DISPOSABLE DOMAINS

Yizheng Chen, Manos Antonakakis, Wenke Lee
Georgia Institute of Technology

Abstract

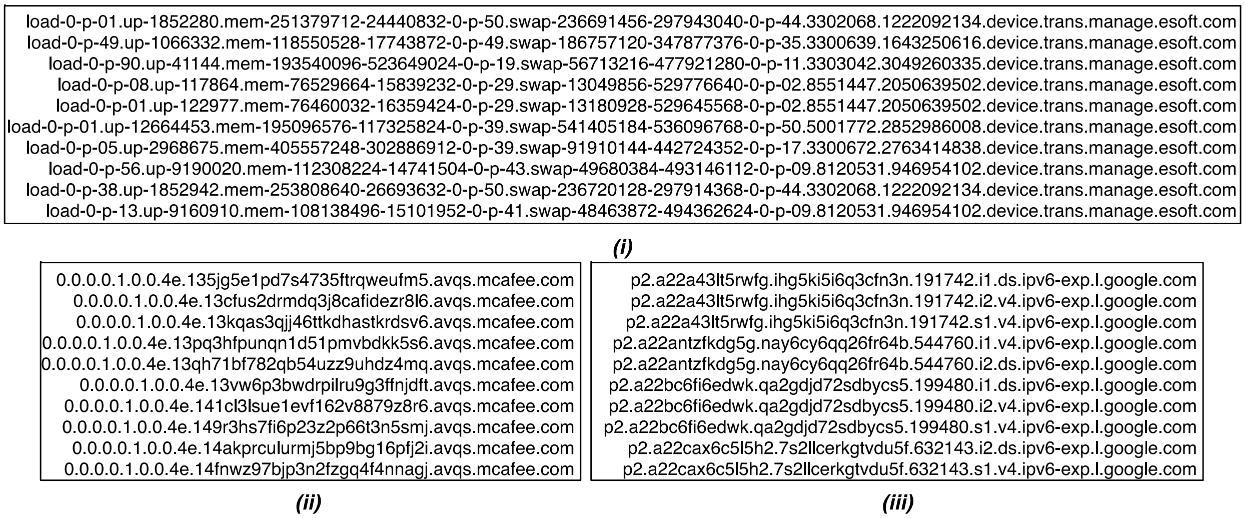
In recent years DNS has been increasingly leveraged to build and scale highly reliable network infrastructures. In this paper, we will introduce and analyze a new class of domains, which we refer to as disposable domains. Disposable domains appear to be heavily employed by common Internet services (i.e., Search Engines, Social Networks, Online Trackers etc.), and they seem to be automatically generated. They are characterized by a "one-time use" pattern, and appear to be used as a way of "signaling" via DNS. While this is yet another "creative" use of the DNS to enable new Internet applications and efficient scaling of services, little do we know about the size and DNS caching properties of this family of domains.

To shed light on the pervasiveness and growth of disposable domains, we present a study of their characteristics based on live DNS traffic observed at Comcast, in a city that serves millions of end users. We found that disposable domains increased from 23.1% to 27.6% in all queried domain names, and from 27.6% to 37.2 % among all resolved domain names daily, and more than 60% of all distinct resource records observed daily in modern DNS traffic are related to disposable domains. We discuss the possible negative implications that disposable domains may have on the DNS caching infrastructure, resolvers validating DNSSEC transactions, and passive DNS data collection systems.

INTRODUCTION

Domain Name System was originally designed for mapping a human-friendly domain name to a machine-readable IP address. Over the years, people have used DNS in new ways to make their services more agile and scalable. However, they all had unanticipated and sometimes negative impact as the following three examples shows.

The first example is using DNS to select a Content Delivery Network (CDN) server that is closest to client. When a CDN sees a DNS request for content, it will return a CDN server IP address that is closest to the requester IP, and with small load at the time. Since what CDN sees is the IP address of the DNS server that user's machine is configured to, not the user's IP address, the effectiveness of such approach depends on how close users are to their local DNS servers. Researchers [1] have shown that 64% of associations of user's and the local DNS server's IP addresses are in the same Autonomous System. However, only 16% of associations are in the same network-aware clusters, from the perspective of BGP routes. The second example is browser prefetching to speed up webpage loading performance [2]. When a user is entering search queries, the browser will look up unfinished search queries as possible domain names and pre-resolve all the domain names before user finishes typing. The design of prefetching is used for web objects as well, to minimize the delay user perceives while browsing. However, an unanticipated negative impact from that is DNS prefetching could potentially leak user's privacy by exposing the search terms in just the DNS queries. The last example is NXDOMAIN redirection for displaying commercials. Parked domains are often redirected to advertisement pages to monetize existing users for the old domain name. The practice of doing that was called "DNS lie" [3] [4]. It has always been controversial of whether ISPs should do that, since advertisement page is not the page users intend to look for.



```
load-0-p-01.up-1852280.mem-251379712-24440832-0-p-50.swap-236691456-297943040-0-p-44.3302068.1222092134.device.trans.manage.esoft.com
load-0-p-49.up-1066332.mem-118550528-17743872-0-p-49.swap-186757120-347877376-0-p-35.3300639.1643250616.device.trans.manage.esoft.com
load-0-p-90.up-41144.mem-193540096-523649024-0-p-19.swap-56713216-477921280-0-p-11.3303042.3049260335.device.trans.manage.esoft.com
load-0-p-08.up-117864.mem-76529664-15839232-0-p-29.swap-13049856-529776640-0-p-02.8551447.2050639502.device.trans.manage.esoft.com
load-0-p-01.up-122977.mem-76460032-16359424-0-p-29.swap-13180928-529645568-0-p-02.8551447.2050639502.device.trans.manage.esoft.com
load-0-p-01.up-12664453.mem-195096576-117325824-0-p-39.swap-541405184-536096768-0-p-50.5001772.2852986008.device.trans.manage.esoft.com
load-0-p-05.up-2968675.mem-405557248-302886912-0-p-39.swap-91910144-442724352-0-p-17.3300672.2763414838.device.trans.manage.esoft.com
load-0-p-56.up-9190020.mem-112308224-14741504-0-p-43.swap-49680384-493146112-0-p-09.8120531.946954102.device.trans.manage.esoft.com
load-0-p-38.up-1852942.mem-253808640-26693632-0-p-50.swap-236720128-297914368-0-p-44.3302068.1222092134.device.trans.manage.esoft.com
load-0-p-13.up-9160910.mem-108138496-15101952-0-p-41.swap-48463872-494362624-0-p-09.8120531.946954102.device.trans.manage.esoft.com
```

(i)

```
0.0.0.0.1.0.0.4e.135jg5e1pd7s4735ftrqweufm5.avqs.mcafee.com
0.0.0.0.1.0.0.4e.13cfus2drmdq3j8cafidezr8l6.avqs.mcafee.com
0.0.0.0.1.0.0.4e.13kqas3qjj46ttkdhastkrdsv6.avqs.mcafee.com
0.0.0.0.1.0.0.4e.13pq3hfpunqn1d51pmvbdkk5s6.avqs.mcafee.com
0.0.0.0.1.0.0.4e.13qh71bf782qb54uzz9uhdz4mq.avqs.mcafee.com
0.0.0.0.1.0.0.4e.13vw6p3bwdrpilru9g3ffnjdft.avqs.mcafee.com
0.0.0.0.1.0.0.4e.141cl3lsue1evf162v8879z8r6.avqs.mcafee.com
0.0.0.0.1.0.0.4e.149r3hs7fi6p23z2p66t3n5smj.avqs.mcafee.com
0.0.0.0.1.0.0.4e.14akprculurmj5bp9bg16pfj2i.avqs.mcafee.com
0.0.0.0.1.0.0.4e.14fnwz97bjp3n2fzgq4f4nnagj.avqs.mcafee.com
```

(ii)

```
p2.a22a43lt5rwfg.ihg5ki5i6q3cfn3n.191742.i1.ds.ipv6-exp.l.google.com
p2.a22a43lt5rwfg.ihg5ki5i6q3cfn3n.191742.i2.v4.ipv6-exp.l.google.com
p2.a22a43lt5rwfg.ihg5ki5i6q3cfn3n.191742.s1.v4.ipv6-exp.l.google.com
p2.a22antzfkdg5g.nay6cy6qq26fr64b.544760.i1.v4.ipv6-exp.l.google.com
p2.a22antzfkdg5g.nay6cy6qq26fr64b.544760.i2.ds.ipv6-exp.l.google.com
p2.a22bc6fi6edwk.qa2gdjd72sdbycs5.199480.i1.ds.ipv6-exp.l.google.com
p2.a22bc6fi6edwk.qa2gdjd72sdbycs5.199480.i2.v4.ipv6-exp.l.google.com
p2.a22bc6fi6edwk.qa2gdjd72sdbycs5.199480.s1.v4.ipv6-exp.l.google.com
p2.a22cax6c5l5h2.7s2llcerkgtvdu5f.632143.i2.ds.ipv6-exp.l.google.com
p2.a22cax6c5l5h2.7s2llcerkgtvdu5f.632143.s1.v4.ipv6-exp.l.google.com
```

(iii)

Figure 1. Three examples of disposable domain names from eSoft, McAfee, and Google.

As the Internet has evolved over the years, more service providers, such as popular search engines, social networks, and online trackers, began to use a new class of domain names, that we call disposable domains. Disposable domains almost seem to be a natural result of people seeking even more agility and scalability for their Internet services. Using disposable domains, service providers don’t need to set up any dedicated infrastructure for their service, but to simply overload DNS with customized protocols. We will discuss the properties of this new class of domain names, specifically focusing on their algorithmically-generated zone structures and their low cache hit rates obtained from a cluster of recursive DNS resolvers operated by Comcast, a large north-American ISP.

This increase in use of disposable domains may have unanticipated negative effects on day-to-day DNS operations for large ISPs. For instance, a large number of DNS requests for disposable domains could fill up the cache of recursive DNS resolvers. Such an event may cause premature cache evictions of non-disposable domains, which would degrade DNS service for the ISP. In turn, these premature evictions may inflate the traffic between the DNS resolvers and authoritative name servers, a phenomenon that could be very costly for ISPs in a DNSSEC-enabled recursive environment. Lastly, disposable domains increase the storage requirement for passive DNS data collection systems, and could potentially degrade database query latency.

In the rest of the paper, we will first show some examples of disposable domains and discuss their properties. Then we will provide supporting evidence on how disposable domains are currently used by large service providers. Lastly, we will discuss possible negative implications that the growth in disposable domains may have on the DNS caching infrastructure, DNSSEC-validating resolvers, and passive DNS data collection systems.

MINING DISPOSABLE DOMAINS

In this section we will define disposable domain names and we will provide some real world examples of their use in case studies. Then, we will discuss the prevalence of disposable domains. We define disposable domain names as successfully resolved domain names that have the following properties:

1). Their name strings are automatically generated.

2). The median cache hit rate for the resource records of child domain names under a zone that facilitates disposable domain names is low or close to zero. In other words, the resource records under that particular zone are only observed once, or a handful of times, when they are in the recursive DNS servers' cache.

Case Studies

Figure 1 shows three examples of what we define as disposable domain names. The eSoft (i) domain names are used as a storage communication channel that reports CPU load, machine up time, memory usage and swap disk usage. The McAfee [5] (ii) domain names are used for file reputation queries on behalf of McAfee's Global Threat Intelligence File reputation Service. This is yet another case of using the DNS as an information storage communication channel. Lastly, Google's IPv6 experiment domains [6] (iii) are queried by browsers of selected users that perform cryptographically signed background requests after receiving their search results. The background requests record IPv4 and IPv6 addresses, image request latency, and User-Agent strings.

Examining the zone structures from Figure 2 shows that 1) disposable domain names tend to have same number of periods ("."), 2) at certain places between two periods, the labels are "random-looking". The structure property reflects how zone operators parse and use different parts of disposable domains for different purposes or transfer different information, by using algorithm-generated strings.

In addition to zone structural properties, disposable domains typically have very low or sometimes zero cache hit rates. Usually, over 90% of cache hit rates from disposable domains are zero. On the other hand, cache hit rates of non-disposable domains follow a closer to linear cumulative distribution, and the median cache hit rate would be around 40%. In general, resource records of disposable domain names are used only once or up to a few times while they are in the recursive cache, which results in the ***overall*** low cache hit rate distribution for domains under disposable zones.

Measurement Results

We built a disposable domain miner system to automatically mine disposable domains. The technical details of our system can be found in [7]. Over the period of a year, we found 14,488 zones that use disposable domains, with a confidence of more than 90%. Disposable domains are used by various industries, including popular websites (e.g., Google, Microsoft), Anti-Virus companies (e.g., McAfee, Sophos, Sonicwall, Mailshell), DNSBLs (e.g., Spamhaus, countries.nerd.dk), social networks (e.g., Facebook, Myspace), streaming services (e.g., Netflix), P2P services (e.g., Skype), cookie tracking services (e.g., Esomniture, 2o7.net), ad networks (e.g., AdSense, Bluelink Marketing), e-commerce business (e.g., Paypal, ClickBank), etc.

Disposable domains are not only widely used currently, but are also increasingly being used. For unique domains being queried by clients, the percentage of disposable domains increased from 23.1% to 27.6%. Also, of the daily resolved unique domains the percentage of disposable domains grew from 27.6% to 37.2% over the year of 2011. From traffic during 11/28/2011 to 12/10/2011, we observe that the number of new disposable domains seen every day is always high, around 5 million to 7 million. However, the number of new non-disposable domains dropped from 13 million to 1.6 million. So after one day, more than 50% of new domains seen daily are disposable, and after 13 days, more than 80%

of new domains seen daily are disposable, since new disposable domains are constantly generated. Moreover, the volume of unique disposable resource records daily increased from 8,111,274 (02/01/2011) to 29,738,493 (12/30/2011), during which 33,704,127 were observed on 11/14/2011. The percentage of daily unique disposable RRs increased from 38.3% to 65.5%.

DISCUSSION

In this section, we will discuss possible negative effects of using disposable domains. We will discuss their impact on DNS caching, DNSSEC-enabled resolvers, and passive DNS databases, so that the operational community can anticipate them and plan ahead in case changes to current DNS operations are needed.

DNS Caching

As disposable domains are increasingly used, the cache of recursive DNS servers may be filled up with entries that are highly unlikely to be reused. Assuming a typical Least Recently Used cache implementation with fixed memory allocation, during periods of heavy load, queries to disposable domains may cause some useful non-disposable domains to be prematurely evicted from the cache. In turn, this may have the effect of unfairly inflating the traffic between the DNS resolvers and the authoritative name servers responsible for the evicted non-disposable domains, thus increasing the query-response latency.

DNSSEC

There will inevitably be more pressure on validating resolvers when DNSSEC becomes more widely deployed. Validating signed responses requires higher CPU usage, and increased memory needs due to DNSSEC specifications [8] [9] [10]. Disposable domains will naturally, and potentially dramatically, increase this pressure on validating resolvers. In fact, each queried disposable domain may require an additional signature validation whose result will never be reused. Also, the cache must not only store the disposable RRs, but also their signatures. This problem may be mitigated in part if the authoritative servers responsible for the disposable zones register disposable domains under a single signed wildcard domain, from which the disposable domains are synthesized.

pDNS-DB

Passive DNS database systems (pDNS-DBs) have recently been adopted by computer security and networking communities as an invaluable tool to analyze security incidents, monitor and troubleshoot DNS operations, and develop dynamic reputation systems [11] [12]. Disposable domains have the effect of increasing pDNS-DB storage requirements, and potentially their query-response latency, depending on the implementation. In fact, we found that after bootstrapping a pDNS-DB with 13 days of resolution traffic, 88% of all unique resource records in the database are disposable, and new RRs related to disposable domains make up more than 94% of all the new distinct RRs observed daily. The problem can be mitigated by filtering disposable domains and storing a single wildcard domain in the pDNS-DB.

REFERENCES

[1] Z. M. Mao, C. D. Cranor, F. Douglis, M. Rabinovich, O. Spatscheck, and J. Wang. A precise and efficient evaluation of the proximity between web clients and their local dns servers. In *Proceedings of the General Track of USENIX ATEC*, 2002.

[2] S. Krishnan and F. Monrose. DNS prefetching and its privacy implications: when good things go bad. In *Proceedings of USENIX Workshop on LEET*, 2010.

[3] P. Vixie. What dns is not. Queue, (10), Nov. 2009.

[4] N. Weaver, C. Kreibich, and V. Paxson. Redirecting DNS for Ads and Profit. In *USENIX Workshop on Free and Open Communications on the Internet (FOCI),* 2011.

[5] McAfee. Faqs for global threat intelligence file reputation. https://kc.mcafee.com/corporate/index?page= content&id= KB53735, 2013.

[6] S. H. Gunderson. Global IPv6 statistics: Measuring the current state of IPv6 for ordinary users. In *Proceedings of the Seventy-third Internet Engineering Task Force*, 2008.

[7] Y. Chen, M. Antonakakis, R. Perdisci, Y. Nadji, D. Dagon, W. Lee. DNS Noise: Measuring the Pervasiveness of Disposable Domains in Modern DNS Traffic. In *proceedings of the 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, 2014.

[8] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. Dns security introduction and requirements. http://www.ietf.org/rfc/rfc4033.txt, March 2005.

[9] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. Protocol modifications for the dns security extensions, rfc 4035. http://www.ietf.org/rfc/rfc4035.txt, March 2005.

[10] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. Resource records for the dns security extensions. http://www.ietf.org/rfc/rfc4034.txt, March 2005.

[11] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster. Building a Dynamic Reputation System for DNS. In *Proceedings of USENIX Security Symposium*, 2010.

[12] M. Antonakakis, R. Perdisci, W. Lee, D. Dagon, and N. Vasiloglou. Detecting Malware Domains at the Upper DNS Hierarchy. In *Proceedings of USENIX Security Symposium*, 2011.