topic. Therefore, we chose the approach to analyze the histories locally instead of

collecting user histories for analysis. Afterward, the data extracted by Rootopia is collected in an opt-in process, i.e., the users are required to actively hand over the data to participate in the study. Before doing so, the collected data was explained to the participants personally to enable them to understand the extend of data collection, and if required to refuse their data from being used.

Besides that, the data is fully anonymized before further analysis. The analysis does not consider the whole browser history but only the hosts accessed via https. And, it is possible for a participant to deny the storage of the found host names, such that only the involved CAs are available in the further analyzed data set.

4.2 Questionnaire

The questionnaire consists of the questions summarized in Table 2. The data is to be used to group the participants in order to analyze differences between user groups. The country of origin and country of residence are interesting, as people from different countries might be interested in different web pages due to language and social background. The interesting question here is if this has influence on the CAs, and in particular, on the countries where the CAs are located. Furthermore, PC and Internet usage give information on how intensively the PC and the web is used. Intensive use may on the one hand lead to a larger set of CAs and on the other hand, may lead to the complete set of required CAs in a much shorter time span. Furthermore, the use of business PCs is often restricted according to security policies of a company. Besides that, people that use services like e-commerce, e-banking and e-government are more likely to often come in contact with secure connections. IT security and general IT expertise may have implications on how people use the Internet in general. The last question refers to https enforcement tools like HTTPS Everywhere [35]. Those tools enforce https connections instead of http in many cases, which might have influence on the set of seen CAs. We analyze the data according to these aspects. The results can be found in Section 5.

Criterion	**Possible Answers**
gender	male / female
country of origin	country name
country of residence	country name
PC usage	private / business / both
Internet usage	
e-commerce	yes / no
e-banking	yes / no
e-government	yes / no
hours per day	# of daily online hours
IT security expertise	expert / knowledgeable / some familarity / no familarity
general IT expertise	expert / knowledgeable / some familarity / no familarity
use of https tools	name of the tool, otherwise "-"

Table 2. Collected metadata per participant

4.3 The Participants

We analyzed the browser histories of 22 persons. Four persons provided two histories, either from different browsers they use in parallel, or different PCs. We ended up with 26 history files. All participants currently live in Germany, but have different cultural backgrounds. 16 of the participants originate from Germany, 2 from Poland, 2 from Morocco, 1 from Iran and 1 from China. The participants reach from IT experts to persons that only occasionally use a PC. The participants are between 25 and 57 years old. All of the participants either use Chrome or Firefox.

5 Findings

The data of the participants is aggregated and analyzed. Thereby we derive user specific information as well as similarities and differences among user groups. Table 3 shows aggregated numbers concerning history lengths and observed CAs. In the analysis we distinguish between true Root CAs and CAs that were seen both as Root and as Sub CAs (Root/Sub CAs). This resulted from cross-certification between Root CAs or the occasional inclusion of superordinate CAs into the certification path, even if one of the intermediate CAs is present in the root store. As both Root and Root/Sub CAs must be present in the root store to be able to validate all observed certification paths, in the following we refer to the sum of them as the "Root CAs" if not explicitly distinguished between the two cases.

Interestingly, none of the users – even those with a huge number of different https hosts – did see more than 22 different Root CAs, which is about 13.4% of the 164 CAs included in the Firefox root store. Furthermore, a maximum of 75 Sub CAs was reached. The absolute maximum of CAs in total seen by a single Internet user was 96, which is 6.4% of the 1,500 trusted CAs of the Web PKI. Even fewer CAs were found when only considering CAs that signed host certificates. These CAs represent the minimum number of CAs that need to be trusted by a user to be able to verify all the certificates of the hosts he connected to. The maximum value of such host signing CAs was 68 or in other words 4.5% of the currently trusted CAs. The ratio of host signing CAs was in the span of 50%-75% of the total CAs found for the respective user and reached 63% on average. Only one of the participants used HTTPS Everywhere. However, apart from the fact that this participant was one of the four users with most https connections, we could not identify special characteristics within the respective set of required CAs.

Criterion	**Average**	**Min**	**Max**
Duration of analyzed period (months):	18	4	38
Total number of https hosts:	168	12	636
Total number of https connections:	18,475	162	159,882
Total number of Root CAs:	10	4	14
Total number of Root/Sub CAs:	4	0	8
Total number of Sub CAs:	36	11	75
Root + Root/Sub CAs:	14	4	22
# CAs that signed host certificates:	33	8	68

Table 3. History sizes and numbers of observed CAs

Considering the total number of different Root and Sub CAs observed by the whole group of participants, namely the union of all sets of CAs, leads to 28 Root CAs and 145 Sub CAs (please find a list of all CAs in Appendix B). The numbers show that there is a high potential in limiting the number of trusted CAs. Furthermore, for certain user groups, there is a high overlap in the CAs (i.e. CAs that were observed by several persons). The overlap is significantly higher for Root CAs than for Sub CAs, which can be seen as the set union of Root CAs is only 27% larger than the maximum number of Root CAs of a single user, while in the case of Sub CAs the set union consists of twice the number of Sub CAs seen by a single user. However, the significant differences in the numbers for different users – reflected in the minimum and maximum values – shows, that true minima for a single user can only be reached by individualization. Yet, grouping the users into dedicated user groups can lead to good results.

One influencing factor leading to a low number of different CAs is surely the fact, that there are few large CA companies with a high market share in the certification business. However, when considering the distribution we observed among those large players, it turns out that it is not according to the market shares from the Netcraft SSL Survey [36]. Most significant, VeriSign, Inc. is involved into more than 20% of the certification paths relevant for our user group, while it has only around 6% of the market share in the Netcraft Survey. In contrast, Go Daddy with more than 20% of market share achieves only a rather low rate in our data. Namely, Go Daddy was a Root CA in less then 4% of the certification paths. This is another indication, that it highly depends on the individual browsing behavior of the users, which CAs are truly relevant for them. For a complete List of the CAs and their respective relevance, we refer the reader to Appendix B.

We also grouped the observed CAs by country. It turned out that – compared to the total of 52 countries – CAs from only 14 different countries were relevant for the considered user set (see Figure 6). The overwhelming majority of CAs is from the US (US) followed by Germany (DE), Great Britain (GB) and Belgium (BE). Considering the other countries, less than 5 CAs were observed from those and often only by very few users (cf. Section 5.2 for details). This shows, that it might even be viable to limit the number of trusted CAs based on the country they reside in.

Now we present detailed results also considering different user groups.

5.1 Temporal Evolution

In the following, we discuss our findings concerning the development of the individual views on the Web PKI over time according to the dates when related hosts were accessed. It turns out that, in general, the number of observed CAs does not grow linear but shows restricted growth with high growth rates in the first few months. Considering Root CAs, the upper bound is reached after several months. However, growth rates depend on the intensity of Internet usage or rather on the number of https hosts a user connects to.

Considering users with high numbers of https hosts the upper bound is reached faster than for users that only connect to https occasionally. For Sub CAs, the development is similar to the Root CAs, however, it is less significant. Thus, the number of Sub CAs tends to keep growing over a long time. The temporal evolution of the numbers of Root and Sub CAs for selected participants of the study are shown in Figures 2 - 5. For the grouping, we used the number of different https hosts averaged

over the length of the analyzed time span. The average was approximately 9 hosts per analyzed month. The first two figures show the data for the four users that reached twice the average of different hosts per month, i.e. use https intensively. In contrast Figures 4 and 5 show the evolution of the view on the Web PKI for the ten users that only reached half the average of https hosts per analyzed month. The number of CAs depicts the sum of different CAs observed until the respective month according to the user's history. Thereby, each line represents the data of one user.

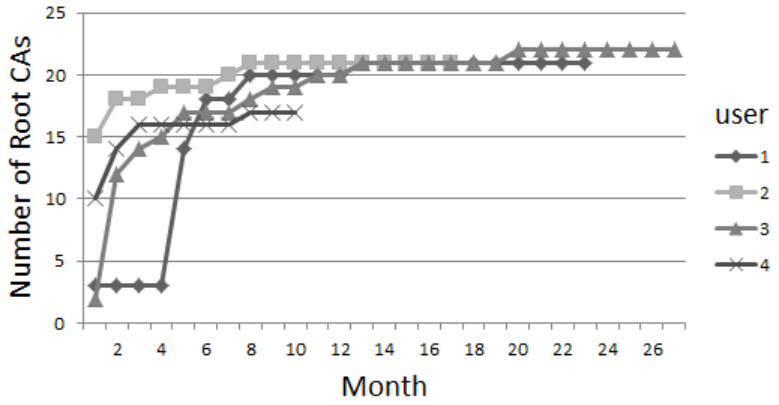

Fig. 2. Temporal Evolution: Root CAs - users with more than 18 different https hosts / analyzed month

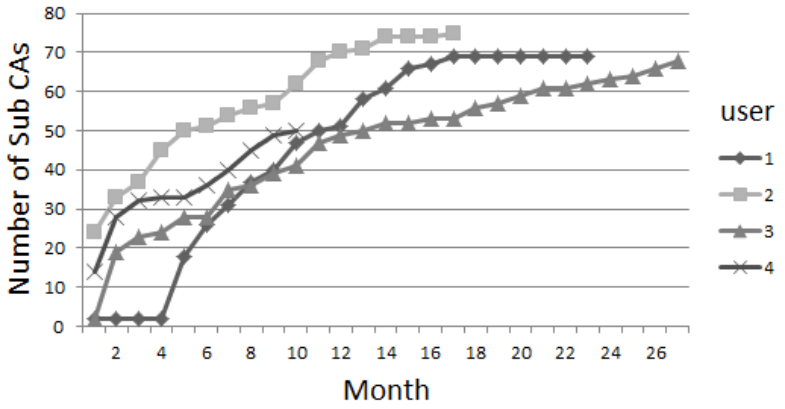

Fig. 3. Temporal Evolution: Sub CAs - users with more than 18 different https hosts / analyzed month

For the users that use https less intensively, it takes a much longer time span until the number of CAs tends towards an upper bound. Yet, the upper bounds lie strictly below the ones observed for users which use https a lot. On the other hand, there also exist users, that only connect to a very limited number of hosts but where the upper

bounds on CAs are reached after very few months. This can be seen best in one data set, where the maximum of 4 Root CAs is reached after 3 months and is constant afterward (16 months). The picture for Sub CAs is nearly the same in that data set. Further investigation showed, that the data belongs to a person using e-banking and e-commerce services, but besides that only occasionally surfs the Internet (a fact, which was identified during a personal discussion).

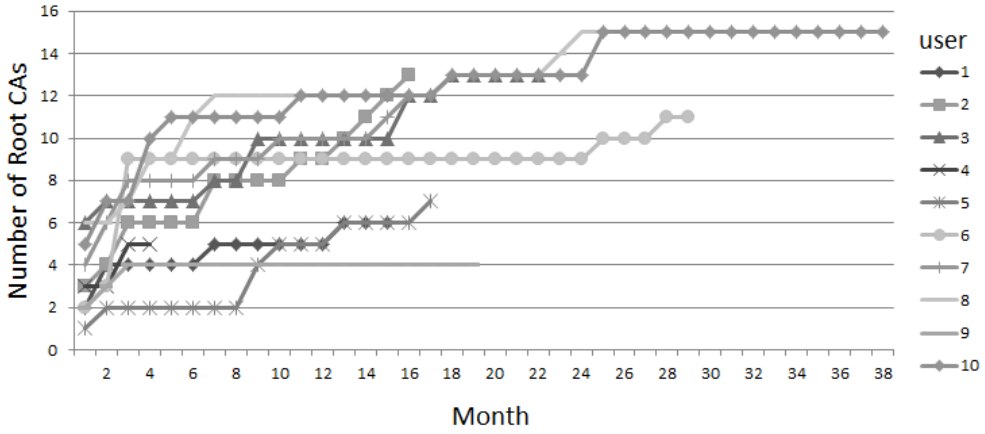

Fig. 4. Temporal Evolution: Root CAs - users with less than 5 different https hosts / analyzed month

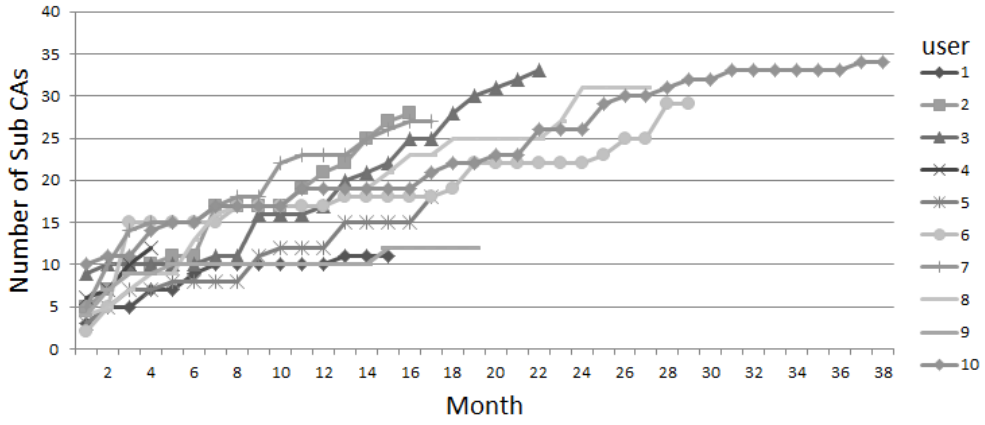

Fig. 5. Temporal Evolution: Sub CAs - users with less than 5 different https hosts / analyzed month

To summarize our findings on the development over time, we state that it is not possible to give a concrete number of months after which all relevant CAs have been seen and the number of CAs stagnates. This is highly depended on the individual browsing behavior. In many cases – due to the regular deletion of the histories – these

are not long enough to derive the upper bound and the set of relevant CAs for the respective user completely. Yet, in general, our observations show that the number of CAs tends towards an upper bound significantly below the total number of existing CAs. This in turn shows the potential for the possible security gain by limiting the number of CAs. For completeness, please find the temporal evolution of the number of CAs for all analyzed data sets in Appendix A.

5.2 CA Countries

As stated above, most of the observed CAs are from the US. Figure 6 shows the observed countries and the number of CAs including all data sets, where each color represents one analyzed history. The second most observed country in our set of participants is Germany. However, this is also a user group dependent outcome and results from the set of analyzed histories. A large number of participants are either from the scientific community or students at a university. Building two groups, the first containing people with academic background and the second one without, shows that German CAs occur much less often in the second group. The percentage of German CAs is on average 18.3% of all observed CAs per user in the first, and only 7.1% in the second group. It results from the fact, that most universities have their own CAs, certified by the DFN Root CA. Those CAs are completely irrelevant for the non-academic users. The distribution of CAs over the other countries did not change significantly.

We also grouped the data into users that originate from Germany and those who do not. Yet, interestingly this did not have significant effects on the distribution over the countries. However, when considering single users, the relevant CA countries can depend on the country of origin as we observed it for a user from Poland (PL). A grouping into different countries of residence would be interesting, yet could not be done with our data set and thus is left for future work.

Considering all data sets, there are some country codes that were observed for most of the participants, yet where the respective CA was always one and the same. These are SE, ZA, NL, and IE. We collected the CAs in Table 4.

Country	DN
ZA	EMAILADDRESS=premium-server@thawte.com, CN=Thawte Premium Server CA, OU=Certification Services Division, O=Thawte Consulting cc, L=Cape Town, ST=Western Cape, C=ZA
NL	CN=TERENA SSL CA, O=TERENA, C=NL
SE	CN=AddTrust External CA Root, OU=AddTrust External TTP Network, O=AddTrust AB, C=SE
IE	CN=Baltimore CyberTrust Root, OU=CyberTrust, O=Baltimore, C=IE

Table 4. Important CAs from rarely observed countries

For the remaining countries (KR, PL, UK, BM, FR, AU) no fix pattern is observable. From these, FR and BM are observed most often.

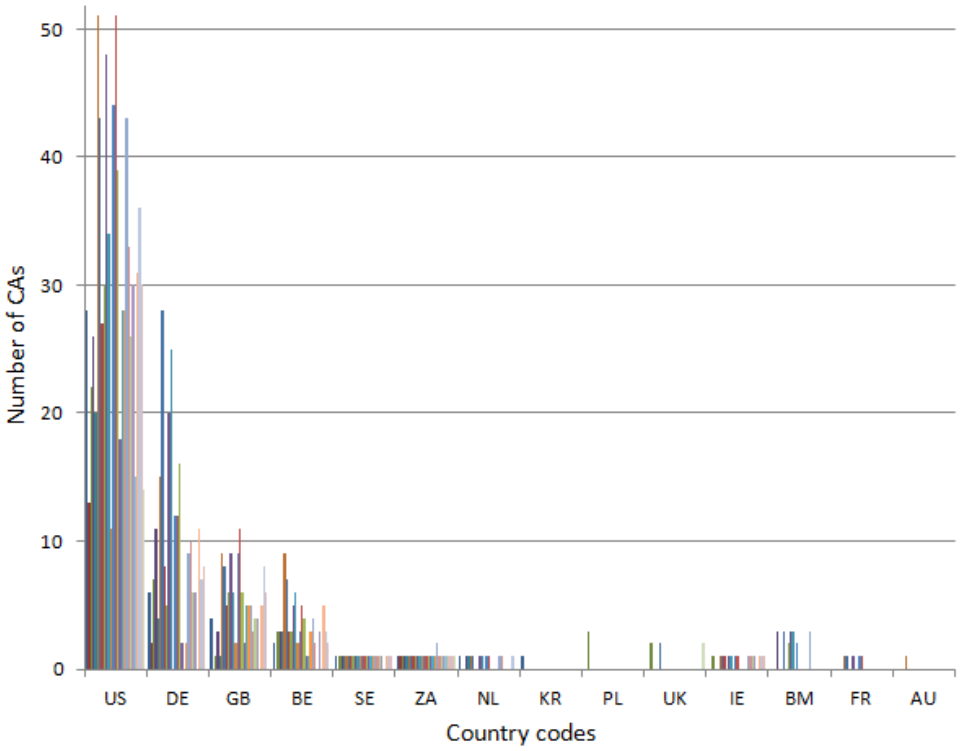

Fig. 6. Distribution of CA countries, different colors represent different users

5.3 Relevance of CAs

To measure the relevance of a CA for a user, we counted the number of hosts related to the respective CA. Interestingly, the number of Sub CAs that are related to only one host lies between 20% and 60% of the total number of Sub CAs found for a user, and is about 43% on average. This shows that Internet users observe many CAs whose relevance is really low. Thus, it is highly questionable if the benefits for the user by fully trusting into those CAs counterbalances the imposed risks, not speaking about the CAs a user never observes.

As it might occur that a single host is accessed extremely often by one user and thus the related CA becomes more relevant to him, we also measured the number of visits, namely taking into account how often a host was accessed. As expected, the number of Sub CAs only observed during a single connection is lower. But still, rates of up to 38% of the total number of CAs for single users are reached and are 17.5% on average. That shows that many of the CAs are only observed by chance. Furthermore, our data shows that a user observes the CAs most relevant for him during the first months, while CAs which are found later are less relevant, both either measured by the relative number of hosts or visits.

For each CA, we also averaged the CA's relevance over all users that observed the respective CA. It turns out, that there is a strong correlation between the number of users that observed a CA, and the averaged relevance of the respective CA. The

numbers can be found in Appendix B. From this findings we conclude, that building user groups and taking the CAs which most users of that group have in common can be a good starting point to set up an individualized set of trusted CAs for e.g. a user where no history data is available.

5.4 Number of CAs and Overlaps

We computed the union set of CAs for different user groups. To identify the similarity of the views on the Web PKI within a group, we computed overlaps in the CA sets, namely how many users have how many CAs in common. If not differently specified, in the following with overlap we mean the ratio of CAs that all group members have in common.

The group of the four users with most https hosts as specified in Section 5.1 jointly observed a total of 25 Root CAs and 108 Sub CAs. With 64% the overlap of Root CAs is twice the overlap of Sub CAs (31%). That shows, that the set of Root CAs relevant to a user is less dependent on the individual browsing behavior. This also holds for the other groupings we analyzed and is as expected, as the total number of existing Root CAs is nearly ten times smaller than the number of Sub CAs. Comparing the 25 Root CAs and 108 Sub CAs with the complete set of CAs jointly observed by all users, it turns out that the CAs seen by the users with most https connections make up for 89% of all Root CAs and 74% of the Sub CAs. Thus, most of the CAs required by the other users are also seen by the users with most https connections.

When comparing the groups of academic and non-academic users, the first observes significantly more CAs (27 vs. 19 Root CAs and 140 vs. 63 Sub CAs). This seems to result from the fact, that all the users with most https connections are also part of the academic group. The overlaps in the academic group are higher than in the non-academic group.

However, to really do a fine grained grouping, more data sets are required. High overlaps were achieved only for the group of users with most https connections, thus an interesting remaining question is if this results from the fact that these users reach a set of CAs that satisfies the requirements of most of the users or if the grouping resulted into a good match of browsing behavior. There are indications for both. The observation, that the remaining users do not need too many additional CAs speaks in favor of the first. On the other hand, the users with most https connections are all from the same scientific working group, and have comparable backgrounds which could indicate a close match of browsing behavior.

6 Discussion

In this section, we discuss limitations of the data collection method and evaluate the influence on our results.

6.1 Collecting Data after the Fact

The problem with rebuilding the view on the part of the Web PKI a user has seen so far, is that the CA data is not directly available from prior interactions. Certification paths obtained to establish an https connection are not stored. Some browsers, like Firefox, cache intermediate CAs from former visits. However, it is not possible to

determine when the CA was first seen and how often. Thus, one cannot examine the development of the user's view and the importance of a CA for a user directly from existing data.

What we actually get from our approach is a current snapshot of the Web PKI seen by the user, i.e. we see all the CAs which are required at the moment of the analysis, to be able to establish all former https connections. If hosts moved from one CA to another in the meantime e.g. when their certificate expired, this is not reflected within our data.

Furthermore, there are hosts that have their certificates issued by several different CAs. Huge server farms such as Google or Facebook, are known for that practice. If we consider Google as an example, the issuers of certificates for Google currently include Verisign, Google Internet Authority, Equifax, GeoTrust, and DigiCert [37], which shows that the number of different CAs can in practice be quite large. One possibility could be to adjust this manually to complete the views on the Web PKI. However, further research is needed to identify these hosts and their actual behavior.

In summary, the assumption that a host sticks to one CA is a simplifying assumption. Subsequently, the sets of CAs we collected for the users may have some inaccuracies. However, as the assumption is true for most of the hosts, our analysis is close to reality. Furthermore, we tested the behavior by connecting to Amazon, Google, Facebook and Dropbox repeatedly over several days. The only host presenting different certificates during our tests was Dropbox.

6.2 Path Validation Errors

We only considered valid certification paths in our evaluation of the user's views to ensure that only publicly visible CAs are counted (i.e. to exclude manually installed CAs). On average, about 10% of the connections failed due to path validation errors, which is strictly below the numbers of failed chains observed by other studies [18]. A further analysis showed that Firefox as well as Chrome do not store URLs in the history in case path validation fails and if no exception is added by the user. Thus, many of the invalid paths are filtered by the browsers and therefore do not occur in our analysis.

On the other hand, we still observed path validation errors. Half of the fails result from Java's smaller root store which contains 79 Root CAs compared to Firefox and Microsoft. One example is the StartCom CA. The other half of the fails resulted from incomplete chains and self-signed certificates.

Thus, the failed path validations lead to the exclusion of several CAs that are actually seen by the users. Thus, our numbers slightly underestimate the total number of observed CAs. On the other hand, if the CAs are only identified within chains that cannot be validated, a removal of those CAs from the set of trusted CAs does not change the user experience as an error is shown anyway. However, for a future study, at least the applied root store needs to be extended.

In summary, when also considering CAs where path validation failed due to the Java root store, this increases the numbers of CAs by about 5% on average per user.

6.3 Future Development

While we can draw a good picture of a user's past view on the Web PKI it is not possible to predict the future. Our data shows, that the number of CAs approaches

a certain upper bound. However, new CAs even occur after long time periods. Thus, derived views might always lack some CAs that are required in the future.

6.4 Further Limitations

Another limitation in the approach evolves from the fact, that many users delete their history – partly or completely – quite often. In such cases, it is not possible to derive the CAs relevant to the user. Furthermore, the browser histories do not contain the CAs relevant to applications installed on the user's system. Thus, we miss those CAs that might be relevant for e.g. software updates.

Besides the unavailability of histories due to deletion, a reservation of Internet users against the approach could be observed. Rootopia analyzes a user's browser history and therewith his browsing behavior. As this is a privacy sensitive task, we were confronted with several privacy and security concerns of the users. On the one hand, users are not willing to hand out their browser histories. This is why, we used the approach of a local analysis and provided the participants with all the data which is sent to us for further research. However, many users also feel uncomfortable with executing unknown programs on their PCs, and convincing people that no privacy sensitive data is extracted is not always an easy task. This has to be considered when conducting a broad scale study.

6.5 Evaluation

As we have discussed, the sets of CAs derived from the browser histories lack some of the CAs a user might actually require. Considering these CAs would change the absolute numbers, however would have small effects on the general results of our study. Namely, the findings about the dynamics with which the user's views on the Wep PKI develop, the general distribution of the CAs and the dependencies between users and user groups do not change when considering several additional CAs.

Also, even the inclusion of the failed certification paths, i.e. the increase of the number of CAs observed by a single user by 5-10% has only minor influence when comparing this to the total number of CAs trusted within the Web PKI. If we consider the extreme case of the user that observed 96 CAs in total and we increase this by 10% we still end up with only 7% of the 1,500 trusted CAs of the Web PKI.

Furthermore, the dynamics of the views on the Web PKI imply, that an absolutely fixed set of trusted CAs is not a good solution. Things change and new CAs will always occur. Thus, a dynamic solution is required, that allows the adaptation of the set of CAs, which renders the exact determination of the CAs observed in the past much less important. The only possibility to exactly determine this set would be to do a long term study, monitoring users over months or even years.

The limited number of participants in the user study makes it impossible to derive results for all Internet users. For example, from our collected data we cannot derive a set of CAs required by a user living in the US and using online services from the US government. As mentioned above to do a more elaborate analysis of dependencies within user groups a broad scale study is required. On the other hand, optimal solutions will only be possible by true individualization. Furthermore, we do not aim to reduce the trusted roots globally. Each one is certainly required in some context and thus a global reduction seems not to be applicable.

We will sketch our approach to locally manage an individualized set of trusted CAs in the following section. The approach allows to individually minimize the attack surface for a user and takes into account our main results. For a detailed description of the approach please refer to [38].

7 Local Trust Management & Future Work

7.1 Minimizing the Attack Surface

We reduce the risk of relying on a malicious CA by reducing the number of trusted CAs to those that are really required on a per user level.

To achieve this, we locally manage what we call a *trust view* that serves as a local and user dependent knowledge base for trust decisions. We also apply public key pinning and notarial reconfirmation of certificates.

A user's trust view contains all the CAs he observed in the past along with additional information about certified hosts, dates and number of observations of the CA. Additionally, among the contained CAs, we introduce variable trust levels to enable more fine grained trust decisions. A core set of CAs in the trust view is then completely trusted. This means certificates from those CAs are considered trustworthy. Which CAs belong to this core set is defined by a local policy. An example for such a policy is, that a CA is fully trusted if it was involved into the certification of ten different hosts the user connected to in the past. Note that trust views do not replace standard path validation but introduce an additional decision mechanism if a valid certification path is to be considered trustworthy.

Whenever a user connects to a host that uses certificates issued by a CA not in the local trust view or issued by a CA which did not yet achieve the status of a fully trusted CA then the obtained certification path is checked with notaries first. If the notaries reconfirm the certificates, the involved CAs are added to the trust view and the related information is updated. By storing the related hosts for each CA in the trust view, CA pinning is realized for known hosts. CAs that issued certificates for a certain host in the past are also trusted to issue certificates for that host in the future.

Thus, CAs that have often been observed – and most probably will also often be observed in the future – achieve a higher level of trust than CAs that are barely observed. CAs with a low relevance for a user are only trusted for the hosts they were reconfirmed for. This provides a trade-off between trusting in (a limited set of) CAs and the costly reconfirmation of certificates. Additionally, privacy problems are mitigated as notary servers are only queried in rare cases and not for every connection establishment, which allows user profiling in the long run. Furthermore, the load for notary servers is reduced.

The trust view is built incrementally over time. Thus, directly after set up, it is empty and requires reconfirmations for each connection. Rootopia provides a possibility for bootstrapping and to initialize the trust view based on the history. As our findings have shown, the frequency with which new CAs are observed shrinks over time, thus reconfirmations will be required less often the longer the system is used. Yet, the set of trusted CAs is not fixed, but adapts to the user's behavior and thus compensates the incomplete information on future user requirements. The dynamic update of the trust view allows us to rely on and only trust a small set of CAs. But still no CA of the Web PKI is completely excluded and can be included if required.

However, while trust views can significantly lower the risk of relying on a malicious CA, they still do not provide perfect protection. If one of the formerly trusted CAs suddenly fails, the user may still falsely rely on a malicious certificate issued by such a CA. Yet, a CA compromise only threatens those users, that trusted in the CA before the compromise, which limits the benefit for attackers.

On the other hand it is also possible, that a certification path is falsely evaluated not to be trustworthy, which relies in the nature of basing decisions on incomplete information. Further research is needed on mechanisms to further improve the accuracy of trust decisions in the Web PKI.

7.2 Future Work

We are currently realizing the concept of trust views along with a Firefox plugin. To manage the trust in the CAs and evaluate trust along certification paths, we use computational trust [39]. Furthermore, we consider the required CAs within certain contexts, e.g. e-banking, e-commerce or general web surfing. This allows to make the decision if a CA is considered trustworthy context dependent. Assigning CAs to a certain context can further reduce the set of trusted CAs, limiting the impact of malfunctions of the system for critical services.

However, many browsers, e.g. Chrome and IE, use the root store of the operating system. The limitation of the root stores that reside within the operating system, is more problematic. A multitude of different applications relies on those root stores and thus, these dependencies must be considered to prevent applications from stopping to work properly. We will examine possibilities to detect CAs required by the installed applications to get a broader view on the required PKI parts and to identify possibilities for a minimization of these root stores. Thereby, interdependencies with root store updates as for example applied by Microsoft need to be considered. From our point of view, several separated sets of trusted CAs dedicated to different purposes have advantages over a central all-purpose root store in respect to the minimization of the attack surface.

While the history analysis provides a good possibility to identify the set of CAs required by a user and thus allows bootstrapping of trust views, this is not possible for users that do not store their browser history. In such cases some kind of group profiles can be interesting for bootstrapping. An interesting question is, if such group profiles can be derived and how they can be applied to define the set of required CAs for users where browser histories are not available. Even if such group profiles might overestimate the set of actually required CAs, this will still lead to a significant reduction of the number of trusted CAs and therewith reduce the risk of relying on a malicious one. For this, a larger study is required to identify potential differences between several user groups. Here, analyzing the requirements of users residing in different countries is of special interest. Furthermore, a large number of participants is required to be able to group users into different categories.

8 Conclusion

In this work we showed that the risk to be affected by CA malfunctions is unnecessarily high. We presented a tool that allows to derive and assess the personal requirements of Internet users based on their browser histories. It turned out, that the individual

views on the Web PKI tend towards a fixed individual set of CAs. The temporal evolution described in Section 5.1 actually shows different courses, thus confirming that the set depends on a user's individual browsing behavior. Our analysis revealed, that a reduction of trusted CAs by more than 90% is possible without restricting the respective user in his daily Internet use. We note, that a global limitation of the trusted CAs is no viable solution. The sets of required CAs are too distinct between different users. Thus, a global minimization of CAs cannot lead to an optimal solution. Furthermore, it would lead to interoperability problems and additional warnings whenever a certificate issued by an unknown CA is presented to the user.

We also found large differences in the relevance of the CAs, which leaves further room for improvement. Also, a limitation based on the countries the CAs reside in is promising. The CAs we observed for different users originate from a rather small set of countries. On the other hand, it turned out that it is a challenging task to completely define the set of relevant CAs for an individual user. One problem is the unavailability of sufficient data about the user's browsing history. In such cases, grouping users and deriving group profiles can help to provide a starting point for the limitation. Further research is needed to define such profiles. On the other hand, mechanisms are needed to deal with CAs that are newly observed and interdependencies between the set of trusted CAs and applications apart from browsers need to be considered. We have sketched our approach to enable individualized trust decisions. We conclude that the individualization of the set of trusted CAs bears huge security improvements. However, the realization of those improvements remains challenging.

References

1. Carl Ellison and Bruce Schneier. Ten Risks of PKI: What You're Not Being Told About Public Key Infrastructure. *Computer Security Journal*, 16(1):1–7, 2000.
2. Peter Gutmann. Pki: it's not dead, just resting. *Computer*, 35(8):41 – 49, aug 2002.
3. Peter Gutmann. *Engineering Security*. 2013. Book draft available online at `http://www.cs.auckland.ac.nz/~pgut001/pubs/book.pdf`.
4. Christopher Soghoian and Sid Stamm. Certified lies: Detecting and defeating government interception attacks against ssl. Technical report, Indiana University Bloomington - Center for Applied Cybersecurity Research, 2010.
5. The EFF SSL Observatory. `https://www.eff.org/observatory`.
6. Martín Abadi, Andrew Birrell, Ilya Mironov, Ted Wobber, and Yinglian Xie. Global authentication in an untrustworthy world. In *Proceedings of the 14th USENIX conference on Hot Topics in Operating Systems*, HotOS'13, pages 19–19, Berkeley, CA, USA, 2013. USENIX Association.
7. Comodo. The Recent RA Compromise. `http://blogs.comodo.com/it-security/data-security/the-recent-ra-compromise/`, visited Nov. 2011.
8. h online. Attack on Israeli Certificate Authority. `http://h-online.com/-1264008`, visited Nov. 2011.
9. FOX IT. Black Tulip - Report of the investigation into the DigiNotar Certificate Authority breach, 2012. `http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2012/08/13/black-tulip-update.html`.
10. h online. Flame – oversights and expertise made for Windows Update worst case scenario. `http://h-online.com/-1614234`, visited July 2012.
11. Peter Eckersley and Jesse Burns. The (Decentralized) SSL Observatory. Invited talk at 20th USENIX Security Symposium, August 2011.
12. Microsoft Security Response Center. Security Advisory 2798897 , 2012. `http://technet.microsoft.com/en-us/security/advisory/2798897`.

13. Cormac Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proceedings of the 2009 workshop on New security paradigms workshop*, NSPW '09, pages 133–144, New York, NY, USA, 2009. ACM.
14. Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. Crying wolf: An empirical study of ssl warning effectiveness. 2009. available online at `http://static.usenix.org/event/sec09/tech/full_papers/sunshine.pdf`.
15. D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley, and W. Polk. RFC 5280 – Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile. RFC 5280 (Proposed Standard), 2008.
16. Microsoft. Microsoft root certificate program, 2009. `http://technet.microsoft.com/en-us/library/cc751157.aspx`.
17. Mozilla. Mozilla ca certificate policy, 2013. `http://www.mozilla.org/projects/security/certs/policy/`.
18. Ralph Holz, Lothar Braun, Nils Kammenhuber, and Georg Carle. The ssl landscape: a thorough analysis of the x.509 pki using active and passive measurements. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, IMC '11, pages 427–444, New York, NY, USA, 2011. ACM.
19. mozilla. Mozilla CA Certificate Store - BuiltInCAs , 2012. `http://www.mozilla.org/projects/security/certs/`.
20. Peter Eckersley and Jesse Burns. An Observatory for the SSLiverse. Defcon 18, July 2010. `https://www.eff.org/files/DefconSSLiverse.pdf`.
21. Peter Eckersley and Jesse Burns. Is the SSLiverse a Safe Place? 27C3, 2010. `https://www.eff.org/files/ccc2010.pdf`.
22. ICSI. The ICSI Certificate Notary, 2013. `http://notary.icsi.berkeley.edu/`.
23. Moxie Marlinspike. Convergence. `http://convergence.io/`, visited July 2012.
24. Carnegie Mellon University. Perspectives Project. `http://perspectives-project.org/`, visited July 2012.
25. Ralph Holz, Thomas Riedmaier, Nils Kammenhuber, and Georg Carle. X.509 forensics: Detecting and localising the ssl/tls men-in-the-middle. In *ESORICS*, pages 217–234, 2012.
26. C. Evans, C. Palmer, and R. Sleevi. Public Key Pinning Extension for HTTP. Internet-Draft, 2013.
27. PSYC. Certificate Patrol. `http://patrol.psyced.org/`.
28. Amir Herzberg, Aviv Sinai, Alexander (Alex) Dvorkin, Itay Sharfi, Ahmad Jbara, Anatoly Vaitsman, and Roie Tov. TrustBar: Re-establishing Trust in the Web, 2006.
29. Ben Laurie, Adam Langley, and E. Kasper. RFC 6962 – Certificate Transparency. RFC 6698 (Experimental), 2013.
30. P. Hoffman and J Schlyter. RFC 6698 – The DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS) Protocol: TLSA. RFC 6698 (Proposed Standard), 2012.
31. netsekure.org. Results after 30 days of (almost) no trusted CAs, 2010. `http://netsekure.org/2010/05/results-after-30-days-of-almost-no-trusted-cas/`.
32. conetrix.com. How-to Limit the Number of Certificate Authorities Your Browser Trusts, 2011.
33. NirSoft. IEHistory View. `http://www.nirsoft.net/utils/iehv.html`.
34. W3C. Browser statistics and trends, 2013. `http://www.w3schools.com/browsers/browsers_stats.aspl`.
35. Electronic Frontier Foundation. HTTPS Everywhere. `https://www.eff.org/https-everywhere`. visited July 2012.
36. Netcraft. Netcraft SSL Survey, 2011. `http://news.netcraft.com/ssl-survey/`.
37. ImperialViolet. Public key pinning, 2011. `http://www.imperialviolet.org/2011/05/04/pinning.html`.
38. Johannes Braun, Florian Volk, Johannes Buchmann, and Max Mühlhäuser. Trust Views for the Web PKI. In *EuroPKI 2013*, LNCS. Springer, September 2013. to appear.

39. Sebastian Ries, Sheikh Mahbub Habib, Max Mühlhäuser, and Vijay Varadharajan. Certainlogic: A logic for modeling trust and uncertainty (short paper). In *TRUST 2011*, pages 254–261. Springer, 2011.

A Temporal Evolution

The Figures 7 and 8 show the temporal evolution of the user's views on the Web PKI for all analyzed histories.

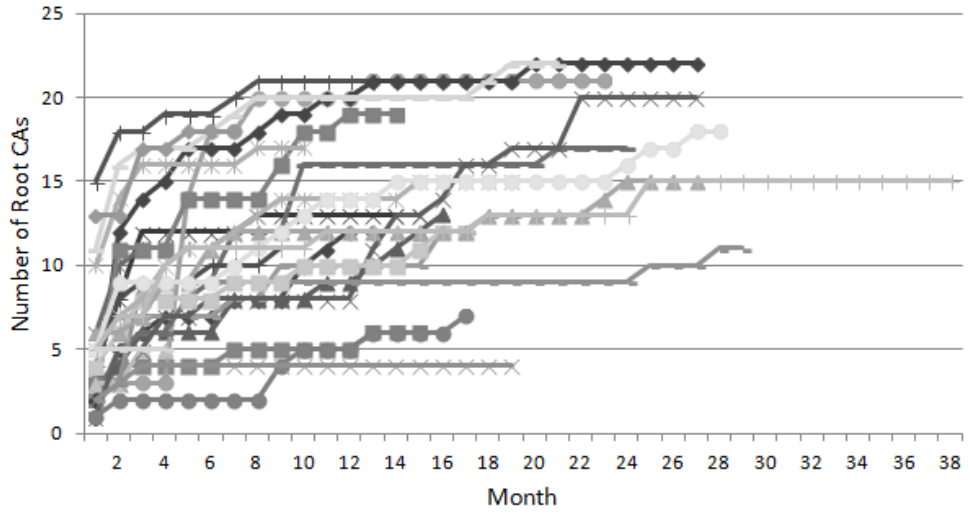

Fig. 7. Temporal Evolution: Root CAs

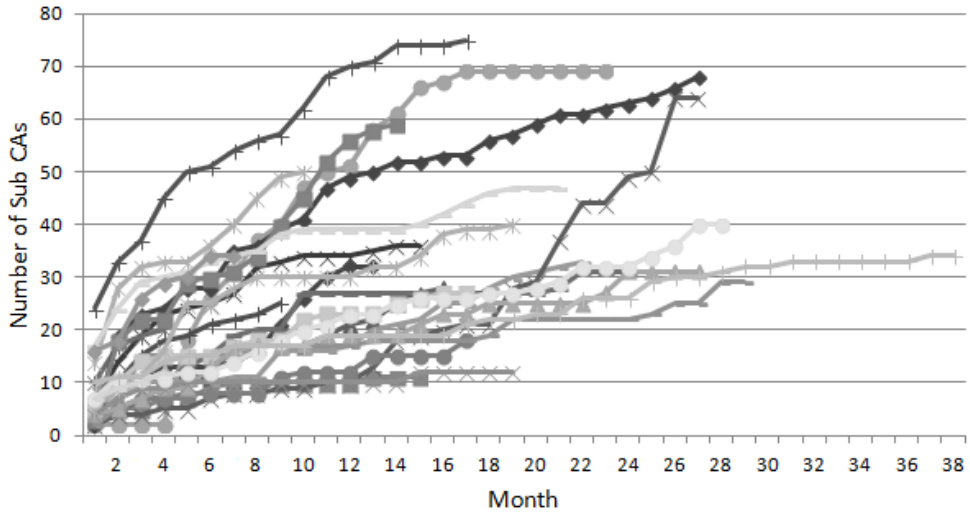

Fig. 8. Temporal Evolution: Sub CAs

B Observed CAs

Table 5 shows all CAs that where observed as Root or Root/Sub CAs in one of the analyzed user data sets. “Relevance by data sets” and “Relevance by percentage of hosts” are indicators for the relevance to the user group whose data was analyzed. The former shows the number of user data sets that contained the respective CA, the

Distinguished Name (DN)	Relevance by no. of data sets	Relevance by percentage of hosts
OU=Equifax Secure Certificate Authority, O=Equifax, C=US	26	16,65%
EMAILADDRESS=premium-server@thawte.com, CN=Thawte Premium Server CA, OU=Certification Services Division, O=Thawte Consulting cc, L=Cape Town, ST=Western Cape, C=ZA	26	11,52%
OU=Class 3 Public Primary Certification Authority, O=”VeriSign, Inc.”, C=US	26	20,54%
CN=AddTrust External CA Root, OU=AddTrust External TTP Network, O=AddTrust AB, C=SE	23	9,62%
CN=GeoTrust Global CA, O=GeoTrust Inc., C=US	23	7,42%
CN=GlobalSign Root CA, OU=Root CA, O=GlobalSign nv-sa, C=BE	22	4,00%
CN=VeriSign Class 3 Public Primary Certification Authority - G5, OU=”(c) 2006 VeriSign, Inc. - For authorized use only”, OU=VeriSign Trust Network, O=”VeriSign, Inc.”, C=US	22	20,09%
CN=GTE CyberTrust Global Root, OU=”GTE CyberTrust Solutions, Inc.”, O=GTE Corporation, C=US	21	4,01%
CN=Entrust.net Secure Server Certification Authority, OU=(c) 1999 Entrust.net Limited, OU=www.entrust.net/CPS incorp. by ref. (limits liab.), O=Entrust.net, C=US	20	3,51%
CN=Deutsche Telekom Root CA 2, OU=T-TeleSec Trust Center, O=Deutsche Telekom AG, C=DE	18	10,02%
OU=Go Daddy Class 2 Certification Authority, O=”The Go Daddy Group, Inc.”, C=US	18	3,82%
CN=TC TrustCenter Class 2 CA II, OU=TC TrustCenter Class 2 CA, O=TC TrustCenter GmbH, C=DE	16	2,10%
CN=DigiCert High Assurance EV Root CA, OU=www.digicert.com, O=DigiCert Inc, C=US	16	4,35%
CN=Entrust.net Certification Authority (2048), OU=(c) 1999 Entrust.net Limited, OU=www.entrust.net/CPS_2048 incorp. by ref. (limits liab.), O=Entrust.net	14	1,89%
CN=Baltimore CyberTrust Root, OU=CyberTrust, O=Baltimore, C=IE	14	2,93%
CN=thawte Primary Root CA, OU=”(c) 2006 thawte, Inc. - For authorized use only”, OU=Certification Services Division, O=”thawte, Inc.”, C=US	12	11,46%
CN=UTN-USERFirst-Hardware, OU=http://www.usertrust.com, O=The USERTRUST Network, L=Salt Lake City, ST=UT, C=US	11	3,18%
OU=VeriSign Trust Network, OU=”(c) 1998 VeriSign, Inc. - For authorized use only”, OU=Class 3 Public Primary Certification Authority - G2, O=”VeriSign, Inc.”, C=US	8	0,65%
CN=QuoVadis Root Certification Authority, OU=Root Certification Authority, O=QuoVadis Limited, C=BM	5	0,56%
EMAILADDRESS=info@valicert.com, CN=http://www.valicert.com/, OU=ValiCert Class 2 Policy Validation Authority, O=”ValiCert, Inc.”, L=ValiCert Validation Network	5	5,88%
OU=Starfield Class 2 Certification Authority, O=”Starfield Technologies, Inc.”, C=US	5	0,71%
CN=QuoVadis Root CA 2, O=QuoVadis Limited, C=BM	4	1,01%
CN=TC TrustCenter Universal CA I, OU=TC TrustCenter Universal CA, O=TC TrustCenter GmbH, C=DE	2	2,58%
CN=UTN - DATACorp SGC, OU=http://www.usertrust.com, O=The USERTRUST Network, L=Salt Lake City, ST=UT, C=US	2	0,78%
CN=Certum CA, O=Unizeto Sp. z o.o., C=PL	1	1,69%
CN=VeriSign Class 3 Public Primary Certification Authority - G3, OU=”(c) 1999 VeriSign, Inc. - For authorized use only”, OU=VeriSign Trust Network, O=”VeriSign, Inc.”, C=US	1	0,31%
CN=GeoTrust Primary Certification Authority, O=GeoTrust Inc., C=US	1	0,42%
CN=America Online Root Certification Authority 1, O=America Online Inc., C=US	1	0,88%

Table 5. Root CAs found for 26 Browser Histories

latter shows with how many hosts the respective CA was observed in percent of the total hosts averaged over all users for which that CA was observed. Tables 6-9 show the same information for all observed Sub CAs.

Distinguished Name (DN)	Relevance by no. of data sets	Relevance by percentage of hosts
CN=Thawte SSL CA, O="Thawte, Inc.", C=US	26	9,40%
CN=VeriSign Class 3 Extended Validation SSL CA, OU=Terms of use at https://www.verisign.com/rpa (c)06, OU=VeriSign Trust Network, O="VeriSign, Inc.", C=US	25	6,09%
CN=VeriSign Class 3 Secure Server CA - G3, OU=Terms of use at https://www.verisign.com/rpa (c)10, OU=VeriSign Trust Network, O="VeriSign, Inc.", C=US	25	5,34%
CN=Google Internet Authority, O=Google Inc, C=US	24	13,43%
CN=VeriSign Class 3 International Server CA - G3, OU=Terms of use at https://www.verisign.com/rpa (c)10, OU=VeriSign Trust Network, O="VeriSign, Inc.", C=US	24	4,30%
CN=VeriSign Class 3 Extended Validation SSL SGC CA, OU=Terms of use at https://www.verisign.com/rpa (c)06, OU=VeriSign Trust Network, O="VeriSign, Inc.", C=US	23	6,67%
CN=RapidSSL CA, O="GeoTrust, Inc.", C=US	23	3,60%
OU=www.verisign.com/CPS Incorp.by Ref. LIABILITY LTD.(c)97 VeriSign, OU=VeriSign International Server CA - Class 3, OU="VeriSign, Inc.", O=VeriSign Trust Network	22	2,04%
CN=GeoTrust SSL CA, O="GeoTrust, Inc.", C=US	21	4,44%
CN=DigiCert High Assurance CA-3, OU=www.digicert.com, O=DigiCert Inc, C=US	20	3,32%
CN=COMODO High-Assurance Secure Server CA, O=COMODO CA Limited, L=Salford, ST=Greater Manchester, C=GB	20	3,02%
CN=Akamai Subordinate CA 3, O=Akamai Technologies Inc, C=US	20	2,22%
CN=COMODO Certification Authority, O=COMODO CA Limited, L=Salford, ST=Greater Manchester, C=GB	19	2,12%
CN=Thawte DV SSL CA, OU=Domain Validated SSL, O="Thawte, Inc.", C=US	19	1,93%
CN=DFN-Verein PCA Global - G01, OU=DFN-PKI, O=DFN-Verein, C=DE	18	9,34%
SERIALNUMBER=07969287, CN=Go Daddy Secure Certification Authority, OU=http://certificates.godaddy.com/repository, O="GoDaddy.com, Inc.", L=Scottsdale, ST=Arizona, C=US	18	3,18%
CN=GlobalSign Organization Validation CA - G2, O=GlobalSign nv-sa, C=BE	18	1,59%
CN=thawte Extended Validation SSL CA, OU=Terms of use at https://www.thawte.com/cps (c)06, O="thawte, Inc.", C=US	18	1,55%
EMAILADDRESS=tud-ca@hrz.tu-darmstadt.de, CN=TUD CA G01, O=Technische Universitaet Darmstadt, L=Darmstadt, ST=Hessen, C=DE	16	7,80%
CN=TC TrustCenter Class 2 L1 CA XI, OU=TC TrustCenter Class 2 L1 CA, O=TC TrustCenter GmbH, C=DE	16	1,78%
CN=DigiCert High Assurance EV CA-1, OU=www.digicert.com, O=DigiCert Inc, C=US	16	1,49%
CN=COMODO Extended Validation Secure Server CA 2, O=COMODO CA Limited, L=Salford, ST=Greater Manchester, C=GB	16	1,35%
CN=thawte Primary Root CA, OU="(c) 2006 thawte, Inc. - For authorized use only", OU=Certification Services Division, O="thawte, Inc.", C=US	14	12,22%
CN=GeoTrust DV SSL CA, OU=Domain Validated SSL, O=GeoTrust Inc., C=US	14	1,40%
CN=GlobalSign Domain Validation CA - G2, O=GlobalSign nv-sa, C=BE	14	1,30%
CN=GeoTrust Extended Validation SSL CA, OU=See www.geotrust.com/resources/cps (c)06, O=GeoTrust Inc, C=US	13	1,89%
CN=PositiveSSL CA 2, O=COMODO CA Limited, L=Salford, ST=Greater Manchester, C=GB	13	1,65%
CN=Entrust Certification Authority - L1C, OU="(c) 2009 Entrust, Inc.", OU=www.entrust.net/rpa is incorporated by reference, O="Entrust, Inc.", C=US	13	1,46%
CN=GeoTrust Primary Certification Authority, O=GeoTrust Inc., C=US	12	2,01%
CN=USERTrust Legacy Secure Server CA, O=The USERTRUST Network, L=Salt Lake City, ST=UT, C=US	12	1,53%
CN=InCommon Server CA, OU=InCommon, O=Internet2, C=US	12	1,51%
CN=COMODO Extended Validation Secure Server CA, O=COMODO CA Limited, L=Salford, ST=Greater Manchester, C=GB	12	0,93%

Table 6. Sub CAs found for 26 Browser Histories

Distinguished Name (DN)	Relevance by no. of data sets	Relevance by percentage of hosts
CN=Microsoft Internet Authority	11	2,41%
CN=TeleSec ServerPass CA 1, OU=Trust Center Services, O=T-Systems International GmbH, C=DE	11	2,08%
CN=TERENA SSL CA, O=TERENA, C=NL	11	2,05%
CN=Microsoft Secure Server Authority, DC=redmond, DC=corp, DC=microsoft, DC=com	11	1,84%
CN=DPWN Root CA R2 PS, OU=IT Services, O=Deutsche Post World Net, DC=com	10	1,52%
CN=DPWN SSL CA I2 PS, OU=I2 PS, O=Deutsche Post World Net	10	1,52%
CN=PositiveSSL CA, O=Comodo CA Limited, L=Salford, ST=Greater Manchester, C=GB	10	0,96%
CN=UTN-USERFirst-Hardware, OU=http://www.usertrust.com, O=The USERTRUST Network, L=Salt Lake City, ST=UT, C=US	9	2,51%
CN=Network Solutions Certificate Authority, O=Network Solutions L.L.C., C=US	9	0,71%
CN=Thawte SGC CA - G2, O="Thawte, Inc.", C=US	8	1,22%
CN=AlphaSSL CA - G2, O=AlphaSSL	8	0,81%
CN=EssentialSSL CA, O=COMODO CA Limited, L=Salford, ST=Greater Manchester, C=GB	8	0,79%
EMAILADDRESS=ca@zivit.de, CN=ZIVIT CA - G01, OU=Betrieb, O=Zentrum fuer Informationsverarbeitung und Informationstechnik, C=DE	8	0,67%
CN=VeriSign Class 3 Secure Server CA - G2, OU=Terms of use at https://www.verisign.com/rpa (c)09, OU=VeriSign Trust Network, O="VeriSign, Inc.", C=US	8	0,65%
CN=Cybertrust Public SureServer SV CA, O=Cybertrust Inc	8	0,64%
CN=COMODO SSL CA, O=COMODO CA Limited, L=Salford, ST=Greater Manchester, C=GB	7	1,00%
EMAILADDRESS=pki@h-da.de, CN=Hochschule Darmstadt, O=Hochschule Darmstadt, L=Darmstadt, C=DE	7	0,70%
CN=GlobalSign Domain Validation CA, O=GlobalSign nv-sa, OU=Domain Validation CA, C=BE	7	0,61%
CN=WebSpace-Forum Server CA, O="WebSpace-Forum, Thomas Wendt", C=DE	7	0,58%
CN=DigiCert High Assurance EV Root CA, OU=www.digicert.com, O=DigiCert Inc, C=US	6	3,42%
CN=COMODO SSL CA 2, O=COMODO CA Limited, L=Salford, ST=Greater Manchester, C=GB	6	1,06%
CN=QuoVadis Global SSL ICA, OU=www.quovadisglobal.com, O=QuoVadis Limited, C=BM	6	0,83%
CN=Cybertrust Global Root, O="Cybertrust, Inc"	6	0,65%
CN=TC TrustCenter Class 4 Extended Validation CA II, OU=TC TrustCenter Class 4 L1 CA, O=TC TrustCenter GmbH, C=DE	6	0,62%
CN=MSIT Machine Auth CA 2, DC=redmond, DC=corp, DC=microsoft, DC=com	5	1,27%
CN=UTN - DATACorp SGC, OU=http://www.usertrust.com, O=The USERTRUST Network, L=Salt Lake City, ST=UT, C=US	5	1,19%
SERIALNUMBER=10688435, CN=Starfield Secure Certification Authority, OU=http://certificates.starfieldtech.com/repository, O="Starfield Technologies, Inc.", L=Scottsdale, ST=Arizona, C=US	5	0,59%
CN=GlobalSign, O=GlobalSign, OU=GlobalSign Root CA - R2	5	0,48%
CN=DFN-Verein-GS-CA - G02, OU=Geschaeftsstelle, O=DFN-Verein, C=DE	5	0,47%
CN=VeriSign Class 3 Public Primary Certification Authority - G5, OU="(c) 2006 VeriSign, Inc. - For authorized use only", OU=VeriSign Trust Network, O="VeriSign, Inc.", C=US	4	26,86%
CN=Entrust Root Certification Authority, OU="(c) 2006 Entrust, Inc.", OU=www.entrust.net/CPS is incorporated by reference, O="Entrust, Inc.", C=US	4	0,61%
CN=Entrust Certification Authority - L1E, OU="(c) 2009 Entrust, Inc.", OU=www.entrust.net/rpa is incorporated by reference, O="Entrust, Inc.", C=US	4	0,61%
CN=Gandi Standard SSL CA, O=GANDI SAS, C=FR	4	0,30%
CN=Vodafone (Corporate Domain 2009), O=Vodafone Group, C=UK	3	1,68%
CN=Vodafone (Corporate Services 2009), O=Vodafone Group, C=UK	3	1,68%
CN=GlobalSign Organization Validation CA, O=GlobalSign, OU=Organization Validation CA	3	0,56%
CN=GlobalSign Extended Validation CA - G2, O=GlobalSign nv-sa, C=BE	3	0,50%

Table 7. Sub CAs found for 26 Browser Histories (cont.)

Distinguished Name (DN)	Relevance by no. of data sets	Relevance by percentage of hosts
EMAILADDRESS=ca-btu@tu-cottbus.de, CN=BTU-CA (G01 2008), OU=Rechenzentrum, O=Brandenburgische Technische Universitaet Cottbus, L=Cottbus, ST=Brandenburg, C=DE	3	0,48%
EMAILADDRESS=ca@pki.tu-dortmund.de, CN=TU Dortmund CA - G01, OU=ITMC, O=Technische Universitaet Dortmund, C=DE	3	0,41%
CN=COMODO High Assurance Secure Server CA, O=COMODO CA Limited, L=Salford, ST=Greater Manchester, C=GB	3	0,39%
EMAILADDRESS=pki@hu-berlin.de, CN=HU-CA, O=Humboldt-Universitaet zu Berlin, C=DE	3	0,36%
CN=Zertifizierungsstelle der TUM, O=Technische Universitaet Muenchen, C=DE	3	0,29%
CN=Trusted Secure Certificate Authority, O=Trusted Secure Certificate Authority, C=US	2	1,70%
CN=TC TrustCenter Class 3 L1 CA IX, OU=TC TrustCenter Class 3 L1 CA, O=TC TrustCenter GmbH, C=DE	2	1,29%
EMAILADDRESS=pki-admin@uni-potsdam.de, CN=Universitaet Potsdam CA - G01, O=Universitaet Potsdam, L=Potsdam, C=DE	2	1,07%
CN=EuropeanSSL Server CA, O=EUNETIC GmbH, C=DE	2	1,02%
CN=Register.com CA SSL Services (OV), O=Register.com, C=US	2	0,94%
EMAILADDRESS=rubca@ruhr-uni-bochum.de, CN=Ruhr-Universitaet Bochum CA, O=Ruhr-Universitaet Bochum, L=Bochum, ST=Nordrhein-Westfalen, C=DE	2	0,80%
EMAILADDRESS=mpg-ca@mpg.de, CN=MPG CA, O=Max-Planck-Gesellschaft, C=DE	2	0,63%
CN=DFN-Verein CA Services, OU=DFN-PKI, O=DFN-Verein, C=DE	2	0,62%
EMAILADDRESS=gwdg-ca@gwdg.de, CN=Universitaet-Goettingen CA, O=Georg-August-Universitaet Goettingen, L=Goettingen, ST=Niedersachsen, C=DE	2	0,59%
EMAILADDRESS=ca@d-nb.de, CN=DNB-CA, O=Deutsche Nationalbibliothek, L=Frankfurt am Main, C=DE	2	0,56%
CN=GlobalSign Extended Validation CA, O=GlobalSign, OU=Extended Validation CA	2	0,47%
CN=QuoVadis Root CA 2, O=QuoVadis Limited, C=BM	2	0,46%
EMAILADDRESS=ca@kit.edu, CN=KIT-CA, OU=Steinbuch Centre for Computing, O=Karlsruhe Institute of Technology, L=Karlsruhe, ST=Baden-Wuerttemberg, C=DE	2	0,44%
EMAILADDRESS=pki@uni-regensburg.de, CN=Uni Regensburg CA - G01, O=Universitaet Regensburg, L=Regensburg, ST=Bayern, C=DE	2	0,33%
EMAILADDRESS=ca@rrze.uni-erlangen.de, CN=FAU-CA, OU=RRZE, O=Universitaet Erlangen-Nuernberg, L=Erlangen, ST=Bayern, C=DE	2	0,33%
EMAILADDRESS=rum-ca@rz.uni-mannheim.de, CN=RUM-CA-G Zertifizierungsinstanz, OU=Rechenzentrum, O=Universitaet Mannheim, L=Mannheim, ST=Baden-Wuerttemberg, C=DE	2	0,32%
CN=GlobalSign Primary Secure Server CA, OU=Primary Secure Server CA, O=GlobalSign nv-sa, C=BE	2	0,28%
CN=GlobalSign ServerSign CA, OU=ServerSign CA, O=GlobalSign nv-sa, C=BE	2	0,28%
CN=Network Solutions DV Server CA, O=Network Solutions L.L.C., C=US	2	0,21%
EMAILADDRESS=pki@tu-dresden.de, CN=TU Dresden CA - G02, OU=ZIH, O=Technische Universitaet Dresden, C=DE	2	0,21%
CN=Experian Root CA, CN=AIA, CN=Public Key Services, CN=Services, CN=Configuration, DC=experian, DC=local	2	0,19%
CN=Experian Issuing CA 1, CN=AIA, CN=Public Key Services, CN=Services, CN=Configuration, DC=experian, DC=local	2	0,19%
CN=Fraunhofer Root CA 2007, OU=Fraunhofer Corporate PKI, O=Fraunhofer, C=DE	2	0,18%
EMAILADDRESS=pki-ca@bundestag.de, CN=Deutscher Bundestag CA - G01, OU=Deutscher Bundestag, O=Deutscher Bundestag, C=DE	2	0,18%
CN=GeoTrust Global CA, O=GeoTrust Inc., C=US	1	25,00%
CN=VeriSign Class 3 Secure Server CA, OU=Terms of use at https://www.verisign.com/rpa (c)05, OU=VeriSign Trust Network, O="VeriSign, Inc.", C=US	1	3,13%
EMAILADDRESS=ca@uni-wuerzburg.de, CN=UNIWUE-CA - G01, O=Universitaet Wuerzburg, C=DE	1	2,94%
CN=Certum Trusted Network CA, OU=Certum Certification Authority, O=Unizeto Technologies S.A., C=PL	1	1,69%

Table 8. Sub CAs found for 26 Browser Histories (cont.)

Distinguished Name (DN)	Relevance by no. of data sets	Relevance by percentage of hosts
CN=Certum Extended Validation CA, OU=Certum Certification Authority, O=Unizeto Technologies S.A., C=PL	1	1,69%
CN=SGTRUST CERTIFICATION AUTHORITY, O=SGssl, C=KR	1	1,45%
EMAILADDRESS=ca@rz.uni-saarland.de, CN=CA Universitaet des Saarlandes, O=Universitaet des Saarlandes, L=Saarbruecken, ST=Saarland, C=DE	1	1,26%
EMAILADDRESS=caadmin@uni-bonn.de, CN=Universitaet Bonn CA, OU=Hochschulrechenzentrum, O=Universitaet Bonn, L=Bonn, ST=Nordrhein-Westfalen, C=DE	1	1,11%
CN=adidas Global Intermediate CA 01, O=adidas AG, C=DE	1	1,03%
CN=adidas EMEA Issuing CA 01, O=adidas AG, C=DE	1	1,03%
CN=Universitaet Bremen CA, O=Universitaet Bremen, L=Bremen, ST=Bremen, C=DE	1	1,03%
EMAILADDRESS=jgu-ca@uni-mainz.de, CN=JGU CA - G01, O=Johannes Gutenberg-Universitaet Mainz, L=Mainz, ST=Rheinland-Pfalz, C=DE	1	0,92%
CN=AOL Member CA, O=America Online Inc., L=Dulles, ST=Virginia, C=US	1	0,88%
EMAILADDRESS=pki@hs-mannheim.de, CN=HS Mannheim CA, O=Hochschule Mannheim, C=DE	1	0,78%
EMAILADDRESS=pki@smi.sachsen.de, CN=Sachsen Global CA, OU=Saechsisches Staatsministerium des Innern, O=Freistaat Sachsen, L=Dresden, ST=Sachsen, C=DE	1	0,46%
EMAILADDRESS=caadmin@fernuni-hagen.de, CN=FernUniversitaet in Hagen Global CA, OU=Zentrum fuer Medien und IT, O=FernUniversitaet in Hagen, L=Hagen, ST=Nordrhein-Westfalen, C=DE	1	0,46%
EMAILADDRESS=zertifizierungsstelle@nw.neclab.eu, CN=NECLAB-CA, OU=NEC Laboratories Europe, O=NEC Europe Ltd., C=DE	1	0,46%
EMAILADDRESS=ca@uni-ulm.de, CN=Global-Uni-Ulm-CA, O=Universitaet Ulm, C=DE	1	0,46%
EMAILADDRESS=pki@uni-marburg.de, CN=Uni Marburg CA - G02, OU=Hochschulrechenzentrum, O=Universitaet Marburg, C=DE	1	0,46%
EMAILADDRESS=ca@rwth-aachen.de, CN=RWTH Aachen CA, O=RWTH Aachen, C=DE	1	0,46%
EMAILADDRESS=pki@uni-kiel.de, CN=Uni Kiel CA - G02, OU=Rechenzentrum, O=Universitaet Kiel, L=Kiel, ST=Schleswig-Holstein, C=DE	1	0,46%
CN=Intel External Basic Policy CA, O=Intel Corporation, C=US	1	0,43%
EMAILADDRESS=hrz-ra@uni-bielefeld.de, CN=CA der Universitaet Bielefeld - G02, O=Universitaet Bielefeld, C=DE	1	0,42%
EMAILADDRESS=camaster@uni-koeln.de, CN=UniKoeln CA, O=Universitaet zu Koeln, L=Koeln, C=DE	1	0,42%
EMAILADDRESS=ca@fh-muenster.de, CN=FH Muenster CA - G01, OU=Datenverarbeitungszentrale, O=Fachhochschule Muenster, L=Muenster, ST=Nordrhein-Westfalen, C=DE	1	0,36%
CN=Thawte SGC CA, O=Thawte Consulting (Pty) Ltd., C=ZA	1	0,36%
CN=Oracle SSL CA, OU=Class 3 MPKI Secure Server CA, OU=VeriSign Trust Network, O=Oracle Corporation, C=US	1	0,31%
CN=Network Solutions EV Server CA, O=Network Solutions L.L.C., C=US	1	0,22%
CN=Cybertrust SureServer Standard Validation CA, O=Cybertrust Inc	1	0,22%
CN=Intel External Basic Issuing CA 3A, O=Intel Corporation, C=US	1	0,22%
CN=Intel External Basic Issuing CA 3B, O=Intel Corporation, C=US	1	0,22%
C=BE, O=GlobalSign nv-sa, OU=RootSign Partners CA, CN=GlobalSign RootSign Partners CA	1	0,21%
CN=Deutsche Telekom CA 5, OU=Trust Center Deutsche Telekom, O=T-Systems Enterprise Services GmbH, C=DE	1	0,21%
CN=Fraunhofer Service CA 2007, OU=Fraunhofer Corporate PKI, O=Fraunhofer, C=DE	1	0,21%
EMAILADDRESS=pki@dagstuhl.de, CN=Schloss Dagstuhl - LZI GmbH CA - G01, OU=IT-Abteilung, O=Schloss Dagstuhl - LZI GmbH, L=Wadern, ST=Saarland, C=DE	1	0,21%
EMAILADDRESS=pki@unibw.de, CN=UniBwM CA-G01, O=Universitaet der Bundeswehr Muenchen, L=Muenchen, ST=Bayern, C=DE	1	0,21%
CN=Cybertrust SureServer EV CA, O=Cybertrust Inc	1	0,21%
EMAILADDRESS=fhw-ca@itc.fh-wiesbaden.de, CN=FHW-CA, OU=IT-Center, O=Fachhochschule Wiesbaden, L=Wiesbaden, ST=Hessen, C=DE	1	0,20%
EMAILADDRESS=pki@fraunhofer.de, CN=Fraunhofer Service CA - G01, OU=Fraunhofer Corporate PKI, O=Fraunhofer, L=Muenchen, ST=Bayern, C=DE	1	0,20%

Table 9. Sub CAs found for 26 Browser Histories (cont.)

Distinguished Name (DN)	Relevance by no. of data sets	Relevance by percentage of hosts
EMAILADDRESS=pki@bsb-muenchen.de, CN=BSB-CA, OU=Bayerische Staatsbibliothek, O=Bayerische Staatsbibliothek, L=Muenchen, ST=Bayern, C=DE	1	0,20%
EMAILADDRESS=ca@uni-frankfurt.de, CN=UNI-FFM CA, O=Johann Wolfgang Goethe-Universitaet, L=Frankfurt am Main, ST=Hessen, C=DE	1	0,20%
EMAILADDRESS=pki@tu-bs.de, CN=Technische Universitaet Braunschweig CA, O=Technische Universitaet Braunschweig, L=Braunschweig, ST=Niedersachsen, C=DE	1	0,16%
CN=Fraunhofer User CA 2007, OU=Fraunhofer Corporate PKI, O=Fraunhofer, C=DE	1	0,16%
CN=Dell Inc. Enterprise CA, O=Dell Inc.	1	0,16%
CN=Dell Inc. Enterprise Issuing CA1, O=Dell Inc.	1	0,16%
CN=SecureTrust CA, O=SecureTrust Corporation, C=US	1	0,16%
CN=AusCERT Server CA, OU=Certificate Services, O=AusCERT, C=AU	1	0,16%

Table 10. Sub CAs found for 26 Browser Histories (cont.)