Measuring the Leakage of Onion at the Root

“A measurement of Tor’s .onion pseudo-TLD in the global domain name system”

Matthew Thomas
Verisign Labs, VA, USA
mthomas@verisign.com

Aziz Mohaisen
Verisign Labs, VA, USA
amohaisen@verisign.com

ABSTRACT

The Tor project provides individuals with a mechanism of communicating anonymously on the Internet. Furthermore, Tor is capable of providing anonymity to servers, which are configured to receive inbound connections only through Tor—more commonly called hidden services. In order to route requests to these hidden services, a namespace is used to identify the resolution requests to such services. A namespace under a non-delegated (pseudo) top-level-domain (TLD) of *.onion* was elected. Although the Tor system was designed to prevent .onion requests from leaking into the global DNS resolution process, numerous requests are still observed in the global DNS. In this paper we will present the state of .onion requests received at the global public DNS A and J root nodes over a longitudinal period of time, a synthesis of Day In The Life of the Internet (DITL) data repository, and potential explanations of the leakage, and highlights of trends associated with global censorship events. By sharing this preliminary work, we wish to trigger further discussions on the matter in the community.

Keywords

DNS; privacy; security; Tor

1. INTRODUCTION

The Domain Name System (DNS) has become a critical and reliable component of the Internet, allowing individuals to quickly match domain names with their corresponding IP-addresses. The DNS is a hierarchical system, in which at the top of the hierarchy is the root domain. Currently, the root consists of a combination of 13 groups of DNS servers located globally around the world. Each of those servers is named in the form X.root-servers.net, where X is a character in the range of A through M. These roots are responsible for the delegation of top-level-domains (TLDs) such as .com [19].

It is well known within the Internet research and engineering community that many installed systems on the Internet query the DNS root for a wide range of TLDs that are not delegated and will ultimately result in an error, or more commonly referred to as a NXDomain [21]. Many of these installed systems depend explicitly or implicitly on the indication from the global DNS that the domain name does not exist. For instance, many internal networks use a domain name suffix that is not currently delegated in the global DNS, such as .corp .home [13]. Due to the recent delegation of new gTLDs within the global DNS [4], several studies have measured the amount of internal name space leakage to the DNS roots [3, 17]. These unintended leaked DNS queries have been shown to expose sensitive private information and present potential new security threat vectors [3, 17, 22]. During the analysis of potential colliding name spaces within the global DNS, queries suffixed in .onion appeared to be one of the more prevalent non-delegated TLDs at the global root DNS.

Tor is an example of a system that exploits the absence of a non-delegated namespace within the global DNS system for its internal use. Hidden services, a unique feature within Tor, provide additional anonymity for users to communicate with servers. To identify these services, Tor uses the .onion name space to identify such requests [26]. While the Tor system was designed to not route requests suffixed in .onion, there exists a clear conflict of interests between internal namespace routing and the global DNS namespace when .onion URLs are shared and or requested [15]. In fact, DNS leakage is a known and well-documented issue within the Tor community. Many tutorials on the Tor website have been published giving users instructions to mitigate the leakage through the use of proxies, disabling DNS pre-fetching within the browser or even installing a local DNS server which rejects .onion addresses [14]. However, non-technical Tor users likely do not practice these mitigation steps due to their complicated nature.

The leakage of .onion requests to the global DNS roots clearly presents some risk to Tor users and also has privacy implications that need to be explored. To this end, in this paper we present a first look at the .onion leakage at the DNS root. We use two root servers, A and J, that are operated by Verisign, and explore .onion resolutions seen at both of them over a period of time close to six months. We complement this measurement with a dataset from all roots over 7 years, with a sample of two days per year. Our findings from both datasets highlight the following:

- A large amount of .onion traffic is observed at both the A and J root servers (as well as other servers in the complementary study) and the requests originate from a diverse set of locations (at the recursive name server level).
- Surges in the amount of .onion traffic and leakage into the public DNS coincide with major global, geopolitical and censorship-related events.
- The.onion’s traffic exhibits a heavy tailed distribution (with respect to the number of queries per .onion), and a very interesting weekly traffic pattern.

WPES’14, November 3, 2014, Scottsdale, Arizona, USA.
Copyright 2014 ACM 978-1-4503-3148-7/14/11 ...$15.00.
http://dx.doi.org/10.1145/2665943.2665951.

- While the exact root causes are not easy to verify, we highlight various potential causes and scenarios of the leakage and call for further investigation into the leakage potential implication on users privacy.

The organization of the rest of this paper is as follows. In section 2 we introduce the DNS profile of the .onion data collected. In section 3, we examine longitudinal patterns of .onion traffic to the A and J root servers operated by Verisign from various network and second-level-domain (SLD) points-of-view, and highlight correlations between global events and increased .onion traffic volumes. In section 4, we explore potential reasons .onion traffic is being leaked to the roots and highlight considerations within the Internet engineering community to address the use of non-delegated TLDs. Finally, in section 5 we will present our conclusions and discuss future directions in which we will further explore the .onion leakage.

2. DATA SET

In this paper we use two datasets. The first dataset is from the resolution at the A and J root servers operated by Verisign, while the second dataset is the "Day In The Life of the Internet" (DITL) managed by the Domain Name System Operations Analysis and Research Center (DNS-OARC). In the following we elaborate on those datasets and their nature.

2.1 Roots A and J Data Set

Verisign operates the A and J root servers in the DNS root zone. NXDomain (NXD) responses for the non-delegated TLD .onion were captured over slightly more than *six months* from both root servers starting on September 10th, 2013 and ending March 31st, 2014. The data set consists of approximately 27.6 million NXD records spanning 81,409 unique SLDs. The DNS requests originated from a wide variety of sources: in total, they are sent from 172,170 IP addresses, 105,772 unique /24 net blocks, and 21,345 distinct Autonomous System Numbers (ASNs).

During the multi-month collection period, numerous NXD TLDs appeared at the roots. Based on the total query volume, we ranked the various TLDs and found that the .onion TLD ranked 461 out of 13.8 billion TLDs. The following section will further depict the traffic patterns and trends observed within the .onion TLD.

Table 1: DITL dataset – root servers characteristics

year	# roots	root servers	# queries
2008	7	(a,c,f,h,k,l,m)	3,710
2009	8	(a,c,e,f,h,k,l,m)	13,343
2010	13	all	2,371,869
2011	11	all except b and g	691,385
2012	10	all except b, d, and g	693,524
2013	11	all except b and g	1,371,650
2014	9	all except b, d, g, and l	1,705,247

2.2 DITL Data Set

The DITL dataset is managed by DNS-OARC, and is a joint effort with CAIDA and ISC. The data captures synchronized and periodic measurements and data collection effort by root name server operators and other organizations (e.g., ISPs). The dataset covers traffic capture of DNS resolution for a period of two days every year. While the dataset captures traffic at the recursive level as seen by various organizations participating in the DITL data collection effort, we only focus on the root traffic. We do that to establish a guideline on how representative the dataset obtained from the A and J root servers is, and to highlight the overall trends of .onion in the DNS over time.

In total, the DITL dataset covers 7 years (from 2008 to 2014), with two days worth of traffic for each year. The dataset captures traffic from all root servers (A through J), however not all root servers are present in all years, as shown in Table 1. For the years of DITL dataset, we found 6,850,728 .onion queries for 18,330 unique .onion SLDs. The various queries are originated from 331,816 IP addresses distributed over 268,616 /24 network addresses.

3. "ONION" DNS CHARACTERISTICS

3.1 Traffic Volume and Diversity Measurements

To better understand the overall traffic pattern, a longitudinal study of query volumes including the total number of requests, number of distinct /24 net blocks and the number of distinct ASNs for a given day was conducted, and the results are represented in Figure 1 for the A and J root data detailed in section 2.1. Overall, we observe that there is a clear upward trend in the total query traffic volume, increasing nearly 300% since the beginning of the collection period. Meanwhile, the diversity of the traffic sources also increased by approximately 20%.

One common characteristic that many DNS researchers and network operators are familiar with is the weekly repeated pattern of the volume of requests, as shown in the ASN and /24 measures in Figure 1. These patterns and trends are clear in ".onion's" /24 and the ASN-level measurements; however, this weekly pattern is surprisingly absent when observing .onion total traffic volume. Many other NXD TLDs at the root have been shown to exhibit a regular weekly query volume pattern [27]. It is unclear to us at this time why .onion does not exhibit this common traffic pattern, and that warrants additional investigation to understand this phenomenon.

The data presented in Figure 1 only represents measurements taken from the A and J root nodes. In order to gauge the total global DNS leakage of ".onion" requests, we can segregate the unique SLDs received at each root node and compare their overlap. This measure will provide us with a SLD root affinity and a simple way of estimating total global DNS leakage if this trend was to be extrapolated over all roots.

Figure 2 depicts the number of unique SLDs observed at the A node, J node, and the combination of A and J nodes. In this figure, we can see that the combined A+J roots, on a daily basis, observe about 3300 unique SLDs; while each of the A and J nodes separately observe roughly 2500 unique SLDs—roughly 75% of the combined A+J root nodes. Prior work studying multi-root distinct SLD overlap [27] has shown that the combined traffic observed at A+J constitutes approximately 40% of all observed distinct SLDs for various TLDs spanning the global DNS roots. The .onion SLD root affinities and overlap between the A and J roots are comparable to the finding in the prior literature concerning other TLDs [27]. Therefore, we postulate that the .onion traffic observed at A+J would continue such a trend and an appropriate sizing of total global .onion leakage could be roughly estimated. Based on the statistics in section 2, we estimate the total number of .onion NXD records at 69 million over the same period of time.

3.2 Hidden Service and SLD Measurements

Figure 2 shows a few days in which the absolute number of distinct SLDs dramatically increases from the average number of daily SLDs observed in the rest of the measurement period. We now turn our attention to the overall distribution of requests for a given SLD within the .onion TLD to better understand the DNS request dy-

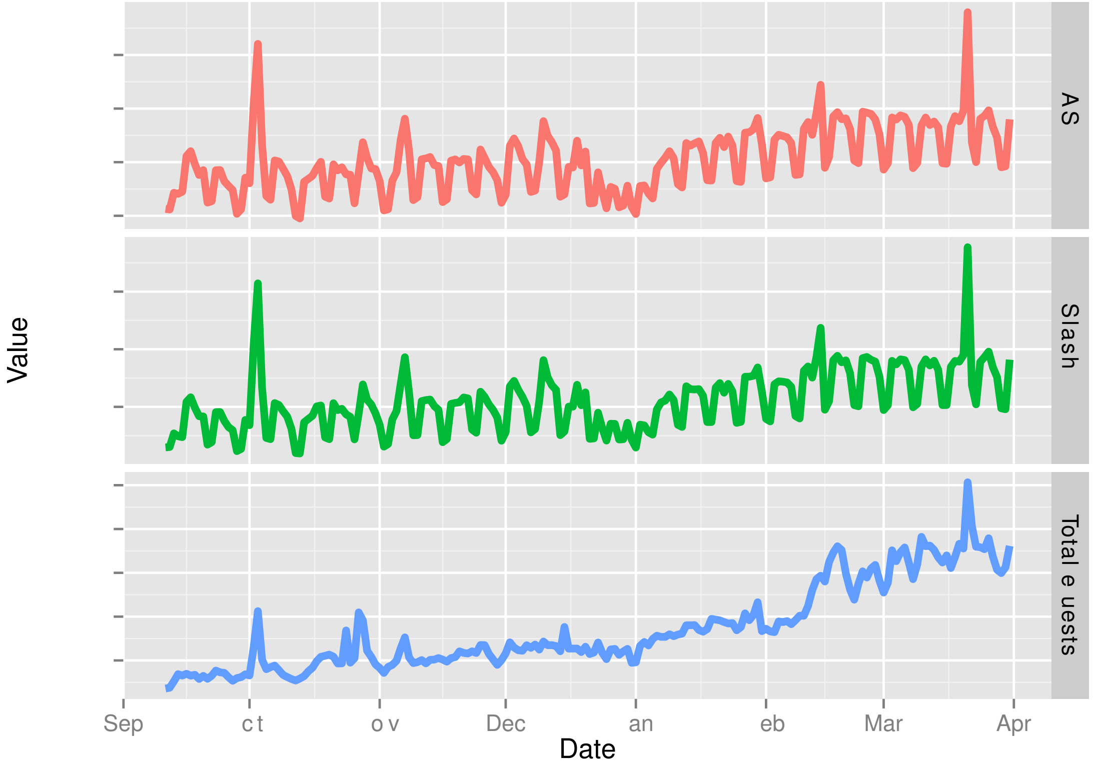


Figure 1: "Onion" Traffic Measurements Observed at A and J Root DNS Nodes

namics of all .onion SLDs. Figure 3 provides three different plots of various traffic diversity measurements, namely the number of total requests, the count of distinct /24 net blocks, and ASNs a distinct SLD received during the collection period. The corresponding cumulative distributions of these measures are reflected in Figure 3.

Clearly, the vast majority of SLDs receive a minimal amount of DNS requests over the six months period covered in our data set, with 50% of the SLDs receiving only one request and nearly 90% of SLDs receiving less than 10 requests. A similar trend of minimal traffic source diversity for the majority of SLDs is displayed, where nearly 95% of the SLDs originate from fewer than 10 distinct ASNs; leaving very few SLDs with large amounts of traffic from a wide variety of network locations. This pattern is in line with the general traffic characteristics and trend for other non-delegated TLDs.

Next we shift our focus to those few but very popular SLDs within the .onion TLD. Table 2 provides a list of the most requested hidden services along with their total percentage of .onion traffic and the type of service provided using them. The mapping of SLDs to their type of service was constructed manually by searching for references of the hidden service online. The SLDs listed in the table have been anonymized (masked) for privacy concerns, where the first and last two characters of each SLD are shown.

From the statistics shown in Table 2, we observe that nearly 27% of all .onion traffic belongs to one hidden service whose focus is on Torrent tracking. The remaining traffic forms a long tailed distribution over the remaining hidden services with an emphasis on services surrounding search, commerce and currency exchange. The top 10 hidden services shown in Table 2 account for more than 38% of the traffic observed over the total period of time of our data set.

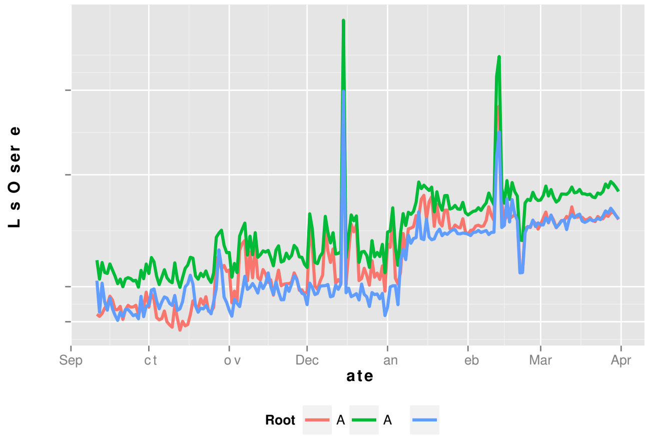


Figure 2: Global DNS Estimation of Onion By Root

3.3 Traffic Source Measurements

In Tables 3, we examine the origination of the .onion DNS requests issued by recursive name servers to the A and J roots from a country perspective[1].

[1] To ensure that publishing those statistics does put the privacy of individual users at risk, we verify that IP allocations for all countries listed herein are large enough.

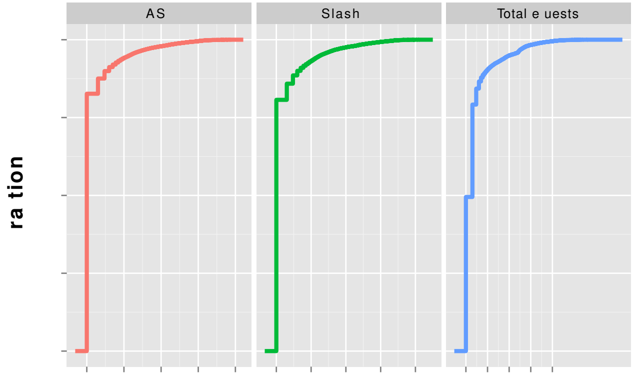


Figure 3: Cumulative Distribution of SLD Traffic Measurements in "Onion"

Table 2: Most Popular SLD Hidden Services and Their Traffic Measurements

Rank	Masked SLD	Type of Service	Traffic (%)
1	Z6------43	Hidden Tracker	26.5
2	DK------II	Silk Road	2.1
3	DP------PC	TorDir	1.7
4	SI------FK	Silk Road	1.4
5	3G------4M	Search Engine	1.3
6	JH------JX	Tor Mail	1.2
7	XM------SL	Search Engine	1.1
8	AG------WW	Agora Marketplace	1.1
9	FO------UI	Bitcoin	0.9
10	TO------NS	TorLinks	0.9

The geographical distribution of .onion requestors deviates from the Top-10 countries by directly connecting users as reported by the Tor project over the same period of time. At nearly 36%, the US is 3 times higher than reported from Tor. Other countries such as Germany, France, and Spain also differed significantly, with 7.7%, 7.23% 6.17% and 4.8% respectively [25]. While clearly leaked .onion queries to the global DNS roots and actual Tor connections are very different (e.g. measuring recursive name servers vs. direct connections), the variance in the distribution of the .onion requests may prove helpful in understanding the root cause of the leaked DNS queries.

With such a large percentage of .onion requests originating in the United States, it is not surprising to observe the major Internet Service Providers (ISP) in Table 4 (AS7922 is Comcast and AS7018 is AT&T)[2]. However, it is interesting to observe that nearly 8% of all .onion traffic originates from AS15169 (Google). We hypothesize that users/advocates of Tor would most likely not use their default ISP name servers and instead would choose to use public DNS providers such as Google Public DNS or OpenDNS (AS36692, which has a share of 2.06%). However, more surprising is to see that many .onion queries originated from AS15169 given that Google Public DNS has an intensive caching policy in use to avoid multiple queries to the root that would potentially result in NXD[3]. Given the nature of .onion queries, and that they are not supposed to be exposed to the DNS infrastructure, a role that such providers may play in addressing the problem can include blocking such requests at the recursive level.[4]

Table 3: Top Geographical Countries and ASNs Requesting "Onion"

Country	Requests	Traffic (%)
US	9878093	35.7
RU	2213691	8.0
DE	1482075	5.3
BR	1258468	4.5
CN	996130	3.6
GB	984059	3.5
KR	980656	3.5
PL	918948	3.3
CA	785184	2.8
FR	670103	2.4
AU	510745	1.8
NL	454441	1.6
ES	448171	1.6
IE	425469	1.5
IT	423550	1.5
AR	387594	1.4
MX	363389	1.3
IN	295122	1.0

3.4 Global Event Correlation

Global events, such as Internet censorship, political reform, and economic shifts, among others, spur the use of privacy enhancing technologies like Tor. The total traffic volume measured on a daily basis in Figure 1 exhibits several spikes in which .onion traffic significantly increases from its moving average. In order to better understand these events, we cross-correlated the spikes with news stories on global events. Table 5 lists the events and their impact on .onion traffic. These events typically manifest themselves in the form of increased traffic from a specific geographical region or the predominance of queries for a particular SLD. Figure 4 plots the events listed in Table 5 against the total daily ".onion" traffic volume, highlighting the spikes in relation with the rest of the traffic volume over the entire period of time observed in our data set.

Certain global events such as the censorship of Internet domains in Turkey may span a longer period of time than a few days. Figure 5 depicts the number of requests for .onion domains originating from Turkey over the multi-month collection period. There is a clear upward trend and a sudden increase in the second half of March 2014 when many DNS-based censorship events took place. The requests originating from Turkey during the censorship spanned hundreds of unique SLDs and were spread over several ASNs.

3.5 Trends from DITL

Now we turn our attention to the DITL dataset in section 2.2. In analyzing this dataset, we concentrate on three aspects:

- **The existence of .onion queries longitudinally:** given that the DITL dataset covers a relatively longer period of time than the dataset we used from the A and J root servers, we

[2] All of the autonomous systems listed in Table 4 have large number of IP addresses allocated to them, thus publishing them does not put the privacy of individual users at risk.

[3] http://bit.ly/UAlAMR

[4] This, however, would not prevent such recursive servers (on any proxies between them and users) from profiling users.

Table 4: Top Geographical Countries and ASNs Requesting "Onion"

ASN	Requests	Traffic (%)
AS15169	2267250	8.2
AS7922	1222955	4.4
AS7018	654680	2.3
AS36692	571609	2.0
AS30607	561349	2.0
AS4766	560739	2.0
AS701	512989	1.8
AS7132	447528	1.6
AS22773	400657	1.4
AS6830	392233	1.4
AS20115	342716	1.2
AS3786	326885	1.1
AS28573	309751	1.1
AS5617	290577	1.0
AS3356	290160	1.0
AS7738	284726	1.0
AS22773	273845	0.9
AS4134	258832	0.9

Table 5: Global Events and Elevated "Onion" Request Correlation

Event	Date	Req.	Event
A	10/03/13	156312	Silk Road Shutdown [2]
B	10/24/13	134236	TorATM Traffic Spike [11]
C	10/27/13	154855	URL Posted on Reddit [1]
D	11/07/13	126398	New Silk Road URL [8]
E	12/15/13	138231	Pirate Bay URL Posted [29]
F	03/21/14	303347	URLs Posted on Reddit [5]

aim to examine whether the .onion leakage is a temporary event or lasting phenomena over that long period of time.

- **Growth trends of .onion leakage:** we aim to examine whether there is a growth trend in the number of .onion requests, SLDs being requested, and IP addresses requesting those TLDs, and whether such trend is consistent over time.
- **Representation:** how representative are the A and J root servers to the total queries at the DNS roots.

3.5.1 Existence of .Onion

Table 1 summarizes the DITL dataset, including the total number of queries observed in each year of the dataset's life. We notice that while the phenomenon starts as a small set of queries in 2008, the total number of queries grows 3 orders of magnitude by the year of 2014, and persists over the years between them.

3.5.2 Growth Trends

Number of queries: The results in Table 1 show a growth trend for the number of .onion requests observed at the root servers over time. This monotonic growth trend is interrupted by a sharp growth in 2010, where the number of queries increased two orders of magnitude more than in the previous year (2009), and dropped by one order of magnitude for the year of 2011. We notice that the sharp increase that interrupted the monotonicity in the growth of the number of queries over years might not be a determining trend. In particular, given the nature of the dataset, a small event may actually cause a sudden surge in the number of queries, as shown in section 3.4, where such surge does not persist as a trend. Indeed, we notice that this interruption of monotonicity is due to a single SLD [5] for a tracker that attracted a large number of queries.

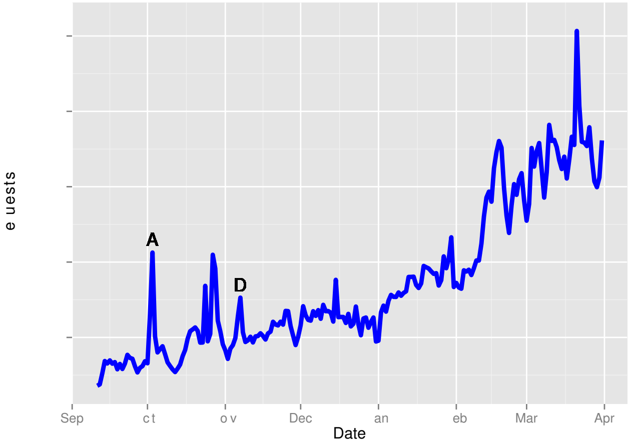


Figure 4: Global Events and Elevated Request Correlation

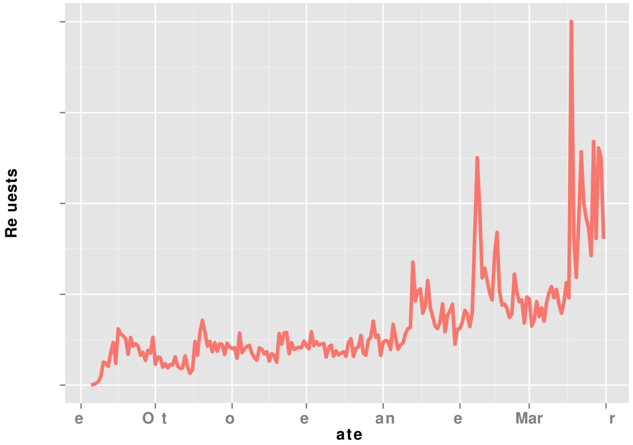


Figure 5: "Onion" Traffic Measurements From Turkey

Number of SLDs: The total number of SLDs that attracted .onion traffic and see at the root for the observation period grows exponentially, as shown in Figure 7. This trend can be used to precisely extrapolate the number of SLDs to be observed at the root unless the root cause of leakage is addressed. Note that, and unlike the interruption in the monotonic growth trend with respect to the total number of queries discussed earlier, no such interruption is introduced at the SLD level, given that the majority of added queries in the surge are due to a single SLD.

IP addresses: Figure 8 shows the total number of IP addresses (and their aggregated counterpart over /24 network addresses) over time. As with the general growth trend with the number of queries, an interruption in the monotonic growth happens in 2010. However, we observe a consistent and persistent trend of growth for the number of IP addresses originating the .onion queries, as well as their diversity of location measured by their /24 association.

[5] z6gw6skubmo2pj43.onion

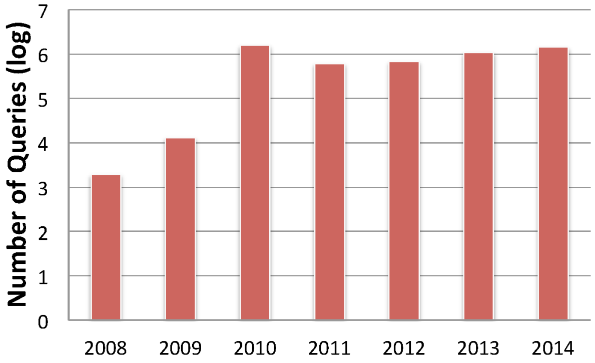


Figure 6: Queries over time.

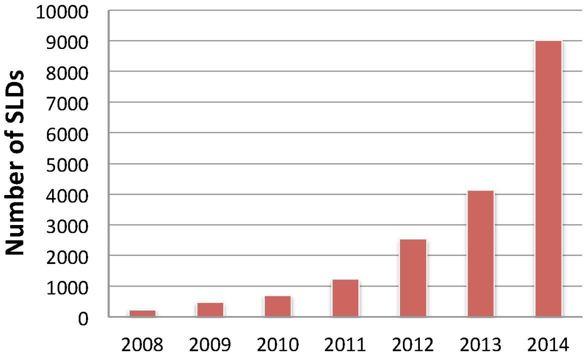
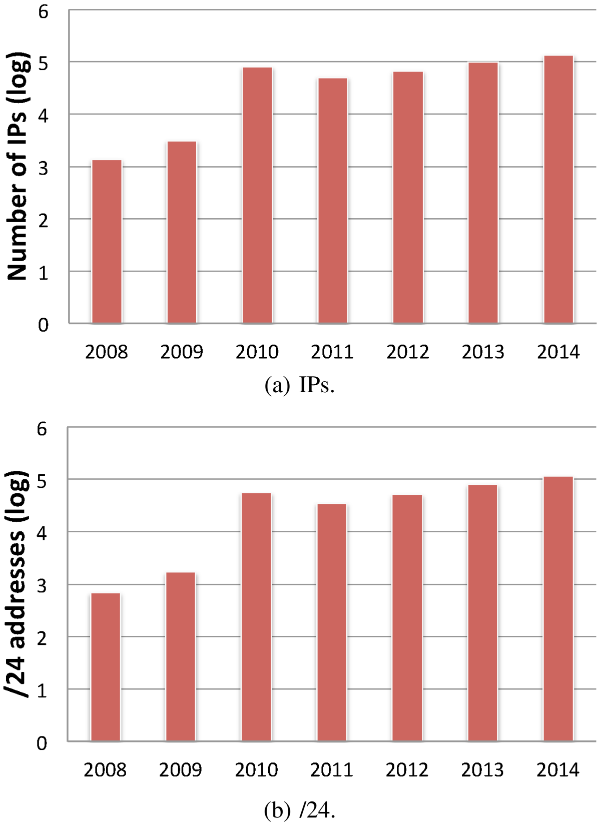


Figure 7: SLDs with .onion queries over time.

3.5.3 Representation

An interesting question that is raised by our reliance on the A and J roots operated by Verisign is "how representative are both root nodes for the population of queries seen at all roots?". Understanding this representation would explain the size of the problem reported in this paper in the Tor system as a whole. Unfortunately, the DITL dataset does not have traffic from all root servers except for the year of 2010, to which we limit our attention to answer the aforementioned question, despite some caveats.

Figure 9 shows the share of queries observed at every root for the year 2010 as a percent of the total number of queries. On this figure, we make two observations. First, the distribution of requests over all root servers is not uniform, with a few servers answering the majority of queries (roots F, I, J, and K answer a combined total of 56% of the queries). Second, and in answering the aforementioned question concerning representation of A and J, we find that they answer 4.9% and 13.3%, respectively, with a combined total number of queries of 18.2%. If such ratios hold over time (an assumption that we were not able to verify for the lack of data), they put the earlier estimates for the total number of queries at the DNS root servers—for the same period of time corresponding to the timeframe where the data described in section 2.1 was collected—to 151.6 million queries (from the previous estimate of 69 million queries). This estimate gives an average query rate of about 840k queries per day. As a result, we conclude that the total number of .onion queries is substantial, and may potentially pose a high risk to the privacy of a large number of users.



(a) IPs.

(b) /24.

Figure 8: Individual IP and /24 addresses originating the .onion traffic. Notice that the same growth trend shown in the number of queries is also reflected on the number of addresses.

3.5.4 Geographical Distribution

Similar to the experiment shown in Table 3, we map the various sources originating traffic to their home countries. With respect to the index of countries shown in Table 3, we choose the top 10 countries, and compute the share of requests originated from them. Table 6, shows the number of queries originated from each given country in the top, along with their percent out of the total queries computed over the 7 years of DITL. We further add a per-country percentage of traffic share for the years 2014 and back until 2012. From this table, we make the following observations:

- Inconsistent representation: while part of the order of the countries is mostly consistent with the ranking provided by the Tor project on its use use [6], we find that the ranking is inconsistent with the previous ranking established for the countries with traffic seen in the A and J root study.
- Inconsistent order: the order of countries as shown in Table 3, which highlights countries in a descending order, is not preserved in Table 6. This highlights dynamics of shares in the traffic, perhaps based on phenomena best seen in those countries through the usage of hidden services.

[6] `http://bit.ly/1ud2CKh`

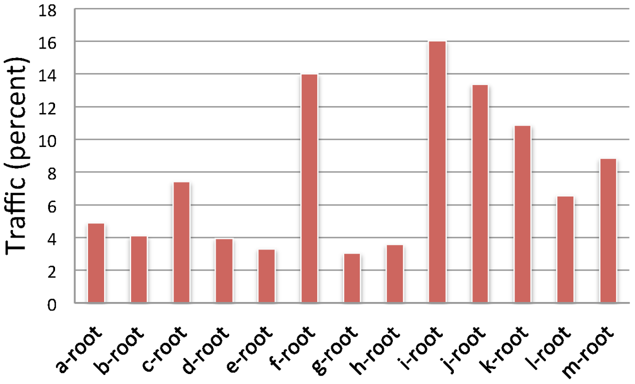


Figure 9: The percentage of queries observed at each root, with A and J having 4.9% and 13.3%, respectively.

Table 6: Top Geographical Countries and Requesting "Onion". The results use the country ranking in Table 3

Country	Requests	(%)	2014	2013	2012
US	1469134	21.45	30.96	24.85	17.98
RU	302222	4.41	5.34	5.74	8.02
DE	228487	3.34	5.06	3.46	3.79
BR	256195	3.74	3.93	8.03	3.44
CN	201318	2.94	2.05	2.27	2.32
GB	320550	4.68	4.14	3.44	3.52
KR	63546	0.93	1.24	1.05	1.33
PL	136693	2	1.34	2.03	1.72
CA	367859	5.37	1.75	2.84	2.01
FR	245159	3.58	1.83	1.59	2.42

4. CAUSES AND IMPLICATIONS

Applications electing to use non-delegated TLDs as a namespace in which they seed their routing and resolution processes face scenarios in which possible DNS leakage may occur. Tor has been specifically designed to prevent .onion requests from leaking within the application into the global DNS infrastructure. However, it is clear from the measurements we presented so far that a significant volume of requests are being issued to the global DNS root servers. Whether they are initiated by users by mistake or caused by a misconfiguration in the underlying application, such as Tor, or the web browsers, leaked DNS queries outside of the Tor network have a significant implication to individuals' privacy and safety. To that end, understanding the causes of the leakage may help reducing the risk at the user side.

4.1 (Potential) Root Causes

4.1.1 User Error

There are many plausible reasons or mechanisms in which .onion queries could be generated and observed in the global public DNS; however, the root cause of how and why these queries are being requested within the global DNS remains unclear. We have seen in this paper numerous global events that spurred additional query volume. One potential explanation associated with surge in the volume of .onion domains in those times is users errors, in which users are not aware that the addresses of hidden services should be run on top of Tor (i.e., by first installing Tor plug-in associated with the browser). Other notable explanations may include browser prefetching, third party application or plug-ins, DNS suffix search lists, web crawlers, and malware.

4.1.2 Malware

Advanced families of malware are also now utilizing Tor within their Command and Control (C&C) infrastructure [16, 9, 12, 10]. Cyber-criminals may use Tor and its Hidden Services in order to avoid detection and prevent takedowns. Several cyber-criminals have now started actively using Tor to host malicious infrastructure via Hidden Services. Variants of Zeus [24], CyrptoLocker [23], ChewBacca [20], CryptorBit [6] and Torec [18] have all been found to use various aspects of the Tor network, including hidden services. Possible misconfigurations within these malware pieces could facilitate a percentage of the leaked DNS requests. To that end, we observed numerous requests for .onion SLDs associated with these malware samples during our analysis.

4.2 Implications and Remedies

4.2.1 Implications

Queries seen at the root come from individual users' IPs, public recursive servers' IPs, ISP's recursive IPs, and open resolvers IPs. The implication depends on where the queries come from as follows. 1) Individual users' IPs and their resolution preserve locality information of the users issuing such information, and may considerably expose users to a high risk, depending on their location and the context of the queried hidden service. 2) Many of the queries issued to the root come from public recursive DNS servers that are responsible for a large number of queries aggregated from potentially multiple users, where the individual users' IP addresses are detached, thus the root does not see those address. However, this still puts the individual users at risk, although their individual IP addresses are not exposed. For example, the DNS queries observed at the root are likely the result of unencrypted traffic that an eavesdropper close by the user can listen to, and associate to the user. Further more, most public recursive service do not preclude the possibility of sharing users' traffic a with a third party in their use agreements. 3) Whereas ISPs might be disincentivized from sharing the individual users information with third parties, eavesdropping while closer to the users may expose them. Furthermore, when ISPs are a government entity (e.g., in Egypt and Turkey), their double function puts users at risk. 4) Unlike ISPs, open resolvers that do not serve a clear business agenda do not have the business relationship with users, and might be willing to share such information with a third party, thus putting users at a great risk.

4.2.2 Remedies

The .onion queries are not supposed to be seen at the root or in the public DNS resolution. However, due to malfunction, bad settings, or even the "ignorance of the crowd", we see a large quantity of such queries at the DNS roots. Furthermore, we found that many of those queries are the result of legitimate Tor usage that may put individual users at risk. To address the problem, several directions are worth exploring.

- Given the nature of .onion, and other privacy or special purpose TLDs, blocking capabilities should be enabled at the edge, including capabilities of blocking in the browser. This is, users are often time not exposed to low level details of connection failures with today's Tor distribution, and blocking may help mitigating the leakage of .onion queries when Tor connection fails for one reason or another.

- Legacy software may expose users' traffic to the outside work. Queries associated with privacy enhancing technologies such as Tor should be controlled as to prevent and notify users if public DNS leakage occurs due to those legal programs.

- Automatic system-level configuration of .onion resolution should be used. The Tor distribution should provide a system-level fix to local DNS configuration and not require users to configure this component manually.

- Many of the queries can be blocked lower in the DNS hierarchy, and be prevented from propagation into the public DNS by deploying techniques such as negative caching [7]. For example, public recursive name servers most close to the users may help by not sending out queries to the root for TLDs that do not exist. Given the (almost) static nature of the TLDs, and the static nature of the TLDs of interest (such as .onion), operators of public DNS services may deploy effective mechanisms in achieving such goal.

4.3 Namespace Management

Focus within the Internet Engineering community has recently increased on ways for applications to properly use non-delegated domains. A recent Internet draft describes several special-use domain names of peer-to-peer name systems and is seeking approval from the Internet Engineering Steering Group (IESG) [15]. Discussions about the proposal on the DNS operators mailing list have brought forth other generic solutions such as proposed .alt alternative TLD in which applications would safe anchor namespace under it [28]. Blurred lines of authority, privacy and security makes such a namespace problem difficult to solve and appease all parties.

5. CONCLUSION AND FUTURE WORK

We looked at a sample of .onion DNS requests issued to the A and J root nodes of the global DNS infrastructure. We examined the unique characteristics of these requests longitudinally as well as the dynamics of requests received from a geographical and network location for unique SLDs. We found that increased traffic spikes within the global DNS for .onion requests corresponded with external global events, highlighting the potential human factor in those leakages (i.e., user error). While the root cause of these leaked DNS queries remains unknown, our preliminary investigation unveiled concerns to the severity of the leakage and to the possibility of more sensitive private information being unintentionally exposed. Our future work will continue the examination of leaked DNS queries to the root but will also extend to other non-delegated TLDs such as i2p and .exit. We will plan to further dissect the impact of global events and the role of malware in the leakage, and investigate the potential privacy consequences of the leakage under the various leakage causes. By sharing this preliminary work, we wish to trigger further discussion in the community.

6. REFERENCES

[1] —. ELI5: What exactly is the “deep web”. Reddit, http://bit.ly/1l7hLbz, Oct. 2013.

[2] —. FBI arrest ‘silk road’ owner Ross William Ulbricht, shut down tor’s most notorious black market. Huffington Post UK, http://huff.to/1fu0tA7, Oct. 2013.

[3] —. New gTLD Security, Stability, Resiliency Update: Exploratory Consumer Impact Analysis. http://bit.ly/QB6ntp, Aug. 2013.

[4] —. New Generic Top-Level Domains. ICANN, http://newgtlds.icann.org/en/, 2014.

[5] —. People who have visited the ‘deep web’ what was it like and why did you do it? http://bit.ly/ROuupk, Mar. 2014.

[6] L. Abrams. Cryptorbit and howdecrypt information guide and faq. Bleepingcomputer, http://bit.ly/1eoKEjh, Feb. 2014.

[7] M. Andrews. Negative caching of dns queries (dns ncache). RFC 2308, 1998.

[8] J. Biggs. Silk road 2.0 rises again. TechCrunch, http://tcrn.ch/QB5HnQ, Nov. 2013.

[9] A. Biryukov, I. Pustogarov, and R. Weinmann. Trawling for tor hidden services: Detection, measurement, deanonymization. In *IEEE Security and Privacy (SP)*, 2013.

[10] A. Biryukov, I. Pustogarov, and R.-P. Weinmann. Content and popularity analysis of tor hidden services. *arXiv:1308.6768*, 2013.

[11] Bitcoin wiki. Toratm. https://en.bitcoin.it/wiki/TorATM, July 2013.

[12] D. Brown. Resilient botnet command and control with tor. *DEF CON*, 18, 2010.

[13] L. Chapin and M. McFadden. Reserved top level domain names. RFC 2606, http://bit.ly/1nIQ5cS, 2011.

[14] R. Garcia. Preventing tor DNS leaks. The Tor Project, http://bit.ly/1royLtU, 10 2014.

[15] C. Grothoff, M. Wachs, H. Wolf, and J. Appelbaum. Special-use domain names of peer-to-peer name systems. IETF Internet Draft, Nov. 2013.

[16] N. Hopper. Challenges in protecting tor hidden services from botnet abuse. In *Proceedings of Financial Cryptography and Data Security (FC’14)*, March 2014.

[17] Interisle Consulting Group, LLC. Name collision in the DNS. ICANN, http://bit.ly/1iQVj5F, 8 2013.

[18] E. Kovacs. Backdoor.androidos.torec.a: First tor-based trojan for android. http://bit.ly/1pte18L, Feb. 2014.

[19] P. Mockapetris and K. J. Dunlap. *Development of the domain name system*, volume 18. ACM, 1988.

[20] M. J. Schwartz. Chewbacca malware taps tor network. Dark Reading, http://ubm.io/1nrFfKY, 12 2013.

[21] I. Security and S. A. C. (SSAC). Invalid top level domain queries at the root level of the domain name system. http://bit.ly/1mDxRJO, Nov. 2010.

[22] A. Simpson. Detecting search lists in authoritative DNS. In *Workshop on Root Causes and Mitigation of Name Collisions (WPNC)*, 2014.

[23] M. Smtih. Cryptolocker crooks charge 10 bitcoins for second-chance decryption service. Network World, http://bit.ly/ROxhPd, 11 2013.

[24] D. Tarakanov. The inevitable move - 64-bit zeus has come enhanced with tor. SecureList http://bit.ly/1mIuAeR, 12 2013.

[25] The Tor Project. Tor metrics portal: Users. The Tor Project, http://bit.ly/1hrHqGp, 2014.

[26] The Tor Project. Tor: Overview. The Tor Project, http://bit.ly/1dZ2zvZ, 2014.

[27] M. Thomas, Y. Labrou, and A. Simpson. The effectiveness of block lists in preventing collisions. In *Workshop on Root Causes and Mitigation of Name Collisions (WPNC)*, 2014.

[28] P. Wouters. DNS operation mailing list. DNSOP, http://bit.ly/1roTIXw, 12 2013.

[29] K. Zournas. Pirate bay relocates to thepiratebay.ac. http://bit.ly/1iQNEEz, 12 2013.