Understanding Evolution and Adoption of Top-Level Domain Names

Thitipong Jarassriwilai, Tiffany Dauber, Nevil Brownlee, and Aniket Mahanti
Department of Computer Science
University of Auckland
Auckland, New Zealand
{tjar408, tdau477}@aucklanduni.ac.nz, {n.brownlee, a.mahanti}@auckland.ac.nz

***Abstract*—The Domain Name System (DNS) is used in the Internet to map Fully Qualified Domain Names to IP addresses. As the Internet continuously grows, there have been challenges in keeping up with the demands on the DNS service. Recently, ICANN announced the introduction of new generic Top-level Domains (gTLDs). Using packet traces collected from a large edge network, we analyze the usage of TLDs between 2008 and 2015. We observe changes in usage of TLDs, and analyze the adoption of the new gTLDs announced by ICANN. We find that while there were no changes in the appearance of most frequently used TLDs, the presence of the new gTLDs in the datasets is growing. The number of different late new gTLDs appearing also doubled or tripled each year, implying that more and more people are starting to use these new gTLDs.**

I. Introduction

The Domain Name System (DNS) is used on the Internet to map Fully Qualified Domain Names (FQDNs) to IP addresses. An IP address is machine-readable and thus allows for easier communication between end systems than a human-readable address, or an FQDN, would. However, to make the Internet easier for humans to use, it is necessary to have mnemonic human-readable addresses. To use both these types of addresses, a mapping scheme is needed. DNS consists of a distributed database in a hierarchy of DNS servers that store mappings between IP addresses and FQDNs, and this distributed database is queried whenever a host needs to retrieve the IP address for an FQDN.

Originally, the solution was a simple text file, called `HOSTS.txt`, which contained the entire mapping and was stored locally on each host. However, as the Internet began to grow in size, it became obvious that this file would grow indefinitely. Thus, it became no longer feasible to store the entire mapping locally at each host. Additionally, it was difficult to administer new names and addresses because the original HOSTS.txt file needed to be updated for each change, and distributed again.

DNS took the place of `HOSTS.txt` to fix these problems. Rather than copying the mapping throughout the Internet, DNS consists of a distributed database and an application-layer protocol that allows hosts to query this database. DNS has a distributed and hierarchical structure to it. It would be unreasonable to expect to store the entire mapping in one instance. Thus, the mapping is distributed among several DNS servers, and parts of the mapping are also stored in local DNS caches throughout the Internet.

Initially, there were seven generic Top-level Domains (gTLDs): `.com`, `.edu`, `.gov`, `.int`, `.mil`, `.net`, and `.org` [1]. Between 2000 and 2004, Internet Corporation for Assigned Names and Numbers (ICANN) released several new gTLDs. These were `.aero`, `.biz`, `.coop`, `.info`, `.museum`, `.name`, `.pro`, `.asia`, `.cat`, `.jobs`, `.mobi`, `.post`, `.tel`, `.xxx`, and `.travel`. Then, in 2010, ICANN began accepting applications for the allocation of new arbitrary gTLDs[1].

In this paper, we use packet traces to perform longitudinal analysis of TLD usage in a large edge network. When a client application issues a query for a FQDN to a resolver, the resolver searches through the DNS hierarchy to find its IP address. If the IP address is not found in the local DNS servers, the query is forwarded to the name servers. The DNS server then sends back its response along with a response code [2]. We observe these responses in packet trace data and use them to analyze what domain names were queried and found by DNS. In this way, we can analyze the extent to which the different types of domain names are being used.

We looked at the usage of both TLDs and second-level domains. We also observed the appearance of unique new gTLDs in our data following the launch of the new gTLD program by ICANN. Furthermore, we ranked the most frequently and least frequently used TLDs. We find that while there were no changes in the appearance of most frequently used TLDs, the presence of the new gTLDs in the datasets is growing. The number of different late new gTLDs appearing also doubled or tripled each year. This implies that more and more people are starting to use these new gTLDs.

Exploring the usage of TLDs especially after the introduction of ICANN's new gTLD program, would benefit both researchers and practitioners. It is important to study the usage of TLDs, their popularity, and adoption rates of newly introduced gTLDs. The results can also assist businesses to decide on choosing an appropriate TLD for their websites as well as deciding on whether investment in a new gTLD is needed.

The rest of the paper is organized as follows. Related work is presented in Section II. Section III describes the objective

[1]http://newgtlds.icann.org/en/about/program

978-1-4673-6773-8/15/$31.00 ©2015 IEEE

of this research. Section IV explains the terminologies used in this paper. Section V describes our data collection and analysis methodology. Results are presented in Section VI. Section VII concludes the paper.

II. Related Work

The launch of the new gTLD has expanded the variety of top-level domains that are being used in global Internet traffic. There have been many efforts in monitoring and analyzing DNS traffic. Past research has primarily focused on understanding technical behaviour and properties of DNS requests and responses, DNS security, and usage of TLDs. We discuss these works next.

A. DNS Traffic

Much research has been done on understanding characteristics of DNS traffic. Callahan *et al.* [3] studied modern DNS behaviour as observed from client vantage point within a small residential network. Brownlee *et al.* [4] studied the performance of one of the root DNS servers. They analyzed the query rates to the server, the errors occurring at the root server, attacks on the server, as well as Microsoft's DNS disaster that happened in January 2001.

Another work by Brownlee *et al.* [5] analyzed the performance of a root server from a network client's viewpoint. They measured the DNS response time, request loss rate, and load on the root server. Xuebiao *et al.* [6] studied the characteristics of traffic on the `.cn` TLD server. They studied traffic distributions of query types, traffic load for each server, and geographical distribution of requests.

Pang *et al.* [7] quantified the degree of responsiveness that can be expected from DNS. They performed several measurements from large content providers and DNS servers to understand DNS-based controls. Jung *et al.* [8] also evaluated the DNS performances by analyzing packet traces. They also performed trace-driven simulations to study the impact of varying Time To Lives (TTLs) and varying degrees of cache sharing on DNS cache hit rates.

B. DNS Security

Literature on DNS research has also covered security issues. The following works proposed mechanisms to detect security threats.

Zdrnja *et al.* [9] analyzed DNS responses to detect unusual behaviour. They proposed a passive DNS anomaly detection scheme based on data captured from a university's Internet gateway. Choi and Lee [10] also observed DNS traffic to identify security threats. They proposed a lightweight mechanism called BotGAD, to detect botnets using their fundamental characteristics such as group activity. They evaluated BotGAD using DNS traces collected from different sources. Ruan *et al.* [11] used DNS traffic to observe anomalous patterns in an effort to detect security threats. They proposed a periodic trend mining method and a traffic prediction method. These methods are used to detect abnormalities in DNS query traffic patterns as a prelude to security breaches.

Other studies have focused on devising solutions to attacks on DNS. Johns *et al.* [12] presented a DNS rebinding attack method and analyzed the holes that allow this kind of attack. They also proposed a solution to this vulnerability. Jackson *et al.* [13] studied rebinding attacks and designed a patch to prevent against these attacks.

C. Top-Level Domain

Limited work has been done on studying TLDs, especially the newly released TLDs. Huang and Zhao [14] analyzed the focus technologies and their developments of each country using the keywords observed in country code TLDs. Solomonides [15] and Rabbi [16] have focused on a specific TLDs. Solomonides [15] analyzed the use of the new gTLD in health sector and evaluated the control of information within the domain as well as other possibilities of the use on the new TLD. Similarly, Rabbi [16] performed a case study and analyzed the use of new gTLD for Muslims, including the management of the proposed TLD.

While previous works have focused on DNS traffic characteristics, DNS security, and usage of specific TLDs, there is no study that analyzes the use of TLDs, especially with the release of the new gTLD program. Our work provides detailed longitudinal analysis of usage of TLDs with a perspective from a large edge network.

III. Objective

Although other research has been done on DNS, there has not been any research into how TLDs are being used in the Internet. TLDs are an important component of DNS. It is important to look into this aspect of DNS because without knowledge of how a technology is being used, it is difficult to know the best ways to maintain and improve it.

Our research is meant to be a starting point, to reveal other potential areas of interest in relation to how TLDs are being used and to highlight any changes in its use due to the allocation of new arbitrary gTLDs. We look at DNS response records seen at the University of Auckland between 2008 and 2015. We specifically look for any changes in the usage of gTLDs and the introduction of new arbitrary gTLDs.

IV. Terminology

We use the following terminologies in the paper:

- The original generic Top-Level Domains are referred to as the generic Top-level Domains (gTLDs). They are the ones that were introduced before 1998. These include `.com`, `.edu`, `.gov`, `.int`, `.mil`, `.net`, and `.org`.
- The earlier round of generic top-level domains is referred to as early new gTLD or gTLD (early). These are the gTLDs that were proposed in 2000 and 2004.
- The later round of gTLDs is referred to as the late new gTLDs or gTLDs (late). The gTLDs in this group have been proposed after 2004.

TABLE I
TRACE OVERVIEW

TLD	2008		2010		2011		2013		2014		2015	
	Query	*%*	*Query*	*%*	*Query*	*%*	*Query*	*%*	*Query*	*%*	*Query*	*%*
Original gTLD	17,892	52.63	40,732	49.92	857,062	35.05	7,442,858	73.05	5,116,158	72.87	4,556,023	66.05
New gTLD (early)	68	0.20	368	0.45	9,074	0.37	21,777	0.21	34,676	0.49	21,419	0.31
New gTLD (late)	-	0.00	2	0.00	42	0.00	40	0.00	362	0.01	3,204	0.05
Country-Code TLD	11,326	33.31	29,073	35.63	994,456	40.67	2,285,501	22.43	1,490,109	21.22	1,831,674	26.55
Infrastructure TLD	4,712	13.86	11,418	13.99	584,330	23.90	438,064	4.30	379,270	5.40	485,380	7.04
Total queries	33,998	100.00	81,593	100.00	2,444,964	100.00	10,188,240	100.00	7,020,575	100.00	6,897,700	100.00
Dataset size (MB)	1.8		16.1		124.1		1,888.30		433.3		464.1	
Date Collected	Tue, 1 Apr		Tue, 29 Jun		Tue, 23 Jun		Thu, 28 Nov		Fri, 4 Jul		Mon, 16 Feb	
Time collected	N/A		11AM		N/A		11AM - 2PM		1PM - 3PM		1PM - 5PM	

- The country-code top-level domains are referred to as ccTLDs. These domains correspond to country root domains. All of them are represented using a two-letter code (e.g., `.ca`, `.jp`, `.au`, and `.ch`).
- The second-level domains are referred as 2LD in this paper. They are directly below the ccTLD (e.g., `.co.uk`, `.com.au`, and `.ac.nz`).
- Hot set is the set of top-ten TLDs based on number of queries. The top TLDs are ranked by counting the number of queries.
- `.arpa` is referred to as infrastructure TLD. The purpose of this TLD is for reverse domain name resolution (lookup) for both IPv4 and IPv6.

V. METHODOLOGY

We use packet traces collected at the border gateway of University of Auckland network. The university has over 40,000 students and staff. Note that this data only represents the DNS usage at the University of Auckland, and is not necessarily reflective of the global Internet. However, it is one sample, and the methods could be repeated on a larger scale to get an understanding of the usage of DNS in the Internet.

We use two Python tools, namely, `python-libtrace`[2] and `pldns`[3] for our work. `Python-libtrace` is a tool to allow the use of `libtrace` with Python. `Libtrace` is written as a C library, and `python-libtrace` translates these tools into a manner that is more in the python style of programming, so that the use of the tool in python does not become overly complex. This tool is useful to analyze packets. We use `python-libtrace` to sift through the packets in our data and find packets that have proper IP and UDP headers. We then extract the UDP payload from these packets.

`Pldns` is similar to `python-libtrace` in that it is also a translation of a C library into a python programming style. `Pldns` is a tool used to analyze DNS records. We use pldns to extract UDP payloads from the packets containing DNS records. `Pldns` also helps us to check the response code in the record to ensure that there were no errors and DNS was able to find the queried name. Lastly, using `pldns` we can extract the domain name being queried.

Once we have the name, we are able to parse it and obtain the TLD. We maintain a list of all the TLDs seen in the data, with the number of appearances of each one. In the case of a ccTLD, we also maintain a list of the second-level domain names seen, with the number of appearances of them. We have obtained a list of all the new gTLDs from ICANN, and we check each TLD seen against this list. We then separate the TLDs by whether or not they are in the list from ICANN, and we are left with a list of ICANN TLDs, a list of non-ICANN TLDs, and a list of second-level domain names and ccTLDs.

Our program runs through each year individually, producing a separate set of results for each year. For each packet in the trace, the program checks if it has IP header and is a UDP datagram. If it is an IP packet and is a datagram, the payload will then be processed. We check for DNS records and look for those with response code = 0. If the packet meets all the criteria then we retrieve the FQDN and subsequently extract the TLD.

This way, we turn our packet trace data into a set of usable lists and statistics. We are able to analyze the ranked lists, and compute the percentage of requests for each name in the list. We are able to determine if any TLDs seem to be climbing the ranks, or if their ranks are dropping from year to year.

The DNS queries were collected at the University of Auckland's border gateway in 2008, 2010, 2011, 2013, 2014, and 2015. Table I presents an overview of the datasets.

VI. RESULTS

Following the analysis of datasets collected in 2008, 2010, 2011, 2013, 2014, and 2015, we have classified each query from the datasets into categories according to its top-level domain type. We categorized the types of top-level domains into these following: the original generic top-level domain, the infrastructure top-level domain, the country-code top-level domain, the earlier rounds of new generic top-level domain, and the later round of the new generic top-level domains. Table I summarizes the total number of queries for each category with regard to the datasets and the percentage to illustrate the fraction it is contributing to each dataset.

[2]https://www.cs.auckland.ac.nz/~nevil/python-libtrace
[3]https://www.cs.auckland.ac.nz/~nevil/python-libtrace/pldns.html

We ranked the most frequently used TLDs and least frequently used TLDs to see if there are changes across the datasets. We then analyzed each of the categories to observe if there are any changes in the traffic behaviour since 2008. We also observed the use of second-level domains used within `.nz` domain.

In the following sections, we will first highlight the rankings from each dataset. After that, we will present the results for each top-level domain type with the focus on the changes in new gTLD. The results of changes in the use of second-level domain will be presented lastly.

A. Top-Level Domain Rankings

We ranked the top-ten most frequently used TLDs and the bottom-five least frequently used TLDs from each of the six datasets with an attempt to see the changes in the use of TLDs. Table II summarizes the top-ten rankings.

TABLE II
TOP-TEN MOST FREQUENT USED TLDS

Rank	*2008*	*2010*	*2011*	*2013*	*2014*	*2015*
1	com (o)	com (o)	nz (cc)	com (o)	net (o)	com (o)
2	nz (cc)	nz (cc)	arpa (i)	net (o)	com (o)	net (o)
3	arpa (i)	arpa (i)	com (o)	nz (cc)	nz (cc)	nz (cc)
4	org (o)	net (o)	org (o)	arpa (i)	arpa (i)	arpa (i)
5	net (o)	org (o)	net (o)	org (o)	org (o)	org (o)
6	de (cc)	ru (cc)	uk (cc)	cn (cc)	cn (cc)	cn (cc)
7	edu (o)	au (cc)	de (cc)	au (cc)	au (cc)	uk (cc)
8	uk (cc)	edu (o)	ru (cc)	ru (cc)	uk (cc)	au (cc)
9	it (cc)	cn (cc)	au (cc)	uk (cc)	edu (o)	de (cc)
10	pl (cc)	jp (cc)	cn (cc)	edu (o)	info (e)	edu (o)

(o): original TLD, (cc): country-code TLD,
(i): infrastructure TLD, (e): early new gTLD

From the table, we find that almost the same set of TLDs is present in every dataset, though their ordering is not the same. TLDs that appear across all datasets include: `.com` (original gTLD), `.nz` (ccTLD), `.arpa` (infrastructure TLD), `.org` (original gTLD), and `.net` (original gTLD). The rest are mostly country-code TLDs. Typical ccTLDs include `.nz`, `.uk`, `.au`, and `.cn`. Please note that there is one early gTLD in our 2014 dataset, which is `.info`. We will refer to the top-ten list in each year as the hot set.

Our finding is consistent with the ranking made by `pcnames.com`[4]. They listed the most common gTLDs and the original purpose of those domains. All of our gTLDs that are listed in the top-ten have been mentioned in their list. We also expected the usage of `.nz` to be very high since the data was collected at a New Zealand university network. The usage of `.au` is also high as Australia is a neighbour country and many Australian businesses also operate in New Zealand. The `.cn` domain is also largely used in the university's network because there are many international students from China. The `.arpa` infrastructure TLD appears in the top-ten list across our datasets. This TLD has not been widely discussed since it is used exclusively for technical infrastructure purposes.

We also ranked the five least frequently used TLDs from our datasets as it is interesting to see if the appearance of least used TLDs has changed. Table III summarizes the results from our data.

TABLE III
BOTTOM-FIVE LEAST FREQUENTLY USED TLDS

Rank	*2008*	*2010*	*2011*	*2013*	*2014*	*2015*
1	cc (cc)	cc (cc)	xyz (n)	ki (cc)	help (n)	network (n)
2	fm (cc)	gd (cc)	jm (cc)	google (n)	an (cc)	directory (n)
3	name (n)	computer (n)	lr (cc)	properties (n)	day (n)	cash (n)
4	ws (cc)	mx (cc)	ng (cc)	foo (n)	az (cc)	zip (n)
5	vn (cc)	eg (cc)	bi (cc)	today (n)	engineering (n)	tools (n)

(n): new gTLD, (cc): country-code TLD

The results from our analysis show that prior to 2013, the least used TLDs were mostly ccTLDs. After 2013, the appearance of the new gTLD (late) started to increase and fill up the ranking. This is implying that the use of the new gTLDs starting to grow accordingly to the launch of the new gTLD program.

The rankings in the least frequently used list are what we expected. Since the launch of the new gTLD program, it is very likely that some of the new gTLDs (late) will appear in the ranking. This has raised another question on the number of the new gTLDs appearing in the datasets and the usage of these new gTLDs. These questions will be addressed in the following sections.

B. Hot Set Behaviour

We observe the changes in the hot set to see if the change is significant. Figure 1 illustrates the percentage of changes in the hot set. The change from year 2008 to 2010 is significant at 40%. It may be due to statistical variations since the size difference is almost ten times. The changes in 2011 to 2015 are fairly constant with 10% to 20% change for each year. The majority of the changes in hot set are ccTLDs. A new gTLD (`.info`) also appeared in the hot set in 2014.

Figure 2 shows the change in queries in relation to the change in TLDs. The changes in queries across our datasets are not substantial. Both Figure 1 and Figure 2 are consistent, implying that the change in TLD count in hot set is corresponding to the change in queries in the dataset. The largest change is between 2008 and 2010 at 6.7% of the queries from the top-ten list. This indicates the TLDs that are changing are not the major TLDs. The top-five alone contribute around 90% of all the queries. These TLDs have never disappeared from the top-ten list. Overall, there is not much change in the hot set across our dataset and most of the changes on the list are ccTLDs.

C. Unique Top-Level Domains

The number of unique TLDs appearing in our datasets for each year is presented in Figure 3. We notice that there is an increasing trend for the number of unique TLDs for both gTLDs

[4]http://www.pcnames.com/articles/common-tlds-and-their-uses

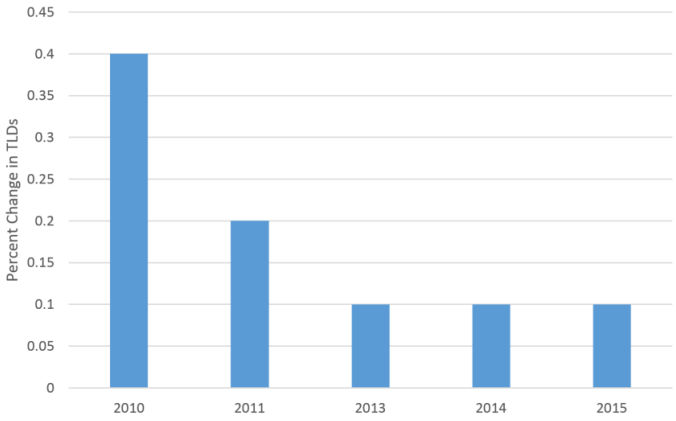

Fig. 1. Change in TLD count in the hot set

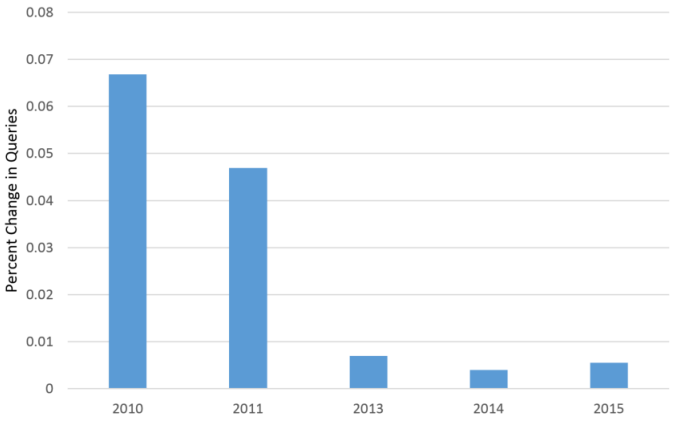

Fig. 2. Change in query count in the hot set

and ccTLDs. The ccTLD vertical bars show that there is a rapid increase from 2010 and 2013. This could be because of the size of the datasets. As the 2011-2015 datasets captured a lot more packets than the 2008 and 2010 datasets, it is expected that more ccTLDs will be identified. The gTLD graph shows that the number is increasing constantly. However, the rate started to grow faster from 2013. This is likely due to the introduction of the new gTLD program.

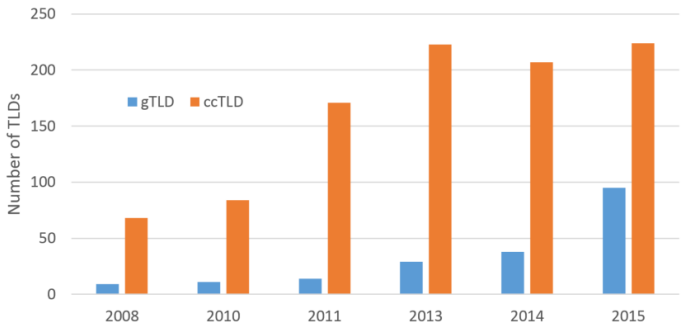

Fig. 3. Number of unique gTLDs and ccTLDs

We extract the data to find the unique gTLDs (including original gTLDs, earlier, and later round of new gTLDs) and the country-code TLDs. We then plot a cumulative graph to illustrate the rate of appearance for unique TLDs. Figure 4 shows the cumulative count of both gTLDs and ccTLDs observed in our datasets.

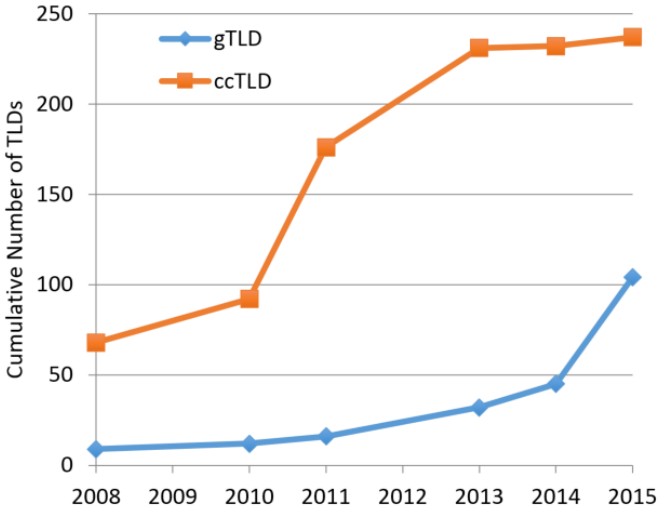

Fig. 4. Cumulative TLDs over the trace period

The ccTLD line plot indicates a constant increase from 2008 to 2013. The growth rate increases in relation to the number of queries in the trace. We see a very sharp increase in 2010 through to 2013. It then slowed down afterward. This indicates that there are a few non-frequently used ccTLDs appearing in our 2013 dataset and onwards. The ccTLDs that appeared between 2013 and 2015 are the following: `.cf` (Central African Republic), `.td` (Chad), `.fk` (Falkland Islands), `.tj` (Tajikistan), `.cw` (Curacao), and `.gq` (Equatorial Guinea). Currently, there are 248 ccTLDs operating, with 237 of them appearing in our datasets.

On the other hand, the gTLD plot shows an increasing exponential trend. The number of unique TLD grows reasonably fast at the start then it increased swiftly in the later years. This is due to the introduction of new gTLD program as there are many unique new gTLD appearing in our datasets from 2011 onwards.

The number of unique TLDs is increasing every year, however, the number for ccTLD will not increase as sharply as the number for gTLD as around 95% of them have already appeared in our datasets. The number of gTLDs is expected to continue to increase following the new gTLD program that offers more than a thousand new gTLDs.

D. New Generic Top-Level Domains

We also considered the number of unique new gTLDs that appeared in our datasets. The summarized results from this analysis are shown in Figure 5.

We will first present and discuss the result of the earlier round of the new gTLD program. The appearance of unique

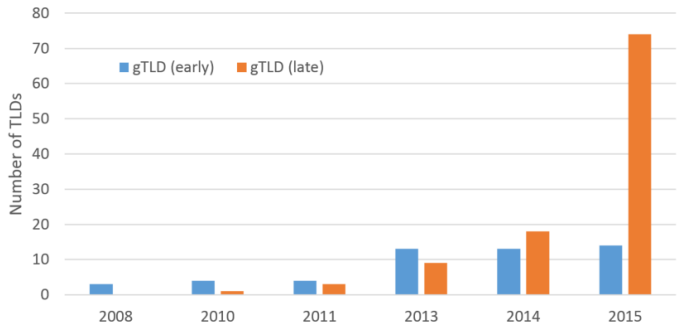

Fig. 5. Number of unique new gTLDs (early and late)

new gTLD (early), as shown in the figure, were steady from 2008 to 2011 with three to four unique early gTLDs appearing. However, the appearance has increased to thirteen in 2013 and 2014, and to fourteen in 2015. One of the reason for this is because some of the earlier round of new gTLDs have only been approved recently, for example, the `.xxx` domain had been approved in April 2011, even though it had been proposed in 2004[5].

The other reason could be because of the increasing popularity of non-original TLDs. The variation in sizes of the sample could also affect the results as our smallest datasets are three hundred times smaller than the largest.

For the later round of the new gTLDs, the appearance of unique TLDs grows very fast across the datasets. The number has doubled or tripled for each year. This reflected the introduction of the later round of the new gTLD program. We expect this number to grow sharply as people are becoming more aware that a TLD could be something else other than the original TLDs. A lot of advertising efforts were used to promote the new gTLD (late). For example, `onlydomains.com`, a company that provides domain name registration services, has offered 5,000 students in Australia and New Zealand to register one of the new gTLD (late), `.xyz`, for free for one year[6].

While the appearance of unique early new gTLDs will not be change much as there are only fifteen of them, the appearance of the late new gTLD will increase as there are more domains registered every day and ICANN expected the number to be potentially more than 1,300 domains[7]. We expect the number of unique domains to grow in the future

E. Changes in New Generic Top-Level Domain Queries

In this section, the result of the change in the new generic top-level domain traffic for both earlier round and later round of gTLDs will be presented and discussed.

The change in the earlier round of new gTLD is relatively insignificant. As presented in Table I, the points are fairly scattered. The range is from 0.20% in 2008 to 0.49% in 2014 dataset. We therefore cannot claim that there is relationship or a trend and hence we cannot conclude that there are changes for this early new gTLD traffic. The traffic behaviour has not been changed. It could be because these TLDs were launched a decade ago and the traffic level for this category has become stable.

For the later round of the new gTLDs, the relationship is moderate. The level of traffic is ranging from 0% in 2008 and 0.05% in 2015. Table I also illustrates the result from our analysis. The trend is slightly positive.

Overall, the usage of the new gTLD, both early and late, is still insignificant. They only contribute less than 1% to the dataset. The level of this traffic may grow in the future but it will still take time to reach the level of those original gTLDs.

F. Changes in Other Top-Level Domain Queries

We also observed the query behaviour of other types of top-level domains, the generic top-level domain, the country code top-level domain, and the infrastructure top-level domain.

Before observing the changes, it is important to see the overall queries of TLDs based on their rankings. Figure 6 is an example of the number of queries received per unique TLD in 2015. It shows how the top few TLDs are responsible for the majority of the queries captured in our datasets, exemplifying the power-law behaviour. The shapes of all other rank plots are similar and consistent across our datasets (Hence, they are excluded in the paper.). We observe that the top-ten TLDs accounted for 90% to 95% of the queries. The highest ranked TLDs are already responsible for around 30% of the queries. The rates dropped more rapidly towards the last few TLD as the number of queries are getting smaller.

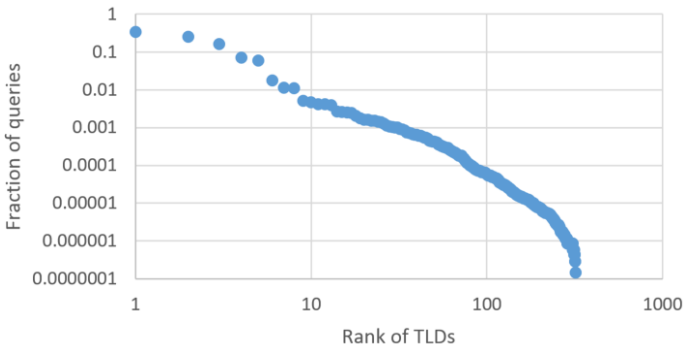

Fig. 6. Log log rank plot of queries to TLDs in 2015

We next discuss the queries for each of the TLD categories.

1) Original gTLDs: The result from our analysis on the queries made for the original gTLDs is shown in Table I. Usage of the original gTLDS reached its lowest value in 2011 at 35.05%, and the highest at 73.05% in 2013. There are some variations in the query count as, again, the datasets were not collected under similar conditions such as the time of data collection. However, the overall trend seems to be increasing.

The usage of this type of TLD is still dominating the traffic. Also, the top-ten websites as per `alexa.com` are 90% original

[5]https://www.icann.org/resources/board-material/resolutions-2011-03-18-en

[6]http://www.onlydomains.com/promotion/xyz/XYZ_Flyer.pdf

[7]http://newgtlds.icann.org/en/about/program/materials/fast-facts-28feb14-en.pdf

gTLDs[8]. We presume that people are more comfortable and inclined to use this type of gTLD because they are familiar with them and hence new sites may still register their domain under the original gTLDs. The usage in our dataset is growing slightly and we expect the original gTLDs to still dominate in the future.

2) Country-Code TLDs: We also observed the changes in traffic of the ccTLDs. The trend seems to be downwards. The query level started at 33.31%, it reached its peak in 2011 at 40.67%, and then reduced to 21.23% in 2014. Note that our ccTLD query results show geographical bias because the data was collected in a New Zealand university network, which is mostly used by students. The traffic of `.nz` domain will be particularly high as students were using the university's intranet systems or visiting the local Web sites. Nonetheless, if we only look at the traffic behaviour from inside the university, we notice that the traffic behaviour is changing. The use of ccTLDs is shrinking slightly. We believe that it is because providers tend to use the original gTLDs instead of ccTLDs.

In general, the ccTLD traffic is reasonably high, with 20% to 40% of the queries were for ccTLD, almost at par with the original gTLD. The use of ccTLD will still be dominant as they are used by regional education institutes, government departments, local companies, as well as other organizations.

3) Infrastructure TLDs: Lastly, the infrastructure top-level domain was observed and analyzed. The infrastructure TLD only consists of `.arpa` domain. It is used for reverse DNS lookup and for verifying email senders, among other things. This type of TLD contributed about 14% queries in 2008, and was reduced to 7% in 2015. The usage of this type of TLD is relatively low.

G. Changes in Second-Level NZ Domain Queries

In addition to the observation of TLDs, we also studied the usage of 2LD. We only focus on five 2LDs for which the query count increased since 2008 or 2010. The five second-level domains we studied are the following: `.geek.nz`, `.gen.nz`, `.health.nz`, `.iwi.nz`, and `.kiwi.nz`. Even though the usage of these domains is small (contributing less than 1% of the queries), we see some changes in uptake of queries for these domains. Figure 7 shows the percentage queries for these five 2LDs.

The `.health.nz` domain seems to grow the fastest according to our data. This domain had just been launched in 2010 and it is growing significantly since. The domain `.kiwi.nz` is also new, being introduced in 2013. The usage for `.kiwi.nz` is not as popular as `.health.nz` but the number is still growing. The other three domains have been introduced prior to 2008. Although they were launched before 2008, we see the growth starting in 2010 or 2013. This could be because more people are aware that they do not have to use the common gTLDs like `.co.nz` or `.org.nz`. The usage of these new or less frequently used 2LDs allows the organizations or individuals to personalize their websites according to the services they provide.

Overall, we see the changes in the use of these new and previously non-frequently used 2LDs. However, the level of traffic of these 2LDs is very small with less than 1% contribution to the total queries.

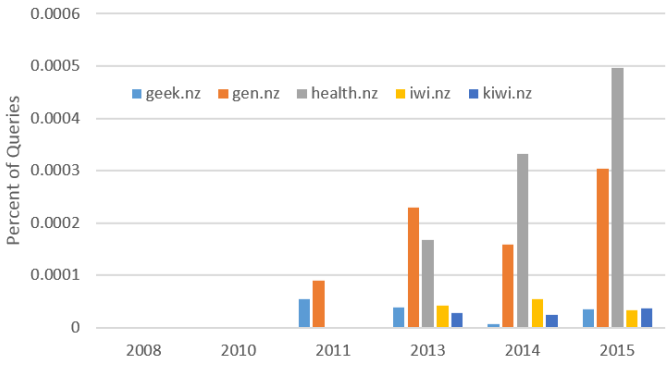

Fig. 7. Fraction of queries for `.nz` 2LDs

VII. Conclusions

We analyzed DNS responses using packet traces collected at the University of Auckland between 2008 and 2015. We looked at the usage of both top-level domains and second-level domains. We also observed the appearance of unique new generic top-level domains in our data following the launch of the new gTLD program by ICANN. Furthermore, we ranked the most frequently and least frequently used top-level domains. All of these were done in an effort to observe the changes in the DNS traffic behaviour and usage of TLDs and 2LDs.

We find that while there were no significant changes in the most frequently used TLDs, the presence of the new gTLDs in the datasets is growing. The number of different late new gTLDs appearing also doubled or tripled each year. This implies that more people are starting to use these new gTLDs.

The usage of other types of top-level domains has changed. We see an increasing trend for the gTLDs and decreasing trend for ccTLDs and infrastructure TLD. We observed increase in queries made for second level `.nz` domains, however, they are still infrequently used.

References

[1] P. Mockapetris, "Domain names - concepts and facilities," Internet Engineering Task Force, RFC 1034, November 1987. [Online]. Available: http://www.rfc-editor.org/rfc/rfc1034.txt

[2] ——, *RFC 1035 Domain Names - Implementation and Specification*, Internet Engineering Task Force, November 1987. [Online]. Available: http://tools.ietf.org/html/rfc1035

[3] T. Callahan, M. Allman, and M. Rabinovich, "On modern dns behavior and properties," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 3, pp. 7–15, Jul. 2013. [Online]. Available: http://doi.acm.org/10.1145/2500098.2500100

[4] N. Brownlee, K. Claffy, and E. Nemeth, "Dns measurements at a root server," in *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, vol. 3, 2001, pp. 1672–1676 vol.3.

[5] N. Brownlee, k. claffy, and E. Nemeth, "DNS Root/gTLD Performance Measurements," in *Usenix LISA*. San Diego, CA: Usenix, Dec 2001.

[8]http://www.alexa.com/topsites

[6] Y. Xuebiao, W. Xin, L. Xiaodong, and Y. Baoping, "Dns measurements at the .cn tld servers," in *Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference on*, vol. 7, Aug 2009, pp. 540–545.

[7] J. Pang, A. Akella, A. Shaikh, B. Krishnamurthy, and S. Seshan, "On the responsiveness of dns-based network control," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '04. New York, NY, USA: ACM, 2004, pp. 21–26. [Online]. Available: http://doi.acm.org/10.1145/1028788.1028792

[8] J. Jung, E. Sit, H. Balakrishnan, and R. Morris, "Dns performance and the effectiveness of caching," *IEEE/ACM Trans. Netw.*, vol. 10, no. 5, pp. 589–603, Oct. 2002. [Online]. Available: http://dx.doi.org/10.1109/TNET.2002.803905

[9] B. Zdrnja, N. Brownlee, and D. Wessels, "Passive monitoring of dns anomalies," in *Detection of Intrusions and Malware, and Vulnerability Assessment*, ser. Lecture Notes in Computer Science, B. M. Hmmerli and R. Sommer, Eds. Springer Berlin Heidelberg, 2007, vol. 4579, pp. 129–139. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-73614-1_8

[10] H. Choi and H. Lee, "Identifying botnets by capturing group activities in {DNS} traffic," *Computer Networks*, vol. 56, no. 1, pp. 20 – 33, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128611002787

[11] W. Ruan, Y. Liu, and R. Zhao, "Pattern discovery in {DNS} query traffic," *Procedia Computer Science*, vol. 17, pp. 80 – 87, 2013, first International Conference on Information Technology and Quantitative Management. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050913001452

[12] M. Johns, S. Lekies, and B. Stock, "Eradicating dns rebinding with the extended same-origin policy," in *Proceedings of the 22Nd USENIX Conference on Security*, ser. SEC'13. Berkeley, CA, USA: USENIX Association, 2013, pp. 621–636. [Online]. Available: http://dl.acm.org/citation.cfm?id=2534766.2534820

[13] C. Jackson, A. Barth, A. Bortz, W. Shao, and D. Boneh, "Protecting browsers from dns rebinding attacks," in *Proceedings of the 14th ACM Conference on Computer and Communications Security*, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 421–431. [Online]. Available: http://doi.acm.org/10.1145/1315245.1315298

[14] L. Huang and P. Zhao, "Empirical study on focus technology based on top-level domain," in *Information Science and Engineering (ICISE), 2010 2nd International Conference on*, Dec 2010, pp. 2042–2045.

[15] A. Solomonides, "Icann, health information and the "dot health" top level domain," in *Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on*, May 2014, pp. 460–462.

[16] R. Rabbi, ".umma - cyber solace in the digital age: Faith-based top level domain extension for a global muslim union," in *Information and Communication Technology for the Muslim World (ICT4M), 2013 5th International Conference on*, March 2013, pp. 1–5.