DNS BULLYING: Análise do comportamento agressivo de recursivos na internet

Anderson Antonio Fontana

Ciência da Computação – Universidade de Passo Fundo (UPF) – Campus I – Passo Fundo – RS

152058@upf.br

Abstract. *It is known that not always the TTL suggested in DNS responses is respected by the recursive servers, causing a number of unnecessary requests to the authoritative servers. In this work, 24 hours of real DNS traffic was used to study the behavior of recursive servers before the root-servers. The analysis was focused on some specific domains and resource records, and results show that most of the recursives disregard the TTL suggested by the root-servers. As a consequence, the number of requests to the root-servers is much higher than theoretically expected. It was observed that 70% of requests are made by only ten companies, generating more than 900 requests per second.*

Resumo. *Sabe-se que nem sempre o TTL sugerido em respostas do DNS é respeitado pelos servidores recursivos, ocasionando um número de requisições desnecessárias aos servidores autoritativos. Neste trabalho foram utilizadas 24 horas de tráfego DNS real para estudar o comportamento de servidores recursivos perante os root-servers. A análise foi focada em alguns domínios e resource records específicos, e resultados mostraram que a maioria dos recursivos desconsideram o TTL sugerido pelos root-servers. Como consequência o número de requisições aos root-servers é muito maior que o teoricamente esperado. Foi observado que 70% das requisições são realizadas por somente dez empresas, gerando mais de 900 requisições por segundo.*

1. Introdução

A experiência de um usuário com determinado serviço na Internet depende parcialmente da disponibilidade e velocidade de algumas etapas da comunicação com esse serviço. Há certas camadas que ficam transparentes aos usuários, mas que são essenciais para o estabelecimento da comunicação entre os envolvidos.

Um desses componentes, muitas vezes desconhecido da maioria dos usuários, é o *Domain Name System* (DNS), ou Sistema de Nomes de Domínio que, conforme descrito nos *Request For Comments* (RFC) 1034 [Mockapetris, 1987a] e 1035 [Mockapetris, 1987b], é um sistema distribuído responsável por, entre outras finalidades, fornecer o mapeamento entre o nome de domínio em endereço *Internet Protocol* (IP), permitindo que o usuário acesse o serviço dispensando a memorização do endereço IP do recurso que deseja estabelecer comunicação [Wicaksana, 2016].

Para o fornecimento desse mapeamento é usado servidores recursivos que fazem o intermédio das requisições, consultando na hierarquia de servidores espalhados pelo

mundo e que contém as informações requisitadas. Para uma melhor performance estes servidores armazenam as consultas, a fim de quando uma nova requisição de mesmo tipo aconteça, este já contenha a informação localmente, bastando simplesmente devolver a resposta.

Este artigo apresenta uma análise sobre o período de tempo que os servidores recursivos armazenam um resultado DNS em seus *caches*. Para tal investigação, foram usados dados obtidos do *Day In The Life of the Internet*[1] (DITL), fornecidos pela *DNS Operations, Analysis, and Research Center* (DNS-OARC). Estes dados contém todas as requisições que passaram nos *root-servers*, originadas dos mais diversos servidores recursivos.

Com isto, é apresentada a atuação dos registros de recurso (RRs) e dos *Top-level-Domains* (TLDs) perante os *root-servers*, além de compreender o mal comportamento de alguns servidores recursivos.

É apresentada também uma simulação do comportamento dos servidores recursivos, com o objetivo de visualizar como seria a carga para os *root-servers*, caso todos os servidores recursivos se comportassem segundo o melhor ou o pior caso, no que diz respeito ao tempo de armazenamento da requisição DNS.

2. Trabalhos Relacionados

Na investigação realizada e descrita detalhadamente a seguir neste artigo, foi apresentado o comportamento de algumas empresas perante os *root-servers*, podendo-se estimar um baixíssimo valor de TTL em alguns servidores recursivos, tendo como base o número de requisições que cada um desses servidores fez em um dia. Diante disso, serão apresentados alguns trabalhos relacionados a seguir.

2.1 TTL

No trabalho de Zyl et al. (2015) é apresentada uma análise do TTL (*Time to Live*) do DNS, com capturas realizadas de janeiro a junho de 2014, identificando e comparando domínios frequentemente consultados. Também apresentam um detalhamento das práticas de TTL pelo tipo *Resource Record*, assim como uma análise de valores anormais de TTL em relação aos dados coletados.

Wills e Shang (2000) examinaram os efeitos do DNS *lookup* para o tempo geral de recuperação dos objetos da *Web*. Eles concluíram que o mecanismo de *cache* do DNS funcionava melhor para servidores *Web* populares do que para servidores *Web* aleatórios, uma vez que servidores *Web* populares são mais frequentemente acessados e, portanto, a probabilidade de estarem em *caches* DNS é maior. Encontraram, também, a partir dos resultados de solicitações de um usuário real, que cerca de 80% de suas requisições encontram-se armazenadas em seu *cache* local. Ainda, destacam que para requisições que não estavam na *cache* local, a média do tempo de resposta foi maior que um segundo.

No trabalho de Vlajic et al. (2012) é examinado o impacto dos valores de TTL do DNS do nível de satisfação do usuário ao acessar um site. Demonstram que um site que utiliza valores TTL de DNS inadequados pode sofrer consequências prejudiciais e

[1] https://www.dns-oarc.net/oarc/data/ditl

custosas, especialmente se for vítima de um ataque DDoS. Posteriormente, analisaram os valores de TTL do DNS de alguns bancos e mostraram que a exposição pública está altamente correlacionado com o nível de sofisticação no gerenciamento do DNS. Apontaram que vários bancos (geralmente de menor porte) escolhem valores de TTL de DNS inadequadamente longos, criando uma vulnerabilidade que pode ser facilmente explorada por um atacante.

Em Bhatti e Atkinson (2011), visando suportar novos serviços e sistemas ágeis, os valores em *cache* precisariam ter valores de TTL muito mais baixos, para que os valores de DNS armazenados em *cache* não fiquem obsoletos à medida que ocorrem alterações no sistema. No entanto, as convenções atuais para configuração do DNS normalmente usam valores TTL altos. Eles realizaram um estudo empírico de uma implementação de DNS ao vivo, onde reduziram a zero os valores de registros de TTL para toda a Escola de Ciência da Computação da Universidade de St. Andrews. Os resultados mostram que o aumento na carga de DNS é muito menor do que o esperado, após uma diminuição não linear em relação ao valor de TTL dos registros de DNS.

2.2. DITL

A base de dados utilizada neste artigo foi obtida pelo projeto DITL. Outros artigos nesta mesma área demonstram algumas possibilidades da utilização desse recurso.

Em Castro et al. (2008), foram analisados os dados coletados pelo DITL de 3 anos consecutivos, extraindo tendências históricas, comparações com outras fontes de dados e interpretações, incluindo crescimento de tráfego, padrões de uso, impacto da distribuição *anycast* e problemas persistentes no sistema de *root-servers* que ameaçam a Internet global. Concluíram que há uma quantidade extraordinária de poluição do DNS, onde ≅ 98% do tráfego não deveria chegar aos *root-servers*.

O trabalho de Freitas et al. (2013), utiliza a base de dados do projeto DITL para prova de conceito de sua ferramenta *MyDnsDump*, onde apresenta uma proposta de arquitetura para monitoramento de tráfego DNS baseado na biblioteca de rede *libtrace*. Concluiu que a sua biblioteca mostrou uma grande facilidade no desenvolvimento de novas aplicações, com grande velocidade de processamento das informações dos arquivos e obtendo resultados superiores ao seu principal concorrente *libpcap*.

3. Terminologias

Neste capítulo serão apresentados os principais conceitos que permeiam o assunto desta pesquisa.

3.1. DNS

O objetivo do inglês *Domain Name Server* (DNS) é fornecer um mecanismo para nomear recursos de forma que os nomes possam ser usados em diferentes hosts, redes, protocolos, Internets e organizações administrativas.

Do ponto de vista de um cliente, com o DNS, é possível obter o mapeamento entre um nome de domínio, derivado de uma requisição, em endereço *Internet Protocol* (IP), possibilitando que o cliente possa acessar diretamente o recurso [Mockapetris, 1987b].

3.1.1. TLD

Na nomenclatura do DNS dos computadores, há uma hierarquia de nomes, chamados de "nomes de domínio de topo" ou *top-level domain names* (TLDs). Podem ser encontrados os TLDs genéricos (gTLD), tais como: EDU, COM, NET, ORG, GOV, MIL e INT, criados para uma categoria geral de organizações. Os domínios de códigos de países ou *country code top-level domain* (ccTLD) inclui por exemplo: BR, FR, NL, KR e US, e são organizados por um administrador para esse país. Esses administradores podem delegar ainda o gerenciamento de partes da árvore de nomes [Postel, 1994].

Sob cada TLD pode ser criada uma hierarquia de nomes. Geralmente, sob os TLDs genéricos, a estrutura é muito plana, ou seja, muitas organizações são registradas diretamente sob o TLD, e qualquer outra estrutura abaixo dessa árvore, depende das organizações individuais [Postel, 1994].

3.1.2. Recursos do DNS

Um nome de domínio identifica um nodo da rede. Cada nodo contém um conjunto de informações sobre seus recursos. O conjunto de informações associados a um nome de domínio específico é composto por diferentes registros de recursos (*resource records*) ou RRs [Mockapetris, 1987a].

Em outras palavras, cada requisição DNS visa buscar uma informação (RR) de um domínio específico. Por exemplo, o endereço IP de um domínio pode ser requisitado usando o RR do tipo A para IPv4 ou do tipo AAAA para o endereço IPv6 da máquina que hospeda o serviço.

O DNS também utiliza da técnica de *Anycast*, que nada mais é a atribuição de um endereço para mais de uma réplica do serviço em questão. Muitas vezes seu uso visa facilitar o roteamento para estas réplicas, melhorando assim a latência do serviço [Partridge; Mendez; Milliken, 1993].

3.1.3. Autoritativo e Recursivo

Com relação a interação entre os resolvedores DNS durante uma requisição, se encontram os resolvedores autoritativos (*authoritative resolvers*) e recursivos (*recursive resolvers*).

Um Resolvedor Autoritativo é um servidor que conhece o conteúdo de uma zona DNS, podendo encontrar o próximo nível definido na hierarquia sem precisar requisitar outros servidores [Mockapetris, 1987a].

Atualmente há 13 *root-servers* autoritativos, sendo eles representados pelas letras de A a M, onde pode ser encontrada várias réplicas *anycast* dentro de cada uma das letras. Entre todos os *root-servers* há 918 instâncias espalhadas pelo mundo [ROOT-SERVERS, 2018].

O Resolvedor Recursivo é aquele que obterá consultas de um grupo de clientes e iterativamente encaminhará a requisição para outro servidor, que retorna o endereço da zona de nível inferior da hierarquia DNS, até que o resolvedor encontre o endereço desejado pelo cliente e o retorne, permitindo que o cliente faça sua requisição para o servidor requisitado [Mockapetris, 1987a].

Geralmente o Resolvedor Recursivo é um serviço prestado pelos *Internet Service Providers* (ISPs) e atende a vários clientes. Pode armazenar respostas em sua memória *cache* por um período de tempo. Se receber uma consulta cuja resposta já esteja armazenada no *cache*, o Resolvedor Recursivo responderá a partir do *cache* [Mockapetris, 1987a].

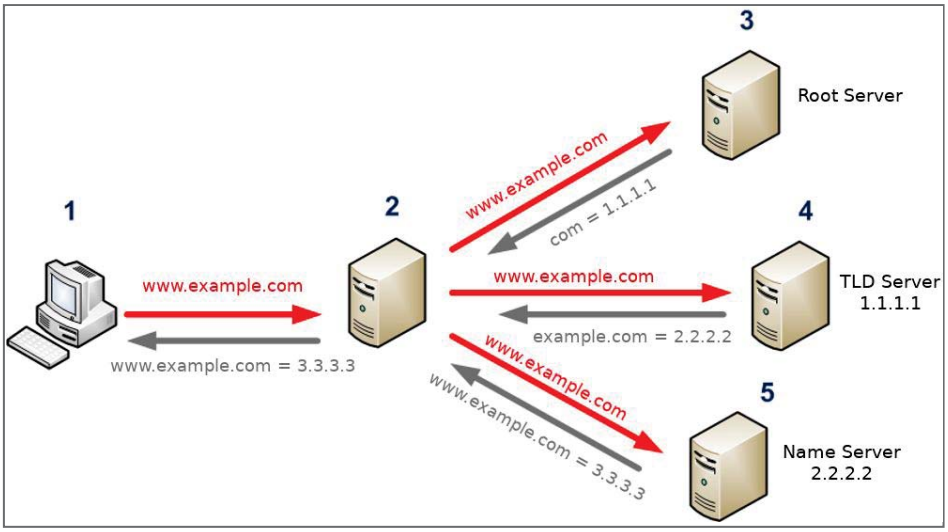

Figura 1. Caminho de uma resolução DNS.

Na Figura 1, é encontrada uma breve explicação de como uma requisição DNS ocorre, partindo do servidor *Stub* (1), que consulta o servidor recursivo (2), onde este passa a requisitar os servidores autoritativos (3, 4 e 5). Cada resposta que o recursivo recebe é parte da resolução, onde este passa a consultar o próximo servidor e assim por diante, até que a resolução se complete. Os servidores *Stub* (1) não serão retratados neste trabalho.

3.1.4. TTL e *Caching*

O *Time-To-Live* (TTL) especifica o intervalo de tempo em que o registro de recurso (RR) pode ficar armazenado na *cache* do resolvedor recursivo. Os valores zero são interpretados para significar que o RR só pode ser usado para a transação em andamento e não deve ser armazenado em *cache*. Valores zero também podem ser usados para dados extremamente voláteis [Mockapetris, 1987b].

A fim de reduzir a latência e diminuir a carga do serviço DNS, os servidores recursivos armazenam as respostas obtidas dos servidores de autoridade em *cache*. Assim quando novos pedidos chegarem no servidor recursivo, este simplesmente retorna o valor armazenado. Quando o TTL da requisição vence, este é eliminado da *cache* e na próxima vez que for requisitado, o resolvedor recursivo irá solicitar novamente o servidor autoritativo e demais servidores envolvidos.

3.2. DNS-OARC e DITL

A DNS *Operations, Analysis, and Research Center* (DNS-OARC) é uma plataforma que reúne principais operadores, implementadores e pesquisadores, para que eles possam coordenar as respostas a ataques, compartilhar informações entre outras preocupações no que diz respeito ao DNS [DNS-OARC, 2018].

Anualmente é realizada por esta plataforma uma coleta passiva do tráfego DNS, através do projeto *Day In The Life of the Internet* (DITL), onde todas as requisições dos servidores recursivos que chegam aos *root-servers* neste dia são coletados e armazenados, para que possam ser organizados e analisados posteriormente.

Eles oferecem, também, acesso a pesquisadores e membros do OARC através do uso de um conjunto de máquinas, objetivando que os pesquisadores tragam suas ferramentas para realizar análises e quaisquer estudos necessários ao seu trabalho [DITL, 2018].

A DNS-OARC não permite a divulgação de dados sensíveis, como endereços IP ou informações que possam ser derivadas desse tipo de dado. Portanto para este trabalho informação de identificação foram abstraídos.

4. Metodologia

Esta seção apresenta a base de dados utilizada, além das etapas de mineração para a extração e análise do tráfego DNS.

4.1. Base de Dados

Para este trabalho foi utilizado o conjunto de dados coletados da DITL de 2018, que ocorreu no dia 11 de abril de 2018. Esta captura contêm o *dump* completo do tráfego que ocorreu dos servidores recursivos para os *root-servers* neste dia em específico, e para cada uma das réplicas *anycast* do *root-server*.

Por razões de segurança e privacidade, estes dados são coletados e mantidos pelo DNS-OARC, e disponibilizados somente para alguns pesquisadores, que obtém acesso remoto a uma máquina via SSH, com acesso somente de leitura às coletas completas, inclusive das demais coletas realizadas de anos anteriores. Esses dados estão organizados por ano, letra do *root-server* e réplica *anycast* do servidor.

Por exemplo, em uma das cópias do *anycast* nyc (Nova Iorque) do root A, há cerca de 300 arquivos .pcap.gz com cerca de 100 MB cada, contemplando cinco minutos de coleta das requisições DNS recebidas por esta réplica neste dia (Tabela 1).

Tabela 1. Lista de arquivos encontrado nos servidores da DNS-OARC.

@DNS-OARC: DITL-2018/a-root/nr1-nyc3-a	
20180411.00:00:00.pcap.gz	120M
20180411.00:05:00.pcap.gz	120M
20180411.00:10:00.pcap.gz	120M
...	
20180411.08:00:00.pcap.gz	102M
20180411.08:05:00.pcap.gz	101M
20180411.08:10:00.pcap.gz	110M
...	
20180411.16:00:00.pcap.gz	115M
20180411.16:05:00.pcap.gz	114M
20180411.16:10:00.pcap.gz	113M
...	

No conjunto de dados de 2018, não estava disponível o G-ROOT. Portanto, os dados apresentados aqui levam em consideração somente as outras 12 letras dos *root-servers*: A, B, C, D, E, F, H, I, J, K, L e M.

4.2. Mineração

Os *scripts* utilizados para a mineração dos dados foram em sua maioria escritos em *bash* (*Bourne-Again SHell*) e utilizando ferramentas como *tshark* (versão do *wireshark* para linha de comando) e *awk* (linguagem utilizada para processamento de texto), uma vez que o acesso concedido às máquinas do DNS-OARC era sobre *ssh* (*Secure Shell*).

Diante da imensa quantidade de requisições DNS capturadas pelos *root-servers* no dia da coleta, o *script* foi desenvolvido para somente separar as requisições com TLDs de nome .com, .nl e .cn, e para os RRs dos tipos 1, 2, 28 e 43, além de selecionar o endereço IP do servidor recursivo e o carimbo de data e hora (*timestamp*) que esta solicitação DNS ocorreu.

O TLD .com foi escolhido por ser o maior domínio gTLD. Os outros dois são domínios ccTLD, .cn é o domínio usado na República popular da China e foi escolhido por ser o maior ccTLD [DENIC, 2018]. O .nl é o domínio usado para os Países Baixos (Netherlands), é um dos ccTLDs mais populares, com mais de 5,8 milhões de nomes de domínio registrados [SIDNLABS, 2018], e também pelo fato que os dados utilizados para este artigo, foi concedido pela equipe que gerencia o ccTLD .nl.

Para os RRs, foi escolhido somente alguns deles, sendo 1, 2, 28 e 43, a saber:

- 1 ou A (*Address record*): é um dos tipos de registro mais usados em qualquer sistema DNS. Usado para mapear um nome de domínio para um endereço IPv4 de 32 bits;
- 2 ou NS (*Name server record*): é usado para delegar um subdomínio a um conjunto de servidores de nomes;
- 28 ou AAAA (*IPv6 address record*): usado para mapear um nome de domínio para um endereço IPv6 de 128 bits;
- 43 ou DS (*Delegation signer*): usado para identificar a chave de assinatura DNSSEC de uma zona delegada.

Para se ter uma visão geral dos dados, apresentados na seção de análise dos dados (Figura 2), foi gerado um *script* para extração da contagem das requisições de cada *root-server*, classificados pelo tipo de consulta através do RR e pela zona desejada, através do TLD, que constam na requisição capturada.

Em seguida os dados passaram por um *script* para separação das requisições por servidores recursivos, sendo que para cada IP foi criado um arquivo contendo todas as requisições ocorridas neste dia. Isso facilitou o posterior processo de análise a fim de encontrar a frequência das requisições por servidor recursivo.

4.3. Resumo dos Dados *Minerados*

Ao final do processo de separação dos dados, o número de total foi de cerca de 120 milhões de requisições para os 12 *root-servers*, os 3 TLDs e os 4 RRs usados. Após coletados, os dados passaram pelas etapas de limpeza e transformação, onde 1.526.723

(1,27% do total) requisições foram removidas por conterem alguns dos campos em branco. Essas consultas foram removidas pois poderiam trazer dados inadequados para a análise.

Foram encontrados também cerca de 833 mil servidores recursivos, que fizeram pelo menos uma requisição para alguma das réplicas *anycast* dos *root-servers* neste dia.

5. Análise dos resultados

Nesta seção será apresentada a análise sobre os dados minerados, analisando a distribuição das requisições e o (inadequado) comportamento de alguns resolvedores recursivos.

5.1. Distribuição dos dados

Dentre todas as requisições ocorridas no DITL de 2018, a Figura 2 nos mostra sua distribuição. Onde perceptivelmente o RR mais requisitado pelos resolvedores recursivos, com exceção do B-ROOT, foi o NS com 65,9%, justamente por ser uma das principais funções dos *root-servers*, que é delegar a requisição para a zona que irá conter a informação. (É possível notar que a quantidade menor de tráfego recebido pelo B-ROOT se dá pelo fato de que esse possui somente duas réplicas *anycast* enquanto outros *root-servers* contém até centenas de réplicas.)

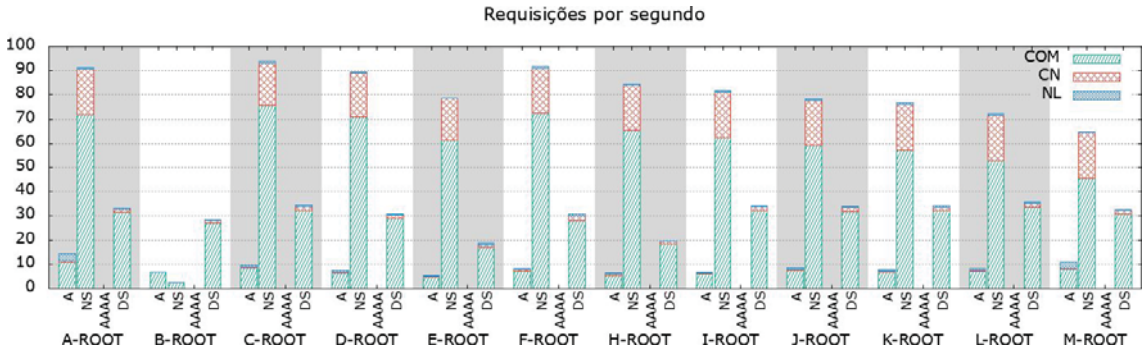

Figura 2. Requisições por segundo para cada um dos *root-servers*, em diferentes RRs e TLDs.

É possível observar que somente o RR do tipo *Name Server* (NS) atingiu cerca de 900 requisições por segundo nos *root-servers*, considerando o conjunto de dados em questão. O segundo RR mais requisitado entre os *root-servers* foi o DS com 26,7%; aproximadamente 367 requisições por segundo. O RR DS é utilizado para o controle de autenticidade de uma zona no DNS.

Quanto a distribuição dos TLDs, foi observado que 82% das requisições foram para o gTLD .com, representando mais de 1.100 requisições por segundo, ou seja, mais de uma por milisegundo. Com relação aos outros TLDs, 16,3% das requisições foram para o ccTLD .cn e 1,7% das requisições para o ccTLD .nl.

Ao relacionar os IPs de servidores recursivos que mais aparecem na lista de requisições a *root-servers*, observou-se que apenas alguns IPs são responsáveis por grande parte de todas as requisições recebidas durante o dia de medições.

5.2. Comportamento do TTL

O tempo sugerido de *caching* pelos *root-servers* para os RRs usados são apresentados na Tabela 2, onde para a maioria dos RRs o valor sugerido é de dois dias. Para o RR do tipo DS o valor recomendado é de um dia, devido ao fato de que este trata de questões de segurança.

Tabela 2. Tempo de *caching* recomendado pelos *root-servers*.

RR	TTL (segundos)	TTL (horas)
A	172800	48
AAAA	172800	48
DS	86400	24
NS	172800	48

Dentro de uma situação ideal, os resolvedores recursivos não deveriam fazer mais de quatro requisições por dia (explicado em detalhes na seção 5.3). Porém isso não é o que foi percebido por alguns dos servidores recursivos encontrados no conjunto de dados usado.

A Tabela 3 apresenta os 10 IPs que enviaram as maiores quantidades de requisições aos *root-servers*. Esses correspondem a cerca de 52% do número total de requisições para os RRs e TLDs usados, totalizando cerca de 720 consultas por segundo. Nesta pesquisa foi definido o comportamento desses recursivos como *agressivo*, e esse termo será utilizado no decorrer deste trabalho.

Tabela 3. Top 10 IPs com maior número de requisições.

IP	Número de requisições (%)	Requisições por segundo
1	19092556 (16,08%)	220,98
2	8057213 (6,79%)	93,25
3	6271289 (5,28%)	72,58
4	5969248 (5,03%)	69,09
5	4404428 (3,71%)	50,98
6	4156020 (3,50%)	48,1
7	3950306 (3,33%)	45,72
8	3890341 (3,28%)	45,03
9	3677436 (3,10%)	42,56
10	2721005 (2,29%)	31,49
SOMA	**62189842 (52,38%)**	**719,79**

Os recursivos que fizeram de uma a quatro requisições durante o dia de medições estão dentro do comportamento esperado, ou seja, estão respeitando o TTL da maioria dos RRs. O termo comportamento esperado também é utilizado no decorrer deste trabalho. Esses correspondem a cerca de 753 mil resolvedores recursivos (90,5% do total de IPs observados), que dispararam cerca de 1 milhão de requisições, correspondendo a menos de 1% do número total de requisições observadas.

Isso mostra que menos de 10% dos resolvedores recursivos fazem 5 ou mais requisições por dia e isso corresponde a mais de 99% das requisições observadas nos *root-servers* neste dia.

Agrupando os IPs de comportamento agressivo, percebeu-se que muitos deles associaram-se em intervalos de endereços que pertencem a algumas empresas. Os IPs listados na Tabela 3 pertencem somente a três empresas. Essa análise foi realizada utilizando a ferramenta *whois* e os resultados são apresentados na Tabela 4.

Tabela 4. Top 10 empresas com maior número de requisições.

Empresa	Número de requisições (%)	Requisições por segundo
1	**51611160 (43,47%)**	**597,35**
2	**15847739 (13,35%)**	**183,42**
3	**10557354 (8,89%)**	**122,19**
4	958469 (0,81%)	11,09
5	858019 (0,72%)	9,93
6	744906 (0,63%)	8,62
7	684132 (0,58%)	7,92
8	665690 (0,56%)	7,7
9	659165 (0,56%)	7,63
10	625249 (0,53%)	7,24
SOMA	**83211883 (70,08%)**	**963,1**

Agrupando os IPs de servidores recursivos das 10 empresas que mais enviaram requisições aos *root-servers*, observou-se que a fatia do total de requisições observadas no dia de medições aumenta de 52,38% (Tabela 3) para 70,08% (Tabela 4). As três empresas cujo recursivos mais enviaram requisições, geraram mais de 900 requisições por segundo.

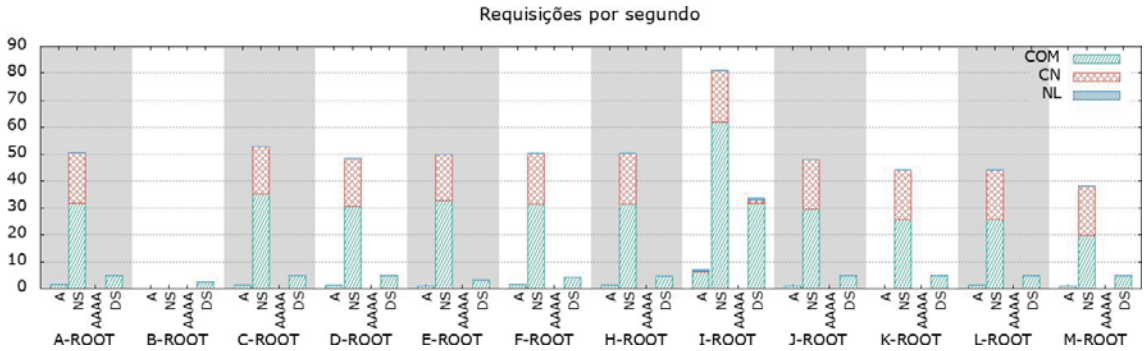

Figura 3. Requisições por segundo das empresas mais agressivas.

Pode-se observar na Figura 3 que a agressividade encontrada por parte das empresas ocorrem majoritariamente no RR do tipo NS com cerca de 557 (84,6%) requisições por segundo. A exceção é o B-ROOT, onde não há registros de requisições originadas de servidores recursivos das empresas mais agressivas.

Também pode ser percebido que o I-ROOT teve uma taxa de requisições por segundo superior das demais, tanto nas requisições do tipo NS quanto nas do tipo DS. Também há uma leve vantagem nas requisições do tipo A do I-ROOT em relação aos demais.

Foi encontrado que mais de 99% das requisições que chegaram no I-ROOT neste dia, foram oriundas dos IPs das dez empresas apresentadas na Tabela 4. Isso pode ser observado, se comparando o número total de requisições (Figura 2), onde é encontrado cerca de 123 requisições por segundo para o I-ROOT, com cerca das 121 requisições por segundo somente das dez empresas mais agressivas encontradas (Figura 3).

5.3. Melhor e pior cenário

Em teoria, cada IP deveria consultar no máximo quatro vezes um *root-server* para cada TLD, pelo fato que na coleta realizada nesta pesquisa foi usado somente quatro RRs, e levando em consideração que o TTL de ambos os RRs fossem de 86400 segundos (1 dia). Logo o número máximo de requisições por dia, em um cenário ideal, seria um valor em torno de 10 milhões, visto que no conjunto de dados foram encontrados cerca de 833 mil IPs, pouco mais de 115,68 requisições por segundo. Porém como percebido na Tabela 3, somente o IP número 9 contém mais requisições do que o máximo esperado por todos os IPs juntos.

Contudo, levando em consideração que todos os IPs se comportassem de acordo com a média dos dez IPs mais agressivos, observado na Tabela 3, então os *root-servers* iriam receber perto de 60 milhões de requisições por segundo, ou seja, mais de 5 trilhões de requisições por dia, somente para estes RRs e TLDs. Para fins de comparação, o conjunto dados utilizado contém cerca de 120 milhões de requisições, recebidas em 24 horas.

6. Conclusão

Nesta pesquisa foi analisado o comportamento dos servidores recursivos perante os *root-servers*, baseado no TTL sugerido nas respostas das resoluções DNS. Para isso, foi usado o conjunto de dados de 2018 capturados pelo projeto DITL, da DNS-OARC, com aproximadamente 120 milhões de requisições. Na análise realizada, a distribuição dos dados mostra que a grande maioria das requisições que ocorreram foram, dentre os RR e domínios estudados, ao tipo NS (2).

Observou-se que servidores recursivos desconsideram o tempo de *cache* sugerido pelos *root-servers*, onde diariamente fazem um número de requisições muito maior do que o teoricamente proposto. Esse comportamento se agrava quando foi agrupado, neste trabalho, as requisições por faixas de IPs, que fazem parte de um pequeno número de empresas. Verificou-se também que a maior parte do tráfego que chegou no I-ROOT, são dessas empresas.

Por fim, foi simulado um cenário em que todos os servidores recursivos se comportassem como o previsto (“bom comportamento”) ou como os dez IPs mais agressivos. No segundo caso constatou-se que poderia chegar em uma situação em que os *root-servers* receberiam até 60 milhões de requisições por segundo.

Como trabalho futuro pode-se analisar as consequências do comportamento agressivo de servidores recursivos aqui apresentado, como agravante durante um ataque de negação de serviço, que tem como alvo os *root-servers*.

Agradecimentos

DNS-OARC pelo acesso concedido aos dados do DITL.

Referências

Bhatti, S., Atkinson, R. (2011). Reducing DNS caching. 2011 IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2011, pages 792 – 797.

Castro, S., Wessels, D., Fomenkov, M., Claffy, K. C. (2008). A Day at the Root of the Internet. Computer Communication Review, 38:41–46.

DENIC (2018). Statistics – Comparison of International Domain Numbers. Disponível: *https://www.denic.de/en/know-how/statistics/monthly-analytics-international-domains/*. Acesso: setembro/2018.

DITL (2018). Day In The Life of the Internet. Disponível: *https://www.dns-oarc.net/oarc/data/ditl*. Acesso: julho/2018.

DNS-OARC (2018). The DNS Operations, Analysis, and Research Center. Disponível: *https://www.dns-oarc.net/*. Acesso: julho/2018.

Freitas, D. L. B., Barbosa, K. R. S., Feitosa, E. (2013). MyDnsDump: Uma ferramenta para medição do tráfego DNS. *ISSN 2238-5096 (CDR).* Disponível: http://marconeds.com.br/eventos/artigos/completos1/112514_1.pdf. Acesso: agosto/2018.

Mockapetris, P. (1987a). Domain Names - Concepts and Facilities. RFC 1034, Internet Engineering Task Force. Disponível: https://www.ietf.org/rfc/rfc1034.txt. Acesso: agosto/2018.

Mockapetris, P. (1987b). Domain names - implementation and specification. RFC 1035, Internet Engineering Task Force. Disponível: https://www.ietf.org/rfc/rfc1035.txt. Acesso: agosto/2018.

Partridge, C., Mendez, T., Milliken, W. (1993). Host Anycasting Service. RFC 1546, Internet Engineering Task Force. Disponível: https://www.ietf.org/rfc/rfc1546.txt. Acesso: outubro/2018.

Postel, J. (1994). Domain Name System Structure and Delegation. RFC 1546, Internet Engineering Task Force. Disponível: https://www.ietf.org/rfc/rfc1591.txt. Acesso: outubro/2018.

ROOT-SERVERS (2018). Root Server Technical Operations Assn. Disponível: *http://www.root-servers.org/*. Acesso: outubro/2018.

SIDNLABS (2018). .nl stats and data. Disponível: https://stats.sidnlabs.nl/en/. Acesso: setembro/2018.

Vlajic, N., Andrade, M., Nguyen, U. (2012). The Role of DNS TTL Values in Potential DDoS Attacks: What Do the Major Banks Know About It?. Procedia Computer Science,10:466–473.

WICAKSANA, Muhammad Arif. IPV4 VS IPV6 ANYCAST CATCHMENT: A

ROOT DNS STUDY. 2016. 77 p. Tese (Doutorado) - Curso de Computer Science, University Of Twente, Enschede, 2016. Disponível: *http://essay.utwente.nl/70921/1/WICAKSANA_MA_EEMCS.pdf*. Acesso: outubro/2018.

Wills, C. E., Shang, H. (2000). The Contribution of DNS Lookup Costs to Web Object Retrieval.

Zyl, I. V., Rudman, L., Irwin, B. (2015). A review of current DNS TTL practices. SATNAC 2015.