Transparency in the New gTLD Era: Evaluating the DNS Centralized Zone Data Service

Jeman Park, *Student Member, IEEE*, Jinchun Choi, Daehun Nyang, and Aziz Mohaisen, *Senior Member, IEEE*

***Abstract*— The centralized zone data service (CZDS) was introduced by the Internet Corporation for Assigned Names and Numbers (ICANN) to facilitate sharing and access to zone data of the new generic Top-Level Domains (gTLDs). CZDS aims to improve the security and transparency of the naming system of the Internet. In this paper, we investigate CZDS's transparency by measurement and evaluation. By requesting access to zone data of all gTLDs listed in the CZDS portal, we analyze various aspects of CZDS, including access status, responsiveness and provided reasons for granting access or denial. Among other findings, we find that while a large percent of the gTLD admins respond within a reasonable time, more than 10% of them have a long request-to-decision waiting time, and sometimes requests go unanswered even after six months of a request. Furthermore, we find that denial cases were for unjustified reasons, where administrators who denied the requests have asked for information that was already provided in the request form. We discuss implications, and how to enforce better outcomes of CZDS using insight from our measurement and evaluation.**

***Index Terms*—New gTLDs, data sharing, domain name system, service transparency.**

I. Introduction

THE DOMAIN Name System (DNS) is the de facto standard protocol, created over 33 years ago, and used today for naming and resolving names of resources and services on the Internet and in private networks [2]. The DNS is responsible for mapping and resolving domain names, names that are easy to recognize and remember by humans, into Internet Protocol (IP) addresses of resources [3]. The DNS is a hierarchical naming system, and at the top of the hierarchy is a Top-level Domain (TLD), where TLDs are stored in the DNS root. Historically, TLDs included a few generic TLDs (gTLDs), such as .com, .net, .edu, and .org, among others, which served Internet users and businesses well for more than 30 years [3].

Manuscript received December 18, 2018; revised July 5, 2019 and August 11, 2019; accepted August 14, 2019. Date of publication September 16, 2019; date of current version December 10, 2019. This work was supported by National Research Foundation of Korea grant 2016K1A1A2912757 (Global Research Lab Project). The seed of this work appeared in the 4th IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb 2016) [1]. The associate editor coordinating the review of this article and approving it for publication was F. De Turck. *(Corresponding author: Aziz Mohaisen.)*

J. Park and A. Mohaisen are with the Department of Computer Science, University of Central Florida, Orlando, FL 32765 USA (e-mail: mohaisen@ucf.edu).

J. Choi is with the Department of Computer Science, University of Central Florida, Orlando, FL 32765 USA, and also with the Department of Computer Science and Information Technology, Inha University, Incheon 22212, South Korea.

D. Nyang is with Inha University, Incheon 22212, South Korea.

Digital Object Identifier 10.1109/TNSM.2019.2941572

Type	%	#
Generic	53%	1,023
Brand	34%	656
IDN	6%	116
Community	4%	84
Geographic	3%	66
$\sum$	100%	1,930

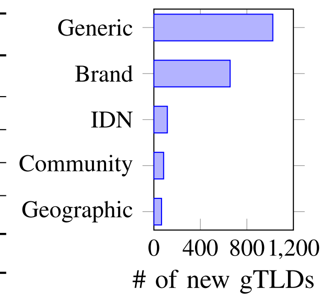

Fig. 1. Statistics of applied for and approved new generic top-level domains (new gTLDs) by type. The application for new gTLDs has been unavailable since 2012 when the first round of application ended. The second round will be determined and announced by the ICANN community [13].

Recently, the Internet Corporation for Assigned Names and Numbers (ICANN), a nonprofit organization that oversees procedures for the coordination and maintenance of the global DNS functions and security, has allowed for the registration of new generic TLDs (gTLDs) [4]. The new gTLDs are intended for improving the competitiveness of the domain name market, and for giving customers more choices for the names they can use, depending on their contexts, interests, among others (language, usage field, location, brand, etc.; more details are shown in Figure 1) [5]. As a result, the first four new gTLDs were delegated on July 15, 2013, and since then the number of registered and delegated new gTLDs has largely increased: to 605 on May 3, 2015 and to 1,228 as of June 2018, as shown and broken down by category in Figure 1 [6].

With the emergence of various security issues [7], [8] due to the introduced new gTLDs, such as domain name collision [9], there has been a need for timely access to lists of domain names registered under those gTLDs [4]. Because of the large number of the introduced new gTLDs, the traditional method of manually requesting access to each gTLD and its associated data is a cumbersome process, thus there has been a need for automating access to data of new gTLDs, including their zone files that list domain names registered in those new gTLDs.

ICANN introduced the Centralized Zone Data Service (CZDS) [10], a service for requesting and distributing zone files of the new gTLDs participating in the CZDS. The objective of CZDS is to facilitate transparency and improve the security of new gTLDs [11]. The zone file requested via CZDS of a given gTLD is a text file that enumerates mappings between the strings (often Second-level Domain names, or simply SLDs) and Internet Protocol (IP) addresses, among other resources with their corresponding types (e.g., name servers, mail exchange servers, etc.) [12] (see Section II-B).

The CZDS provides a centralized tool for requesting access to zone files of various new gTLDs introduced in the program, as well as other traditional gTLDs. CZDS allows interested parties to register for the service for authenticated access by providing various pieces of information about themselves, including personal and network information (e.g., IP address to be whitelisted for later access). CZDS also allows Application Programming Interface (or simply API) level access to the service by providing access tokens. Using the authenticated CZDS, a user may request access to zone data of gTLDs of his choosing. The "request access" form requires requesting parties to specify the reason for requesting the zone data, where "one or two sentences will do." In conclusion, CZDS is a system designed to conveniently manage access rights to gTLDs. Given that governance of the Internet is a critical issue in modern society, it is important to look at whether CZDS, which was created for the governance of gTLDs, is operating in accordance with its original purpose. This represents a key motivation of this study, and we try to understand the extent of the current management by comprehensive measurements.

Contributions: The contributions of this work are two. First, we provide a comprehensive analysis of CZDS, a service launched to facilitate access to data associated with new gTLDs. Second, we provide recommendations to improve CZDS's operation.

More precisely, and with the Internet's transparency and security being two central goals at the heart of CZDS's operation and intended design, the goal of this work is twofold: to investigate and evaluate the CZDS's transparency by measurement, and to provide recommendations for improving CZDS. For the first goal, our contribution includes an in-depth analysis to understand the following aspects of CZDS:

Cooperation: By analyzing this feature, we would like to understand how cooperative new gTLDs operators are with users interested in zone data and applying through CZDS.

Time: In analyzing CZDS temporally, we are interested in understanding how timely is the response of operators for requests to access their zone data, and whether a legitimate request would result in approval or denial.

Denial Justification: When a new gTLD operator denies a request to access his zone data, we are interested in understanding the stated reasons for denials, and whether they are justified or not in light of the conditions and agreements provided by ICANN for CZDS operation.

Access Period: In analyzing the access period, we are interested in understanding the openness of new gTLDs, measured by the period of access granted by the various zone owners.

Correlation and Analysis: In performing such analysis and correlations, we are interested in understanding if there is any implicit or explicit relationship between acceptance, denial, or the pending status of an access to zone data and other features of the new gTLD of the zone, including spatial features of the owner (country, city, etc), size of the owner organization, size of the gTLD, other reputation information, among others.

Mindful of the nature of CZDS through analysis, including the cooperation of participants, timeliness of responses, justifications, and access periods, among others, we then propose various recommendations to improve CZDS. To the best of our knowledge, the research reported in this paper is the first attempt to evaluate CZDS with the goal of understanding its operation and improving the transparency and quality of DNS operation for the emerging new gTLDs.

A. Summary of the Main Findings

The core contribution of this paper is a measurement-oriented study of CZDS, leading to various interesting behavioral and characteristic features of new gTLD owners. We cover those findings at length in the subsequent sections and summarize them below.

- We found that while the majority of the new gTLD owners granted access to their zones, the time it took each of them to respond varies significantly, and may take more than four months for a permission or denial response to be given. In an extreme case, the requests to access some zones were still pending even 11 months after initiation, indicating that the participation in CZDS by the given gTLD is not well managed by the owners.
- We found that when certain gTLD operators deny access to their zone, the reason provided for the denial is often unjustified given that the requested information is already provided on the zone request form submitted by CZDS (e.g., invalid request or no IP address provided).
- For denied access cases, we found that most of the gTLDs are registered and operated by the same few owners, although operated under different front organizations.
- We have found various strong and interesting correlations. By mapping the various gTLDs to their country of origin of the owner (i.e., where the new gTLD's operating company is registered), we found *an almost binary* decision at the country level *in many cases*, where operators in certain countries either always deny or always granted access for all gTLDs in that country.
- We found that some new gTLDs are both timely and categorically granting access to zone files despite being registered in countries that are traditionally known for their record of Internet censorship.
- We review possible outsourcing of gTLDs operation, explaining some of the discrepancy between the outcomes of CZDS records and the aforementioned characteristics of groups of new gTLDs registered in certain countries.

With the various findings in mind, we conclude with recommendations on improving the quality of CZDS.

Organization: The organization of this paper is as follows. In Section III, we introduce our dataset and data acquisition method. In Section IV, we introduce our measurement criteria and measurement results. In Section V, we address various issues related to our study. In Section VI, we discuss the related work. In Section VII, we provide a summary and concluding remarks, including recommendations.

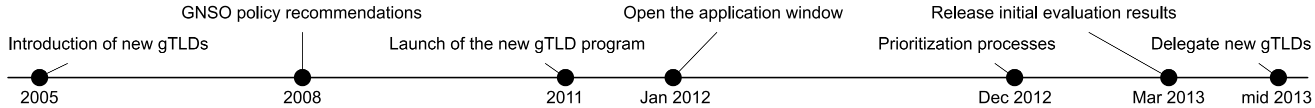

Fig. 2. Historical background and timeline of the events associated with the introduction of the new gTLDs.

II. PRELIMINARIES

In this section, we introduce preliminaries of this work. In particular, we review the historical background and activities in the IETF leading to the new gTLD program in Section II-A. We provide an overview of CZDS in Section II-B.

A. An Overview of the New gTLDs

For the completeness of the treatment of the subject of this work, in the following we provide an overview of the new gTLD focusing on historical progression of its introduction, and events in the standardization community (namely, in the IETF) leading to their introduction.

1) Historical Background: In 2005, the Generic Names Supporting Organization (GNSO), an organization within ICANN, started the development of policies to introduce new gTLDs, as shown in Figure 2. The two-year policy process development included consultations with the domain name stakeholders globally, including governments, civil society, business and intellectual property stakeholders, and technologists. As a result, ICANN adopted 19 specific GNSO policy recommendations for implementing the new gTLDs, in 2008, while working on addressing some of the concerns raised by stakeholders, including intellectual property and consumer protection, as well as DNS stability, resulting in the "Applicant Guidebook" of new gTLDs which was finally approved in 2011, thus authorizing the launch of the new gTLD program. The stated goals of the program are "enhancing competition and consumer choice, and enabling the benefits of innovation via the introduction of new gTLDs, including both new ASCII and internationalized domain name (IDN) top-level domains".

• *New gTLD Applications:* Applications were accepted starting on 12 January 2012, resulting in 1,930 applications as of 17 December 2012, leading to the introduction of prioritization processes to determine the order in which applications would be processed. On March 22, 2013, ICANN released the initial evaluation results of the applications to new gTLDs that were applied for. The published evaluation was the result of experts reviews. Applications that passed the initial review and received no public objections then proceeded to the contractual phase with ICANN by the middle of 2013, upon which those new gTLDs were delegated to the root zone.

2) DNS Developments: Closely related to the new gTLDs are the Internationalized Domain Names (IDNs), where there have been several works on "multilingual" or "internationalized" TLDs to languages not written in a Roman-based script (e.g., Chinese, Arabic, Japanese, Korean, Hebrew, etc). In 2005, and around the same time that GNSO started policy discussions for the new gTLDs (see above), Klensin [14] reviewed the motivations for IDN, made several suggestions to provide a needed functionality, and constraints that the DNS imposes on the use of IDNs, all in RFC 4185. One of Klensin's suggestions was an alternative local translation, which avoids significant DNS protocol changes and deployment delays. In 2010, Klensin [15] revised his earlier RFC draft and introduced RFC 5894, the Internationalized Domain Names for Applications (IDNA) protocol, which addresses several developments, including new Unicode versions. Following that, in 2010, Resnick and Hoffman [16] further specified in RFC 5895 the steps taken between receiving user input and passing permitted code points to the new IDNA protocol. In 2010, Sharikov *et al.* [17] provided registries and registrars guidelines on registering IDNs based on several languages, including Bosnian, Bulgarian, Macedonian, Russian, Serbian, and Ukrainian languages, in RFC 5992. He also described conversion methods from Greek to Latin scripts.

```
030.zone. 86400   in   ns   ns3.myhostadmin.net.
030.zone. 86400   in   ns   ns4.myhostadmin.net.
031.zone. 86400   in   ns   ns3.myhostadmin.net.
031.zone. 86400   in   ns   ns4.myhostadmin.net.
032.zone. 86400   in   ns   ns5.myhostadmin.net.
032.zone. 86400   in   ns   ns6.myhostadmin.net.
033.zone. 86400   in   ns   ns5.myhostadmin.net.
033.zone. 86400   in   ns   ns6.myhostadmin.net.
034.zone. 86400   in   ns   ns1.myhostadmin.net.
034.zone. 86400   in   ns   ns2.myhostadmin.net.
```

Fig. 3. A snippet of a zone file from the "zone" new gTLD.

Several other standards have been published to address various issues in IDNs, including the use of multiple encoding schemes by Thaler *et al.* [18] (RFC 6055, 2011), practices of the used provision of IDN variants by Levine and Hoffman [4] (RFC 6927, 2013), special use domains, including policy considerations associated with such domains by Cheshire and Krochmal [19] (RFC 6761, 2013), the special use TLD of .onion by Appelbaum and Muffett [20] (RFC 7686, 2015), and DNS blocking and associated techniques by Barnes *et al.* [21] (RFC 7754, 2016). In introducing the new gTLD program in 2011, and delegating various gTLDs in the root since the middle of 2013, ICANN also authorized various IDN-based new gTLDs (a total of 116 IDN-based TLDs as of 13 June 2012, as shown in Figure 1).

B. An Overview of CZDS

The zone data files of a given TLD contain the information needed to resolve domain names to IP addresses, domain names, their associated name server names and possibly IP addresses for those name servers. More precisely, as shown in Figure 3 for the "zone" new gTLD example, a typical zone file consists of tab-separated records of names of resources, time to live (TTL) in seconds, resource class (can be either IN, which stands for Internet resource, or CH, which stands for chaos), record type (NS stands for name servers in this example), and the record data (name servers in this case). Traditionally, an interested party in zone data of a given TLD would file a request to the owner or operator of such TLD, who would review the request and grant (or deny) such party access to the files upon verifying the requester's identity and legitimacy of his access reasons. Zone files are oftentimes are used for resolving conflict, or collision in the domain name system, performing analysis for security intelligence on bad use of domains, intellectual property infringement, among others.

With the large number of new gTLDs, such a manual request of zone file data from multiple owners and operators is impractical. Thus, the CZDS is a service to simplify the process of requesting zone files and entering into standardized agreements with registries who own the new gTLDs to provide TLD zone data files. Registries update their zone data files for their TLDs, and the CZDS enables interested parties to request access to the zone files provided by participating TLDs. This service is the method for scaling zone data transfer as new gTLDs are added to the Internet.

• *The Procedure of Accessing CZDS:* To use CZDS, a user has to create an account at https://czds.icann.org by providing credentials for login (username and password), actual name, organization, organization address, actual contact information; fax and telephone numbers (used for verification), and customer IPs (IP addresses from which the user is to access the service, which will be whitelisted, upon the creation of a user account). Upon the approval of the account and its creation, the user will be able to simply login to CZDS using the username and password provided at the time of creating the account.

III. Dataset and Data Acquisition

In this section, we described our dataset and how we obtained it. We start by requesting access the new gTLDs' zone data using CZDS as highlighted in Section III-A. We augmented our dataset to have a detailed understanding of the owners of new gTLDs and to study any relationships between acceptance, denial, or the pending status, including spatial features of the owners, and the number of gTLDs each owner has, as shown in Section III-B. We highlighted the features of the augmented CZDS data by providing details of the top gTLDs owners in Section III-C. We provided high-level statistics of the zone files in Section III-D.

A. Data Acquisition

We proceeded to ask for access to all gTLDs listed in the CZDS portal. In order to fill the forms, we registered to the CZDS system using factual and accurate information about the participant (last author). For an accurate and representative collection of the data, we included our research mission as wanting "to analyze the use, utilization, and security of new gTLDs" for the reason for requesting the zone data.

Once the request is approved by the operator, the user will be able to download zone data for the gTLD. As default, CZDS allows the user to download files using HTTPS. In addition, CZDS allows access to those files with SFTP, if needed by the operators. However, while conducting the measurement, we did not find any new gTLD which requires SFTP credentials.

TABLE I
THE SUMMARY OF THE STATUS AFTER SIX (NOV 2015)/THREE MONTHS (JUN 2018) FROM REQUESTING ACCESS TO THE LISTED NEW gTLDS IN CZDS. NOTICE THAT T CORRESPONDS TO THE TOTAL NUMBER OF DELEGATED NEW gTLDS, A CORRESPONDS TO THE NUMBER OF APPROVED REQUESTS, P MEANS THE NUMBER OF PENDING, AND D MEANS THE NUMBER OF DENIED REQUESTS

Date	T	A	$\%_A$	P	$\%_P$	D	$\%_D$
Nov. 2015	787	708	90.0	9	1.1	70	8.9
Jun. 2018	1,228	1,164	94.8	58	4.7	6	0.5

A useful tool that implements a client to the CZDS REST API to automate the access to CZDS is in [22]. This library helps users to easily trace updates in zone file and download multiple zone files in parallel.

Dataset and Initial Characterization: Table I shows the summary of the collected data through CZDS. We performed data collection via CZDS twice, by requesting access to all listed new gTLDs on November 11, 2015 and on June 22, 2018. In the first time, 787 new gTLDs out of the 1,930 were listed as participants in CZDS. In the second time, that grew to and 1,229 new gTLDs. In our first data collection attempt, in November 2015, 708 out of 787 requests were approved, 70 were rejected, and no decision was made on the final 9 new gTLDs' access requests after six months. That outcome of approved, rejected, and no-decision translates to 89.96%, 8.89%, and 1.15%, respectively. On the other hands, the second data collection resulted in 1,165 approvals, 6 denials, and 58 no-response, which translate to 94.8%, 0.5%, and 4.7%, respectively. As an extension of the second data collection, we checked the changes in pending requests on December 4, about five months and two weeks after the requests. Among the 58 pending requests, five requests were approved, while one request was denied. In total, we have 1,170 approvals, 52 pending, and 7 denials as of December 4, 2018. The further analysis below was based on the data collected in September 2018 (in Table I), but we also included a brief description about the changes (6 approvals and 1 denial) at the end of the paragraphs.

From the summary of the results above, and by comparing the new gTLD ecosystem over less than 3 years, we can see that the number of delegated new gTLDs increased by 442 which is about 56% of growth rate. In the meanwhile, the ratio of the approval to the total increased from about 90% to about 94.8%, while the ratio of the denial dropped from 8.9% to 0.5%. On the other hand, the no-response category has increased from 1.1% to 4.7%, which is substantial.

B. Dataset Augmentation

In order to have a better understanding of the request results, we collected subsidiary information related to new gTLDs. ICANN provides a search engine for new gTLDs at https://gtldresult.icann.org. Application status is consistently updated to reflect the new gTLD delegation processes. A user can look into the current application status of new gTLDs by applying filters through this search interface, using type (geographic, community, IDN), application status (delegated,

TABLE II
A SNIPPET OF OUR DATASET WITH VARIOUS FEATURES, INCLUDING DECISION, NOTES, TIMESTAMPS, ETC.

TLD	Status	Zone_data_files	Date & Time	Response
bharti	Pending	Not approved yet.	11 Nov 2015, 18:07:45 UTC	Request submitted
club	Approved	zone_file/club.gz	12 Nov 2015, 14:46:53 UTC	Access granted.
cooking	Denied	Denied request.	13 Nov 2015, 18:02:12 UTC	Invalid request.
corsica	Pending	Not approved yet.	11 Nov 2015, 18:07:58 UTC	Request submitted
moscow	Denied	Denied request	12 Nov 2015, 08:31:05 UTC	Your request is invalid.
win	Approved	zone_file/win.gz	13 Nov 2015, 07:21:53 UTC	Approved
xn–80adxhks (.москвао))	Denied	Denied request.	12 Nov 2015, 08:31:05 UTC	Your request is invalid.
xyz	Approved	zone_file/xyz.gz	16 Nov 2015, 10:40:32 UTC	Approved.

TABLE III
TOP 20 gTLDS WITH APPLICANT AND THE NUMBER OF SLDS (#) IN MAY 2016

gTLD	Applicant	#
xyz	xyz.com LLC	2,673,614
win	First Registry Ltd.	885,996
club	.club domains, LLC	737,800
link	Uniregistry, Corp.	357,908
site	DotSite Inc.	345,113
science	dot Science Ltd.	332,616
red	Afilias plc	304,324
party	Blue Sky Registry Ltd.	234,918
online	DotOnline Inc.	231,363
loan	dot Loan Ltd.	224,366
click	Uniregistry, Corp.	207,530
date	dot Date Ltd	175,172
website	DotWebsite Inc.	152,756
space	DotSpace Inc.	137,859
kim	Afilias plc	116,598
xin	Elegant Leader Ltd.	106,777
xxx	ICM Registry LLC	100,044
tech	Dot Tech LLC	93,360
lol	Uniregistry, Corp.	92,049
nyc	The City of New York	75,171
Total	-	7,585,334

TABLE IV
THE TOP 20 gTLDS WITH THEIR APPLICANTS AND THE NUMBER OF SLDS (#) AS OF SEPTEMBER 2018

gTLD	Applicant	#
top	Jiangsu Bangning Science & Tech.	13,562,949
loan	dot Loan Ltd.	4,883,061
xyz	xyz.com LLC	4,649,097
club	.club domains, LLC	2,740,999
online	DotOnline Inc.	2,246,867
vip	Minds + Machines Group Ltd.	1,519,649
app	.APP REGISTRY INC.	1,516,353
win	First Registry Ltd.	1,433,253
shop	GMO Registry, Inc.	1,282,997
site	DotSite Inc.	1,095,918
mobi	Afilias Technologies Limited	1,065,546
work	Minds + Machines Group Ltd.	1,062,016
ltd	Donuts Inc.	1,029,376
ooo	INFIBEAM INCORP. LIMITED	807,894
men	Famous Four Media	792,402
pro	Afilias Technologies Limited	697,329
bid	dot Bid Ltd.	637,413
stream	Famous Four Media	632,298
space	DotSpace Inc.	531,748
website	DotWebsite Inc.	529,024
Total	-	42,716,189

evaluation complete, in auction, in contracting, on-hold, withdrawn), updates, objections, etc. The result shows the new gTLD's application information, such as the prioritization number, string, applicant, and location. When a user selects a certain gTLD, the user can download a public portion of application of the gTLD, and unveil the owner's information and the gTLD's purpose in details.

We gathered the application information of the new gTLDs by scraping the search interface above. The features added include prioritization number, string, applicant, location, application ID, public portion of application (applicant information, secondary contact, proof of legal establishment, applicant background), A-label (A-label is the ASCII-compatible encoded representation of an IDN. A-label is transmitted internally within the DNS protocol. A-labels start with the prefix "xn–" and this representation is called *punycode*.), U-label (U-label is the Unicode representation of an IDN. U-label is shown to the end-user.), address, website, primary contact, phone number, email, attachments, application status, evaluation result, and registry agreement. After we collected the additional information, we augmented the dataset of the new gTLDs and access decision with these additional features.

C. Example of Augmented Data

In the following, we highlight the features we obtained for the various operators of the gTLDs using the augmented dataset by providing details on the top gTLDs owners. We sorted the various gTLDs according to their zone size (the number of SLDs in that zone).

Table III lists the top 20 gTLDs by the number of SLDs in each zone in 2016. As of May 2016, the most populated new gTLDs was `.xyz` with 2,673,614 SLDs (although much of the `.xyz`'s registrations might be due to giveaway domain names as part of the operator's marketing campaign [23]). As an updated exploration on new gTLDs' trends, we also investigated the recent top 20 gTLDs according to the number of SLDs as listed in Table IV, as of September, 2018. In 2018, CZDS provided the zonefile of `.net`, but we excluded it for a fair comparison, as `.net` is one of the traditional gTLDs. `.top` ranked in the first place with about 13 million domains, while it did not appear in the top 20 in 2016 [24].

The total number of SLDs belonging to the top 20 gTLDs has increased from 7.6M to 42.7M over two years. We notice that all gTLDs in the top 20 are non-exclusive, which are sold commercially. It is not surprising that exclusive gTLDs,

TABLE V
THE OWNER OF TOP 20 gTLDs, WITH EXAMPLE gTLDs, POINT OF CONTACT (PoC)—DOMAIN NAME IS AMONG OTHER INFO

Example gTLDs	PoC
xyz, link, click, lol	uniregistry.com
red, kim	afilias.info
win, science, party, loan, date	famousfourmedia.com
site, online, website, space, tech	radixregistry.com

such as .nike, .axa, etc. are not heavily used. In other words, the increase in the number of SLDs means that more users (possibly companies) are purchasing gTLDs through domain registration services upon becoming familiar with new gTLDs.

We note that the number of SLDs of a particular new gTLD is generally influenced by the consumers' choice, and it is difficult to understand the motives of the choice from statistics, such as those presented in this study. The most obvious way to understand the motives of registration is through surveys. However, in practice this becomes impractical as the number of SLDs is in the order of millions.

Based on the historical trends and growth patterns of the number of SLDs under a given old or new gTLD, we believe that the sudden growth is strongly correlated with the pricing and promotion campaigns. With the pricing, we notice that certain new gTLDs, such as website, while generic, have a consistently higher price than new gTLDs in the top of the list. For example, while .website has a price of about \$12 (\$20 for mohaisen.website, for example), a .top domain is sold at \$1.38 for the first year and for \$6.88 for subsequent years (renewal) [25]. We believe that, while not the only factor, such promotions and low price play a major role in determining the size of the zone for a given new gTLD.

While the augmented dataset indicates that the various gTLDs are owned by different owners, thus belonging to different companies, we unveiled that some of the top gTLDs (as well as other less used gTLDs) are related to each other. In some cases, we found that the various seemingly unrelated organizations are the same company running multiple front organizations (for branding purpose). By looking into the application forms of the various new gTLDs, we were able to consolidate the various gTLDs in Table III into a smaller set of organizations with the same set of points of contact in Table V (with the exception of four gTLDs; namely `.club`, `.xin`, `.nyc`, and `.xxx`). We used this augmented dataset later to understand the denied cases for requests to zone files.

D. New gTLD Utilization

Before outlining the CZDS measurements and results, we review some findings on the utilization of the new gTLDs for which we were granted access by providing high-level statistics of the zone files. We examined the zone files of the new gTLDs and find out the number of SLDs of each gTLD.

Figure 4 shows the cumulative distribution of the number of the SLDs in gTLDs in 2016 and 2018. We found that, in 2016, 2.92% of the gTLDs to which we were given access have more than 100,000 SLDs, 79.59% with less than 10,000 SLDs, 44.77% with less than 1,000 SLDs, 38.77% with less

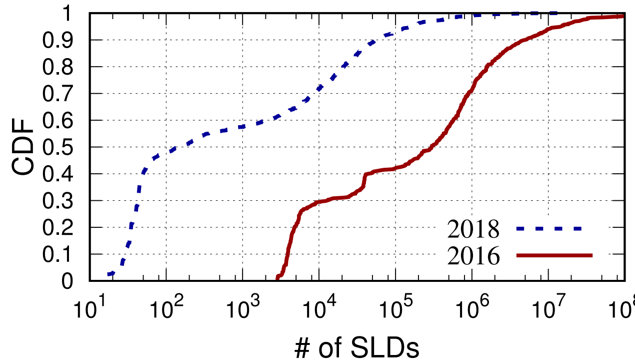

Fig. 4. CDF of the number of SLDs associated with gTLDs. Notice that the total numbers of gTLDs are 787 and 1,229 in 2016 and 2018, respectively.

than 100 SLDs, 25.73% with less than 10 SLDs, and 2.92% with one SLD. On the other hands, in 2018, we found an overall increase in the number of SLDs compared to 2016. In the measurement in 2018, about 1.04% of new gTLDs have more than 1 million SLDs, and 6.56% with more than 100,000 SLDs (3.64% increase from the previous 2.92%). Moreover, about 71.85% have less than 10,000 SLDs, 57.43% have less than 1,000 SLDs, and 47.75% have less than 100 SLDs. These ratios are significantly higher than those of 2016. There is no new gTLD with less than 10 SLDs. Moreover, about 2.42% of new gTLDs have 17 SLDs, which is the minimum number.

These findings highlight 1) various levels of maturity of the new gTLDs and 2) the variant fundamental purposes different new gTLDs serve (e.g., brand vs. generic).

IV. EVALUATION AND RESULTS

In this section, we introduce the evaluation and results of this study. First, we outline the evaluation criteria to understand various aspects of the CZDS by utilizing the access request process for the new gTLD zone files in Section IV-A. We provide the evaluation results and findings on the transparency of CZDS by considering the cooperation of participants, timeliness of responses, and justifications for denied access in Section IV-B.

A. Evaluation Criteria

In the following, we outline and define the various evaluation criteria used in the rest of this paper for understanding various aspects of the CZDS. We first also identify two basic criteria associated with our requests: the **request status** (approved, denied, expired, and pending) and the **reasons behind denial** (if denied). With these criteria, we are interested in understanding whether individual administrators of TLDs are receptive to and collaborative in granting access for reasonably justified requests. We are also interested in the *reasoning behind the denial cases* (where provided) in isolation and along with other gTLDs requested and associated information. This information is important since we use it to understand whether the denial is justifiable or not.

1) Request Status: We use the request status to evaluate CZDS and the listed gTLDs through the CZDS portal. Four possible status values exist: approved, denied, expired,

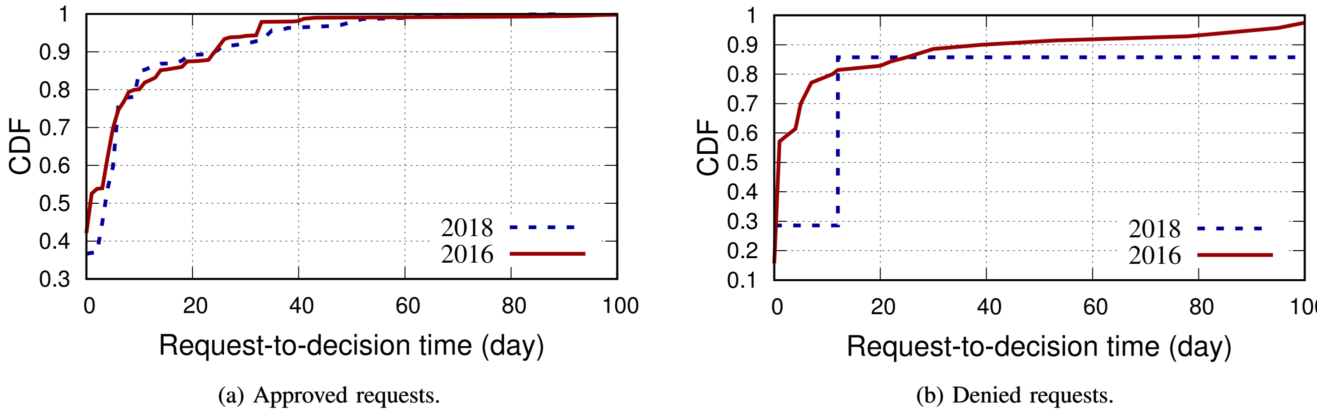

Fig. 5. CDF of request-to-decision time.

or pending. Upon requesting access using CZDS, the first status allocated to each request is "pending", which indicates that no decision is taken by admins on whether to grant access to a request to their zone files. Upon receiving the request, the administrators of the new gTLD can act upon the request, by approval or denial, thus changing the status into approved or denied. When making a decision concerning a request, administrators often provide a remark, highlighting reasons for denial, when access is denied, or notes on allowed use, where access is approved. Some administrators of the gTLDs further limit the period of access to the zone data of their gTLDs. When the access period lapses, the service changes the current status into "expired".

2) Temporal Criteria: Responsiveness and cooperation in our study are best captured by time. To this end, we identify two timing criteria for evaluation: request-to-decision and period to access.

Request-to-Decision (RtD): For RtD, we are interested in measuring and characterizing the time (in days) between the request to access a given zone data of a new gTLD and the approval (or denial, respectively) of that zone data.

Period-to-Access (PtA): For PtA, we are interested in the time allotted by the zone owner to the requesting users (i.e., us) to access the zone. This time is often given by zone owners who granted access and is indicated in the response form at the grant of access time.

3) Reasons for Denial: We also explore reasons of the denial cases in isolation and along with other gTLDs requested and associated information as an evaluation criterion. The CZDS portal provides the status of the request with a response comment. The response comments vary by status. For approved or denied cases, the response comments are logged by the administrators of the gTLDs. By inspecting this field, we have found that when the status is approved, the response comment is "Access granted," "Approved," "Your request has been approved," "For the non-commercial purpose only," or "Thanks for your request. 90 days of access is granted." When the access request is denied, the response comment is "Invalid request," "No IP address provided," "Incomplete user information," "Please re-submit a new request with a valid IP address," among others. Upon expiration of access, the response comments automatically change into "Request expired," where "Request submitted" is associated with pending cases. In order to comprehend the denial reasons, we group and count similar reasons together and highlight them in the results below.

B. Results and Findings

Now we introduce the results, by outlining measurements of the criteria provided in the previous section. We investigate the various characteristics of the new gTLDs including the request result, the RtD and PtA time, the correlation between approval and countries, and reasons for denied access through the CZDS.

1) Request Status: Using the criterion above, we study the behavior of various gTLD owners using their final decisions concerning our request to access the zone files via CZDS. As of May 9, 2016 (6 months after the request to all zones has been filed), we found that 708 requests to zone files of different gTLDs were approved, 70 were denied, and nine were still pending. In total, we found that we were either denied or no answered for roughly 10% of all requested gTLDs' zones.

In September, 2018, we had 1,165 approvals and 6 denials. The number of pending requests as of late September 2018 is 58. The rough percentage of denied and pending is about 5.2%. We can see that the ratio of approved requests significantly increased, from 90% in 2016 to 95% in 2018. As of December, 2018 the number of approvals increased to 1,170, but the rate is still around 95%.

2) Request-to-Decision (RtD) Time – Approvals: First, we measure the RtD time, which represents the responsiveness of the various new gTLD owners to our requests. The CZDS portal provides the history of each status with the exact date and time information. We calculate the time difference between the time of the request and the time of approved/denied status for each gTLD. Figure 5(a) shows the CDF of request-to-decision for the approved access and granted accesses in 2016 and 2018 as highlighted in Section IV-A3. Figure does not include the pending gTLDs, nor the denied gTLDs. The results in this figure highlight various interesting findings and aspects of CZDS as a collective service of various gTLD operators. First, the median time to grant access to a zone file of a gTLD was just under two days, indicating a somewhat timely response

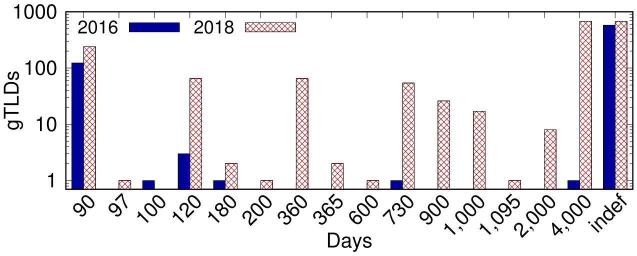

Fig. 6. Distribution of expiration dates given to the approved access. Indef means indefinite access. The total number of approved requests is 708 in 2016 and 1,165 in 2018.

(given the different time zones of the gTLDs and their operators). Second, among the gTLDs requested in this study, 80% took less than 10 days, which is somewhat large in 2016 (this is, 30% of requests took between 2 and 10 days). In 2018, we can see that about 36.7% of approvals were made on the same day as the request date. During a 10 days period, about 84.9% of approvals were given, which is an increase by about 5% from two years earlier. We note that such time is reasonably acceptable with respect to reviews for granting access.

However, we also notice that in 2016 about 20% of the requested gTLDs took more than 10 days, of which about 13% took more than 20 days, about 6% took 30 or more days, and about 2% took more than 40 days. In 2018, about 11.2% took over 20 days, 7.5% took over 30 days, and 3.6% took more than 40 days. The ratio of the approvals which took over 2 months is about 0.7%, indicating that the response time tended to become faster, especially for approvals. As of December 2018, we got five more approvals than the dataset in September. The RtDs of these late approvals are 94, 103, 111, 133, and 138 days, and there are still 52 pending requests.

3) Request-to-Decision (RtD) Time – Denials: Figure 5(b) shows the CDF of request-to-decision time for the gTLDs with denied requests. Compared with the approval time, in 2016 we found that denial of requests came faster: within 24 hours, more than 57% of all denied cases were decided. Moreover, it took less than 5 days for 70% of the denials, 39 days for 90%, 78 days for about 93% of the denials, 95 days for about 96%, and 118 days for all of the denials to be decided. In 2018, we got only 6 denials during 3 months period of observation. Among them, two denials were made on the day of the request, while the four other denials took exactly 12 days (for new gTLDs by the same operator). Given the 58 requests that are still pending, the average RtD time might be longer. As of December, we got one more denial with RtD of 111 days.

4) Period of Access: When approving access to a zone file of a gTLD, administrators can set a "period of access" by indicating an expiration date. During this period of access one may access and download the zone files within the limits of the use policy of CZDS, or as indicated in the note provided by the administrator. If a period of access is given, one has to request access again upon its expiration to continue using the zone. Figure 6 shows the summary of the expiration dates given to the approved access in 2016 and 2018.

Expiration date: We found that the expiration field is not always utilized. Out of the 708 gTLDs requested and approved, in 2016 only 125 gTLDs utilize the expiration field. However, not all the rest (583) provide indefinite access to their zones: we found 12 cases of gTLDs where the administrators also indicated in the comments a period of access (or expiration date) using the response comment for approval, such as "Thank you for your requests. We approved 4,000 days."

By discarding the overlapping cases, we found that 130 out of 708 gTLDs provided an expiration date, where the rest (578, or about 81.6%) provided indefinite access to the zone files, while 676 permissions out of 1,165 approvals in 2018 were given with an indefinite period of access. In 2016, we also found that 123 out of the 130 gTLDs with an expiration date (about 95%) had 90 days of access period, 1 had 100 days, 3 had 120 days, 1 had 180 days, 1 had 730 days, and 1 had 4000 days (`.sakura`). We did not find any specific trend among the gTLDs who specify the expiration date and allow more than the 90 default period utilized by others.

In 2018, 240 approvals were given 90 days, one approval was given 97 days, 6 approvals were given 120 days, 65 were given 180 days, two were given 200 days, one was given 360 days, 65 were given 365 days, two where given 600 days, one was given 730 days, 54 were given 900 days, 26 were given 1,000 days, 17 were given 1,095 days, one was given 2,000 days, and eight were given 4,000 days of access. We can see that the ratio of allowing indefinite period decreased from 81.6% to 58.0%. On the other hands, among the approvals with specified period, the ratios of new gTLDs which allow 90 days (default) decreases from 95.1% (123 out of 130) to 49.1% (240 out of 489). Moreover, the new gTLDs that allowed access for more than 1 year have increased from 2 (1.5%) to 174 (35.6%).

5) Correlation: Approval and Country of Origin: Now we focus on studying the correlation between approval and other characteristics and features of the applicants of gTLDs, viz. spatial information. We use the augmented dataset highlighted in Section III for this analysis. Among other features, we highlight the correlation between the country of the applicant (as explicitly stated in ICANN fillings) and the approval or denial of access.

We use the location feature we obtained from ICANN's resources for new gTLDs as shown in Figure 7. The location information is acquired from the applicant of the new gTLD as the principal place of business. The country code is based on ISO 3166-1 [26]. Figure 7 shows the approval per country, where we start with countries with the highest number of gTLDs applied for and delegated. Before delving into further details, we emphasize that the results provided herein include several countries known as "tax havens"; countries that are more "friendly" to new business, such as Cayman Islands (KY), Gibraltar (GI), etc., which makes it very plausible that the characteristics provided herein are more associated with the group of new gTLDs registered in the said countries than the countries themselves or their policies. In the following, and for brevity, we use the country to refer to new gTLDs with PoC in them as a shorthand form.

In 2016, we found that admins who registered their PoC's address in Japan (JP), Cayman Islands (KY), Ireland (IE), Gibraltar (GI), and Spain (ES) provided access to all their gTLDs. Moreover, those in France (FR) and China (CN) have

Rank	cc	#	A	D	P	R
1	US	365	340	25	0	0.932
2	JP	45	45	0	0	1.000
3	KY	43	43	0	0	1.000
4	DE	38	34	3	1	0.895
5	CH	33	32	0	1	0.970
6	GB	24	21	3	0	0.875
7	AU	22	19	2	1	0.864
7 (tie)	FR	22	16	4	2	0.727
9	IE	21	21	0	0	1.000
10	CN	20	15	4	1	0.750
11	GI	15	15	0	0	1.000
12	ES	13	13	0	0	1.000
13	HK	12	10	2	0	0.833
14	BR	9	8	1	0	0.889
15	SE	8	7	1	0	0.875

(a) 2016 data

Rank	cc	diff	#	A	D	P	R
1	US	-	585	575	0	10	0.983
2	LU	*new*	60	60	0	0	1.000
3	JP	▼ 1	58	58	0	0	1.000
4	DE	-	50	49	0	1	0.980
5	KY	▼ 2	48	44	0	4	0.917
6	GB	-	45	40	0	5	0.889
7	CH	▼ 2	41	37	0	4	0.902
8	FR	▼ 1	33	31	2	0	0.940
9	CN	▲ 1	32	20	0	12	0.625
10	AU	▼ 3	31	30	0	1	0.968
11	HK	▲ 2	28	24	4	0	0.857
12	IE	▼ 3	23	23	0	0	1.000
13	GI	▼ 2	17	17	0	0	1.000
14	IN	*new*	15	11	0	4	0.733
15	ES	▼ 3	14	11	0	3	0.786

(b) 2018 data

Fig. 7. Approval per country (cc), sorted by the number of new gTLDs from the highest. Comparison criteria are the number of applied for (#), approved (A), denied (D), and pending (P) TLDs, as well as the ratio of the approved to total applied for TLDs (R).

Rank	cc	#	A	D	P	R
1	VG†	8	0	8	0	0.000
1 (tie)	PT	2	0	2	0	0.000
1 (tie)	TR	2	0	2	0	0.000
1 (tie)	TW	2	0	1	1	0.000
1 (tie)	FI†	1	0	1	0	0.000
1 (tie)	QA	1	0	1	0	0.000
7	RU†	7	2	4	1	0.286
8	CA	4	2	2	0	0.500
9	LU†	7	4	3	0	0.571
10	IN	7	5	1	1	0.714
11	FR	22	16	4	2	0.727
12	CN	20	15	4	1	0.750
13	HK	12	10	2	0	0.833
14	AU	22	19	2	1	0.864
15	GB	24	21	3	0	0.875

(a) 2016 data

Rank	cc	diff	#	A	D	P	R
1	PT	-	2	0	0	2	0.000
2	TR	▼ 1	7	2	0	5	0.286
3	TW	▼ 2	2	1	0	1	0.500
4	CN	▲ 8	32	20	0	12	0.625
5	IN	▲ 5	15	11	0	4	0.733
6	ES	*new*	14	11	0	3	0.786
7	FI	▼ 6	5	4	0	1	0.800
8	IT	*new*	6	5	0	1	0.833
9	HK	▲ 4	28	24	4	0	0.857
9 (tie)	RU	▼ 2	7	6	0	1	0.857
11	GB	▲ 4	45	40	0	5	0.889
11 (tie)	CA	▼ 3	9	8	0	1	0.889
13	CH†	*new*	41	37	0	4	0.902
14	BR	*new*	11	10	0	1	0.909
15	KY†	*new*	48	44	0	4	0.917

(b) 2018 data

Fig. 8. Approval per country (cc), sorted by the approval ratio from the lowest. Comparison criteria are the number of applied for (#), approved (A), denied (D), and pending (P) TLDs, as well as the ratio of the approved to total applied for TLDs (R). † indicates a country with more than 50% of the applications with inconsistent information.

relatively a lower approval ratio. In 2018, those in LU also provided access to all gTLDs as well as additional three countries; JP, IE, and GI, which had previously allowed access to all in 2016. On the other hands, CN, IN, and ES showed the relatively lower ratio.

Fig. 8 shows the approval ratio according to the country code, where we listed them from the lowest approval ratio to the highest for the tail of the country distribution. Virgin Islands (VG), Portugal (PT), Turkey (TR), Taiwan (TW), Finland (FI), and Qatar (QA) do not provide any access permissions to their data. Interesting too, we also notice that administrators from the same countries make up the majority of the pending requests (in total, 7 out of 9 requests are for gTLDs in those countries). Table VI shows these pending gTLDs. Taiwan and Russia, which are the owners' principal place of business of the new gTLD `.xn-mxtq1m` (.政府) and `.gdn`, have especially a very low approval ratio. In conclusion, we found a strong spatial correlation of denial, where administrators in certain countries tend not to open their zone data files categorically.

In 2018, only 6 countries showed a ratio smaller than 0.8. Compared with 12 countries in 2016, this is a significant

TABLE VI
9 gTLDs BASED ON PENDING REQUESTS, WITH APPLICANT, AND COUNTRY CODE IN 2016

gTLD	Applicant	cc
xn–mxtq1m (政府)	Net-Chinese Co., Ltd.	TW
gdn	Joint Stock Co. "Navigation-info. sys."	RU
bharti	Bharti Enterprises Private Ltd.	IN
bzh	Association www.bzh	FR
corsica	Collectivité Territoriale de Corse	FR
xn–vuq861b (信息)	Beijing Tele-info Network Tech. Co., Ltd.	CN
iinet	Connect West Pty. Ltd.	AU
ifm	ifm electronic gmbh	DE
voting	Valuetainment Corp.	CH

TABLE VII
COUNTRIES GRANTED ACCESS TO THEIR ZONE DATA, WITH COUNTRY CODE (CC) AND THE NUMBER OF APPLIED FOR (#). † INDICATES A COUNTRY WITH MORE THAN 50% OF THE APPLICATIONS WITH INCONSISTENT INFORMATION (PERHAPS DUE TO OUTSOURCING REGISTRATION)

cc	#	cc	#	cc	#	cc	#	cc	#
JP	45	KY†	43	IE†	21	GI	15	ES	13
AE†	7	IT	5	NL	5	DK	4	KR	4
SA†	4	SG	4	BE	3	ZA	3	AT	2
BM†	2	MX†	2	NO†	2	UY†	2	IQ	1
NZ†	1	TH	1						

(a) 2016 data

cc	#	cc	#	cc	#	cc	#	cc	#
LU†	60	JP	58	IE†	23	GI	17	NL	12
SE	10	AT	6	SG	5	DK	5	VA	4
KR	4	SA†	4	UY	4	ZA	4	VG†	4
BE	3	EG	2	KW	2	BH†	2	MX†	2
BM†	2	NO	2	TH	1	NZ†	1	IQ	1
IM	1	QA	1	IL	1	CO†	1	PA†	1
PH	1								

(b) 2018 data

decrease. We can deduce that these changes are likely related to an increase in the average ratio seen in Table I. Among those countries, PT, TR, TW, CN, IN, and FI were below 0.8 in both 2016 and 2018. From this table, we can see that those countries do not promptly respond to the requests or even given limited access to the gTLDs.

In Table VII, we present the list of all countries where admins provided access to all of their gTLDs regardless to the number of gTLDs being applied for. From this list, it is very difficult to draw any conclusion: counter-intuitively, admins in countries in Western Europe approved requests in their entirety, where one would speculate that administrators would not share zone files for privacy concerns, as well as other countries in the Middle East (e.g., Iraq, Saudi Arabia, United Arab Emirates; notoriously known for their Internet censorship) approved access categorically; contrary to the presumption that access to data in such countries is very tightly controlled. Part of such behavior is perhaps due to the "outsourcing" of operation, where some applicants use off-shore companies to register and operate their new gTLDs, as discussed below.

TABLE VIII
REASONS FOR REJECTION, WITH REASON, THE NUMBER OF APPLIED FOR (#), AND THE PERCENT OF APPLIED FOR (%)

Reason	#	%
Invalid request	25	35.71%
No IP address provided	21	30%
Incomplete user information	14	20%
Lack of the required credentials	10	14.29%

Registration Outsourcing: We consider the consistency of the information provided by applicants in each of those countries listed in Table VII. We identify a country to be inconsistent if a piece of information provided on the application for the majority of the applicants is inconsistent with the country of the application. For example, having a phone number of a foreign country (other than the applicant) that is inconsistent with the rest of the contact information (e.g., address) is an indicator of inconsistency. We mark those countries with a special mark (†), where we found the inconsistent applications for those countries to constitute more than 50% of their total applications. We use such (in)consistency to further understand the approved access in certain countries.

We found that the inconsistent pieces of information (PoC, phone number, etc.) are for registrars and admins from the United States, United Kingdom, Hong Kong, China, Finland, Switzerland, South Africa, Germany, Japan, Malaysia, Australia, and Argentina. Accordingly, half of them are domain holding companies and agents acting on behalf of registrants for their new gTLD registration needs. Despite the differences in the information and features of administrators, it is an interesting phenomenon that an almost binary decision occurs at the country level *in some cases*.

We notice that establishing outsourcing directly, with the limited data we have, is almost impossible. We use indirect measurements, including the use of inconsistency, as a clue. This is a best-effort attempt at the problem. In the future, we will seek to improve our findings via further data augmentation, which may require manual collection.

6) Reasons for Access Denial: We look into the reasons for the denial of access to new gTLDs to identify whether they are justified, in light of the provided information at the time of the request for access, and in accordance with the goal and terms of CZDS. Table VIII shows the reasons for denied access. Out of the 70 gTLDs, 25 were denied for "invalid request", 21 were denied for "no IP address provided", 14 were denied for "incomplete user information", and 10 were denied for "the lack of the required credentials".

We categorized the strings of invalid request cases among these gTLDs into location, activity, drink, brand, and general words as shown in Table IX. We found that the majority of these gTLDs are everyday terms. Furthermore, we investigated the owners of the 70 gTLDs based on their reasons for rejection. We found that many of the gTLDs that rejected the request are owned by a single owner, as shown in Table X.

TABLE IX
THE STRINGS OF INVALID REQUEST CASES, WITH CATEGORY, EXAMPLE gTLDs, AND COUNTRY CODE (CC)

Category	Example gTLDs (cc)
Location	bayern (DE), budapest (US), london (GB), miami (VG), nrw (DE)
Activity	cooking (VG), fishing (VG), rodeo (VG), surf (VG), work (US), yoga (KY)
Drink	beer (US), casa (US), vodka (US)
Brand	bradesco (BR)
General word	abogado (US), country (US), fashion (VG), fit (VG), garden (VG), gop (US), horse (ZA), law (ZA), luxe (ZA), wedding (ZA)

TABLE X
THE OWNER OF INVALID REQUEST CASES, WITH EXAMPLE gTLDs, WEB SITE, AND POINT OF CONTACT (PoC)

Example gTLDs			Web site	PoC
bayern			bayernconnect.com	≈
budapest	miami	nrw	tldh.org	Antony Van Couvering
cooking	fishing	rodeo		
surf	work	yoga		
beer	casa	vodka		
abogado	country	fashion		
fit	garden	horse		
law	luxe	wedding		
gop			rslc.com	≈

While the various TLDs use different websites for the registering organization and for administrative purposes, by looking into other pieces of information provided on the TLDs applications we found that the primary point of contact (contact name) for the various TLDs is the same, thus establishing the connection between the various TLDs. In particular, we found that most denied access cases are for TLDs owned by a company that is currently chaired by the former chairman of the board of ICANN, Peter Dengate Thrush [27], a "barrister specializing in Internet law". Exceptions of TLDs not owned by the same company are the new gTLD `.london` and `.yoga`.

In 2018, We received 6 denials during the 3 months of observation after requesting access to all new gTLDs. Among them, four denials from the gTLDs (wang, xn–45q11c (八卦), xn–czru2d (商城), and xn–hxt814e (网店) whose PoC are *zodiac-corp.com*, stated the reason as "Please offer your website so we can take a look at it, and we usually refuse those who publish the detailed domain list to the public." The remaining two denials from .alstom and .bnpparibas, have different PoCs, *alstom.com* and *bnpparibas.com*, respectively, but with the same reason that "You failed to provide the complete required credentials. The telephone number you have provided is not valid format and could not be reached. Please be also aware that .alstom (.bnpparibas) is a closed and restrictive TLD limited to the Registry Operator and its affiliates". We note that the provided phone number was valid (the lead author's phone number), and was reachable. Between September and December 2018, we got one more denial from mtn. The reason of denial was "The .mtn not currently in use."

TABLE XI
TOP 13 OWNERS, WITH EXAMPLE gTLDs, POINT OF CONTACT (PoC; DOMAIN NAME–ASSOCIATED ATTRIBUTES ARE OMITTED), AND THE NUMBER OF DELEGATED AND APPLIED FOR TLDs (DEL/APP)

Example gTLDs	PoC	del/app
team, camera, lawyer, band, family, ...	donuts.co	215/307
yahoo, book, buy, yamaxun, nokia, ...	valideus.com	54/101
android, gmail, google, hangout, play, youtube, app, ...	google.com	43/101
surf, work, cooking, garden, ...	tldh.org	26/77
poker, bet, black, lotto, pink, ...	afilias.info	24/57
auto, game, hiphop, link, lol, tattoo, help, property, ...	uniregistry.com	23/54
viking, ally, origins, travelers, ...	cscinfo.com	22/95
toyota, honda, hyundai, kia, epson, softbank, playstation, sony, ...	brights.jp	18/25
social, actor, airforce, army, navy, ...	unitedtld.com	17/26
racing, party, date, loan, review, science, trade, ...	famousfourm[..].com	16/56
xerox, lifeinsurance, aig, baby, marriott, pharmacy, ...	fairwindspart[..].com	15/54
bmw, mini, allfinanz, dvag, zuerich, ...	thomsentram[..].com	12/15
IDN of TLDs (com, net, ...)	verisigninc.com	11/12

Out of 7 denials we collected in 2018, we found that 6 new gTLDs (wang, xn–45q11c (八卦), xn–czru2d (商城), xn–hxt814e (网店), bnpparibas, and mtn) also rejected in 2016; the other new gTLD (alstom) was delegated in Jun. 2016, which is later than our first data collection.

7) Dominant Operators in New gTLD Ecosystem: Table XI shows the top 13 owners with more than 10 delegated gTLDs, having collectively 409 out of 968 (42.25%) gTLDs. The `donuts.co` has as many as 215 gTLDs, more than 20% of the total number of gTLDs. Interestingly, such a large number of gTLDs, while indeed increases the number of choices available to consumers, also highlights the presence of monopolies and central players in the evolving ecosystem.

V. DISCUSSION

The previous sections provided a comprehensive analysis of CZDS as a service. In this section, we introduce a couple of suggestions for the improvement of CZDS.

Policy on Registry's Response: In the previous sections, we found that many of registries for new gTLDs responded to our requests in timely manner. However, we also could see that about 10% of decisions took more than 20 days in the measurement in 2018 and even some of them did not respond as of November 2018, 5 months after the initial requests.

To figure out the reason of the delayed responses (or even no-response), we inquired to ICANN about the following: 1) whether there is policy in place in relation to the response time of the registry, 2) and (if not) whether there is a plan to establish the relevant policy. We received a response with a detailed answers to the inquiries. The followings are the snippets of the ICANN's response to the above questions.

> "There is no defined time period that a Registry Operator must reject or approve a request. In general, a request is usually responded to in less than one month, however, a response may take longer due to the number of requests the Registry receives or

> other business factors. You are encouraged to check each Registry's zone file policy or procedure, where it exists, for more information."
>
> "Regarding changing the Registry Agreements to include a time period by which Registries must abide to approve/deny requests, please note that ICANN policy is developed by the community in a bottom-up, multistakeholder fashion. There are multiple ways in which you can participate, from attending a meeting in person, joining a working group or webinar to contributing to the Public Comment··· "

In conclusion, and despite the explicit goals of CZDS as an enabler of DNS operation and transparency, there is no policy related to the response time in place. However, in principle, there is nothing, in theory, that prevents establishing such as a policy as an article in the Registry Agreement. As a result of this dialogue with ICANN, we are considering presenting the findings of this work to the appropriate work group so that they can have more informed discussion the incorporation of the policy to facilitate timely sharing in CZDS.

Limited Time of Access to Zone File: In our measurement, we also found that while half of all registries allow users to access their zone files indefinitely, the other half assigned a specified period of permission (90 days, 1 month, 1 year etc.) This result shows that many registries still restrict the period of access. Considering that analyzing zone files can be part of a continuous monitoring effort; e.g., domain name security research [28], [29], the limited access to zone files may hinder such an application of continuous monitoring.

We also conducted a survey to understand the reason of various periods given to the user. First, we contacted ICANN to inquiry about the policy on the length of permission.

> "Generally, the Registry Operator will default requests to minimum of 90 days. However, the request may have a longer life span. To gain access, you may log back into the CZDS system and re-request the zone file. Additionally, you may contact the Registry Operator regarding the length of the request."

From the response of ICANN, we found that the expiry date is determined by the registries and ICANN only has a guideline of the minimum period of 90 days. It is thus at the discretion of the registries to determine the length of access.

To follow up on this issue by exploring root causes at the operators, we also contacted 15 among those who allowed access only for a limited period of time by emailing the addresses providing in the CZDS as a PoC. However, most of them (12 registries) did not respond to our inquiries. Where we had responses, they were limited to generic explanations, such as "the default setting of the system" or "no specific reason".

We recognize it would be natural the registry that owns a TLD has an autonomy to decide the execution of the policy. However, we still believe that it is better to the transparency of the Internet operation to establish a guideline on the period of access. It would be helpful for the improvement of CZDS to ensure that a sufficient length of access to researchers, especially academic researchers with confirmed affiliation.

Privacy in CZDS: In CZDS, a user is required to provide his/her IP address which will be used for downloading the zone file. Interestingly, however, we found that there is no restriction on IP address to download the zone file. More importantly, we could download the zone files for approved TLDs regardless of the IP address of the machine downloading the files. Both through CZDS portal and the API provided by CZDS, a machine—which does not have the IP address we provided initially—successfully downloaded the zone file. At this point, we have a concern about the user's privacy. Even if the specified IP address is not used to restrict the access, CZDS still requires the user's information, which is considered Personally Identifiable Information (PII). Unless such information is needed for the restricted access, it would be better not to collect the user's information in CZDS.

VI. Related Work

While there has been a lot of work on general DNS measurements [30], [31], transparency [32], operation [33], and security [34], there are only a few studies considering the introduced new gTLDs. The launch of new gTLDs has expanded the top-level domains used in global Internet. Past research has primarily focused on understanding the trademark-related issues, evolution, adoption, and usage of the new gTLDs. Less work is on understanding the transparency of sharing information of new gTLDs and zones (e.g., CZDS).

New gTLDs: Previous work was done on the trademark-related issues of the negTLDs. Nagel and Sandner [35] introduced the positive branding effects and the trademark-related risks of new gTLDs. They conducted interviews to test whether branding theories are applicable to new gTLDs. Kuipers [36] studied how domain names relate to trademarks and trade names for commercial and non-commercial purposes. Joseph [37] found that the majority of brand owners do not welcome new gTLDs since they require brand owners to spend a significant amount of energy and resources on them. He also found that new gTLDs would likely harm trademark law by leading to widespread trademark dilution and consumer confusion.

Prahl and Null [38] points out that ICANN has not yet provided adequate protection to gTLD trademark owners. There are no measures other than monitoring gTLDs directly against attacks such as cybersquatting, indicating that they could have a negative risk of harming their brands or brands.

Limited work has been done on studying evolution, adoption, and usage of the new gTLDs. Jarassriwilai *et al.* [39] analyzed TLDs usage in 2008 to 2015 and found that while there were no changes in top talkers [40], new gTLDs use was growing. They found that while there were no changes in the appearance of most frequently used TLDs, the presence of the new gTLDs in the dataset is growing. Another work by Halvorson *et al.* [24] analyzed the types of domain registrations in the new TLDs to determine registrant behavior in the brand new world of naming abundance. They investigated the cost structures and monetization models for the new TLDs to identify which registries were profitable. They found that only 15% of domains in the new TLDs show characteristics consistent with primary registrations, while the rest are promotional, speculative, or defensive in nature; 16% of domains with NS records do not even resolve yet, and 32% are parked. Osterweil *et al.* [41] evaluated the risks that Internet users are about to face by the introduction of as many as one thousand new gTLDs. They proposed a set of measures that represent risks to end users, and illustrated risk incidents by measuring operational threat vectors that could be used to orchestrate failures and attacks. Chen *et al.* [42] demonstrated a man-in-the-middle attack by name collision using new gTLDs.

Malicious TLDs: As usage of new gTLDs has climbed consistently and many abusive activities in several new gTLDs have occurred. However, limited work has been done on this issue. Previous work has focused on malicious domain detection. Abbasi and Chen [43] proposed a fraudulent site detection system using a support vector machine classifier and a rich feature set derived from website text, linkage, and images to protect Internet users from fake website developers. Bilge *et al.* [44] presented a malicious domain detection system. They used features from DNS traffic of local recursive DNS servers to characterize properties of DNS names and the query pattern. They performed a real-life deployment with a large and real dataset consisting of 100 billion DNS requests in an ISP. They can automatically identify unknown malicious domains that are misused in various malicious activity, including botnet, spamming, and phishing. Antonakakis *et al.* [45] developed a malware-related domain name detection system. The proposed system passively monitors DNS network traffic at the upper DNS hierarchy. They can detect malware domains through analyzing global DNS query resolution patterns. They experimented with the real dataset and showed high detection rates and low false-positive rates. Moreover, the system can detect new malware domains before they appear in public. Moura *et al.* [46] proposed a warning system of malicious domains for TLD registries. The system monitors an entire DNS zone and singles out newly added suspicious domains. They used domain registration and global DNS lookup patterns of a TLD to monitor and classify an entire DNS zone. The proposed system can detect some types of domain abuse, including malware, phishing, and allegedly fraudulent Web shops.

Internet Governance: Several proposals raise issues with the current Internet governance paradigm. Clark [47] described the current Internet governance model and the process towards a future mode of operation. Larsen [48] provided a historical overview of major events implicating TLD ownership. He considered cases and statutes relating to domain names and TLDs and identified two strains of judicial interpretation. He argued for an extralegal practice using four normative considerations: stability, predictability, descriptive accuracy, and respect for the interests of the Internet community at large. Matsumoto *et al.* [49] addressed the problem of scaling authentication for naming, routing, and end-entity certification to a global environment in which authentication policies and users' sets of trust roots vary widely. They proposed a Scalable Authentication Infrastructure for Next-generation Trust (SAINT), which partitions the Internet into groups with common, local trust roots, and isolates the effects of a compromised trust root. They presented that SAINT makes trust root management a central part of the network architecture, enabling trust root updates within seconds and allowing users to make flexible trust decisions. Loutocky [50] found that the Uniform domain name Dispute Resolution Policy (UDRP) delimits procedural rules to solve disputes regarding the registration of domain names and to create non-binding decisions which approve and reject the transfer of the domain to new owners. He focused on the question of whether the decision-making process is ready for such a big change or it is easily malleable. He also investigated disputes to show defects of the rules, which are pointing at problems not only recently but already occurring for quite a long time.

Domain Name Transparency: Limited work has been done on domain name transparency. Weber [32] found that ICANN has recognized the need to improve the transparency framework within its structures; the ongoing attempts should be strengthened by scholar research supporting the efforts of the ICANN bodies in the present consultation phase. While previous works have focused on evolution, adoption, and usage of TLDs, there is no study that analyzes CZDS, especially with the new gTLD. Our research provides a detailed analysis of CZDS with a perspective from Internet transparency and service quality. McGovern *et al.* [51] presented a working definition of transparent governance and described its importance for governments and explores key enabling transparency forces and mechanisms. He provided some suggestions for increasing transparency in all countries and at all levels of governance.

VII. Conclusion and Future Work

The CZDS is a centralized service for requesting access to the zone data of new gTLDs and its objective is to facilitate transparency and improve the security of the new gTLDs. In this paper, we investigated and evaluated the CZDS's transparency by measurement and provided recommendations for its improvement. By requesting access the new gTLDs' zone data using CZDS, we measured various characteristics of the new gTLDs and their operation, including the request result, the Request-to-Decision and Period-to-Access time, the correlation between approval and countries, and reasons for access denial. We found that it took a long time for some gTLD administrators to respond by approval or denial of access, whereas some administrators denied access with unjustified reasons. This study highlights the need for a well-maintained process and policy by ICANN for the stable operation of the CZDS, by specifying a reasonable time for response, the guideline for expiration dates, and clear conditions for denial. We will explore influencing the discussing around CZDS in ICANN's working groups as our future work.

References

[1] A. R. Kang, S. H. Jeong, S. Y. Ko, K. Ren, and A. Mohaisen, "Transparency in the new gTLD era: Evaluating the DNS centralized zone data service," in *Proc. IEEE HotWeb*, Washington, DC, USA, 2016, pp. 54–59.

[2] P. Mockapetris, "Domain names-implementation and specification," Internet Eng. Task Force, RFC 1035, 1987.

[3] J. Postel, "Domain name system structure and delegation," Internet Eng. Task Force, RFC 1591, 1994.

[4] J. Levine and P. Hoffman, "Variants in second-level names registered in top-level domains," Internet Eng. Task Force, RFC 6927, 2013.

[5] A. Cormack, "Do generic gTLDs require their own ex ante regulation?" *Int. Rev. Law Comput. Technol.*, vol. 30, no. 3, pp. 150–173, 2016.

[6] CNN. (2016). *List of Proposed New Top-Level Domains*. [Online]. Available: http://money.cnn.com/infographic/technology/new-gtld-list/

[7] S. Wang, Z. Zheng, Z. Wu, M. R. Lyu, and F. Yang, "Reputation measurement and malicious feedback rating prevention in Web service recommendation systems," *IEEE Trans. Services Comput.*, vol. 8, no. 5, pp. 755–767, Oct. 2015.

[8] ICANN. *Domain Abuse Activity Reporting (DAAR)*. Accessed: Jun. 10, 2019. [Online]. Available: https://www.icann.org/octo-ssr/daar

[9] ICANN. (2013). *Name Collision Resources & Information*. [Online]. Available: https://www.icann.org/resources/pages/name-collision-2013-12-06-en

[10] ICANN. (2014). *Centralized Zone Data Service*. [Online]. Available: https://czds.icann.org/en

[11] ICANN. (2014). *ICANN Targeted in Spear Phishing Attack | Enhanced Security Measures Implemented*. [Online]. Available: https://www.icann.org/news/announcement-2-2014-12-16-en

[12] A. Mohaisen, M. Bhuiyan, and Y. Labrou, "Name server switching: Anomaly signatures, usage, clustering, and prediction," in *Proc. 15th Int. Workshop Inf. Security Appl. (WISA)*, 2014, pp. 202–215.

[13] ICANN. *About the New gTLD Program*. Accessed: Nov. 15, 2018. [Online]. Available: https://newgtlds.icann.org/en/about/program

[14] J. Klensin, "National and local characters for dns top level domain (TLD) names," Internet Eng. Task Force, RFC 4185, 2005.

[15] J. Klensin, "Internationalized domain names for applications (IDNA): Background, explanation, and rationale," Internet Eng. Task Force, RFC 5894, 2010.

[16] P. Resnick and P. Hoffman, "Mapping characters for internationalized domain names in applications (IDNA) 2008," Internet Eng. Task Force, RFC 5895, 2010.

[17] S. Sharikov, D. Miloshevic, and J. Klensin, "Internationalized domain names registration and administration guidelines for European languages using Cyrillic," Internet Eng. Task Force, RFC 5992, 2010.

[18] D. Thaler, J. Klensin, and S. Cheshire, "IAB thoughts on encodings for internationalized domain names," Internet Eng. Task Force, RFC 6055, 2011.

[19] S. Cheshire and M. Krochmal, "Special-use domain names," Internet Eng. Task Force, RFC 6761, 2013.

[20] J. Appelbaum and A. Muffett, "The '.onion' special-use domain name," Internet Eng. Task Force, RFC 7686, 2015.

[21] R. Barnes, A. Cooper, O. Kolkman, D. Thaler, and E. Nordmark, "Technical considerations for Internet service blocking and filtering," Internet Eng. Task Force, RFC 7754, 2016.

[22] I. Foster. *CZDS Access Automation API*. Accessed: Aug. 1, 2019. [Online]. Available: https://github.com/lanrat/czds

[23] A. Allemann. *XYZ Domain Names Now Cost One Penny*. Accessed: Jun. 13, 2019. [Online]. Available: https://domainnamewire.com/2016/05/31/xyz-domain-names-now-cost-one-penny/

[24] T. Halvorson, M. F. Der, I. D. Foster, S. Savage, L. K. Saul, and G. M. Voelker, "From .academy to .zone: An analysis of the new TLD land rush," in *Proc. ACM IMC*, 2015, pp. 381–394.

[25] NameCheap. *Namecheap Domain*. Accessed: Jun. 13, 2019. [Online]. Available: https://bit.ly/2z8H6in

[26] ICANN. (2016). *New gTLD Current Application Status*. [Online]. Available: https://gtldresult.icann.org

[27] A. Raja, "ICANN's new generic top-level domain program and application results," *Intellectual Property Brief*, vol. 4, no. 2, p. 3, 2013.

[28] R. Chandramouli and S. Rose, "An integrity verification scheme for DNS zone file based on security impact analysis," in *Proc. ACSAC*, 2005, pp. 312–321.

[29] T. Halvorson, M. F. Der, I. Foster, S. Savage, L. K. Saul, and G. M. Voelker, "From. academy to. zone: An analysis of the new TLD land rush," in *Proc. ACM IMC*, Tokyo, Japan, 2015, pp. 381–394.

[30] K. Schomp, T. Callahan, M. Rabinovich, and M. Allman, "On measuring the client-side DNS infrastructure," in *Proc. IMC*, Barcelona, Spain, 2013, pp. 77–90.

[31] R. van Rijswijk-Deij, M. Jonker, A. Sperotto, and A. Pras, "A high-performance, scalable infrastructure for large-scale active DNS measurements," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 6, pp. 1877–1888, Jun. 2016.

[32] R. H. Weber, "Transparency and the governance of the Internet," *Comput. Law Security Rev.*, vol. 24, no. 4, pp. 342–348, 2008.

[33] V. Ramasubramanian and E. G. Sirer, "The design and implementation of a next generation name service for the Internet," *SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 4, pp. 331–342, 2004.

[34] A. Kountouras *et al.*, "Enabling network security through active DNS datasets," in *Proc. RAID*, Paris, France, 2016, pp. 188–208.

[35] A. Nagel and P. Sandner, "Evaluation of new generic top level domains from a brand and a trademark perspective," *Int. J. Manag. Marketing Res.*, vol. 8, no. 1, pp. 93–112, 2015.

[36] S. Kuipers, "The relationship between domain names and trademarks/trade names," M.S. thesis, Lund Univ., Lund, Sweden, 2015. [Online]. Available: http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=5470120&fileOId=5470125

[37] A. M. Joseph, "I cann't believe it's not better: Why new gTLDs are bad for brand owners and trademark law," *J. Intellectual Property Law*, vol. 20, no. 1, p. 149, 2012.

[38] D. S. Prahl and E. Null, "The new generic top-level domain program: A new era of risk for trademark owners and the Internet," *Trademark Rep.*, vol. 101, no. 6, p. 1757, 2011.

[39] T. Jarassriwilai, T. Dauber, N. Brownlee, and A. Mahanti, "Understanding evolution and adoption of top-level domain names," in *Proc. Local Comput. Netw. Conf. Workshops (LCN Workshops)*, 2015, pp. 687–694.

[40] E. Osterweil, D. McPherson, S. DiBenedetto, C. Papadopoulos, and D. Massey, "Behavior of DNS top talkers: .com/.net view," in *Proc. PAM*, 2012, pp. 211–220.

[41] E. Osterweil, M. Thomas, A. Simpson, and D. McPherson. (2013). *New gTLD Security, Stability, Resiliency Update: Exploratory Consumer Impact Analysis*. [Online]. Available:http://bit.ly/QB6ntp

[42] Q. A. Chen, E. Osterweil, M. Thomas, and Z. M. Mao, "MitM attack by name collision: Cause analysis and vulnerability assessment in the new gTLD era," in *Proc. IEEE Security Privacy Symp.*, 2016, pp. 675–690.

[43] A. Abbasi and H. Chen, "A comparison of tools for detecting fake websites," *Computer*, vol. 42, no. 10, pp. 78–86, Oct. 2009.

[44] L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi, "Exposure: Finding malicious domains using passive DNS analysis," in *Proc. NDSS*, 2011, pp. 1–17.

[45] M. Antonakakis, R. Perdisci, W. Lee, N. Vasiloglou, II, and D. Dagon, "Detecting malware domains at the upper DNS hierarchy," in *Proc. USENIX Security Symp.*, vol. 11, 2011, pp. 1–16.

[46] G. C. M. Moura, M. Müller, M. Wullink, and C. Hesselman, "nDEWS: A new domains early warning system for TLDs," in *Proc. NOMS*, Apr. 2016, pp. 1061–1066.

[47] N. Clark, "Internet governance: Is it finally time to drop the training wheels?" *Aust. J. Telecommun. Digital Econ.*, vol. 3, no. 1, pp. 16–30, 2015. [Online]. Available: http://telsoc.org/ajtde/index.php/ajtde/article/view/3

[48] W. Larsen, "A 'stern' look at the property status of top-level domains," *Univ. Chicago Law Rev.*, vol. 82, no. 3, pp. 1457–1509, 2015. [Online]. Available: http://www.jstor.org/stable/43575202

[49] S. Matsumoto, R. M. Reischuk, P. Szalachowski, T. H.-J. Kim, and A. Perrig, "Designing a global authentication infrastructure," *CoRR*, vol. abs/1506.03392, pp. 1–14, Jun. 2015. [Online]. Available: http://arxiv.org/abs/1506.03392

[50] P. Loutocky, "Are we getting good decisions by top-level domain name dispute resolution providers," *Masaryk Univ. J. Law Technol.*, vol. 9, no. 1, pp. 111–128, 2015.

[51] M. H. McGovern, T. C. Beierle, T. Jandl, and N. E. Harrison, "Transparent governance: The role of nongovernmental organizations and the Internet," in *National, Regional Institutions & Infrastructures*, UNESCO—EOLSS, 2005.

Jeman Park (S'15) received the B.Sc. degree in computer and communication engineering from Korea University, Seoul, South Korea, in 2016. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, University of Central Florida. His research has been focused on privacy, and computer security and systems.

Jinchun Choi received the B.Eng. and M.S. degrees in computer and information engineering from Inha University, Incheon, South Korea, in 2011 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Science, University of Central Florida and the Department of Computer Information Science, Inha University (joint Ph.D. program). He is a member of the Global Research Lab on Big Data Security and is conducting research in the field of information security. In particular, his interests include biometrics, network security, user authentication, and IoT security.

Daehun Nyang received the B.Eng. degree in electronic engineering from the Korea Advanced Institute of Science and Technology in 1994, and the M.S. and Ph.D. degrees in computer science from Yonsei University, South Korea, in 1996 and 2000, respectively. He has been a Senior Member of the Engineering Staff with Electronics and Telecommunications Research Institute, South Korea, from 2000 to 2003. Since 2003, he has been an Associate Professor with the Computer Information Engineering Department, Inha University, South Korea, where he is also the Founding Director of the Information Security Research Laboratory. His research interests include cryptography and network security, privacy, usable security, biometrics and their applications to authentication, and public key cryptography. He is a member of the board of directors and an editorial board of Korean Institute of Information Security and Cryptology.

Aziz Mohaisen (SM'15) received the M.Sc. and Ph.D. degrees from the University of Minnesota in 2012. He was an Assistant Professor with SUNY Buffalo from 2015 to 2017, and a Senior Research Scientist with Verisign Labs from 2012 to 2015. In 2017, he joined the University of Central Florida, where he is currently an Associate Professor and directs the Security and Analytics Lab. His research interests are in the areas of networked systems and their security, online privacy, and measurements. He is an Associate Editor of the IEEE TRANSACTIONS ON MOBILE COMPUTING. He is a Senior Member of ACM in 2018.