Análise Longitudinal de Dados do DNSpot

Tiago Heinrich

Joinville

2017

Tiago Heinrich

Análise Longitudinal de Dados do DNSpot

Relatório de Trabalho de Conclusão de Curso (TCC) apresentado ao Curso de Graduação Bacharelado em Ciência da Computação, da Universidade do Estado de Santa Catarina (UDESC), como requisito parcial da disciplina de Trabalho de Conclusão de Curso.

Orientador: Prof Rafael Rodrigues Obelheiro

Doutor

Joinville

2017

Tiago Heinrich

Análise Longitudinal de Dados do DNSpot

Relatório de Trabalho de Conclusão de Curso (TCC) apresentado ao Curso de Bacharelado em Ciência da Computação da UDESC, como requisito parcial para a obtenção do grau de BACHAREL em Ciência da Computação.

BANCA EXAMINADORA

Prof Rafael Rodrigues Obelheiro

Doutor

Prof Charles Christian Miers

Doutor

Prof Guilherme Piegas Koslovski

Doutor

Agradecimentos

Agradeço a todos os amigos e professores que estiveram presente nesta jornada acadêmica, principalmente ao Dr. Rafael Rodrigues Obelheiro por todo o apoio e motivação. Agradeço ainda ao meu irmão Rodrigo Heinrich, pelo apoio e amizade, e aos meus pais, Catia Arlene Hoeller Heinrich e Arlindo Heinrich por sempre estarem ao meu lado.

“dig @200.19.107.235 udesc.br. ANY +edns=0 +notcp +ignore” - Bash

Resumo

Devido ao crescimento e popularização da Internet nos últimos anos, ataques como *Distributed Denial of Service* (DDoS) estão cada vez mais frequentes, sendo difícil encontrar usuários que ao realizar algum serviço *online* não tenha sofrido com um ataque. O monitoramento do tráfego de um serviço de *Domain Name System* (DNS) recursivo, pode ser realizado para a descoberta de ataques DDoS. Uma ferramenta com estas características é o DNSpot, sendo responsável por registrar o tráfego de envio a um serviço de DNS recursivo aberto. Este trabalho de conclusão de curso apresenta uma análise sobre os dados coletados pelo DNSpot em um período de monitoramento de 250 dias, realizando uma análise de tendências sobre os ataques de DDoS.

Palavras-chave: *Domain Name Systeam (DNS)*, *Segurança em redes*, *DNSpot*.

Abstract

Due to the growth and popularization of the Internet in recent years, attacks such as Distributed Denial of Service (DDoS) are increasingly frequent, and it is difficult to find users who have not suffered an attack when performing an online service. Traffic monitoring of a recursive Domain Name System (DNS) service can be performed for the discovery of DDoS attacks. A tool with these characteristics is the DNSpot, being responsible for registering the sending traffic to an open recursive DNS service. This course completion paper presents an analysis of the data collected by DNSpot over a 250 day monitoring period, performing a trend analysis on DDoS attacks.

Keywords: Domain Name Systeam (DNS), Network Security, DNSpot.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

C&C	Comunicação e Controle
DNS	*Domain Name System*
DoS	*Denial of Service*
DDoS	*Distributed Denial of Service*
GTI	*Global Terrorism Index*
HDD	*Hard Disk Drive*
IDS	*Intrusion Detection System*
IP	*Internet Protocol*
KAIST	Korea Advanced Institute of Science and Technology
MIT	Massachusetts Institute of Technology
NAT	*Network Address Translation*
RR	*Resource Record*
SIP	*Session Initiation Protocol*
TTL	*Time to Live*

1 Introdução

O DNS (*Domain Name System*) (MOCKAPETRIS, 1987a) é um sistema distribuído de resolução de nomes que desempenha um papel fundamental na Internet. Sua principal funcionalidade é traduzir nomes de domínio mais facilmente memorizáveis (como `www.udesc.br`) em endereços IP (como 200.19.105.194), que são usados pelos protocolos subjacentes de rede para localizar e identificar nós na Internet.

Em vista de sua ampla utilização, o DNS também é tanto um alvo quanto um vetor de ataques. As principais ameaças envolvendo o DNS são resumidas por (CONRAD, 2012), que as divide em duas classes: aquelas em que o DNS é o alvo e aquelas que são oportunizadas pelo DNS. A classe de ameaças ao DNS inclui:

- Negação de serviço: impedir o acesso de usuários ao DNS, com isso prejudicando ou mesmo bloqueando o seu acesso à Internet;
- Corrupção de dados: modificar dados publicados no DNS de forma não autorizada, o que pode, por exemplo, levar usuários a acessar sites ilegítimos (como páginas falsas de bancos ou comércio eletrônico); e
- Exposição de informação: revelar informações sobre o comportamento dos usuários, como histórico de sites web acessados.

O DNS também pode ser usado como um veículo de ataques. A classe de ameaças oportunizadas pelo DNS abrange:

- Ataques de amplificação: servidores DNS mal configurados podem ser usados para realizar ataques de negação de serviço contra terceiros (CERT.BR, 2014);
- *Fast flux* DNS: servidores usados para propósitos nefastos, como propagação de *software* malicioso ou controle remoto de *botnets*, podem ter diversos endereços IP distintos associados. Uma fração desses endereços são associados a um nome DNS específico e trocados com alta frequência, de modo a dificultar a localização dos servidores e a identificação dos seus responsáveis, e até mesmo balancear carga entre servidores (SALUSKY; DANFORD, 2007); e

- Exfiltração de dados: como o tráfego DNS geralmente não é barrado ou modificado por *firewalls*, ele é usado com frequência para transmitir dados sensíveis (capturados no curso de uma invasão) sem que isso seja percebido pelos mecanismos de defesa da rede.

Para observar o comportamento de atacantes contra servidores DNS, foi desenvolvido o DNSpot (LONGO, 2015), um *honeypot* DNS com o propósito de monitorar e registrar o tráfego enviado a um serviço de DNS recursivo aberto. *Honeypots* são recursos computacionais de segurança, cujo objetivo é serem sondados, atacados ou comprometidos em um ambiente controlado (STEDING-JESSEN; VIJAYKUMAR; MONTES FILHO, 2008).

Em um primeira etapa de desenvolvimento do DNSpot (LONGO, 2015), foi implantado na rede da UDESC durante 49 dias, entre 09/09/2015 e 28/10/2015. Nesse período, o *honeypot* processou mais de 4 milhões de consultas DNS, mais de 99% das quais relacionadas a ataques distribuídos de negação de serviço (DDoS, *Distributed Denial of Service*). A análise dos dados coletados revelou a existência de nomes DNS projetados para maximizar a amplificação de tráfego nesse tipo de ataque. Conforme mostrado na Tabela 1.1, nove dos 10 nomes ou registros de recursos (RRs, *resource records*) que apareceram com maior frequência não podem mais ser aproveitados em ataques DDoS, seja porque não estão mais ativos ou porque agora geram respostas consideravelmente menores; uma nova verificação dos dados foi realizada em 19/08/2016, cerca de 10 meses após o fim do estudo original. Um outro fato observado no estudo original foi o desaparecimento de domínios usados em ataques DDoS; isso foi constatado especificamente para o domínio l3x.ru, que aparece na Tabela 1.1.

Este trabalho de conclusão de curso realizou uma coleta de dados com o DNSpot durante um período de 250 dias, colaborando para uma análise detalhada dos ataques. A análise evolutiva das atividades maliciosas foi realizada ao longo de 250 dias, possibilitando solucionar as seguintes questões:

- Qual a diferença entre as transações realizadas com o DNSpot neste período de 250 dias com o trabalho (LONGO, 2015)?
- Quais os volumes de dados recebidos neste período?
- Os clientes apresentaram algum tipo de comportamento ou alguma diferença em relação ao trabalho (LONGO, 2015)?
- Os Domínios e RRs apresentam alguma diferença em relação ao trabalho anterior?

RR	Ativo?	Resposta (bytes)	
		2015	2016
hehehe.ru. ANY	sim	3850	221
mototrazit.ru. ANY	não	3853	–
vp47.ru. ANY	sim	3959	151
l3x.ru. A	não	3875	–
. ANY	sim	1503	1790
gransy.com. ANY	sim	3591	594
vp47.ru A	sim	3892	91
lifemotodrive.ru. ANY	não	3969	–
nhl.msk.su. ANY	sim	3965	341
oi69.ru. A	sim	3637	91

Tabela 1.1: Situação em 19/08/2016 dos 10 RRs observados com maior frequência por Longo (2015)

Fonte: O próprio autor

- Qual o comportamento dos ataques DoS observados?
- Correlacionar o tráfego de ataques DDoS com alguns fatores externos, como questões geopolíticas ou econômicas?
- Quais anomalias foram observadas?

Devido ao período de coleta, algumas características foram identificadas e mudanças em certos comportamentos, não antes observados em análises realizadas em períodos pequenos de tempo.

1.1 Objetivos

Objetivo geral: Fazer uma análise da evolução temporal de dados coletados pelo DNSpot. Visando um período de coleta entre oito e nove meses, que contribuiu com resultados que não conseguiram ser observados no primeiro estudo, devido ao tempo.

Objetivos específicos: Segue uma lista dos principais objetivos abordados neste estudo:

- Realizar uma revisão bibliográfica abrangendo segurança do DNS, *honeypots* e trabalhos relacionados;
- Operacionalizar armazenamento de longo prazo no DNSpot;
- Fazer uma coleta de longa duração;
- Comparar os resultados deste estudo com os trabalhos anteriores; e
- Analisar a evolução temporal dos dados observados.

1.2 Metodologia

Este trabalho de conclusão de curso consiste em uma pesquisa aplicada, tendo como principais métodos a pesquisa bibliográfica e o estudo de caso. Primeiramente, foi realizado um estudo sobre o DNSpot e uma revisão bibliográfica sobre aspectos de segurança do DNS e *honeypots*. Foi realizando uma análise para avaliar a necessidade de adaptações no DNSpot para coleta de dados de longo prazo. Como o DNSpot realiza o armazenamento dos dados diretamente no banco de dados, algumas medidas para diminuir o volume do banco de dados foram tomadas, para garantir o correto funcionamento do DNSpot.

Após, foi iniciada a coleta de dados para a realização da análise, juntamente com o estudo. Por fim, foram realizadas a comparação com os dados originais (LONGO, 2015) e análise evolutiva dos ataques observados pelo DNSpot.

1.3 Organização do Texto

Este trabalho está dividido em cinco capítulos. O Capítulo 2 apresenta a fundamentação teórica, abrangendo DNS, *honeypots* e DNSpot. O Capítulo 3 discute trabalhos encontrados na literatura envolvendo análises sobre o tráfego DNS. O Capítulo 4 apresenta a análise dos resultados encontrados. O Capítulo 5 discute limitações de escalabilidade do DNSpot. O Capítulo 6 apresenta as considerações finais deste trabalho e perspectivas de trabalhos futuros.

2 Fundamentação Teórica

Este capítulo introduz os conceitos necessários para o entendimento deste estudo e as conclusões encontradas. Buscando destacar os principais tópicos para o entendimento deste trabalho.

O capítulo está divido em cinco seções. A Seção 2.1 apresenta uma definição e breve contextualização do *Domain Name System* (DNS), discutindo a sua funcionalidade, a sua importância para a Internet e alguns aspectos de segurança da sua estrutura. A Seção 2.2 discute o *honeypots* e a sua utilização. Na Seção 2.3 é descrito o DNSpot, um *honeypot* específico para servidores DNS recursivos, que foi utilizado para a realização deste estudo. A Seção 2.4 apresenta os resultados encontrados na primeira análise realizada pelo DNSpot. Por fim, na Seção 2.5 é apresentado as considerações do capítulo.

2.1 Domain Name System

O *Domain Name System* (DNS) desempenha uma funcionalidade essencial para a operação da Internet, sendo responsável por, entre outras funcionalidades, realizar a associação de um nome de domínio com um endereço IP. O sistema é implementado como uma estrutura hierárquica, possuindo servidores raiz, que são responsáveis por atualizar a lista de nomes e endereços IPs (GAO, 2013).

2.1.1 Definição

DNS é um sistema distribuído de banco de dados, cujo objetivo original era permitir que recursos de rede fossem identificados por nomes em vez de endereços de baixo nível (MOCKAPETRIS, 1987a). Em particular, o DNS permite que usuários refiram-se a nós da rede usando nomes (como `www.udesc.br`) no lugar dos endereços IP (como 200.19.105.51) efetivamente usados para a comunicação com esses nós. Ao longo do tempo, o escopo do DNS foi ampliado, basicamente devido à sua ampla disseminação, passando a associar vários tipos diferentes de dados a nomes de domínio (MOCKAPETRIS, 1987a). Dada a sua estrutura com abrangência global e a sua ubiquidade, o DNS necessita escalabilidade e desempenho, oferecendo para os

usuários baixa latência em redes de larga escala (JUNG, 2002). O DNS é crítico para o funcionamento da maioria dos serviços encontrados na Internet: embora sempre seja possível referir-se a um nó (por exemplo, um servidor web) usando seu endereço IP, os usuários buscam resolver os endereços utilizando o seu nome (ZDRNJA; BROWNLEE; WESSELS, 2007). Além disso, o DNS introduz uma camada de indireção, permitindo, por exemplo, que um nó mude seu endereço IP de forma transparente para os usuários e aplicações que desejem se comunicar com ele.

2.1.2 Hierarquia

O espaço de nomes do DNS segue uma estrutura em árvore (MOCKAPETRIS, 1987a). A cada nó da árvore, seja um nó interno ou uma folha, corresponde um conjunto de registros de recursos (*resource records*, RRs), que pode ser vazio. Cada nó possui um rótulo com até 63 bytes de comprimento. Nós irmãos devem ter rótulos distintos, como por exemplo na Figura 2.1 o nó `gov` não pode conter dois `treasury` como filhos, o rótulo `treasury` só pode ser utilizado para nós que não são irmãos. A raiz da árvore possui um rótulo com comprimento zero, tipicamente representado por um ponto (“.”). O nome de domínio de um nó é a lista de rótulos que formam o caminho do nó até a raiz da árvore. Por exemplo, na árvore DNS ilustrada na Figura 2.1, o nó mais à esquerda possui o nome de domínio `irs.treasury.gov.`; em geral, o ponto final é omitido, quando isso não causar ambiguidade. Na Figura 2.1 é apresentado a coexistência do rótulo “irs” em duas subárvores diferentes. Outro rótulo que pode ser encontrado em diferentes subárvores é o “www”, como por exemplo `www.google.com` e `www.udesc.br`.

Na árvore DNS, cada subárvore é um domínio. Um conceito chave no DNS é a administração descentralizada, que consiste em delegar a administração de domínios a entidades autônomas (MOCKAPETRIS; DUNLAP, 1988). A administração de um domínio engloba a criação de nós nesse domínio e a definição dos recursos associados a nomes pertencentes ao domínio. Cada domínio possui um ou mais servidores responsáveis pelos seus nomes, chamados de servidores autoritativos. Em geral, esses servidores são configurados de forma que um deles (servidor primário ou mestre) é o repositório central de dados para o domínio, e os demais (servidores secundários ou escravos) apenas replicam os dados do servidor primário para oferecer balanceamento de carga e tolerância a falhas (ALBITZ; LIU, 2006). A divisão entre servidores mestres e escravos é visível apenas para os administradores de domínios, sendo transparente para os usuários do DNS.

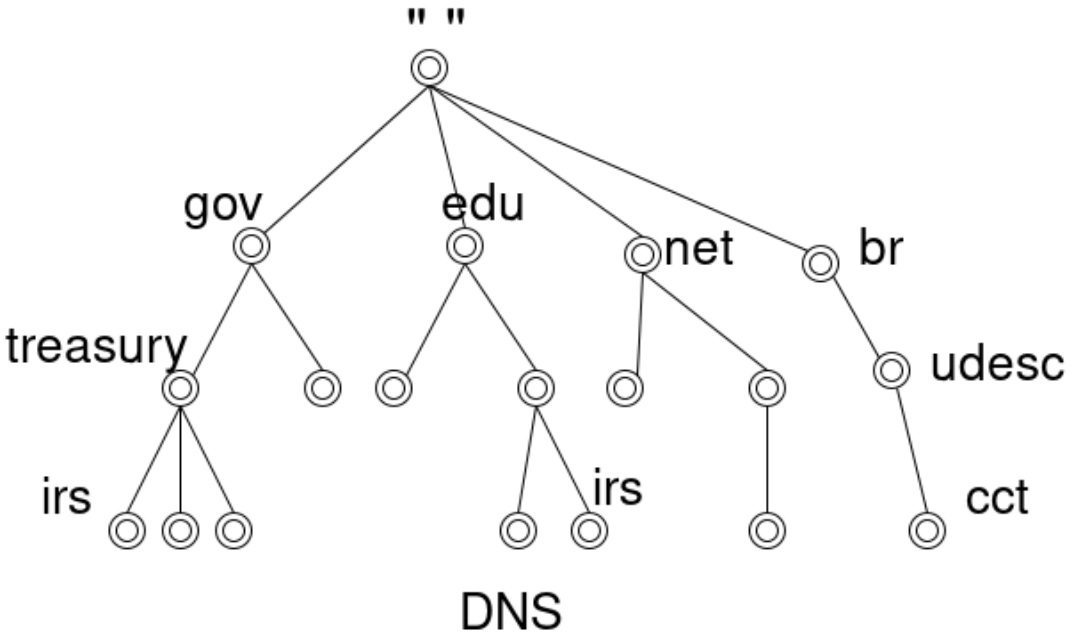


Figura 2.1: Árvores representando a estrutura de domínios do DNS

Fonte: O próprio autor

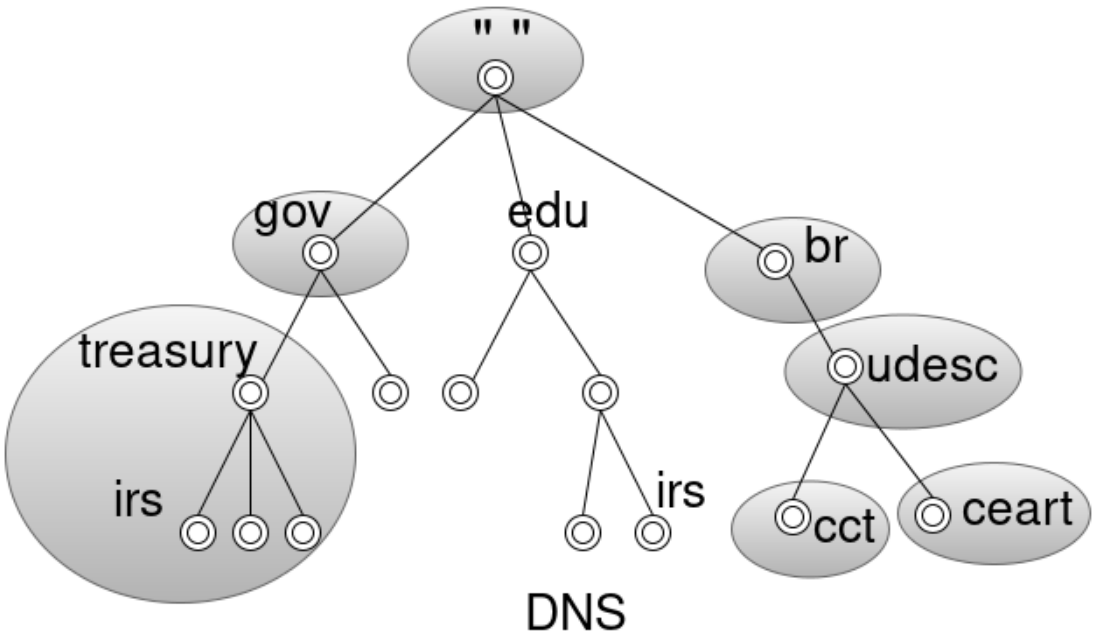


Figura 2.2: Representação de zonas na estrutura de domínios do DNS

Fonte: O próprio autor

A Figura 2.2 ilustra a delegação do subdomínio `treasury.gov` pelo domínio `gov`. A delegação é efetivada atribuindo-se um conjunto de servidores autoritativos que irão administrar os nomes em `treasury.gov` de forma autônoma. Um conceito associado ao de domínio é o conceito de zona, que abrange todas as partes de uma subárvore que não estão delegadas; por exemplo, na Figura 2.2, a zona *gov* contém nomes de domínios que forem terminados em *gov* e não estejam em nenhuma zona de delegação. O domínio `treasury.gov` é um exemplo de domínio de segundo nível. Na Figura 2.2, `cct.udesc.br` e `ceart.udesc.br` é apresentado o domínios de terceiro nível, sendo ambos zonas separadas.

2.1.3 Dados armazenados no DNS

Como mencionado na Seção 2.1.2, a cada nó da árvore DNS pode ser associado um conjunto de registros de recursos. Cada registro de recurso é identificado por uma tripla $< nome, tipo, classe >$. O *nome* determina a sua localização na árvore DNS. O *tipo* determina qual a espécie de informação codificada no RR; os principais tipos estão listados na Tabela 2.1.A *classe* possibilita que RRs de mesmo tipo tenham formatos distintos caso se refiram a famílias de protocolos diferentes. Na Internet é usada apenas a classe `IN`. Cada RR possui ainda um TTL (*time-to-live*), que determina por quantos segundos um RR pode ser armazenado em *cache*. A alocação de tipos, classes e outros valores do DNS é feita pela IANA (*Internet Assigned Numbers Authority*). A tabela de alocação corrente pode ser obtida em (RIEDEL, 2016).

Tipo	Definição	Utilização
`A`	endereço	mapeia um nome em um endereço
`NS`	servidor de nomes	designa o servidor de nomes com autoridade sobre uma zona
`SOA`	início de autoridade	define diversos parâmetros administrativos de uma zona
`PTR`	ponteiro	mapeia um endereço em um nome
`CNAME`	nome canônico	define um *alias* (apelido) para um nome
`MX`	servidor de *mail*	especifica o servidor que deve receber *mail* endereçado a um *host* ou domínio

Tabela 2.1: Principais tipos de RRs

2.1.4 Resolução de nomes

Quando um usuário realiza uma consulta como `www.google.com`, ocorre um processo conhecido como "resolução de nomes". Este processo é responsável por realizar a recuperação de dados armazenados no DNS (ALBITZ; LIU, 2006).

A resolução de nomes pode ser realizada de duas formas, iterativa e recursiva:

1. Iterativa: o servidor de DNS pode retornar uma resposta parcial ao cliente, informando o servidor de nomes mais próximo do nome consultado na hierarquia (MOCKAPETRIS, 1987a). Na Figura 2.3 é possível observar uma resolução DNS iterativa.

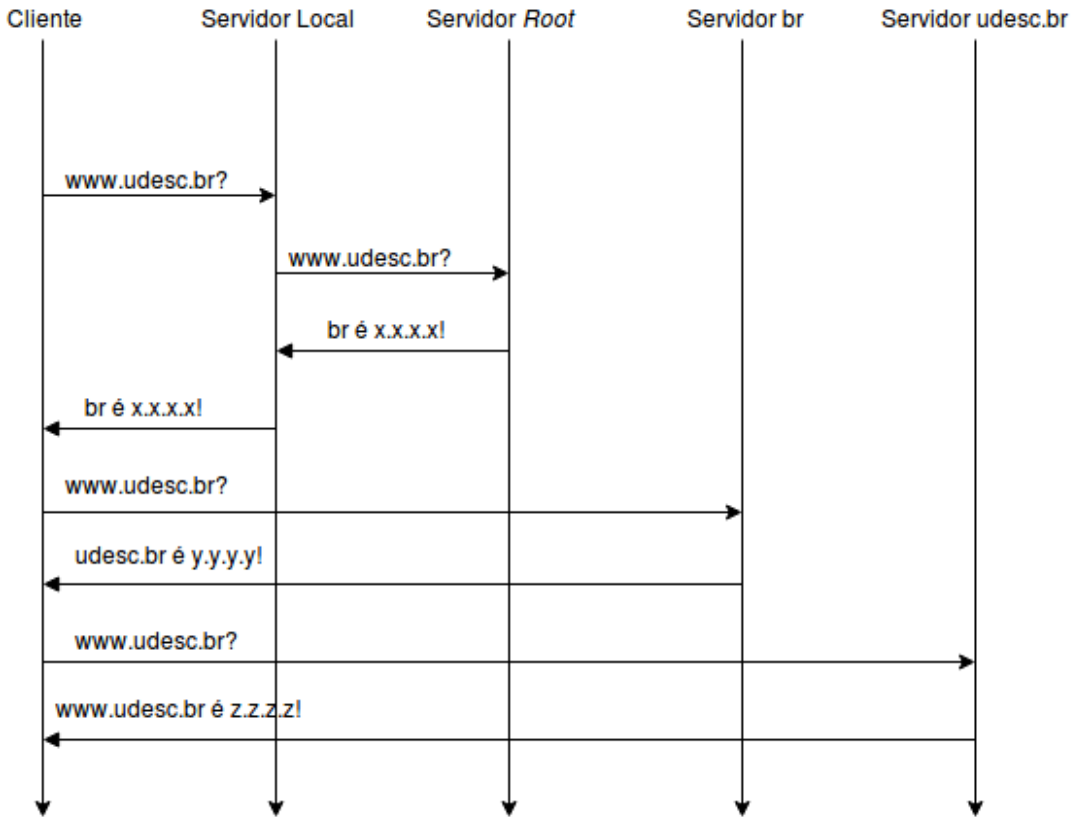


Figura 2.3: Resolução DNS iterativa

Fonte: O próprio autor

2. Recursiva: uma servidor DNS local (conhecido como *resolver*) encaminha a consulta do cliente para todos os servidores DNS necessários até que uma resposta seja encontrada (MOCKAPETRIS, 1987a). Na Figura 2.4 é possível observar uma resolução DNS recursiva.

Como pode ser visto, em ambos os casos (Figura 2.3 e 2.4), a resolução de um único nome pode envolver diversas consultas a servidores DNS. A principal diferença entre a resolução iterativa

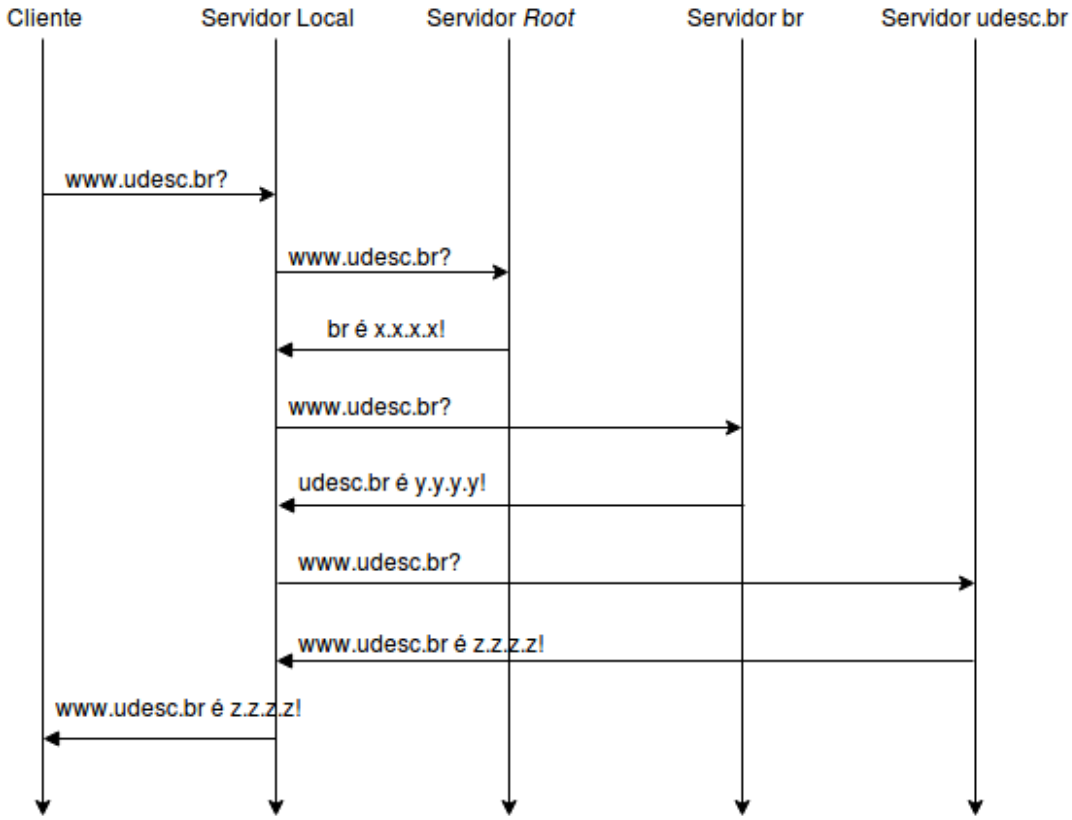


Figura 2.4: Resolução DNS recursiva

Fonte: O próprio autor

e a resolução recursiva é que, na segunda, um servidor DNS local assume a responsabilidade de contactar os servidores DNS autoritativos necessários para resolver o nome, devolvendo apenas a resposta final para o cliente.

Para melhorar o desempenho, os *resolvers* DNS armazenam em *cache* as respostas recebidas, respeitando o TTL de cada RR. Por exemplo, na Figura 2.4, se a consulta por `www.udesc.br` fosse seguida por uma consulta por `expresso.udesc.br`, o servidor DNS local já possui em *cache* o endereço dos servidores DNS responsáveis pela zona `udesc.br`, e pode contactar um deles diretamente, sem precisar passar pelos servidores DNS raiz (global e do domínio `.br`). Percebe-se, que esse uso de *cache* diminui o número de passos envolvidos na resolução de um nome, reduzindo assim a latência para o cliente e a utilização da rede.

2.1.5 Formato das Consultas no DNS

As consultas são realizadas através da troca de mensagens entre o usuário e o serviço DNS. O cliente envia uma requisição para o servidor, usando transporte UDP (caso mais

comum) ou TCP, com porta padrão 53. A mensagem de resposta enviada pelo servidor pode conter a resposta para a pergunta formulada (isto é, o nome consultado), indicar um servidor mais apropriado para fornecer essa resposta (*referral*) ou sinalizar alguma condição de erro (MOCKAPETRIS, 1987a). Tanto a requisição quanto a resposta possuem o mesmo formato de mensagem, ilustrado na Figura 2.5. Em uma mensagem, apenas o cabeçalho é obrigatório; a pergunta e as seções de resposta, autoridade e adicional podem estar vazias (embora a pergunta raramente esteja) (MOCKAPETRIS, 1987b).

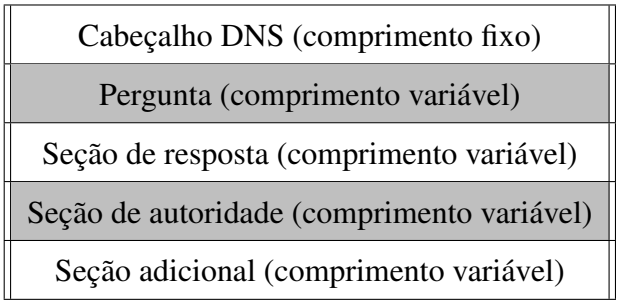


Figura 2.5: Formato de uma mensagem DNS

Fonte: adaptado de (MOCKAPETRIS, 1987b)

O cabeçalho possui os seguintes campos (MOCKAPETRIS, 1987b):

- ID (16 bits): responsável por identificar uma transação DNS, o cliente usar este ID para casar uma resposta com a consulta que a originou;
- QR (1 bit): *flag* que indica se a mensagem contém uma consulta (QR=0) ou uma resposta (QR=1);
- OpCode (4 bits): tipo de consulta (tipicamente 0, indicando consulta padrão);
- AA (1 bit): *flag* válida apenas em respostas, indica se o servidor que responde é autoritativo para o domínio consultado;
- TC (1 bit): *flag* que indica se a mensagem foi truncada em função do tamanho máximo permitido pelo canal de comunicação;
- RD (1 bit): *flag* que indica se o cliente solicita resolução recursiva (em uma resposta, a *flag* recebe o mesmo valor da consulta correspondente);
- RA (1 bit): *flag* setada em uma resposta caso o servidor ofereça resolução recursiva;
- Z (3 bits): campo reservado, deve ser zero em todas as mensagens;

- RCODE (4 bits): código de resposta, usado para sinalizar sucesso ou erro de operações. A RFC 1035 (MOCKAPETRIS, 1987b) define os seguintes valores para RCODE:

RCODE	significado
0	sucesso
1	erro de formatação (FORMERR)
2	falha transiente do servidor (SERVFAIL)
3	domínio inexistente (NXDOMAIN)
4	não implementado (NOTIMP)
5	operação recusada (REFUSED)

  Consultas sempre têm RCODE=0, os demais valores são usados apenas em respostas.

- QDCOUNT (16 bits): número de entradas na seção de pergunta;
- ANCOUNT (16 bits): número de RRs na seção de resposta;
- NSCOUNT (16 bits): número de RRs na seção de autoridade; e
- ARCOUNT (16 bits): número de RRs na seção adicional.

A seção de pergunta tem o objetivo de descrever a consulta que deseja-se realizar no serviço DNS. Consiste em uma tripla $<$ *nome*, *tipo*, *classe* $>$; o tipo 255 (`ANY`) pode ser usado em uma pergunta para obter todos os RRs de um dado *nome* e *classe*.

As seções de resposta, autoridade e adicional possuem a mesma estrutura, uma lista contendo zero ou mais registros de recursos (RRs).

- A seção de resposta contém RRs que respondem à consulta enviada pelo cliente.
- A seção de autoridade contém RRs que indicam os servidores de nomes autoritativos para o domínio consultado.
- A seção adicional contém RRs que não respondem diretamente à consulta formulada, mas são relacionados a ela (por exemplo, os registros `A` contendo os endereços IP correspondentes aos servidores de nomes – registros `NS` – contidos na seção de autoridade).

Cada RR em uma dessas seções tem os seguintes campos:

- Nome: apresenta o nome do domínio DNS;

- Tipo: tipo de RR (`A`, `NS`, `CNAME`, ...);
- Classe: classe do RR (tipicamente `IN`);
- TTL: tempo que o RR pode ser mantido em *cache*, expresso em segundos;
- Comprimento (RDLENGTH): comprimento dos dados em bytes; e
- Dados (RDATA): valor, com formato específico para cada tipo de RR.

2.1.6 Aspectos de Segurança

Dada a centralidade do papel desempenhado pelo DNS na Internet, é natural que ele seja um protocolo bastante visado do ponto de vista de segurança. A maioria das redes possui um ou mais servidores DNS, sejam recursivos ou autoritativos. Além disso, diferente de outros protocolos de uso menos disseminado, o DNS sofre menos com filtragem em *firewalls* – eventuais interrupções no serviço de resolução de nomes são sentidas por praticamente todas as aplicações. Conrad (2012) divide as ameaças envolvendo o DNS em dois grupos, as ameaças contra o próprio serviço e as ameaças que usam o DNS como um vetor de ataques. Esses dois grupos são discutidos na sequência.

2.1.6.1 Ameaças ao DNS

Embora o DNS possa usar tanto UDP quanto TCP como protocolo de transporte, o primeiro é por ampla margem o mais usado (THOMAS; WESSELS, 2014), tanto por questões de desempenho – especialmente evitar a latência associada ao estabelecimento de conexões TCP e melhorar a vazão dos servidores DNS (MEDEIROS, 2011) – quanto pelo fato de muitos servidores não responderem a consultas usando TCP. O UDP não oferece nenhum meio de validação das informações trafegadas, existe a possibilidade de um atacante forjar informações, ou até mesmo a existência de injeção de informações na mensagem DNS.

Corrupção de dados A corrupção de dados ocorre quando um usuário altera dados do DNS de maneira não autorizada. Um servidor DNS recursivo pode ter informações corrompidas quando existe a inserção de respostas incorretas em seu *cache*, ou quando mensagens DNS são alteradas em trânsito (CONRAD, 2012). Um tipo de ataque de corrupção de dados é o *pharming*, que tem o objetivo de encaminhar tráfego para endereços controlados pelo atacante. Um exemplo é inserir falsos registros `A` para o nome `www.banco.com.br`, fazendo com que usuários

que tentem acessar a página web do banco acabem acessando uma página falsa que captura as suas credenciais.

Exposição de informação Devido à ausência de criptografia no canal de comunicação, um atacante que seja capaz de observar o tráfego DNS de um cliente passa a ter acesso ao histórico de *hosts* com os quais esse usuário interage, por exemplo a lista de *sites* web que o usuário acessou (ou tentou acessar). Embora o DNS não tenha a confidencialidade como um requisito – na verdade, as especificações são explícitas em afirmar que os dados DNS são públicos (ATKINS; AUSTEIN, 2004) –. Essa premissa faz bem menos sentido em uma época em que diversas atividades da vida cotidiana são realizadas via Internet do que na época em que a rede era basicamente acadêmica (BORTZMEYER, 2015). De acordo com (CONRAD, 2012) a exposição de informações pode ser danosa, afetando a confiança dos usuários.

Negação de serviço contra servidores DNS Ataques de negação de serviço têm o objetivo de impedir que os usuários consigam utilizar o serviço DNS e, por conseguinte, prejudicar o acesso a outros serviços na Internet. Um exemplo recente e dramático foi o ataque distribuído de negação de serviço (*Distributed Denial of Service*, DDoS) contra o provedor DNS Dyn em 21 de outubro de 2016, que impediu o acesso de usuários a *sites* populares como Twitter, Netflix, Spotify, Reddit, CNN, PayPal e Pinterest (THIELMAN; JOHNSTON, 2016). Ataques de negação de serviço, especialmente os distribuídos, estão entre as ameaças contra as quais é mais difícil de defender-se (SCHNEIER, 2016).

2.1.6.2 Ameaças oportunizadas pelo DNS

A seguir é listado algumas das ameaças oportunizadas pela estrutura do DNS;

***Fast flux* DNS** Uma rede *fast-flux* é uma *botnet* usada tipicamente para servir conteúdo ilegal ou malicioso (SALUSKY; DANFORD, 2007). Como os *bots* que compõem uma *botnet* são intrinsecamente instáveis, podendo ser desligados ou reparados a qualquer momento, para garantir acesso a uma rede *fast-flux* são usados nomes DNS *fast-flux*. Esses nomes são mapeados em um subconjunto dos endereços IP usados pela *botnet*, sendo que o conjunto de IPs retornado como resposta muda com alta frequência, até mesmo a cada consulta. Além de garantir a resiliência da *botnet* diante da instabilidade dos *bots*, os nomes *fast-flux* também ajudam no balanceamento de carga entre os nós da rede.

Ataques de amplificação O DNS pode ser usado como refletor em ataques DDoS baseados em amplificação (CERT.BR, 2016). Esses ataques têm base em dois critérios, a assimetria e o uso de UDP (CONRAD, 2012), e são ilustrados na Figura 2.6. Um atacante envia a um servidor DNS recursivo um número elevado de consultas com o endereço IP de origem forjado como sendo o da vítima do ataque (passo 1); o servidor recursivo interage com servidores autoritativos para obter a resposta para a consulta (passos 2 e 3); a resposta é enviada para a vítima (passo 4). O ataque é facilitado pelo fato de consultas DNS pequenas poderem gerar respostas grandes; por exemplo, Longo (2015) observou consultas de 40 bytes que geraram respostas da ordem de 4.000 bytes, um fator de amplificação de 100:1 (um tráfego de consulta de 1 Mbps gera um tráfego de resposta de 100 Mbps). Se forem usados diversos clientes e servidores simultaneamente (por exemplo, uma *botnet* de clientes usando servidores DNS distintos), é relativamente fácil gerar uma intensidade de tráfego suficiente para sobrecarregar a vítima, indisponibilizando sua conectividade.

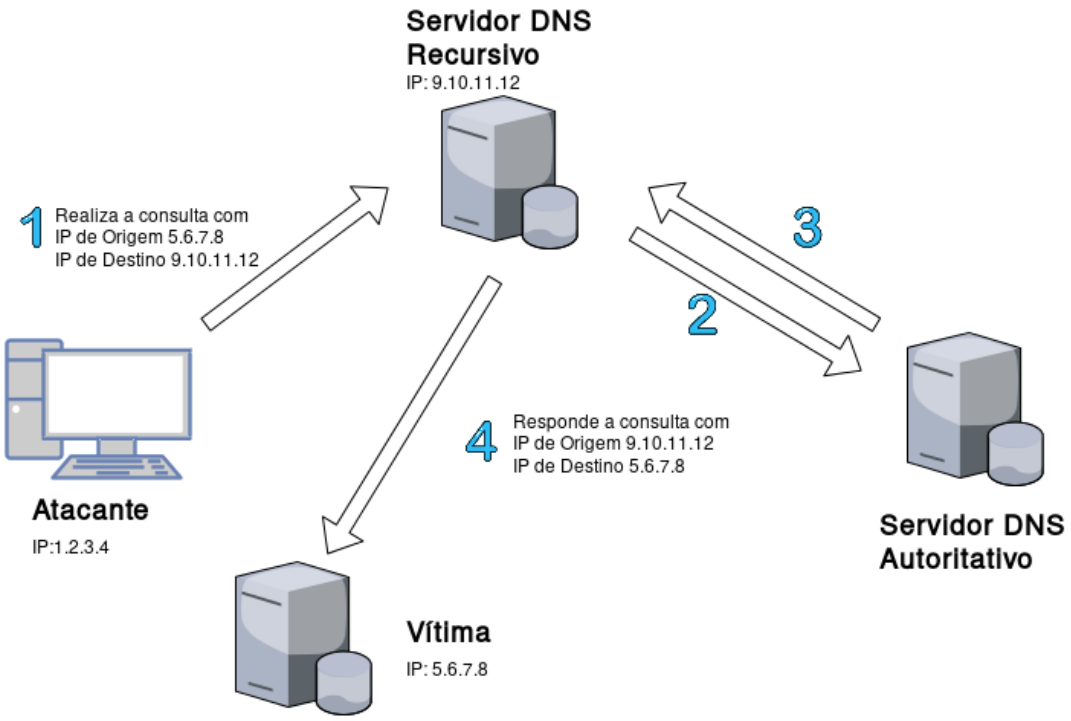


Figura 2.6: Ataque de amplificação, utilizando um servidor de DNS recursivo

Fonte: O próprio autor

Canais cobertos O uso de canais cobertos tem o objetivo de esconder mensagens dentro de outros protocolos, para realizar a comunicação sem ser detectado. Requisições e respostas DNS podem ser utilizadas para realizar uma comunicação dissimulada. Como resultado, o atacante acaba burlando sistemas de monitoramento de segurança e evitando a detecção de sua atividade

maliciosa (CONRAD, 2012).

2.2 Honeypots

Um *honeypot* é um recurso computacional com o objetivo de ser sondado, atacado ou até mesmo comprometido (HOEPERS; STEDING-JESSEN; CHAVES, 2007). Geralmente o *honeypot* é um *host* Internet, possuindo um endereço IP público (isto é, um endereço IP roteável globalmente, que não pertence a nenhuma faixa de endereços reservados[1]), mas que não hospeda nenhum serviço anunciado publicamente. Qualquer interação realizada com UM *honeypot* já pode ser considerada suspeita, já que é necessária a realização de uma varredura para a descoberta do endereço IP do *honeypot*. O sistema deve ser monitorado de forma discreta, para evitar que um atacante desconfie de estar sendo observado (HOEPERS; STEDING-JESSEN; CHAVES, 2007).

Spitzner (2003) apresenta os principais pontos positivos e negativos da utilização de *honeypots*. Os pontos positivos incluem:

1. Redução do conjunto de dados: o *honeypot* só recebe dados que não sejam legítimos, contribuindo no momento que é necessário a realização de uma análise.

2. Diminuição nos falsos positivos: como toda interação com o *honeypot* é considerada maliciosa, a chance de uma interação legítima ser mal classificada é mínima.

3. Recursos mínimos: alguns tipos de *honeypots* requerem recursos mínimos, e um *hardware* simples tem a capacidade de realizar o monitoramento em milhões de endereços IPs.

Os pontos negativos do uso de *honeypots* abrangem:

1. Riscos: o objetivo de um *honeypot* é permitir que usuários maliciosos interajam com o sistema, o que introduz um certo nível de risco (o qual pode e deve ser mitigado).

2. Campo de visão limitado: um *honeypot* pode observar apenas as interações de um usuário malicioso dentro do próprio sistema, e aquelas em que o sistema é origem ou destino

[1] As faixas de endereços reservadas são definidas pela IANA; a lista atualizada de endereços IPv4 reservados pode ser encontrada em `http://www.iana.org/assignments/iana-ipv4-special-registry`.

de tráfego de rede. Interações envolvendo outros sistemas na mesma rede, por exemplo, tipicamente não são observadas pelo *honeypot*. Por outro lado, as informações capturadas por ele são muito detalhadas.

A Figura 2.7 ilustra um *honeypot* instalado na rede de uma organização. No caso apresentado o *honeypot* é instalado em um segmento de rede segregado, reduzindo assim os riscos para a rede de produção. O *firewall* é responsável por controlar o tráfego que entra e sai do *honeypot*, evitando a comunicação com a rede interna e limitando a comunicação com a rede externa.

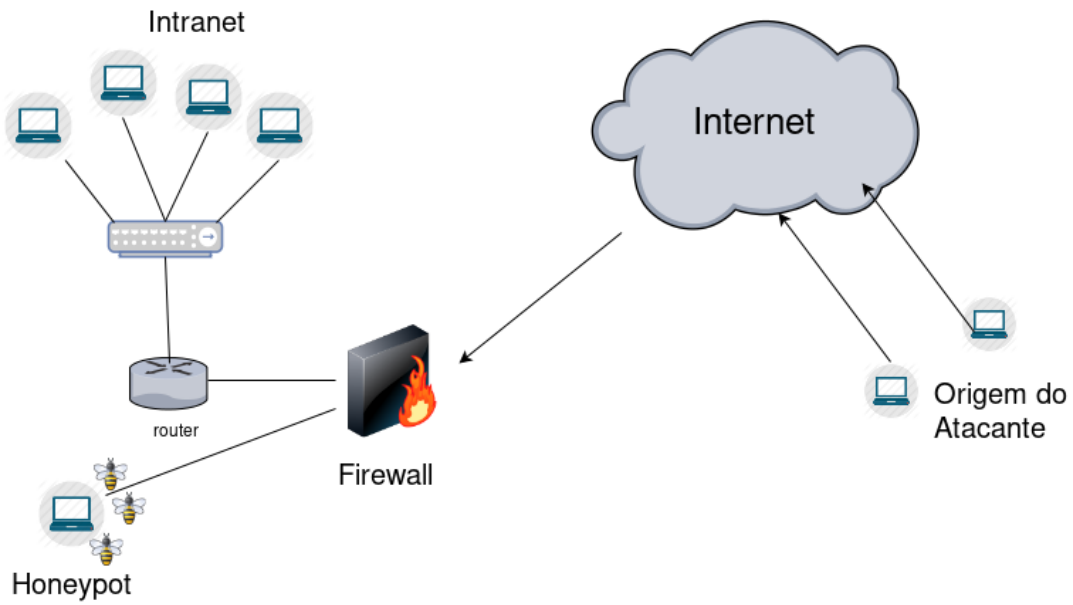


Figura 2.7: Exemplo da localização de um *honeypot* na rede de uma organização

Fonte: O próprio autor

Os *honeypots* podem ser classificados de acordo com o seu nível de interatividade (SPITZNER, 2003). Um *honeypot* de baixa interatividade basicamente emula algumas funcionalidades de um sistema vulnerável, permitindo uma observação mais restrita do comportamento dos atacantes mas oferecendo um risco menor. Um *honeypot* de alta interatividade, por outro lado, permite que atacantes interajam com aplicações e serviços reais, o que oferece uma visão mais detalhada de suas atividades mas introduz um nível maior de risco. Entre esses extremos se situam os *honeypots* de média interatividade, que oferecem níveis intermediários de visibilidade e risco.

Um conceito associado ao de *honeypot* é o de uma *honeynet*, que é um segmento de rede contendo diversos *honeypots* (SPITZNER, 2003). Tipicamente esses *honeypots* usam

diversos sistemas operacionais e hospedam serviços de rede distintos, oferecendo uma gama mais ampla de oportunidades de interação para atacantes.

2.3 DNSpot

O DNSpot é um *honeypot* projetado especificamente para monitorar e analisar o tráfego DNS (LONGO, 2015). Seu objetivo é permitir a observação das interações de usuários potencialmente maliciosos com servidores DNS recursivos.

A arquitetura do DNSpot pode ser observada na Figura 2.8. Ele possui um *proxy* que escuta na porta 53/UDP, que é a porta padrão do serviço DNS. Ao receber uma consulta, o *proxy* será responsável por armazenar a consulta em um banco de dados e logo em seguida repassá-la a um servidor DNS recursivo real. Esse servidor real, que aceita apenas consultas originadas na própria máquina, vai interagir com servidores autoritativos na Internet para a resolução desta consulta. Por último, o *proxy* irá receber a resposta do servidor recursivo, armazená-la no banco de dados e encaminhá-la para o cliente que enviou a consulta (LONGO, 2015).

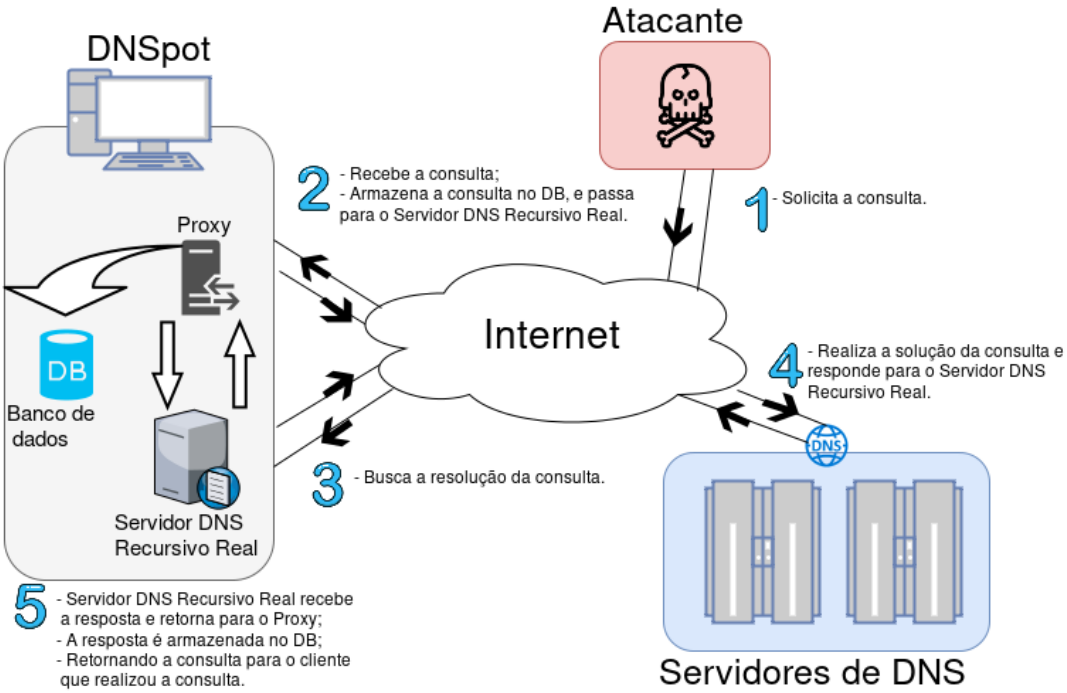


Figura 2.8: Arquitetura do DNSpot

Fonte: O próprio autor

Embora o DNSpot geralmente responda às consultas recebidas, ocasionalmente ele retorna uma mensagem falsa de erro para o cliente (a frequência dessas mensagens de erro é

configurável). O objetivo é simular um servidor DNS inconfiável, tentando não levantar suspeitas caso o DNSpot seja desligado ou não consiga processar algumas consultas.

Para reduzir o impacto caso o DNSpot seja usado como refletor em um ataque de negação de serviço distribuída, um limite diário é determinado para o número de consultas atendidas para cada endereço IP de origem. Caso esse limite seja atingido o *proxy* passa a ignorar novas consultas do mesmo endereço, até o dia seguinte (onde a atividade é retornada novamente) (LONGO, 2015).

2.4 Resultados do DNSpot

O DNSpot foi colocado em operação na rede da UDESC e foi realizada uma coleta de dados durante 49 dias. O serviço, mesmo não sendo anunciado externamente, foi sondado e descoberto no seu segundo dia de execução. Foram observadas um total de 4 milhões de requisições DNS, sendo que 99% destas estavam relacionadas a ataques DDoS (LONGO, 2015). Dentre os resultados mais relevantes obtidos com o DNSpot incluem-se os seguintes:

- O volume de requisições maliciosas é significativo, tendo sido recebida uma média de 298 MB de tráfego por dia.
- Apesar do volume de tráfego ser significativo, ataques em geral apresentam um período de duração curto (93,3% duram até 5 minutos), e envolvem poucas requisições (75% dos ataques têm até 402 requisições).
- Somente 5% dos *hosts* estiveram relacionados com mais de 6 ataques, demonstrando que a estratégia abordada para os ataques visa um tráfego variado, com diversos alvos recebendo uma pequena quantidade de tráfego de um único refletor – presumivelmente um número grande de outros refletores estão engajados em um ataque para gerar volume suficiente para causar uma negação de serviço na vítima.

2.5 Considerações do Capítulo

Ao longo de seus aproximadamente 30 anos de adoção, o DNS se consolidou como um serviço essencial na Internet, posto que é um dos pilares da infraestrutura usada pela imensa maioria das demais aplicações. Devido ao DNS poder ser ao mesmo tempo alvo e vetor de

ataques, o seu monitoramento com propósitos de segurança torna-se importante para oferecer visibilidade a diferentes tipos de atividades maliciosas na rede.

O DNSpot é um *honeypot* específico para servidores DNS recursivos. A sua implantação na rede da UDESC permitiu uma caracterização de tráfego DNS malicioso durante um período de 49 dias. Essa caracterização, no entanto, oferece um instantâneo do período de análise. Para compreender a evolução da atividade maliciosa relacionada ao DNS, seria importante uma análise longitudinal, considerando um período mais longo de observação. O Capítulo 3 traz uma revisão de diversos trabalhos da literatura que envolvem a monitoração de servidores DNS, buscando entender em que medida esses trabalhos examinam a evolução do tráfego ao longo do tempo.

3 Trabalhos Relacionados

Estudos envolvendo análise de tráfego DNS vêm sendo conduzidos há vários anos, sendo que um dos pioneiros foi (DANZIG; OBRACZKA; KUMAR, 1992). Este capítulo discute trabalhos que usam análise de tráfego para realizar caracterização de carga de trabalho (em servidores e/ou clientes) e detecção de tráfego DNS anômalo e/ou malicioso. A revisão considera apenas referências publicadas desde o ano 2000.

3.1 Caracterização de tráfego DNS

A literatura registra estudos de caracterização de tráfego DNS tanto do ponto de vista de servidores DNS raiz (BROWNLEE; CLAFFY; NEMETH, 2001; CASTRO, 2010; BARBOSA; PEREIRA, 2009) quando do ponto de vista de *resolvers* (JUNG, 2002; GAO, 2013).

O trabalho (BROWNLEE; CLAFFY; NEMETH, 2001) teve como foco caracterizar tráfego em um servidor DNS raiz, durante 18 dias (em períodos fracionados). O tráfego DNS de/para o servidor raiz F[1] foi monitorado passivamente, em diversos intervalos, chegando a atingir picos de 5000 consultas por segundo. Em média, 14% da carga do servidor foi causada por consultas inválidas, que violavam as especificações DNS; também foram identificadas grandes quantidades de consultas por domínios de primeiro nível inexistentes, de consultas repetidas emitidas pelo mesmo *host* de origem, e de consultas malformadas.

Castro (2010) analisou o tráfego DNS de/para servidores DNS raiz em quatro períodos distintos de um dia, entre 2006 e 2009. O artigo apresentou a evolução da intensidade de tráfego e da distribuição dos tipos de registros consultados ao longo dos quatro anos de monitoramento, e caracterizou a incidência de consultas inválidas. Foram analisadas também características dos clientes, como suporte a aleatoriedade de porta de origem, IPv6 e DNSSEC.

Barbosa e Pereira (2009) buscaram caracterizar o tráfego nos servidores DNS raiz

[1] Existem 13 servidores DNS raiz, nomeados de A a M, operados por 12 organizações independentes; o servidor F é operado pelo *Internet Systems Consortium* (ISC) (IANA, 2016). Esses servidores são replicados ao redor do globo para propiciar baixa latência e alta disponibilidade. No Brasil existem atualmente 24 réplicas, sendo 15 do servidor L, sete do servidor J, uma do servidor F e uma do servidor I (KARRENBERG, 2016).

do domínio `.br`. O tráfego DNS de cinco servidores raiz do domínio `.br` foi monitorado passivamente durante 48 h, totalizando 5,4 bilhões de consultas; os dados foram fornecidos pelo projeto *Day in the Life of the Internet* (CASTRO, 2008). Foram analisadas a distribuição dos tipos de registros consultados e a validade das consultas recebidas pelos servidores. Além disso, foi identificado tráfego DNS relacionado a atividades maliciosas, como envio de SPAM e varreduras de rede.

Jung (2002) focam em caracterizar o comportamento de *resolvers* DNS. O tráfego DNS em duas universidades (MIT e KAIST) foi monitorado de forma passiva nos enlaces de acesso à Internet. Foram três períodos de monitoração, os dois primeiros de 7 dias (MIT, janeiro e dezembro de 2000) e o terceiro de 6 dias (KAIST, maio de 2001). Foram analisadas a distribuição de tipos de registros consultados, a latência e a taxa de sucesso na resolução de nomes, a incidência de falhas e erros. Simulações foram usadas para avaliar a eficácia do uso de *caching* em função do TTL e o grau de compartilhamento dos *caches*.

Gao (2013) caracterizou o tráfego DNS gerado por centenas de *resolvers* locais distribuídos pela Internet durante duas semanas. O trabalho caracterizou a intensidade de tráfego, a distribuição de tipos de registros consultados, a taxa de sucesso na resolução de nomes, a distribuição de TTLs e a incidência de consultas inválidas e consultas repetidas. Além disso, aplicaram-se técnicas de aprendizagem de máquina para identificar domínios maliciosos com base na correlação temporal entre consultas DNS.

3.2 Detecção de tráfego DNS anômalo/malicioso

O trabalho (ZDRNJA; BROWNLEE; WESSELS, 2007) tem o objetivo de detectar tráfego DNS anômalo. O tráfego DNS foi monitorado passivamente ao passar pelo roteador de borda da Universidade de Auckland (Nova Zelândia), durante um período de 8 meses. Apenas o tráfego de resposta foi armazenado, uma vez que uma resposta DNS inclui a consulta correspondente. De um ponto de vista mais quantitativo, o trabalho caracterizou a intensidade de tráfego e a distribuição de tipos de registro DNS, e estimou a quantidade de tráfego DNS decorrente do uso de listas de bloqueio de SPAM. Sob um prisma mais qualitativo, buscou-se identificar anomalias, incluindo domínios similares a domínios conhecidos (*typo squatter domains*), que podem ser usados para enganar usuários, nomes *fast-flux*, e nomes associados a campanhas de envio de SPAM.

Perdisci (2009) buscam detectar de nomes DNS *fast-flux*. A abordagem de detecção proposta utiliza um sensor para monitorar passivamente consultas a um servidor DNS recursivo e as respostas correspondentes. Os dados são filtrados de modo a descartar nomes que não possuem características de *fast-flux*; sobre o conjunto de dados restantes são aplicadas técnicas de aprendizagem de máquina para identificar nomes DNS com característica de nomes *fast-flux*. O trabalho se destaca por não realizar uma análise somente de domínios suspeitos extraídos de uma lista de *emails* de SPAM ou uma lista de domínios (*blacklist*).

Zhao (2015) têm o foco na detecção de nomes usados para canais de comunicação e controle (C&C) de *malware*, que são usados para o controle remoto do *malware*, especialmente de *botnets*. O tráfego DNS que entra e sai de uma rede é monitorado de forma passiva, e filtrado para eliminar domínios conhecidos. Sobre o conjunto de dados restantes são aplicadas técnicas de aprendizagem de máquina para identificar nomes que podem estar sendo usados para hospedar servidores C&C. Os endereços IP associados a esses nomes são passados para sistemas de detecção de intrusões baseados em assinatura e em anomalia, que analisam o tráfego envolvendo tais endereços procurando por atividade característica de servidores C&C. Os resultados da detecção baseada em DNS e dos IDSs são combinados para fornecer um escore de reputação, que indica se um nome/endereço é ou não um servidor C&C. O artigo descreve um experimento que durou um período de 2 meses, coletando um volume de aproximadamente 400 milhões de consultas DNS durante este período.

3.3 Considerações do Capítulo

A Tabela 3.1 resume os trabalhos discutidos, agregando também o DNSpot. Na coluna *Tempo de monitoração*, (F) indica que o tempo citado foi na verdade fracionado em múltiplos períodos.

Os trabalhos de caracterização de tráfego em servidores DNS raiz e em *resolvers* buscam comparar os resultados observados com os anteriormente publicados, apontando similaridades e diferenças entre medições realizadas em períodos distintos. No geral, essa comparação se fixa em aspectos quantitativos, como intensidade do tráfego e distribuição dos tipos de consultas. Dos trabalhos com tempo de monitoração de pelo menos 45 dias, (PERDISCI, 2009) e (ZHAO, 2015) utilizam técnicas de aprendizagem de máquina, que requerem conjuntos de dados relativamente extensos para treinamento e avaliação dos algoritmos. No geral, percebe-se que

não existem trabalhos que buscaram analisar tendências de mais longa duração.

Trabalho	**Tempo de monitoração**	**Foco do trabalho**
Caracterização de tráfego DNS		
(BROWNLEE; CLAFFY; NEMETH, 2001)	18 dias (F)	caracterização do tráfego em um servidor DNS raiz
(CASTRO, 2010)	4 dias (F)	caracterização do tráfego em um conjunto de servidores DNS raiz
(BARBOSA; PEREIRA, 2009)	2 dias	caracterização do tráfego em um conjunto de servidores DNS raiz do `.br`
(JUNG, 2002)	20 dias (F)	caracterização do comportamento de *resolvers* DNS
(GAO, 2013)	14 dias	caracterização do comportamento de *resolvers* DNS
(LONGO, 2015)	49 dias	caracterização do tráfego em um servidor DNS recursivo aberto
Detecção de tráfego DNS anômalo/malicioso		
(ZDRNJA; BROWNLEE; WESSELS, 2007)	8 meses	detecção de tráfego DNS anômalo
(PERDISCI, 2009)	45 dias	detecção de nomes DNS *fast-flux*
(ZHAO, 2015)	2 meses	detecção de nomes usados para canais de comunicação e controle (C&C) de *malware*

Tabela 3.1: Resumo dos trabalhos relacionados. Na coluna *Tempo de monitoração*, os tempos assinalados com (F) foram fracionado em diversos períodos menores.

4 Análise de Resultados

Para a realização da coleta de dados, uma máquina foi disponibilizada pela UDESC. A máquina foi implantada na rede interna da universidade, onde está realizando a coleta de dados, que iniciou no dia 17/09/2016.

Este período de captura corresponde a 250 dias; na sequência será apresentado as comparações estatísticas levando em consideração o trabalho desenvolvido (LONGO, 2015), e algumas considerações que foram observadas ao longo deste período. Este capítulo apresenta a implantação do DNSpot, as estatísticas de tráfego após o período de coleta, os ataques DoS observados durante este período, uma análise temporal da coleta, a discussão dos resultados encontrados e o detalhamento de anomalias de tráfego encontradas.

4.1 Implantação

A rede do DCC utiliza endereços privados, por este motivo é necessário a utilização do NAT (*Network Address Translation*) para realizar o redirecionamento do tráfego para porta 53/UDP. O NAT é um método utilizado para remapear espaços de endereços IPs, realizando a modificação no cabeçalho dos pacotes no momento em que estes se encontram no dispositivo de roteamento (NAUGLE, 1998).

A configuração do *hardware* e *software* da máquina utilizada é a seguinte:

- Modelo: Dell Inc. OptiPlex 755
- Disco: WDC WD1600AAJS-75PSA0
- Sistema Operacional: OpenBSD 5.7 i386;
- Processador: Intel Core2 Duo CPU E6550 @ 2.33GHz (“GenuineIntel” 686-class);
- Memória RAM: 1 GB;
- Servidor DNS recursivo: Unbound, versão 1.5.2;
- Python, versão 3.4.2;

- DNSLib, versão 0.9.4; e
- SQLite3, versão 3.8.6.

A máquina está localizada em um dos laboratórios da UDESC, onde realiza a coleta de dados 24/7. O serviço é verificado todos os dias para a garantia do seu funcionamento. Durante o período em que o sistema esteve executando a coleta, ocorreram algumas interrupções do serviço devido a queda de energia e interrupção do *firewall* que dá acesso à Internet (consequência da queda de energia). A indisponibilidade total é estimada em 1% a 2% dos 250 dias de coleta.

O DNSPot permite definir uma taxa de transações DNS que serão respondidas com falha de servidor ao cliente (*ServFail*), visando simular um servidor DNS de implementação instável e com falhas (LONGO, 2015). A quantidade de *ServFail* foi definida em 20%.

Observações de tráfego DNS revelaram a ocorrência de varreduras (*scans*), este pode ser classificado em dois diferentes grupos, o primeiro grupo que realiza estes *masscan* para fim de pesquisa. Como por exemplo alguns destes testes são realizados por (TENTLER, 2017), visando encontrar vulnerabilidades específicas em redes, ataques DDoS ou de amplificação que buscam tirar proveito de serviços de DNS mal configurado, *botnets* entre outras. O segundo grupo busca encontrar estas vulnerabilidades para explorá-las em diversos ataques.

Estas consultas tem a característica de possuir um domínio fixo, com um sufixo bem definido (LONGO, 2015). O DNSPot foi configurado para ignorar estes tipos de consultas, não permitindo a sua identificação nestas varreduras. A Tabela 4.1 apresenta a lista de sufixos de domínios ignorados.

	DOMAIN	Referência
1	DNSSCAN.SHADOWSERVER.ORG.	`http://dnsresearch.cymru.com`
2	OPENRESOLVERTEST.NET.	`http://openresolverproject.org`
3	SYSSEC.RUB.DE.	`http://scanresearch.syssec.rub.de`
4	OPENRESOLVERPROJECT.ORG.	`http://openresolverproject.org`
5	SATELLITE.CS.WASHINGTON.EDU.	`http://satellite.cs.washington.edu`
6	DNSRESEARCH.CYMRU.COM.	`http://dnsresearch.cymru.com`
7	SYSSEC-RESEARCH.MMCI.UNI-SAARLAND.DE.	`http://syssec.mmci.uni-saarland.de`

Tabela 4.1: Sufixos ignorados pelo DNSpot.

4.2 Estatísticas de Tráfego

O período de monitoramento do DNSpot gerou um total de 32.358.928 requisições. Esta seção realiza uma análise de todo o tráfego capturado, buscando apresentar algumas características antes não observadas em (LONGO, 2015), junto com uma análise geral do volume e distribuição das transações.

4.2.1 Período de monitoramento

O DNSpot foi iniciado na rede da UDESC no dia 17/09/2016 e está em funcionamento até o dia atual, mas para a análise de dados foi realizada uma cópia do banco de dados até o dia 25/05/2017. As análises são geradas automaticamente por um código em R (R Development Core Team, 2008), que realiza a conexão com o banco, para realizar as operações matemáticas e consultas diretamente com o banco do DNSpot.

Devido ao volume do banco de dados, foi adotada uma estratégia para o armazenamento das informações. Para o ganho de desempenho o banco do DNSpot é dividido em versões: quando o banco de dados atinge um determinado volume de informações entre 7 GB e 10 GB, as tabelas que não possuem chaves ou informações necessárias para a continuação das operações internas do DNSpot são removidas, gerando uma versão do banco em um intervalo de tempo. Para a realização de análises parciais estes bancos podem ser utilizados. No caso foram geradas quatro versões do banco:

- Versão 1: 09-09-2016 até 06-03-2017;
- Versão 2: 06-03-2017 até 27-03-2017;
- Versão 3: 27-03-2017 até 20-04-2017; e
- Versão 4: 20-04-2017 até 25-05-2017.

Ao final para realizar o estudo, todas as tabelas antes removidas do banco foram adicionadas novamente, gerando um banco com todas as informações coletadas pelo DNSpot com um volume total de 30 GB.

A Tabela 4.2 apresenta o período que o DNSpot esteve em produção. Durante este período de monitoramento algumas interrupções ocorreram, devido a:

- Queda de energia afetando o *host* onde o DNSpot é executado;
- Interrupção do acesso à Internet;
- *Lock* no banco de dados: devido ao volume que o banco de dados atingiu, algumas *threads* recebiam *timeout* no momento que estava realizando interações no banco de dados, impossibilitando o funcionamento do DNSpot até que o sistema fosse reiniciado. Este problema vai ser discutido na Seção 5.2.3.

	Período	Tempos
1	Início	17/09/2016 8:00
2	Fim	25/05/2017 20:47
3	Total (segundos)	21.644.892
4	Total (minutos)	360.748,2
5	Total (horas)	6.012,47
6	Total (dias)	250,52

Tabela 4.2: Período do DNSpot em produção.

A duração do período de análise foi de 250 dias, aproximadamente cinco vezes o período analisado por (LONGO, 2015), que foi de 49 dias.

4.2.2 Transações

Durante o período que o DNSpot ficou ativo, foram processadas 32.358.928 transações DNS, conforme mostrado na Tabela 4.3. Ao final foram respondidas 4,9% de todas as consultas recebidas, destas 80% eram consultas válidas para as quais o DNSpot enviou uma resposta para o cliente, e 19,9% destas receberam *ServFail* falso gerado pelo próprio DNSpot. As transações não respondidas somaram um total de 95% do volume de requisições, destas 1,6% apresentaram erro e 98,3% foram ignoradas pelo DNSpot. 98,7% de todas as transações recebidas utilizam o protocolo *EDNS(0)*, este permite a utilização de consultas maiores que 512 bytes, por consequência gerando uma amplificação entre a requisição realizada pelo cliente e a resposta enviada.

A Tabela 4.4 apresenta as transações ignoradas por regra de filtragem de domínio. Observa-se que quase todas são referentes a clientes que excederam o limite diário de 30 consultas.

	Transações	Quantidade	Porcentagem
1	Respondidas	1.600.386	4,90
2	Válidas	1.280.425	80,00
3	*ServFail*	319.961	19,99
4	Não respondidas	30.758.542	95,05
5	Ignoradas	30.243.241	98,32
6	Erros	515.301	1,67
7	Total	32.358.928	100,00
8	EDNS(0)	31.943.625	98,71

Tabela 4.3: Resumo das transações DNS. Porcentagem representa a proporção dentro de uma categoria.

	Transações	Quantidade
1	Ignoradas	30.241.560
2	Blacklist de domínios	1489
3	Blacklist de IPs	192
4	Máximo diário atingido	30.243.241

Tabela 4.4: Transações ignoradas por regras.

A Tabela 4.5 apresenta a análise das transações respondidas ao cliente (4,9%). O estado de cada resposta DNS é indicado pelo seu RCODE. No total 74% das consultas não apresentaram nenhum tipo de erro, 19,9% foram respondidas com *ServFail* falso gerado pelo próprio DNSpot, 5,93% com *ServFail* gerados pelo próprio *Unbound* e 0,02% das consultas não foram interpretadas, pois apresentavam formato inválido. *NXDomain* caracteriza um domínio inexistente, somente 28 domínios inexistentes foram buscados e *Notlmp* caracteriza um tipo de requisição não suportado (total de 8 requisições). Nenhuma requisição foi negada (*Refused*) ou desconhecida (*Unknown*).

Em comparação com as requisições processadas por (LONGO, 2015), destacam-se a queda na porcentagem de requisições respondidas (de 12,1% para 4,9%) e o aumento nas requisições ignoradas por erro (de 0,1% para 1,6%). Considerando apenas as requisições que obtiveram resposta, a diminuição das requisições com erro de formatação (*FormErr*), de 0,5% para 0,02%.

Considerando o período de observação, a Tabela 4.6 apresenta a taxa média de tran-

	RCODE	Quantidade	Porcentagem
1	*NoError*	1.185.083	74,04
2	*FormErr*	370	0.02
3	*ServFail* falso	319.961	19,99
4	*ServFail* real	94.936	5.93
5	*NXDomain*	28	0,00
6	*Notlmp*	8	0,00
7	*Refused*	0	0,00
8	*Unknown*	0	0,00
9	Total	1.600.386	100,00

Tabela 4.5: RCODEs enviados ao cliente.

sações por segundo. A primeira linha destaca os valores observados neste estudo, com médias de 1,50 transações por segundo e 129.955,61 transações por dia. A segunda linha apresenta o número de transações do estudo original (LONGO, 2015), com médias de 0,96 transações por segundo e 81.353,02 transações por dia. Observa-se que a taxa de requisições deste estudo foi 59,7% superior à registrada no estudo original.

	Estudo	Transações por segundo	Transações por dia	Período (dias)
1	Atual	1,50	129.955,61	250
2	(LONGO, 2015)	0,96	81.353,02	49

Tabela 4.6: Taxa de transações processadas.

4.2.3 Volume de dados em bytes

A Tabela 4.7 apresenta o volume de tráfego processado pelo DNSpot durante este período. Ao todo foram processados 5.241,51 MB de tráfego pelo DNSpot, sendo 1.196,44 MB (22,8%) de consultas e 4.045,07 MB (77,2%) de respostas. Para um cálculo do volume total de tráfego que seria produzido durante este período, foi levado em consideração o tamanho médio das respostas válidas para cada RR, assim gerando o volume de respostas que acabaram sendo ignorados pelo DNSpot. Foi encontrado um valor de tráfego de resposta esperado igual a 33.579,49 MB, o que somado ao tráfego de consulta processado resulta em um tráfego esperado total de 34.775,93 MB. Portanto, é possível constatar que o DNSpot deixou de enviar 29.534,42 MB de tráfego de resposta, proporcionando uma redução de 84,93% em relação ao

tráfego esperado total (87,95% considerando apenas o tráfego esperado de resposta).

	Tipo	Volume (MB)	Porcentagem
1	Tráfego processado	5.241,51	100,00
2	Consultas	1.196,44	22,83
3	Respostas	4.045,07	77,17
4	Tráfego esperado	34.775,93	–
5	Respostas	33.579,49	–
6	Redução de tráfego de resposta	29.534,42	84,93

Tabela 4.7: Volume de tráfego processado e esperado.

Longo (2015) conseguiu uma redução de 90,5% do tráfego esperado (um total de 14.775,6 MB). Esta diferença é dada principalmente pela duração de cada análise: com o crescimento no tempo do experimento, o sistema tem a tendência de ser descoberto por um número maior de usuários, gerando uma quantidade relativamente maior de clientes a serem atacados ao decorrer do tempo. O crescimento no número de clientes será abordado na Seção 4.2.4.

A Tabela 4.8 apresenta as estatísticas dos tamanhos de consultas e respostas. O tamanho das consultas variou entre 1 e 343 bytes, com média de 38,77 bytes e um desvio padrão de 4,57 bytes. O tamanho das consultas realizadas apresenta uma distribuição assimétrica positiva (concentração de valores pequenos); somente 43 das consultas apresentaram um tamanho maior que 100. Para as respostas o volume variou entre 12 e 4.096 bytes, com média de 2.008,65 bytes e um desvio padrão de 1.812,22 bytes; a variação foi maior que a do tamanho das consultas. A variação dos tamanhos pode ser observada na Tabela 4.9, que apresenta os dez tamanhos mais frequentes nas consultas e respostas, junto com a sua respectiva frequência.

O tamanho das consultas possui uma distribuição assimétrica positiva, enquanto o tamanho das respostas possui uma distribuição bimodal, conforme pode ser visto nos histogramas da Figura 4.1. A Tabela 4.9 permite observar a predominância de tamanhos pequenos para as consultas, e a ocorrência de respostas pequenas e grandes: considerando os 81,4% dos tamanhos de respostas representados na tabela, 38,1% têm até 46 bytes, enquanto que 40,6% são tamanhos maiores ou iguais a 3.230 bytes.

Os tamanhos de consultas encontrados por Longo (2015) apresentam uma concentração maior, pois os pacotes com tamanho de 44 e 39 obtiveram uma frequência de 96,14% de todas as consultas realizadas; no presente trabalho, nem somando os seis tamanhos de consulta mais frequentes se consegue chegar a esse valor (os seis tamanhos mais populares correspondem

	Estatísticas	Consultas (bytes)	Respostas (bytes)
1	Média	38,77	2008,65
2	Desvio padrão	4,57	1812,22
3	Mínimo	1	12
4	1 percentil	28	24
5	5 percentil	28	26
6	1 quartil	37	31
7	Mediana	37	3155
8	3 quartil	42	3893
9	95 percentil	46	3893
10	99 percentil	50	4078
11	Máximo	343	4096

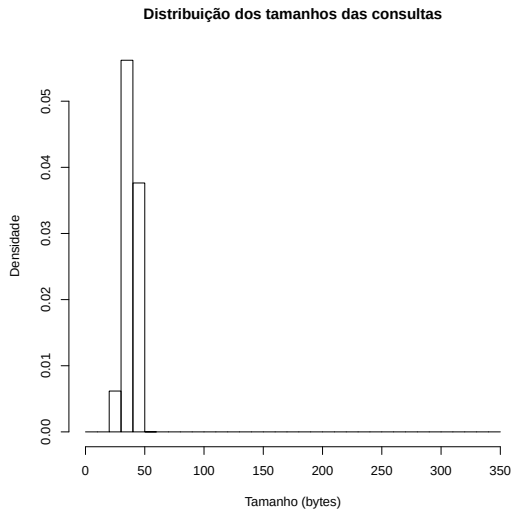
Tabela 4.8: Estatísticas de consultas e respostas.

	Tamanho	Freq. Consultas(%)	Tamanho	Freq. Respostas(%)
1	37	39,88	31	26,08
2	42	24,85	3893	22,61
3	46	10,28	3846	9,04
4	36	8,69	26	5,41
5	28	6,17	42	4,38
6	35	5,24	3230	4,17
7	50	1,91	1328	2,71
8	40	1,43	3409	2,69
9	48	0,55	46	2,21
10	39	0,52	3248	2,11
11	Frequência total %	99,55	Frequência total %	81,45

Tabela 4.9: Dez tamanhos mais frequentes de consultas e respostas.

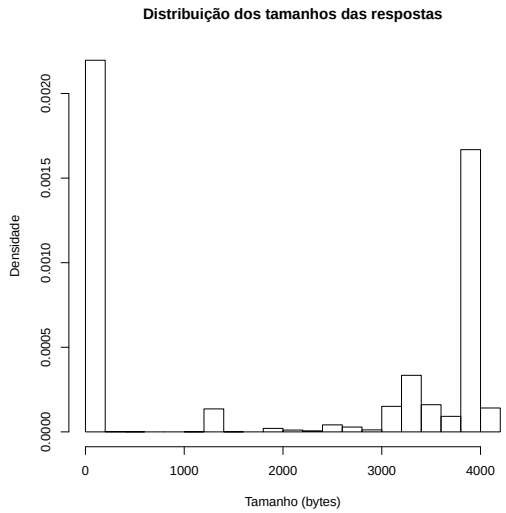
a 95,11% das consultas). Cabe observar que a distribuição dos tamanhos observada no estudo original foi muito influenciada por um único RR com 39 bytes de tamanho (`hehehey.ru ANY`), que respondeu por 97% de todas as consultas.

Para o tamanho das respostas é possível observar a influência do RR `hehehey.ru ANY`, que correspondia a 90,38% das respostas, com um tamanho de 3.850 bytes. Este tamanho



(a)

Fonte: O próprio autor



(b)

Fonte: O próprio autor

Figura 4.1: Distribuição dos tamanhos de; (a) consultas. (b) respostas.

não aparece na Tabela 4.9, e não foi sequer encontrado durante o período desta análise. Em (LONGO, 2015), os cinco tamanhos de respostas mais frequentes abrangem 96,8% das respostas, e o tamanho mínimo dentre esses cinco foi de 1.503 bytes. Em comparação, neste estudo as dez respostas mais frequentes abrangem apenas 81,4% das respostas, e o tamanho mínimo dentre essas dez foi de 26 bytes. Percebe-se, portanto, que do estudo original para este houve um aumento na variabilidade dos tamanhos tanto de consultas quanto de respostas.

4.2.4 Análise dos clientes (IP)

O DNSpot encerrou com um total de 184.564 clientes, ou seja, 184.564 endereços IP distintos. A Tabela 4.10 apresenta a distribuição destes endereços de acordo com a sua geolocalização (MAXMIND, 2017). Devido ao tempo que o DNSpot ficou em funcionamento, foi encontrado um número significativo de clientes. Na Tabela 4.10 é apresentada a distribuição destes países. Os países com maior aparição são China com 15%, Estados Unidos com 8,4% e Brasil com 5,3%. A distribuição de endereços por IPs também pode ser observada na Figura 4.2, que leva em consideração a Tabela 4.10.

Longo (2015) encontrou um total de 4.287 clientes, sendo 30% consultas da China e 27,2% consultas dos Estados Unidos, apresentando uma distribuição menor entre as geolocalizações. Como pode ser observado na Tabela 4.10, um total de 53,4% dos países não foi apresentado na tabela, em comparação a 3,6% encontrados no outro estudo; isto acontece de-

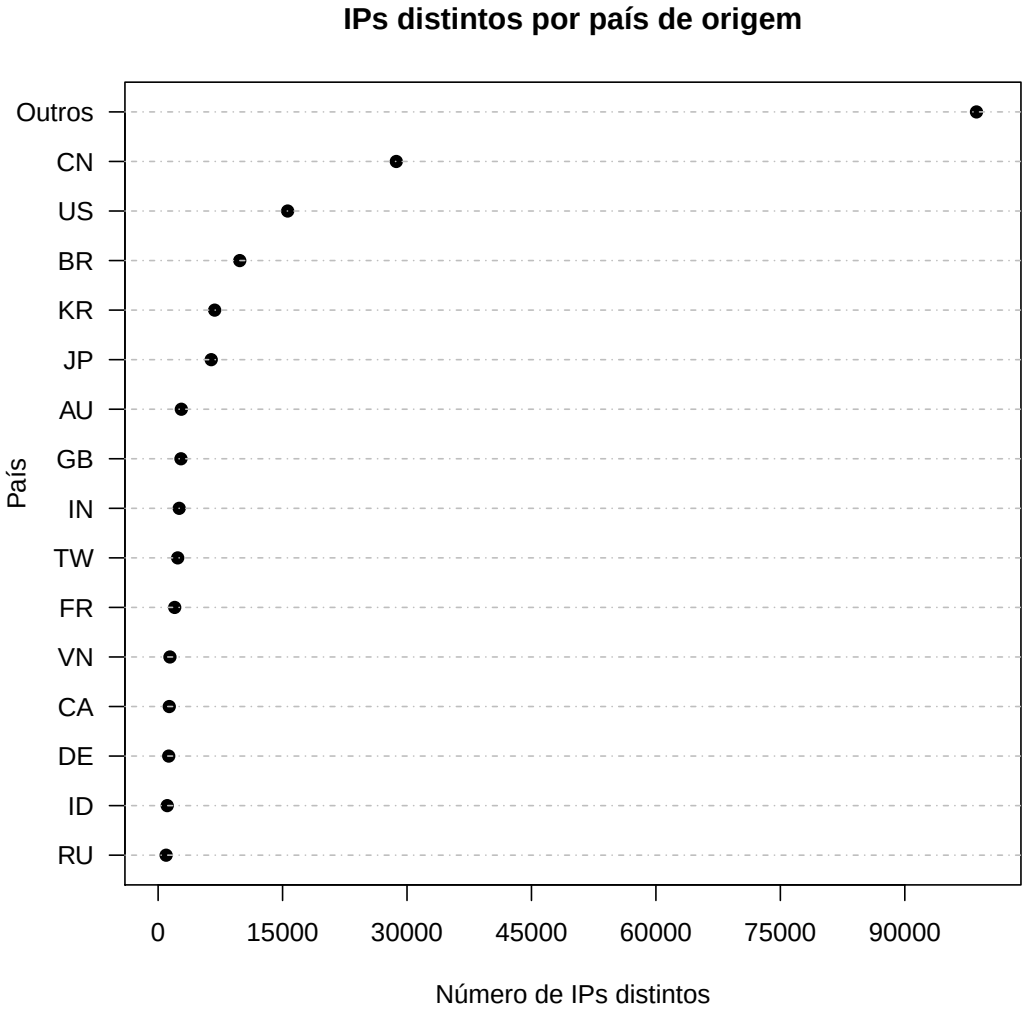


Figura 4.2: IPs distintos por país de origem.

Fonte: O próprio autor

vido ao elevado número de países encontrados (ao todo 161 países).

Khalimonenko (2017) apresenta uma análise anual de ataques DDoS, é possível observar um domínio da China, Estados Unidos e Coreia do Sul nas últimas análises, mesmo com uma diminuição de 23,8% dos ataques gerados na China de 2016 para 2017. Esta predominância também é observada neste estudo.

A Figura 4.3 mostra a evolução do número de clientes distintos observados pelo DNSpot. É possível perceber um aumento notável no mês de maio de 2017, o qual não é acompanhado por um aumento do número de requisições: foram 4,8 milhões de requisições em abril e 5,2 milhões em maio (até o dia 25). Considerando que o mês de março teve 13,7 milhões de requisições, é possível afirmar que a incidência de novos clientes e a quantidade de requisições

	País	Número de IPs distintos	Porcentagem
1	China (CN)	28.705	15,55
2	Estados Unidos (US)	15.613	8,46
3	Brasil (BR)	9.838	5,33
4	Coreia do Sul (KR)	6.815	3,69
5	Japão (JP)	6.398	3,47
6	Austrália (AU)	2.797	1,52
7	Grã-Bretanha (GB)	2.754	1,49
8	Índia (IN)	2.540	1,38
9	Taiwan (TW)	2.361	1,28
10	França (FR)	1.998	1,08
11	Vietnã (VN)	1.425	0,77
12	Canadá (CA)	1.342	0,73
13	Alemanha (DE)	1.280	0,69
14	Indonésia (ID)	1.096	0,59
15	Rússia (RU)	964	0,52
16	Outros	98.638	53,44
17	Total	184.564	100,00

Tabela 4.10: Países de origem das consultas ao DNSpot.

não apresentam uma correlação significativa.

A Tabela 4.12 apresenta o número de requisições por IP considerando apenas o mês de maio. Observa-se que a distribuição possui predominância de valores pequenos: 95% dos clientes efetuaram apenas uma requisição, e apenas 1.683 clientes (ou seja, 1,10% de um total de 153.556) estão vinculados a pelo menos cinco requisições, limiar usado para caracterizar ataques DDoS (vide Seção 4.3).

A explicação encontrada para o aumento no número de novos clientes em maio de 2017 é que, nessa época, usuários finais passaram a usar o DNSpot como servidor DNS recursivo, como corroborado por dados discutidos na Seção 4.5.2. A hipótese é que o endereço IP do DNSpot tenha sido incluído em alguma lista de servidores DNS recursivos abertos; essa hipótese é melhor investigada na Seção 4.3.2.

A Tabela 4.11 apresenta os dez clientes com a maior quantidade de aparições. As informações encontradas na Tabela 4.11 foram retiradas de (WHOIS, 2017b). Considerando a

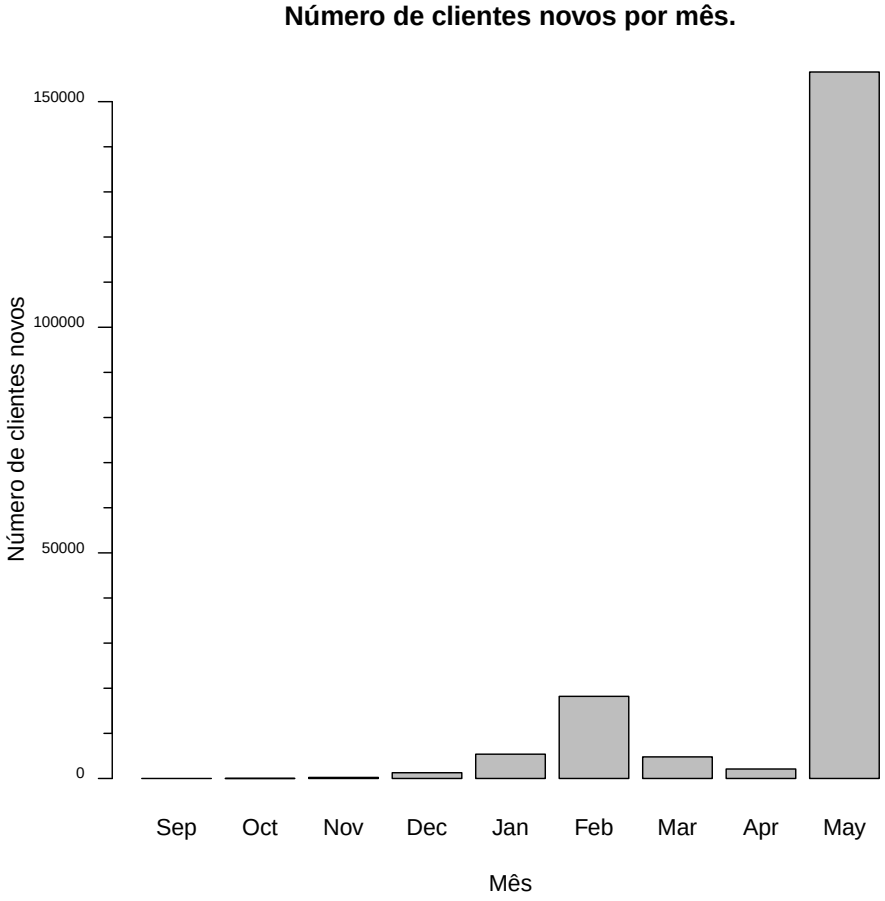


Figura 4.3: Número de clientes novos por mês.

Fonte: O próprio autor

geolocalização dos dez clientes mais atacados, somente os Estados Unidos e França tem aparição na Tabela 4.11, demonstrando que os endereços IP que sofreram mais ataques DDoS não necessariamente estão nos países que aparecem com frequência na Tabela 4.10.

4.2.5 Transações por IP

O número de transações por IP é apresentado na Tabela 4.13, com um valor médio de 221,85, e variação mínima de 1 e máximo de 764.846. Os dados têm uma distribuição assimétrica positiva, que pode ser confirmado pela Figura 4.4 com escala logarítmica. Ao final 13,2% dos clientes realizaram mais que 30 consultas, a maioria dos IPs possui um baixo número de consultas, poucos IPs possuem um volume relativo de consultas 1,5% dos IPs com mais de 2.000 consultas. Esta diferença entre os números de consultas já foi abordada por (LONGO, 2015), mas existe a deterioração dos resultados devido ao resultado apresentado na Tabela 4.12,

	Endereço cliente	País	Volume (requisições)
1	23.234.34.15	US	764.846
2	103.55.25.148	CH	605.073
3	23.234.5.43	US	597.044
4	103.39.19.98	HK	569.338
5	104.160.183.35	US	544.978
6	198.44.251.140	US	478.527
7	78.227.159.180	FR	425.523
8	81.57.209.101	FR	358.601
9	5.254.111.199	RO	329.775
10	23.234.34.24	US	282.298

Tabela 4.11: Dez clientes mais atacados.

	Estatísticas	Perguntas
1	Média	15,28
2	Desvio padrão	676,94
3	Mínimo	1
4	1 percentil	1
5	5 percentil	1
6	1 quartil	1
7	Mediana	1
8	3 quartil	1
9	95 percentil	1
10	99 percentil	13
11	Máximo	143.413

Tabela 4.12: Estatística de número de transações por IP para o mês de maio

onde é possível observar a distribuição no mês de maio, com um elevado número de clientes que realizaram somente uma consulta.

(LONGO, 2015) obteve uma distribuição média de 941,4 com variação mínima de 1 e máxima de 98.217. Com o 1 percentil, 5 percentil e 3° quartil (com valores 1) é possível demonstrar o baixo número de consultas por IPs.

	Estatísticas	Número de consultas
1	Média	221,85
2	Desvio padrão	4664,39
3	Mínimo	1
4	1 percentil	1
5	5 percentil	1
6	1 quartil	1
7	Mediana	1
8	3 quartil	1
9	95 percentil	474
10	99 percentil	3.125
11	Máximo	764.846

Tabela 4.13: Estatísticas de número de transações por IP.

4.2.6 Domínios e RRs

O DNSpot finalizou com um total de 4.982 RRs distintos. Este valor leva em conta os diferentes QNAME e QTYPE de cada consulta, onde um QNAME pode ter sido usado com diferentes QTYPEs. A Tabela 4.14 apresenta os quinze RRs mais populares; nessa listagem é possível encontrar apenas um RR em comum com (LONGO, 2015), que é `gransy.com ANY`. O valor total é inferior ao encontrado no Tabela 4.3 devido a requisições malformadas.

Os RRs `fema.gov`, `nccih.nih.gov` e `wapa.gov` são responsáveis por 79,42 % de todas as consultas. Levando em consideração todos os 15 RRs apresentados é encontrado um valor igual a 99,01% do total, apresentando uma distribuição maior em relação a (LONGO, 2015), que encontrou um total de 97% das consultas para um único RR. Mas ainda é possível destacar a concentração de consultas em poucos RRs verificada em ambos os estudos.

O nome `nccih.nih.gov.pkt ANY` é inválido, pois não existe o domínio de primeiro nível `.pkt`. Considerando a frequência com que ele aparece nas consultas, é provável que tais consultas sejam geradas por uma ferramenta de ataques DDoS mal configurada, que pretendia usar o RR `nccih.nih.gov ANY`.

A Tabela 4.15 apresenta a distribuição dos RRs, de acordo com os domínios de primeiro nível mais populares encontrados. Os cinco domínios apresentados foram encontrados em 65,3% de todos os RRs.

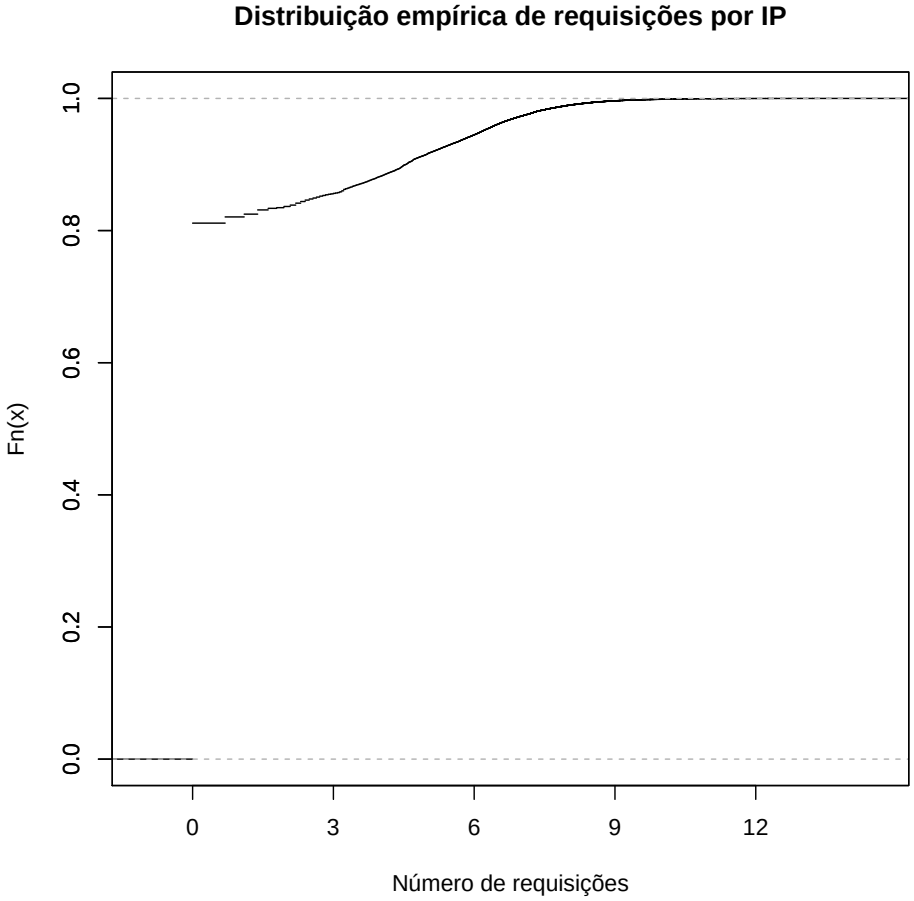


Figura 4.4: Distribuição empírica de requisições por IP.

Fonte: O próprio autor

A Tabela 4.16 apresenta o número de transações por RR. Foi encontrada uma média de 7.931,24, com uma variação de mínimo 1 e máximo 14.996.411 transações por RR, como apresentado na Figura 4.5. Em torno de 99% dos RRs tiveram menos de 57 requisições, 95% com menos de 6 requisições. Este resultado também pode ser observado em (LONGO, 2015) onde 70,0% dos RRs tiveram uma única consulta.

O número de consultas por QTYPE é apresentado na Tabela 4.17. Foram observados três tipos distintos (A,TXT e ANY), com uma ampla predominância de ANY (94% do total). Essa distribuição também pode ser observada em (LONGO, 2015), onde existe a predominância do tipo ANY (99,2% do total).

A Tabela 4.18 apresenta o tamanho das consultas e respostas para os 15 RRs mais consultados (apresentados na Tabela 4.14), e o respectivo fator de amplificação (razão entre tamanho da resposta e tamanho da consulta). Os resultados da Tabela 4.18 apresentam uma diferença significativa em relação aos valores encontrados por Longo (2015): quatro dos quinze

	RR	Número de consultas	Porcentagem
1	fema.gov. ANY	14.996.411	46,94
2	nccih.nih.gov. ANY	6.139.702	19,22
3	wapa.gov. ANY	4.237.711	13,26
4	NRC.GOV. ANY	1.693.615	5,30
5	1x1.cz. ANY	1.176.287	3,68
6	. ANY	1.135.121	3,55
7	nccih.nih.gov.pkt. ANY	873.652	2,73
8	commerce.gov. A	684.237	2,14
9	learnengs.com. ANY	335.241	1,04
10	cpsc.gov. A	147.089	0,46
11	gransy.com. ANY	93.475	0,29
12	cdn147.megaporno.se. ANY	30.293	0,09
13	cdn142.megaporno.se. A	27.978	0,08
14	cdn152.megaporno.se. A	27.918	0,08
15	cdn157.megaporno.se. ANY	27.520	0,08
16	Outros	317.375	0,99
17	Total	31.943.625	100,00

Tabela 4.14: Distribuição das consultas observadas pelo DNSpot.

RRs não possuem um fator de amplificação significativo para ataques DDoS, sendo que um destes é inexistente (`nccih.nih.gov.pkt ANY`). Considerando os RRs `*.megaporno.se`, que possuem 21 ocorrências ao todo, dez deles possuem tamanho de resposta maior que 4.000 bytes, e onze possuem respostas menores que 300 bytes.

O tráfego esperado para as respostas dos 15 RRs mais consultados é apresentado na Tabela 4.19. Juntos, os cinco RRs mais frequentes que são responsáveis por aproximadamente 91 GB de tráfego de resposta.

Ao longo do estudo, foi observado o desaparecimento de 17 dos RRs. Ao total de todas as transações realizadas pelo DNSpot, 3.734 clientes passaram a ter uma resposta de nome inexistente (RCODE=3, *NxDomain*) no lugar de uma resposta válida (RCODE=0, *NoError*). Desses clientes, 126 fizeram mais de 50 consultas por um RR desaparecido, e três chegaram a fazer mais de 100 requisições. Isso sugere que ferramentas de ataque DDoS estão usando algum mecanismo para identificar que um RR não está mais sendo obtendo resposta.

	Domínio	Quantidade presente	Porcentagem
1	.gov	22	0.4415897
2	.ru.	44	0.8831794
3	.info.	199	3.99438
4	.net.	392	7.868326
5	.com	2600	52.18788
6	Total	4982	65.37535

Tabela 4.15: Popularidade de domínios.

	Estatísticas	Número de consultas
1	Média	7.931,24
2	Desvio padrão	260.432,37
3	Mínimo	1
4	1 percentil	1
5	5 percentil	1
6	1 quartil	1
7	Mediana	1
8	3 quartil	2
9	95 percentil	6
10	99 percentil	56,19
11	Máximo	14.996.411

Tabela 4.16: Estatísticas de número de transações por RR.

	QTYPE	Número de consultas	Porcentagem
1	ANY	30.040.775	93,7
2	A	1.902.832	6,3
3	TXT	18	0,0
4	Total	31.943.625	100,0

Tabela 4.17: Tipos (QTYPE) usados nas consultas.

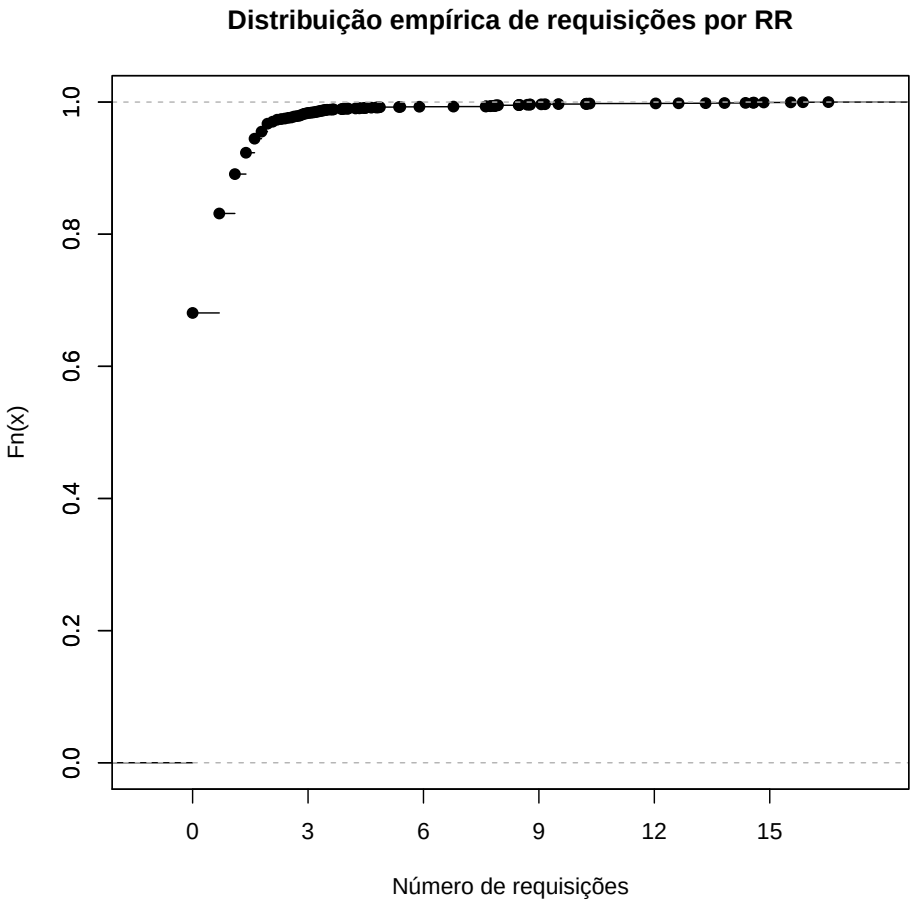


Figura 4.5: Distribuição empírica de requisições por RR.

Fonte: O próprio autor

	RR	Consulta (bytes)	Resposta (Bytes)	Fator de Amplificação
1	fema.gov. ANY	42	3893	92,6904
2	nccih.nih.gov. ANY	42	1892	45,0476
3	wapa.gov. ANY	42	4096	97,5238
4	NRC.GOV. ANY	42	2157	51,3571
5	1x1.cz. ANY	37	3834	103,6216
6	. ANY	42	2237	53,2619
7	nccih.nih.gov.pkt. ANY	37	121	3,2702
8	commerce.gov. A	42	57	1,3571
9	learnengs.com. ANY	35	51	1,4571
10	cpsc.gov. A	39	53	1,3589
11	gransy.com. ANY	42	4096	97,5238
12	cdn147.megaporno.se. ANY	42	4032	96,0000
13	cdn142.megaporno.se. A	42	4032	96,0000
14	cdn152.megaporno.se. A	42	112	2,6666
15	cdn157.megaporno.se. ANY	42	250	9,9523

Tabela 4.18: Tamanho de consulta e resposta e fator de amplificação para os 15 RRs mais populares.

	RR	Tráfego de consulta (MB)	Tráfego esperado de resposta (MB)
1	fema.gov. ANY	600,67	55.676,34
2	nccih.nih.gov. ANY	245,92	11.078,18
3	wapa.gov. ANY	169,73	16.553,56
4	NRC.GOV. ANY	67,83	3.483,89
5	1x1.cz. ANY	41,50	4.300,96
6	. ANY	45,46	2.421,63
7	nccih.nih.gov.pkt. ANY	30,82	100,81
8	commerce.gov. A	27,40	37,19
9	learnengs.com. ANY	11,18	16,30
10	cpsc.gov. A	5,47	7,43
11	gransy.com. ANY	3,74	365,13
12	cdn147.megaporno.se. ANY	1,21	116,16
13	cdn142.megaporno.se. A	1,12	107,52
14	cdn152.megaporno.se. A	1,12	2,98
15	cdn157.megaporno.se. ANY	1,10	10,94

Tabela 4.19: Tráfego esperado de resposta para os 15 RRs mais consultados.

4.3 Ataques DoS

Atualmente, é esperado que uma parte do tráfego direcionado a um servidor DNS recursivo aberto seja devido a ataques DDoS em que o servidor esteja sendo usado como refletor. Nesta seção são caracterizados os ataques DDoS em que o DNSpot foi usado como refletor. Para essa caracterização, foi adotada a definição de ataque DoS proposta por Longo (2015):

> Um ataque DoS é formado por um conjunto com no mínimo 5 consultas com o mesmo IP de origem e com espaçamento máximo de 60 segundos entre consultas consecutivas, e pelas respostas a essas consultas.

De acordo com a definição, é possível afirmar que o DNSpot sofreu um total de 23.788 ataques DoS distintos. O primeiro ataque ocorreu no primeiro dia em que o DNSpot foi colocado em funcionamento. Um total de 23.745 IPs estavam envolvidos com ataques DoS (pela natureza dos ataques, é provável que estes sejam as vítimas), e 840 RRs foram utilizados nas consultas correspondentes a esses ataques.

	Métricas	Envolvido em DoS	Total	Envolvidos	Envolvidos (LONGO, 2015)
1	IPs	23.745	184.564	12,87%	81,6%
2	RRs	840	4.982	16.86%	64,0%
3	N° de consultas	30.661.228	32.358.928	94,75%	99,9%

Tabela 4.20: Porcentagem do envolvimento em DoS de métricas comparadas aos totais computados no DNSpot.

Comparando as proporções da Tabela 4.20 com os dados de (LONGO, 2015), observa-se que todas as métricas tiveram redução expressiva. Duas explicações para o fenômeno são possíveis. Uma delas, já discutida, é o uso do DNSpot como servidor recursivo propriamente dito, e não como refletor em ataques DDoS. Outra explicação seria uma possível mudança na natureza do tráfego DDoS, que faz com que a definição proposta em (LONGO, 2015) não seja mais adequada. Essa segunda hipótese precisaria ser melhor investigada na continuação deste trabalho.

A Figura 4.6 apresenta a distribuição das requisições com uma mesma porta de origem. 850 ataques DoS (3,6% do total) mantiveram a mesma porta de origem nas consultas. Este fenômeno já foi observado por (LONGO, 2015), apontando a existência de uma ferramenta automatizada para a realização dos ataques.

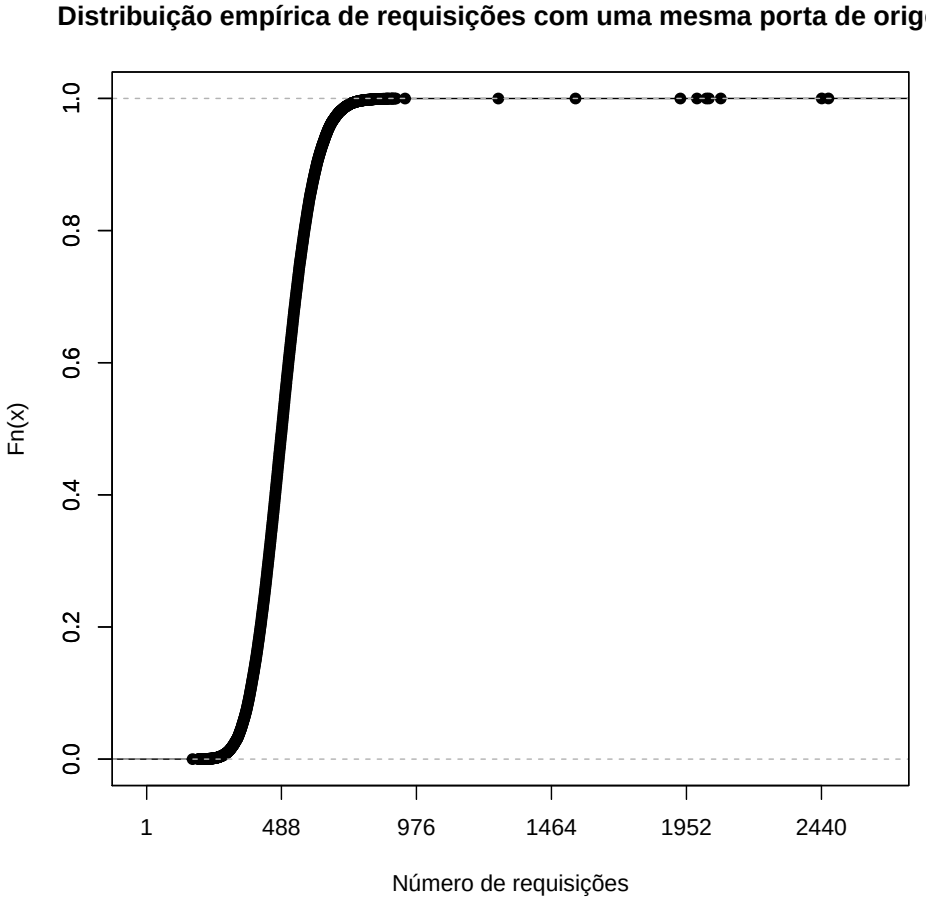


Figura 4.6: Distribuição empírica de requisições com uma mesma porta de origem associadas a ataques DoS.

Fonte: O próprio autor

4.3.1 Especificação dos ataques DoS

A duração dos ataques DoS é apresentada na Tabela 4.21, e sua distribuição empírica é mostrada na Figura 4.7. Os ataques tiveram uma duração mínima menor que 1 segundo e máxima de 59 minutos, com uma média de 13 minutos e 46 segundos.

Os ataques observados tem uma duração pequena: 25% das consultas observadas duram até 3 minutos, 75% duram até 19 minutos. Este tempo de duração dos ataques DoS é relativamente maior que o tempo encontrado por (LONGO, 2015), onde 95% das consultas observadas obtiveram uma duração menor que 9 minutos.

Na Figura 4.7 é apresentado a distribuição empírica da duração de ataques DoS.

Foram observados um valor mínimo de 5 e máximo de 203.474 requisições por ataque DoS, com média de 6.029,7 e desvio padrão de 13.374,18. A distribuição empírica das

	Estatísticas	Duração (minutos)
1	Média	13,46
2	Desvio padrão	15,08
3	Mínimo	< 1
4	1 percentil	< 1
5	5 percentil	< 1
6	1 quartil	3
7	Mediana	8
8	3 quartil	19
9	95 percentil	53
10	99 percentil	56
11	Máximo	59

Tabela 4.21: Estatísticas da duração de ataques DoS.

requisições pode ser observada na Figura 4.8; em (a) o eixo x é linear, o que torna possível observar alguns dos ataques com o maior volume; em (b) o eixo x está em escala logarítmica, o que permite visualizar a distribuição para os ataques com poucas requisições.

É possível observar que 25% dos ataques realizados tiveram até 1.010 requisições e 75% dos ataques observados tiveram até 5.024 requisições. Em relação a (LONGO, 2015), onde 75% dos ataques tiveram até 402 requisições, constata-se que houve um aumento no volume de requisições por ataque, de forma geral.

As estatísticas de número de ataques DoS por IP, é apresentado na Tabela 4.23 e a Figura 4.9 mostra a respectiva distribuição para a tabela. Com uma média de 1,01 por IP, sendo o mínimo 0 e máximo 41 com um desvio padrão de 0,2, é possível observar que 95% dos IPs esteve envolvido em somente um ataque. Os cinco clientes que apresentaram o maior número de ataques na Figura 4.23 são apresentados na Tabela 4.24, com a sua geolocalização, IP e número de ataques.

4.3.2 *Open DNS Server*

Buscando um melhor entendimento de como foi utilizado o endereço 200.19.107.235 (endereço do DNSPot) e por quais meios estava sendo realizado a divulgação do serviço de DNS, foi realizado uma busca para tentar encontrar o meio de divulgação

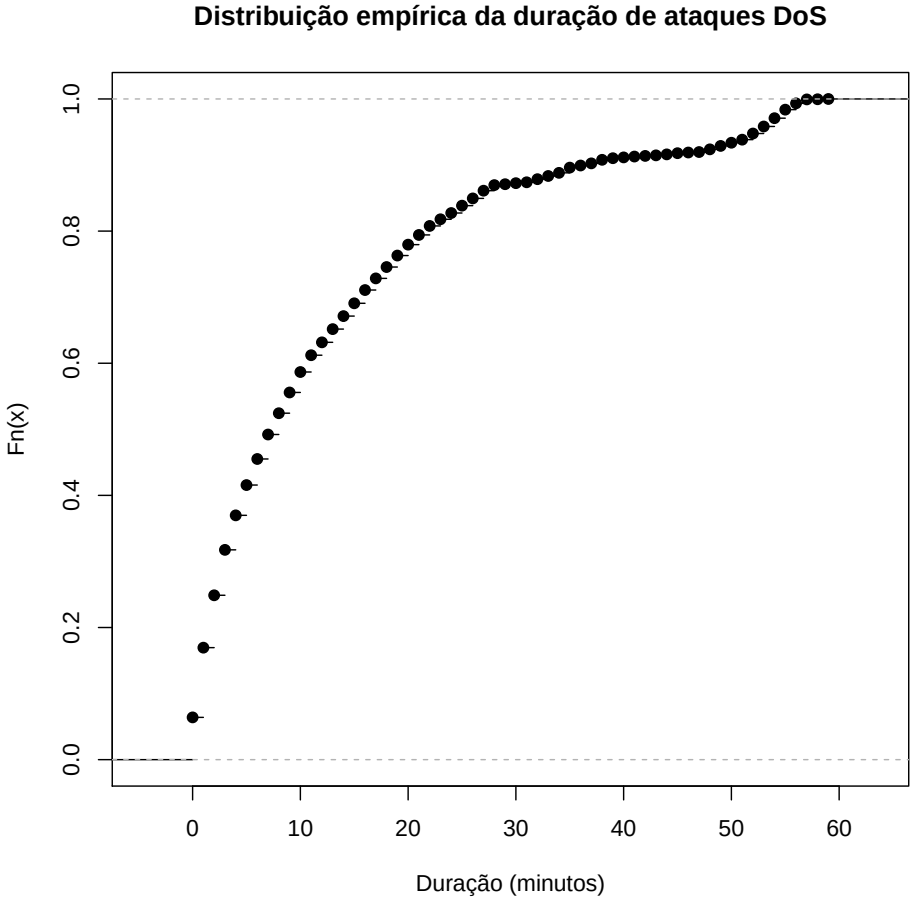
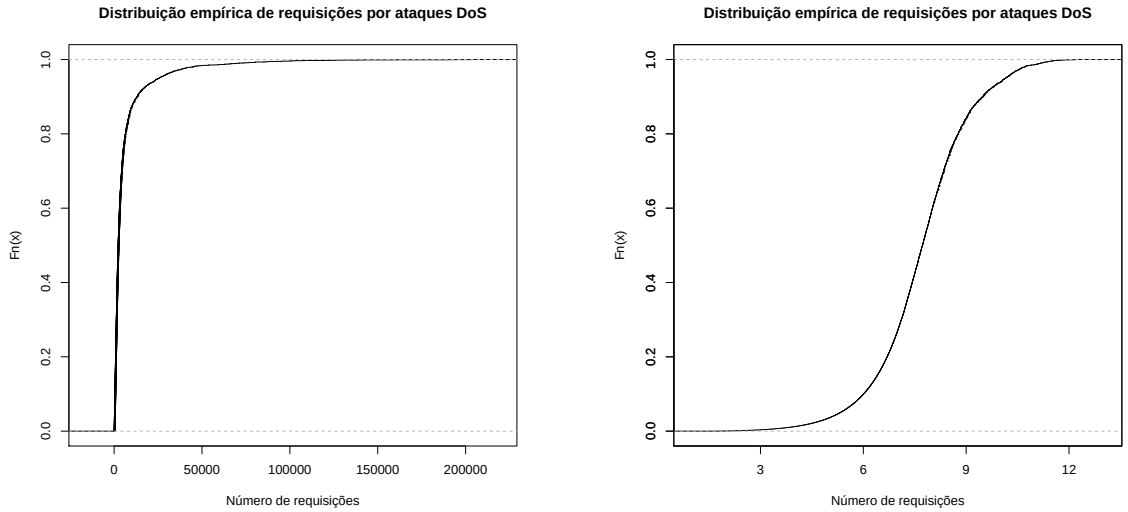


Figura 4.7: Distribuição empírica de duração de ataques DoS.

Fonte: O próprio autor



(a)

Fonte: O próprio autor

(b)

Fonte: O próprio autor

Figura 4.8: Distribuição empírica de requisições por ataques DoS. (a) Eixo x linear. (b): Eixo x escala logarítmica.

	Estatísticas	Número de requisições
1	Média	6.029,71
2	Desvio padrão	13.374,18
3	Mínimo	5
4	1 percentil	41,12
5	5 percentil	201,60
6	1 quartil	1.010
7	Mediana	2.264
8	3 quartil	5.024
9	95 percentil	25.594,40
10	99 percentil	70.355,88
11	Máximo	203.474

Tabela 4.22: Estatísticas de requisições por ataque DoS.

	Estatísticas	Quantidade de ataques DoS
1	Média	1,01
2	Desvio padrão	0,20
3	Mínimo	0
4	1 percentil	1
5	5 percentil	1
6	1 quartil	1
7	Mediana	1
8	3 quartil	1
9	95 percentil	1
10	99 percentil	2
11	Máximo	41

Tabela 4.23: Estatísticas de número de ataques DoS por IP.

destes endereços e qual o seu objetivo.

A lista observada foi a *Public DNS info*, que oferece uma tabela com uma média de 8,5 milhões de endereços IPs de serviço de DNS aberto. A lista contém um total de IPs de 220 países, destes o Brasil possui listado um total de 1.430 endereços IPs que podem ser utilizados para ataques DDoS, o endereço do DNSPot não foi encontrado nesta lista.

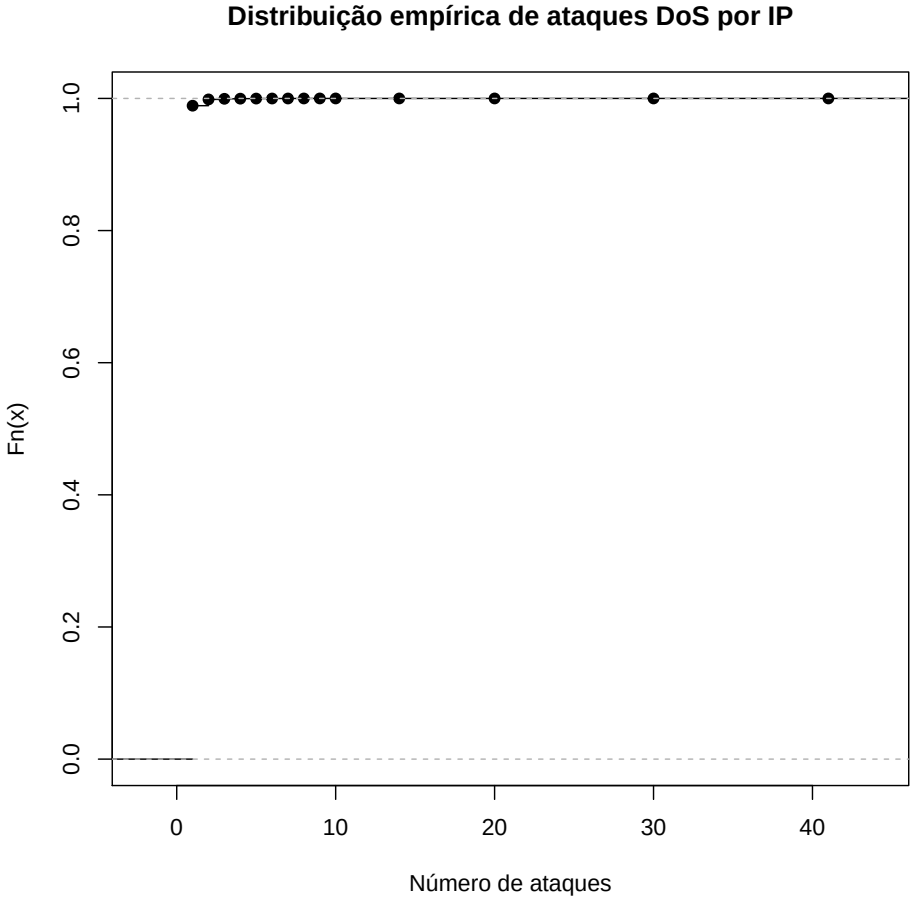


Figura 4.9: Distribuição empírica de ataques DOS por IP

Fonte: O próprio autor

	Número de ataques	IPs	País
1	41	223.205.167.20	TH
2	20	185.94.111.1	RU
3	30	183.56.172.145	CN
4	14	208.74.190.211	US
5	10	5.254.111.199	RO

Tabela 4.24: Cinco IPs mais atacados.

Outras listas de acesso público também foram observadas, mas não foi possível encontrar o endereço do DNSPot. Acredita-se que este resultado é consequência dos domínios ignorados, como apresentando na Tabela 4.1. Existe a possibilidade do serviço estar sendo listado em uma lista que não foi encontrada pelo autor do trabalho.

4.4 Análise Temporal

Esta seção analisa o comportamento do tráfego observado pelo DNSpot ao longo do tempo. Para melhor entendimento desse comportamento foram analisadas as quantidades de requisições recebidas e de respostas enviadas por dia e por mês.

As Figuras 4.10 e 4.11 apresentam o número de requisições processadas e de respostas enviadas por dia, respectivamente. Observa-se um comportamento altamente irregular, com bastante variação no volume de tráfego processado. Também é possível constatar que a proporção entre requisições recebidas e respostas enviadas não é constante. Embora as respostas sejam provocadas pelas requisições, a ocorrência de dias com muitas requisições e um número consideravelmente menor de respostas evidencia a atuação dos mecanismos de contenção de tráfego DDoS do DNSpot.

Os dados mensais, mostrados na Figura 4.12, complementam os dados diários. Percebe-se que o volume de tráfego entre set/16 e jan/17 é baixo, aumentando sensivelmente a partir de fev/17. Em parte, isso pode ser explicado pelo fato do DNSpot ter ficado inativo entre dez/15 e set/16, o que pode ter causado sua exclusão de listas de servidores recursivos abertos e a consequente necessidade de ser redescoberto mediante varreduras de rede.

O tráfego recebido é distribuído irregularmente. Na maioria dos meses, ele fica concentrado em alguns poucos períodos de três a cinco dias do mês. Com exceção dos meses de novembro, março e maio que apresentam períodos de tráfego sustentado, chegando a 19 dias em maio. O tráfego de consulta também é irregular e na maioria dos meses acompanha o tráfego recebido, com exceção em março (dias 7 a 11), abril (21 e 22) e maio (18 e 19). Não é possível realizar alguma previsão ao observar a quantidade de requisições e respostas diárias.

O mês com maior volume de tráfego foi março de 2017, com 13.737.569 consultas e 1.036.079 respostas (respectivamente 42,5% e 49,1% do total), mais que o dobro de mai/17, que aparece em segundo lugar. Não foi identificada nenhuma causa específica para esse fenômeno.

A maior diferença entre consultas recebidas e respostas enviadas foi verificada em abr/17. Essa discrepância pode ser explicada pela perda de informações no banco de dados, que se deveu ao crescimento do próprio banco.

Ao juntar todas as versões geradas do banco de dados, foi constatado que algumas versões possuíam tabelas mal formadas ou o banco estava corrompido. Para contornar este problema algumas linhas de tabelas específicas foram ignoradas, desta forma possibilitando a união

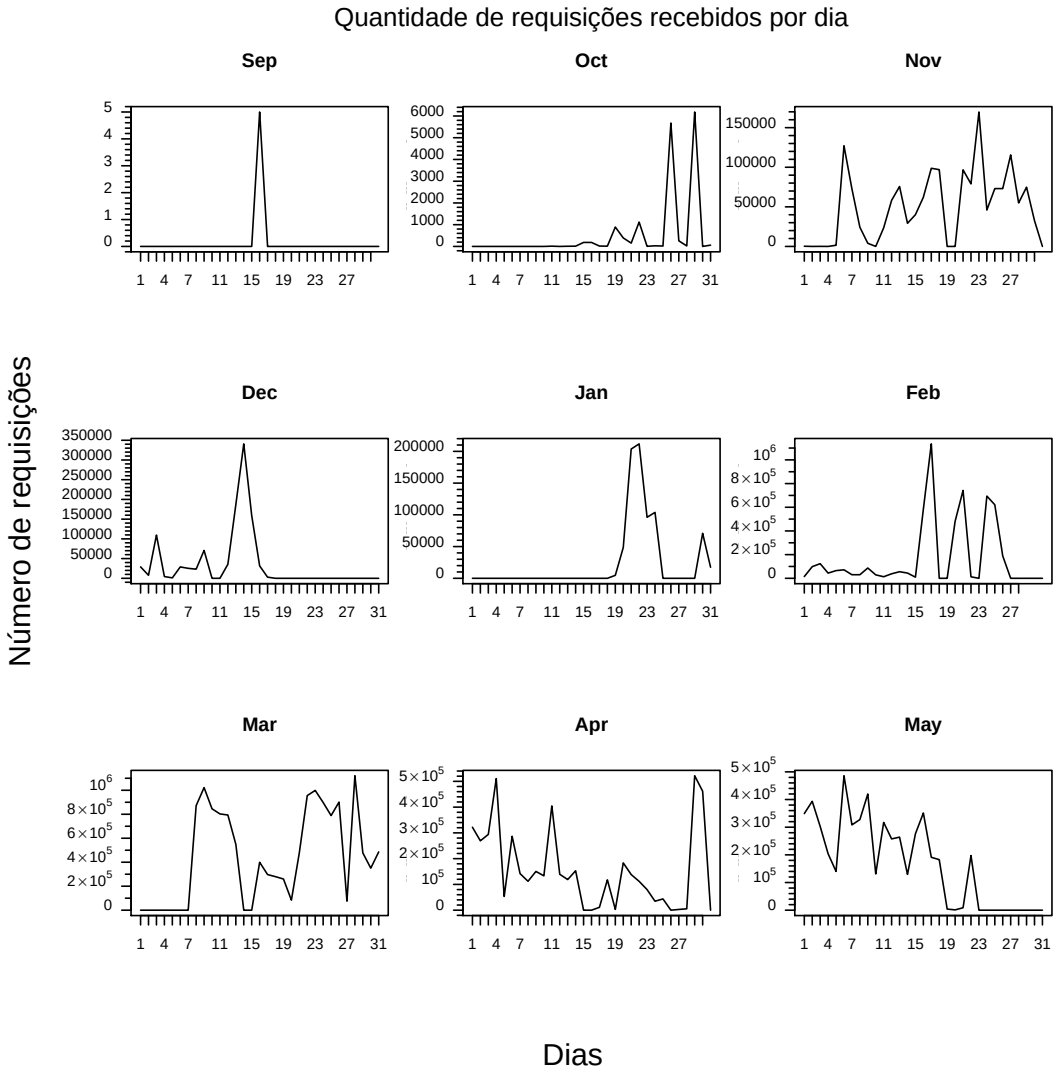


Figura 4.10: Quantidade de requisições recebidos por dia.

Fonte: O próprio autor

de todas as versões geradas do banco. Algumas tabelas possuíram uma perda maior, a tabela de transações chegou a ter uma perda de 20.000 linhas (o que corresponde a 20.000 transações perdidas).

A Tabela 4.25 apresenta a variação nos tamanhos dos pacotes recebidos pelo DNS-pot ao decorrer dos meses que o sistema ficou online. O crescimento acabou ocasionando uma leve perturbação no tamanho dos pacotes recebidos em 2017. Um fator a ser destacado é a distribuição do tamanho médio entre os meses de fevereiro, março e abril, que apresenta uma distribuição do valor médio muito próxima, e outubro, novembro e dezembro, com outra distribuição relativamente próxima. Esta distribuição pode ser explicada por um elevado número de ataques de um mesmo cliente.

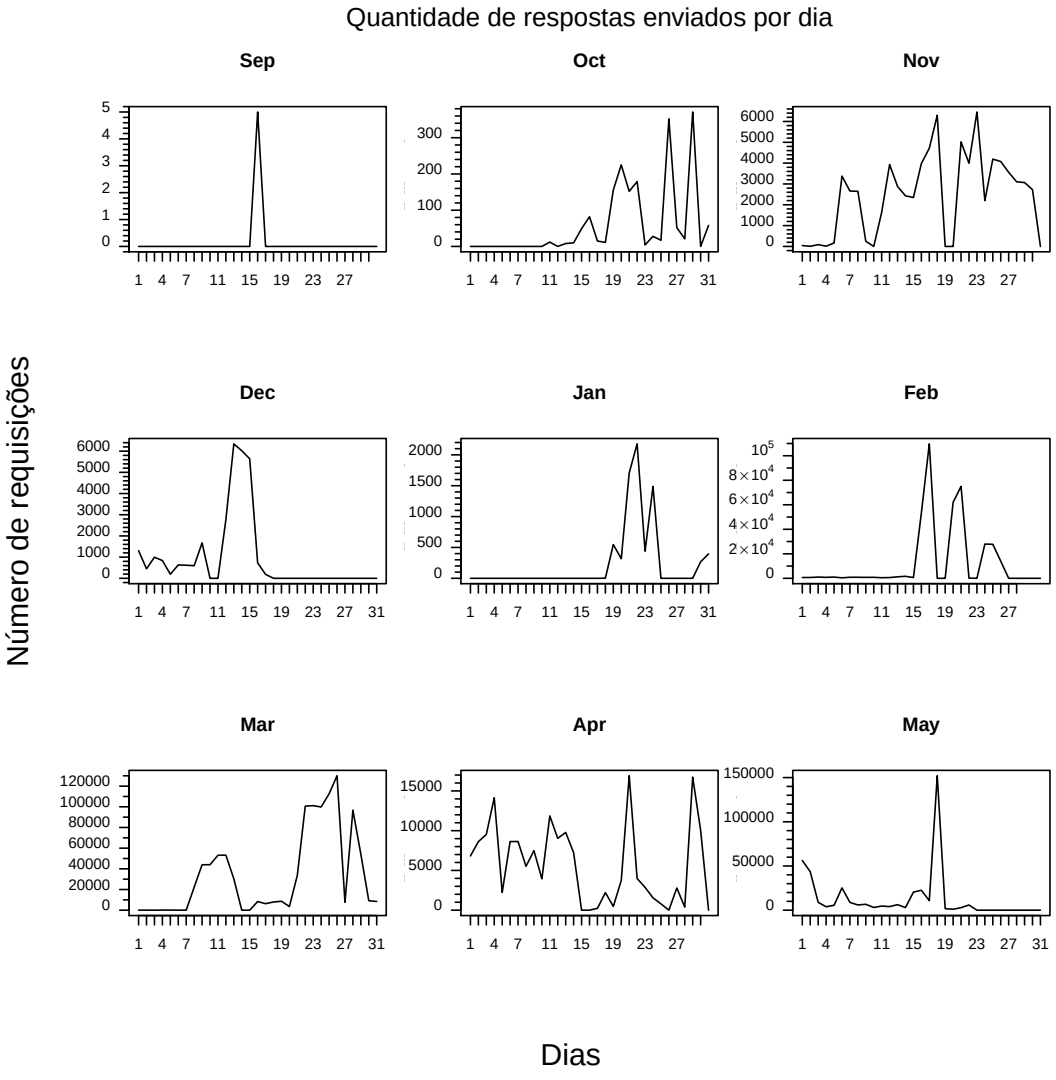


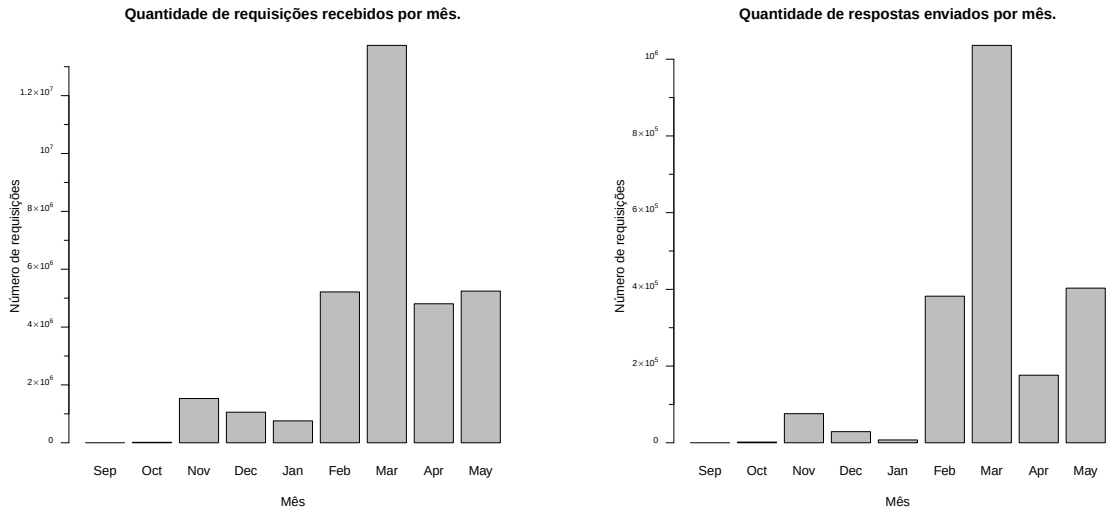
Figura 4.11: Quantidade de respostas enviados por dia.

Fonte: O próprio autor

A porcentagem apresentada revela a distribuição dos pacotes recebidos em cada mês. É possível destacar que o mês de março recebeu 42,4% de todas as requisições analisadas neste período de nove meses. Esta variação não pode ser observada na Tabela 4.8, e contribui com a distribuição de tamanhos encontrados na Tabela 4.9. O mês de janeiro apresentou as consultas com o menor tamanho médio dentre todos os meses, e maio apresentou o maior tamanho médio.

A Tabela 4.26 apresenta a variação nos tamanhos de pacotes enviados. É possível observar no valor médio uma distribuição de tamanhos maiores em relação aos valores apresentados na Tabela 4.9.

O mês de maio apresenta a distribuição com as consulta de menor tamanho em re-



(a) Requisições

Fonte: O próprio autor

(b) Respostas

Fonte: O próprio autor

Figura 4.12: Quantidade de requisições recebidas e respostas enviadas.

	Tamanho máximo	Tamanho mínimo	Tamanho médio	Mês	Porcentagem
1	36	25	33.60	Sep	0.0000
2	72	24	35.15	Oct	0.0471
3	106	17	34.61	Nov	4.7281
4	50	21	35.41	Dec	3.2647
5	46	8	28.01	Jan	2.3394
6	89	17	38.36	Feb	16.1155
7	56	17	38.76	Mar	42.4537
8	343	17	36.88	Apr	14.8486
9	50	17	44.36	May	16.2029

Tabela 4.25: Análise dos tamanhos de consultas recebidas por mês.

lação aos nove meses observados. Este valor aparentemente é consequência do uso do DNSpot como resolvedor padrão por usuários finais, como discutido nas Seções 4.5 e 4.2.4.

Estas características são apresentadas na Tabela 4.8, onde 25% das consultas possuem até 28 bytes de tamanho e 95% destas tamanho de até 46 bytes. Para as respostas enviadas 75% possuem um tamanho de até 3.893 e 25% das consultas apresentam valor maior que 3.893 bytes. O mês de março apresenta a maior variação em relação a quantidade de pacotes enviados, com um total de 49% de todos os pacotes enviados nesta análise.

	Tamanho máximo	Tamanho mínimo	Tamanho médio	Mês	Porcentagem
1	3548	25	2100,00	Sep	0.0002
2	4096	12	2170.01	Oct	0.0853
3	4096	12	2332.83	Nov	3.5915
4	3959	17	2674.88	Dec	1.3724
5	3763	12	2065.38	Jan	0.3477
6	4085	12	2949.81	Feb	18.0993
7	4085	12	1979.13	Mar	49.0650
8	4094	12	2485.57	Apr	8.3439
9	4094	12	873.36	May	19.0946

Tabela 4.26: Análise dos tamanhos de consultas enviadas por mês.

4.5 Anomalias de Tráfego

Ao decorrer das análises algumas anomalias foram identificadas, esta seção busca destacar estas anomalias e demonstrar qual o entendimento destas anomalia de acordo com as informações apresentadas pelo DNSpot.

4.5.1 Diminuição no tamanho dos pacotes

Esta análise busca verificar se alguma medida foi tomada para reduzir o impacto de ataques DDoS durante este período. Para verificar estas alterações foi gerada uma lista com todos os clientes que enviaram pacotes com tamanho maior que 3.550 bytes (sendo o valor com maior popularidade entre os dez RRs mais utilizados em (LONGO, 2015)), e verificado se houve alguma modificação no tamanho do pacote ao decorrer do tempo (com um foco nos pacotes menores que 30 bytes).

Entre todas as transações realizadas pelos clientes, 17.677 clientes tiveram respostas com tamanho igual ou superior a 3.550 bytes. Desses, somente 1.307 clientes obtiveram mensagens de resposta com o tamanho reduzido no decorrer do período de análise. O fator de redução destas consultas pode ser observado na Tabela 4.27. Ao todo estes clientes foram responsáveis por 7.382.209 requisições; considerando o fator médio de redução encontrado, pode-se estimar que a mudança nas respostas tenha produzido uma redução no tráfego de aproximadamente 25 GB.

	Estatísticas	Fator de Amplificação
1	Média	99.42
2	Desvio padrão	0.13
3	Mínimo	99.28
4	1 percentil	99.30
5	5 percentil	99.32
6	1 quartil	99.33
7	Mediana	99.36
8	3 quartil	99.41
9	95 percentil	99.70
10	99 percentil	99.70
11	Máximo	99.70

Tabela 4.27: Fator de redução.

4.5.2 Nomes utilizados pelos domínios

Ao analisar os nomes de domínios utilizados, foi possível identificar algumas características, estas são apresentadas nesta Seção.

Domínios com múltiplos nomes consultados

Na maioria dos RRs presentes nas consultas processadas pelo DNSpot, havia apenas um nome em cada domínio (sufixo) distinto, desconsiderando domínios de primeiro nível globais (como .com e .net) e nacionais (como .com.br e .net.br). No entanto, foram observados domínios com um grande número de nomes consultados. Os três domínios com maior número de nomes distintos foram `test003.com` (com 292 nomes), `aegins-dns.info` (178 nomes) e `megaporn.se` (21 nomes). Esses nomes apresentaram características peculiares, que são discutidas a seguir.

`test003.com`. Consultas para qualquer nome no domínio `test003.com` com QTYPE A ou ANY são resolvidas para o mesmo endereço IP, 204.11.56.48. Uma consulta ao (WHOIS, 2017a) revela que esse domínio foi registrado em 18/05/2017 por uma empresa que atua com *domain parking* (ICANNWIKI, 2017). Alguns exemplos de nomes observados pelo DNSpot são mostrados abaixo; eles sugerem algum tipo de canal coberto, possivelmente um canal de comando e controle de *malware* (PCWORLD, 2017). Como houve a mudança no registro do

domínio desde que o tráfego foi observado pelo DNSpot (período de 17/10/2016 a 20/02/2017), não é possível inferir mais sobre esse domínio em particular.

```
xpq.test003.com ANY
9etbf41.test003.com ANY
ql3rdfc.test003.com ANY
```

`aegins-dns.info`. Aegins é um provedor de DNS filipino (AEGINS, 2017). Consultas para qualquer nome no domínio `aegins-dns.info` retornam RCODE=0 (*NoError*) e uma resposta vazia. De acordo com a RFC 2308 (ANDREWS, 1998), essa resposta indica que não existe um RR com o QNAME e QTYPE pesquisados, mas que existem dados para o mesmo QNAME e um ou mais QTYPEs diferentes. Em comparação, uma resposta com RCODE=3 (*NxDomain*) indica que não existe nenhum RR com o QNAME consultado, de nenhum QTYPE. A anomalia identificada é que respostas vazias são retornadas mesmo quando o QTYPE é ANY, algo que não faz sentido. Alguns exemplos de nomes observados pelo DNSpot são mostrados abaixo; embora eles também possam estar sendo usados como canais cobertos, a estrutura irregular e o fato de serem nomes curtos reduz essa probabilidade em comparação ao caso anterior.

```
hi.aegins-dns.info ANY
e0.aegins-dns.info ANY
sdbur.aegins-dns.info ANY
upey9.aegins-dns.info ANY
```

`megaporno.se`. Os nomes nesse domínio seguem o padrão `cdn`*`NNN`*`.megaporno.se`, onde *`NNN`* é um número de dois ou três dígitos. Alguns desses nomes (como `cdn147.megaporno.se`) contêm mais de 250 registros com QTYPE A, apontando para endereços IP pertencentes à mesma sub-rede; as respostas correspondentes têm aproximadamente 4 KB de tamanho, o que torna esses nomes interessantes para ataques DDoS devido ao seu alto fator de amplificação (de aproximadamente 96,0). Em contrapartida, outros nomes nesse domínio retornam respostas pequenas, da ordem de 250 bytes, que são ineficazes para ataques DDoS.

```
cdn153.megaporno.se. ANY
cdn651.megaporno.se. ANY
cdn241.megaporno.se. ANY
```

Outro ponto notável é que, segundo o WHOIS, os endereços IP contidos nas respostas grandes estão alocados a provedores de Internet residencial (cabo, DSL, etc.), em diferentes localizações geográficas, como Estados Unidos, França e Holanda. Isso parece ser altamente improvável para uma CDN (*Content Delivery Network*, ou rede de fornecimento de conteúdo (VAKALI; PALLIS, 2003)), como sugerido pelo nome (`cdn`*`NNN`*`.*`).

A variação de nomes no domínio `megaporno.se` pode estar sendo usada para evitar a concentração de requisições em um único nome e assim chamar menos atenção. Outra hipótese é que a diversificação seja usada como precaução contra a reconfiguração do DNS para reduzir o tamanho das respostas caso os nomes sejam envolvidos em ataques DDoS.

Protocolos associados a Zeroconf

Foram identificadas consultas geradas por *DNS Service Discovery* (DNS-SD) (CHESHIRE; KROCHMAL, 2013), um protocolo associado a *Zeroconf* (GROUP, 2017). O *Zeroconf* (*Zero Configuration Networking*) é um conjunto de técnicas que permitem que clientes configurem automaticamente uma rede IP e descubram serviços disponíveis nessa rede sem a necessidade de intervenção manual. O *Zeroconf* é suportado por Apple MacOS X, Unix e GNU/Linux. Os nomes abaixo correspondem a consultas para descobrir o domínio de rede *default* recomendado para registro de serviços (`dr.*`) e para *browsing* (`db.*`) nas respectivas redes (192.168.1.0 e 192.168.177.0) (CHESHIRE; KROCHMAL, 2013). Como essas faixas de endereços pertencem a redes privadas (REKHTER, 1996), sendo frequentemente usadas para atribuir endereços a dispositivos em redes domésticas, tais consultas aparentam ser decorrência de configuração incorreta, especificamente do DNSpot configurado como servidor recursivo dessas redes. No entanto, é possível que se trate de varreduras usando DNS-SD em busca de alvos para ataques (ATLASIS, 2017).

```
dr._dns-sd._udp.0.1.168.192.in-addr.arpa. ANY
db._dns-sd._udp.0.177.168.192.in-addr.arpa. ANY
```

Nomes típicos de usuários finais

Além das consultas DNS-SD descritas acima, uma lista de nomes tipicamente associados a usuários finais, como por exemplo `google.com`, `facebook.com` e `amazon.com`

foi observada, indicando que usuários estão configurando suas máquinas para utilizar o DNSpot como resolvedor padrão das consultas DNS. Este nomes possuem maior aparição no mês de maio (vide Seção 4.2.4 na página 45), reforçando a hipótese de que o aumento no número de clientes distintos em mai/2017 esteja associado ao uso do DNSPot como resolvedor regular.

O `local.cloud.mcafee.com` corresponde ao servidor de nomes dos subdomínios encontrados no DNSpot como `avgs.mcafee.com` são utilizados pelo GTI (*Global Terrorism Index*) (MCAFEE, 2017). Caso algum arquivo suspeito seja encontrado na máquina, e o antivírus não possua a sua assinatura, uma consulta DNS é enviada para um servidor da *McAfee Labs*. Ao receber a consulta, o servidor determina se o arquivo suspeito é algum tipo de ameaça. Ao total foram encontradas 21 variações deste nome na lista de domínos do DNSpot. Este é mais uma característica que sugere fortemente que o DNSpot está sendo configurado para ser utilizado por usuários finais.

```
0.11-a3092481.20483.1518.18a4.3ea1.410.0.ezg6u89nikkw32n9ssr1pcd8ei.avqs.mcafee.com. ANY
b-0.19-22094008.61081.1518.19d4.3ea1.410.0.cfqf8cl1gw829ff7ctuvqcsszi.avqs.mcafee.com. ANY
a-0.19-a309d000.2170092.1518.19d4.3ea1.210.0.mjcu9aaeb23cpducliwh1wip8j.avqs.mcafee.com. ANY
```

Varreduras de segurança

O DNSpot foi escaneado pela (RUHR UNIVERSITY BOCHUM, 2017) com um total de 16 consultas distintas, o objetivo do *scan* é buscar serviços que podem ser utilizados para ataques de amplificação. Após sua identificação esse domínio foi adicionado à lista de domínios ignorados, como observado na Tabela 4.1.

```
6t80.eb6b13c8.wc.syssec.rub.de. ANY
59xn.eb6b13c8.wc.syssec.rub.de. ANY
```

4.6 Discussão dos resultados

Ao final destas análises, é possível extrair várias considerações, dentre as quais destacam-se as seguintes:

- A utilização dos servidores de DNS recursivos abertos permite a realização de ataques DDoS com um fator de amplificação considerável. Ao limitar em 30 o número de con-

sultas diárias por endereço IP, o DNSpot consegue interagir com clientes na Internet e observar tais ataques, ao mesmo tempo em que reduz significativamente o tráfego enviado em caso de ataque.

- O DNSpot acabou recebendo no total 32,3 milhões de requisições, 94% das quais consistiam de ataques DDoS. Considerando os resultados encontrados por (LONGO, 2015), é possível afirmar que houve uma diminuição na proporção de ataques no tráfego recebido pelo DNSpot, de 99,9% para 94%.

- 25% dos ataques realizados duraram até 3 minutos. Deste total 75% dos ataques realizados tiveram uma duração de até 19 minutos, demonstrando um crescimento no tempo de alguns ataques, em relação ao tempo encontrado por (LONGO, 2015), com um tempo de até 9 minutos para 95% dos ataques realizados.

- Reforçando a análise de (LONGO, 2015), é importante destacar o elevado número de requisições em uma quantidade relativamente pequena de clientes.

- O mês de mar/2017 apresentou uma quantidade de 42,4% de todos os ataques realizados, não tendo sido encontrada nenhuma explicação lógica para essa quantidade de ataques.

- O *downtime* pode ter prejudicado a análise de algumas datas, principalmente feriados quando o acesso a máquina não era possível (problemas com falta de energia).

- No decorrer da análise foi possível observar uma mudança na intensidade de tráfego observado em relação a (LONGO, 2015). Essa característica é fortemente observada na evolução do tráfego ao longo dos meses em que o estudo aconteceu.

- O crescimento no número de requisições recebidas por dia em relação ao (LONGO, 2015), que possui registrado um total de 0,94 transações por segundo, gerando um total de 81.353,02 transações por dia, já neste trabalho foi registrado um crescimento neste valor com um total de 1,50 transações por segundo e 129.955,61 transações por dia.

- Observando o mês de mai/2017, é possível apontar um crescimento de 81,5% no número de clientes em um único mês, um crescimento 35 vezes maior que o período de observação de 49 dias, com um total de 4.287 clientes.

- O fator de amplificação dos quinze RRs mais populares revela quatro consultas com um fator de amplificação menor que os registrados por (LONGO, 2015), demonstrando a existência de RRs que não estão sendo utilizados para ataques de amplificação.

- Diversos indícios – crescimento no número de clientes, quantidade de respostas enviadas, queda no tamanho médio das respostas, característica dos nomes consultados – apontam que a partir de maio o DNSpot passou a ser usado como servidor recursivo por usuários finais, e não apenas para realizar ataques. A grande quantidade de clientes que realizaram apenas uma consulta nesse mês (vide Seção 4.2.5) levanta a hipótese de que o endereço IP do DNSpot não seja usado diretamente por esses usuários finais, mas por algum *proxy*, que possivelmente espalhe consultas por diversos servidores recursivos abertos.

- O uso de nomes DNS anômalos, já apontado por Longo (2015), cujos propósitos são no mínimo suspeitos, também pode ser observado neste estudo.

4.7 Considerações Parciais

Neste capítulo foram apresentados os resultados do período de análise de 250 dias, durante o qual foram processadas quase 43 milhões de requisições DNS. Foi possível definir um comportamento para alguns dos ataques, junto com uma nova visão em relação aos resultados apresentados por (LONGO, 2015), caracterizando uma maior variedade nos resultados. É possível observar a grande predominância em relação as requisições entre poucos clientes, e a utilização de ferramentas otimizadas para a realização dos ataques. Alguns dos comportamentos apresentados não foram encontrados em (LONGO, 2015), mostrando novas características antes não observadas em um estudo de curta duração.

A solução para as perguntas apresentadas na Seção 1 são listadas a seguir:

- Qual a diferença entre as transações realizadas com o DNSpot neste período de 250 dias com o trabalho (LONGO, 2015)?

  As diferença são apontadas na Seção 4.2.2, destacando o resumo das transações, transações ignoradas, RCODEs enviados ao cliente e taxa de transações processadas ao final do período de coleta de dados.

- Quais os volumes de dados recebidos neste período?

  O volume de dados é apresentado na Seção 4.2.3, junto com uma análise da distribuição dos tamanhos de pacotes de respostas e consultas.

- Os clientes apresentaram algum tipo de comportamento ou alguma diferença em relação ao trabalho (LONGO, 2015)?

A análise dos clientes é apresentada na Seção 4.2.4, e na Seção 4.2.5 é apresentado uma análise sobre as transações realizadas por IP.

- Os Domínios e RRs apresentam alguma diferença em relação ao trabalho anterior?

  Um estudo sobre os domínios e RRs é apresentado na Seção 4.2.6. Identificando os 15 RRs mais populares, seu respectivo fator de amplificação e tráfego esperado de resposta.

- Qual o comportamento dos ataques DoS observados?

  O comportamento dos ataques DoS é abordado na Seção 4.3, com uma especificação dos ataques DoS na seção 4.3.1.

- Correlacionar o tráfego de ataques DDoS com alguns fatores externos, como questões geopolíticas ou econômicas?

  Ao realizar uma análise sobre os dias com o maior número de requisições (dias com mais de 500.000 requisições), foi possível observar algumas datas com relações a eventos específicos (como feriados e datas políticas), mas não foi possível confirmar a participação do DNSpot nestes ataques em específico.

- Quais anomalias foram observadas?

  As anomalias são apresentadas na Seção 4.5, sendo retratado a diminuição no tamanho dos pacotes e a análise dos nomes utilizados pelos domínios.

No próximo capítulo são discutidas alguns aspectos de escalabilidade do DNSpot, notadamente limitações que foram percebidas no decorrer do estudo.

5 Aspectos de Escalabilidade do DNSpot

A coleta de dados efetuada neste trabalho, que foi de longa duração (250 dias), ajudou a revelar limitações de escalabilidade do DNSpot que não foram identificadas em (LONGO, 2015), que teve um período 80% menor de observação (49 dias). Este capítulo discute essas limitações, que estão associadas tanto à intensidade de tráfego (Seção 5.1) quanto ao banco de dados (Seção 5.2).

5.1 Intensidade de tráfego

Durante a coleta de dados, devido a problemas identificados com o tempo de inserção no banco de dados (que serão descritos na Seção 5.2.3), constatou-se que o DNSpot estava perdendo requisições, ou seja, não estava conseguindo processar todas as requisições recebidas. Para quantificar essa perda, foi utilizado o PF (*Packet Filter*), que é o *firewall* nativo do OpenBSD (OPENBSD, 2017), para contar os pacotes recebidos na porta 53/UDP; o contador é zerado quando a máquina é iniciada, e incrementado toda vez que um novo pacote é recebido pelo sistema na referida porta. O valor do contador foi amostrado uma vez por minuto, usando um *job* disparado via *crontab*.

A Figura 5.1 apresenta o número de requisições recebidas (contabilizadas pelo PF) e o número de requisições efetivamente processadas pelo DNSpot, no decorrer de 49 dias (de 05/04 até 25/05). O gráfico demonstra que o DNSpot consegue processar apenas uma fração das requisições DNS recebidas, e que o índice de perdas aumenta nos períodos em que ocorrem ataques de maior intensidade, quando a curva do PF apresenta uma inclinação mais acentuada. Isso fica melhor evidenciado nos períodos entre os dias 0 e 10 e entre os dias 40 e 47, aproximadamente.

A Tabela 5.1 compara a quantidade diária de requisições contabilizadas pelo PF e processadas pelo DNSpot. Em média, o DNSpot conseguiu processar 41,4% do tráfego DNS recebido. O tráfego máximo diário registrado pelo PF foi de quase 2 milhões de requisições, enquanto que o máximo processado pelo DNSpot foi inferior a 522 mil requisições. Durante o período de 49 dias foram recebidas 22.572.257 requisições, das quais foram processadas

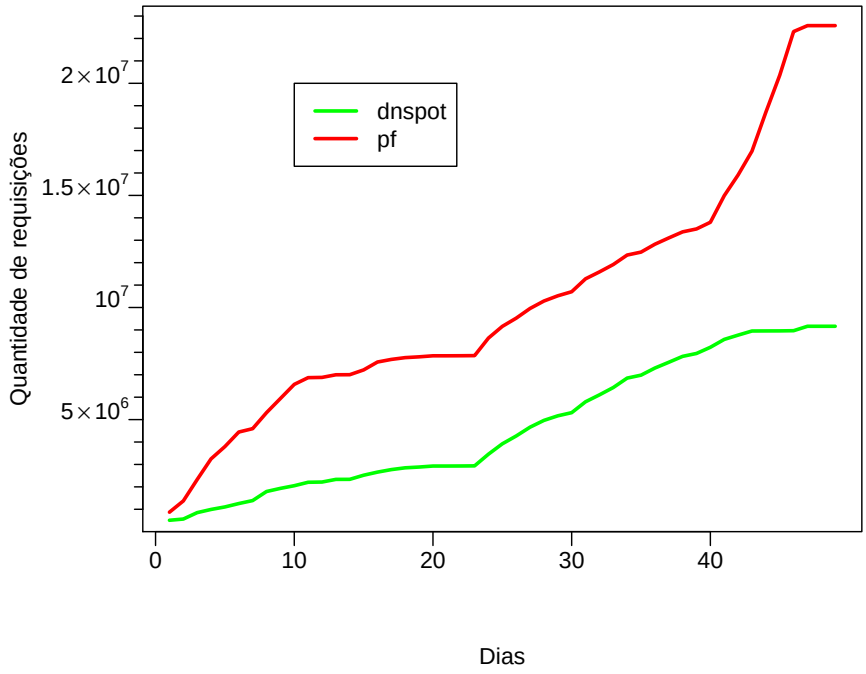


Figura 5.1: Número de requisições recebidas e processadas pelo DNSpot.

Fonte: O próprio autor

9.613.893, ou 42,6% do total. Isso corresponde a 12.958.364 requisições recebidas e não processadas, uma perda de 57,4%.

A análise dos dados coletados pelo PF permitiu também identificar alguns ataques notáveis que ocorreram no período de observação. A intensidade e duração de tais ataques, que são descritos a seguir, dão uma dimensão dos desafios de escalabilidade que afetam o DNSpot.

Ataque 1

O primeiro ataque (Figura 5.2) foi observado no dia 13/04, entre 04:38 e 11:55, e teve duração de 7 h 17 min. Ao todo foram recebidas 527.801 requisições, com uma taxa de chegada máxima de 88,35 requisições por segundo e média de 20,12 requisições por segundo; como o contador de pacotes do PF foi amostrado com intervalo de 1 min, as taxas de requisições por segundo são dadas pela diferença entre minutos consecutivos dividida por 60, correspon-

		Requisições por dia	
	Estatísticas	PF	DNSpot
1	Média	451.453,04	186.996,63
2	Desvio padrão	461.418,09	153.032,85
3	Mínimo	6,00	0,00
4	1 percentil	7,96	0,96
5	5 percentil	187,10	440,20
6	1 quartil	120.818,75	53.331,00
7	Mediana	319.150,00	141.940,00
8	3 quartil	631.181,75	302.732,00
9	95 percentil	1.439.653,70	476.246,40
10	99 percentil	1.845.973,35	516.396,80
11	Máximo	1.951.414,00	521.720,00

Tabela 5.1: Distribuição do número de pacotes por dia no DNSpot e pf.

dendo à média durante aquele minuto. Considerando o mesmo período, o DNSpot processou um total de 118.817 transações, ou seja 408.984 transações (77,5% do total) foram perdidas.

Ataque 2

O segundo ataque, mostrado na Figura 5.3, iniciou às 09:51 do dia 16/05 e terminou às 08:44 do dia 17/05, uma duração de 22 h 53 min. Chama a atenção que esse ataque (ou conjunto de ataques) foi ininterrupto, com um mínimo de 14,15 req/s. A intensidade máxima foi de 34,13 req/s, com média de 22,46 req/s, e um total de 1.851.343 requisições. Durante esse ataque, o DNSpot processou 542.290 requisições, ou seja, 1.309.053 requisições (70,7% do total) não foram registradas no banco de dados.

Ataque 3

A Figura 5.4 mostra um ataque com duração de 3 dias, 14 horas e 53 minutos (das 09:04 do dia 18/05 às 23:20 do dia 21/05) e um total de 6.275.430 requisições. Esse ataque também foi ininterrupto, com um mínimo de 10,23 req/s. A intensidade máxima foi de 39,03 req/s, e a média foi 20,52 req/s. Durante este mesmo período o DNSpot registrou um total de 196.596 requisições, correspondendo a uma perda de 6.078.834 requisições (96,9% do

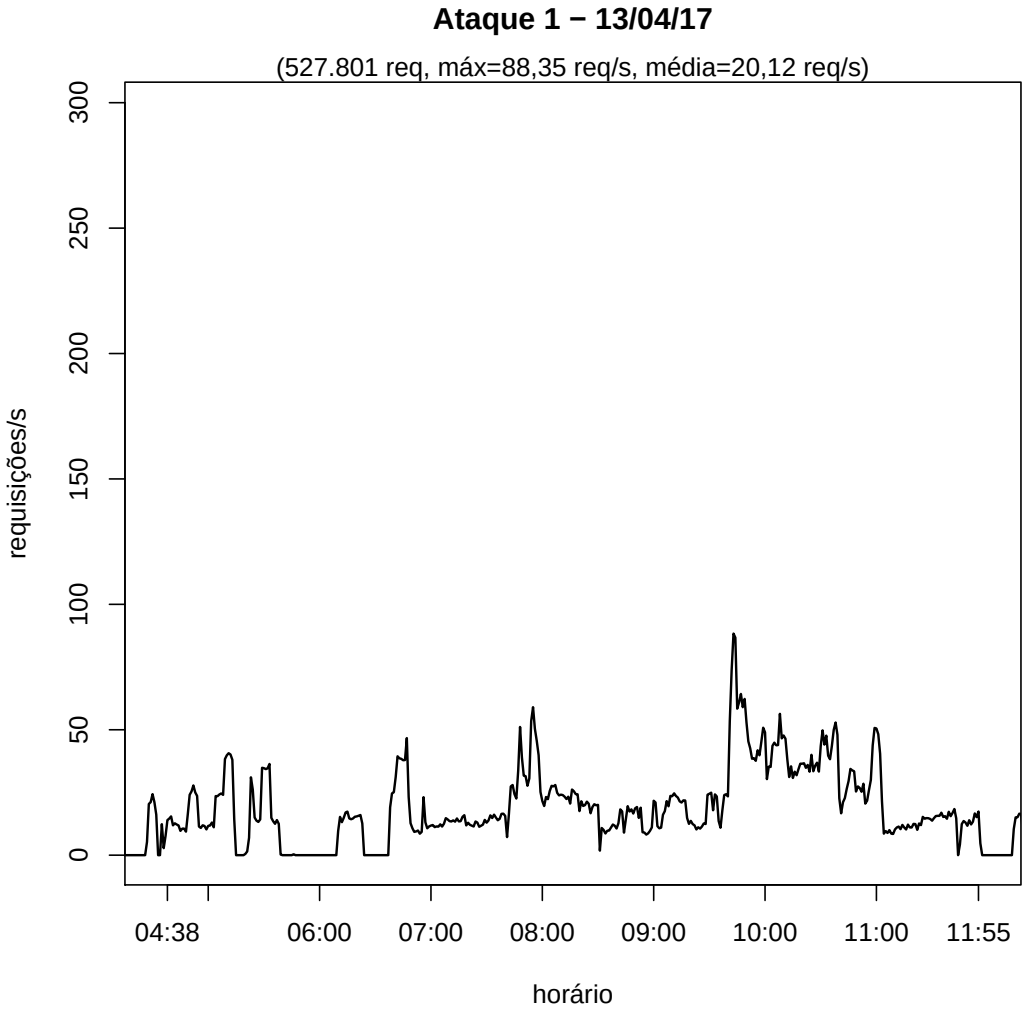


Figura 5.2: Número de requisições recebidas (Ataque 1).

Fonte: O próprio autor

total). Considerando os três dias do ataque, o DNSpot só registra um dia com mais de 8.500 requisições no banco de dados.

Ataque 4

A Figura 5.5 apresenta um ataque que ocorreu do dia 01/06 à 01:48) até o dia 06/06 às 07:14, o que corresponde a uma duração de 5 d 05 h 26 min. Esse período apresentou um total de 20.919.162 requisições, com uma intensidade máxima de 296,33 req/s e uma média de 46,32 req/s. Embora esse ataque recaia fora do período analisado neste capítulo, entendeu-se que seria importante relatá-lo para enfatizar a intensidade de tráfego que o DNSpot deve ser capaz de lidar.

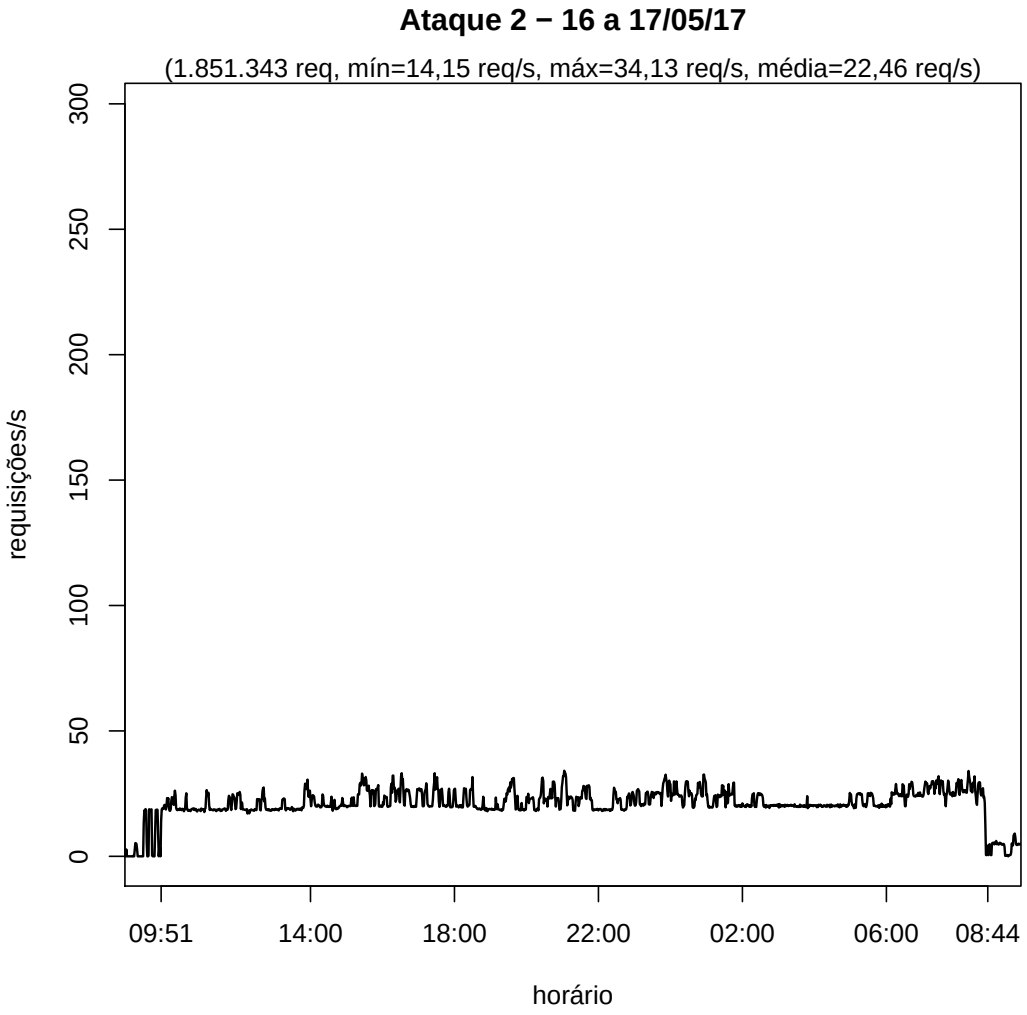


Figura 5.3: Número de requisições recebidas (Ataque 2).

Fonte: O próprio autor

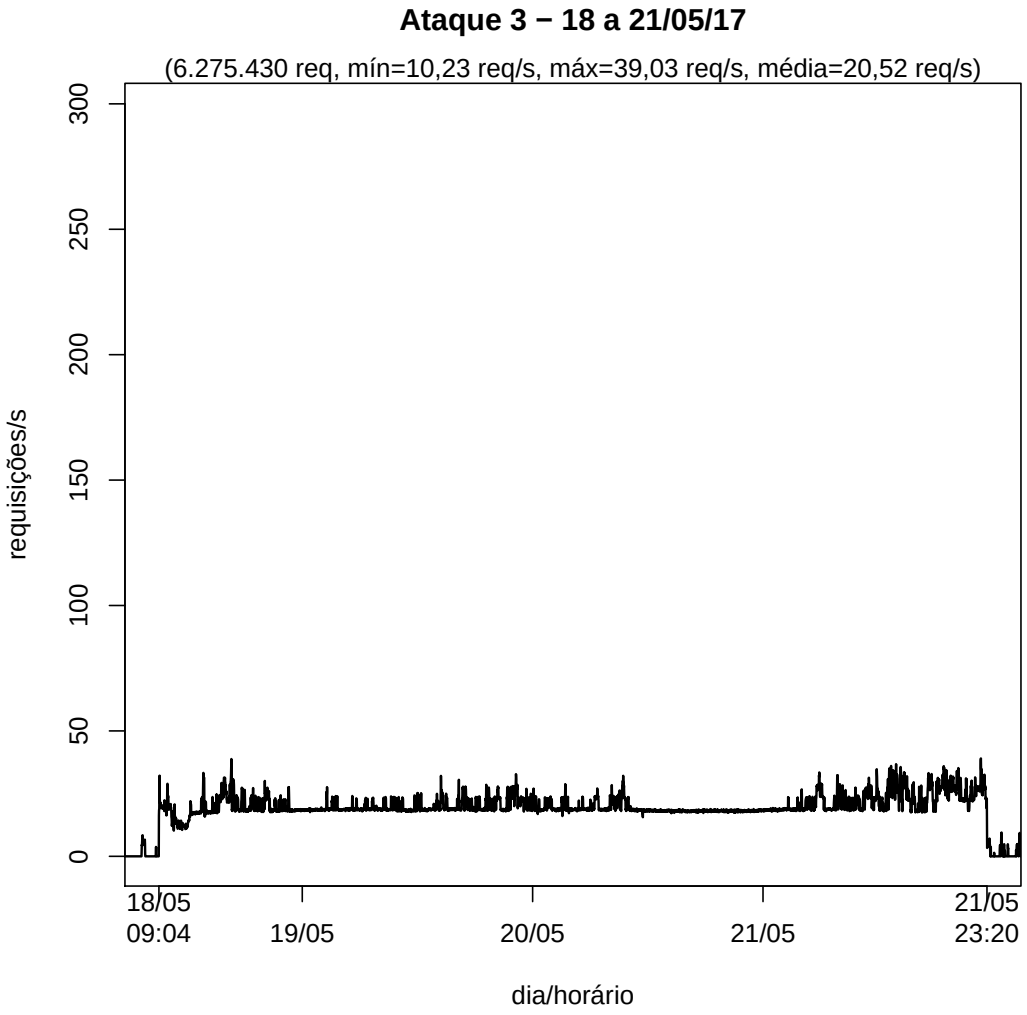


Figura 5.4: Número de requisições recebidas (Ataque 3).

Fonte: O próprio autor

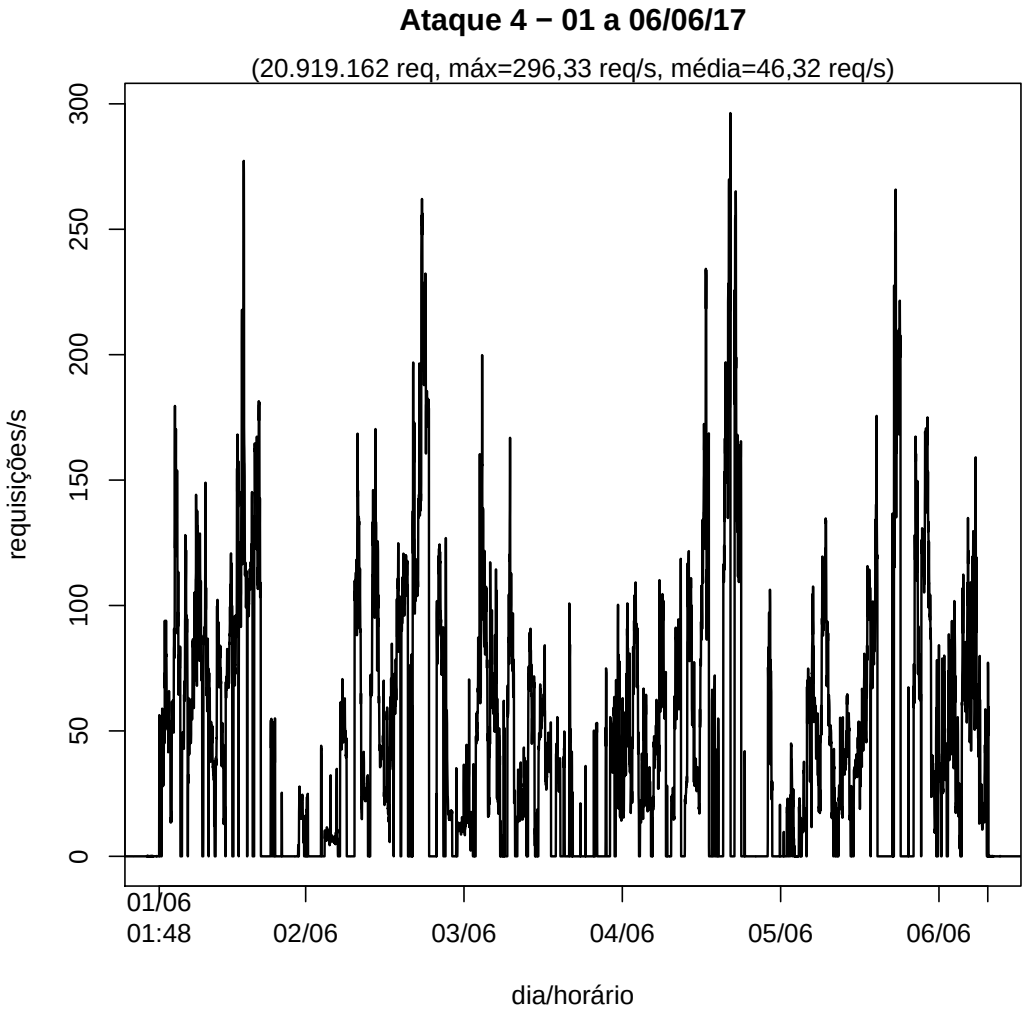


Figura 5.5: Número de requisições recebidas (Ataque 4).

Fonte: O próprio autor

5.2 Banco de Dados

Nesta Seção será apresentado a análise do banco de dados do DNSpot, apontando alguns fatores relevantes ao banco de dados da ferramenta que foram observados durante este trabalho.

5.2.1 Utilização de espaço pelas tabelas

Conforme mencionado na Seção 4.2.1 (página 39), o banco de dados contendo todos os dados analisados no trabalho atingiu um volume de 30 GB. Para entender como esse volume está distribuído entre as tabelas do banco, foi usado o *sqlite3_analyzer* (SQLITE, 2017d), que é uma ferramenta que descreve a utilização de espaço em disco de um banco de dados SQLite. Uma das informações fornecidas é a contagem de páginas para cada tabela no banco de dados. No SQLite, o armazenamento de um banco de dados é organizado internamente em páginas (SQLITE, 2017a), que no caso do DNSpot são de 4 KB.

A Tabela 5.2 apresenta a distribuição de páginas entre as tabelas do banco de dados. Observa-se que 81,6% do total é encontrado em somente sete tabelas do banco, sendo que DNS_SENT_RAWDATA e DNS_RECV_RAWDATA sozinhas respondem por 65,6% de todo o volume, ou aproximadamente 19,6 GB. Essas duas tabelas são responsáveis por armazenar as mensagens DNS recebidas pelo DNSpot em formato binário, de modo a permitir seu eventual reprocessamento (LONGO, 2015). Tendo em vista o espaço ocupado por essas tabelas, faz sentido reavaliar a pertinência de armazenar esse conteúdo, ao menos no banco de dados. Eliminá-las do banco reduziria o espaço ocupado e a carga de operações de escrita a cada requisição processada, o que pode melhorar o desempenho da aplicação.

5.2.2 Tamanho do arquivo

O tamanho do arquivo contendo o banco de dados é um fator que contribui para o desempenho das operações com o banco: quanto maior o arquivo, mais lentas tendem a ser as operações. Para uma análise em um período grande de tempo, como a realizada neste trabalho, é recomendável uma avaliação do banco de dados, evitando manter grandes volumes de dados históricos no banco usado para realizar inserções. A solução adotada neste trabalho foi mover periodicamente os dados históricos para um banco de dados separado, dando início a

	Tabela	Porcentagem
1	DNS_SENT_RAWDATA	49,3
2	DNS_DATABASE_ERROR	17,3
3	DNS_RECV_RAWDATA	16,3
4	DNS_TRANSACTION	5,9
5	DNS_RECV_HEADER	3,6
6	DNS_RECV_QUESTION	3,2
7	DNS_TRAN_PROBLEM	2,3
	Total	81,6

Tabela 5.2: Distribuição de páginas (4 KB) entre as principais tabelas do banco de dados.

uma nova versão do banco. Cabe destacar que, em relação a (LONGO, 2015), algumas tabelas que continham dados que foram considerados redundantes já não estavam sendo usadas; essa modificação foi proposta por Diatel (2016), no contexto de um TCC em desenvolvimento que visa introduzir melhorias no sistema de armazenamento do DNSpot.

O principal problema dessa abordagem de versionamento é o tempo gasto para criar uma tabela adequada para o DNSpot. A remoção das tabelas não é a única operação necessária. O SQLite não remove efetivamente dados do banco ao executar um DROP ou DELETE de uma tabela, ele apenas marca os dados como removidos (e, eventualmente, páginas como livres) (SQLITE, 2017a). Para realmente liberar espaço em disco, ainda é necessário executar o comando VACUUM, que desfragmenta o arquivo contendo o banco de dados (SQLITE, 2017c). Durante a execução do comando, o banco de dados fica indisponível para a aplicação, a qual consequentemente também permanece indisponível. O tempo para executar estas operações está ligado diretamente com o volume do banco de dados, que é grande em coletas de longa duração. Outra limitação do versionamento é que, atualmente, uma nova versão já inicia com aproximadamente 2,3 GB, volume esse que aumenta monotonicamente.

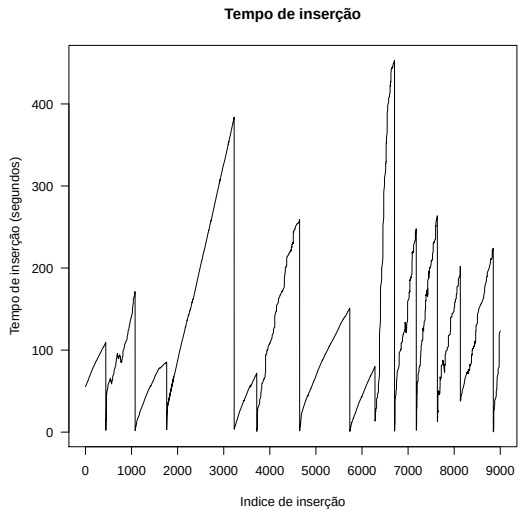
5.2.3 Tempo de inserção

Outro problema encontrado foi o tempo de inserção no banco de dados do DNSpot, que pode ser atribuído em parte ao tamanho do banco e em parte à intensidade de tráfego (Seção 5.1). A demora na inserção de dados é prejudicial para o DNSpot, já que, no momento em que um ataque DDoS está sendo realizado, o sistema apresenta um significativo número de

threads ativas tentando realizar inserções. Como inserções exigem *locks* exclusivos (SQLITE, 2017b), isso gera uma grande contenção no BD, pois as escritas devem ser serializadas, o que alonga o tempo de inserção. Uma vez que as *threads* demoram a finalizar, durante ataques severos o número de *threads* no sistema tende a atingir o limite (o número de *threads* varia de acordo com a intensidade do ataque, tenho chegado ao limite várias vezes durante o período de observação), o que inibe o processamento de novas requisições e acaba provocando o *lock* do banco de dados. Isso ocorre pois quando o DNSpot alcança o número máximo de 1.450 *threads* definido por Longo (2015), o sistema para de criar novas *threads* por um determinado tempo (60 s, especificamente), para permitir o escoamento das requisições pendentes e evitar estouros de memória. Quando as inserções se tornam excessivamente demorada, *threads* mais antigas podem sofrer *timeout* antes de conseguirem inserir seus dados, e com isso o banco de dados pode ficar bloqueado caso uma *thread* que estava realizando modificações não tenha liberado o *lock* antes do seu *timeout*. Esse problema foi observado em períodos em que o volume no banco de dados era maior que 5 GB.

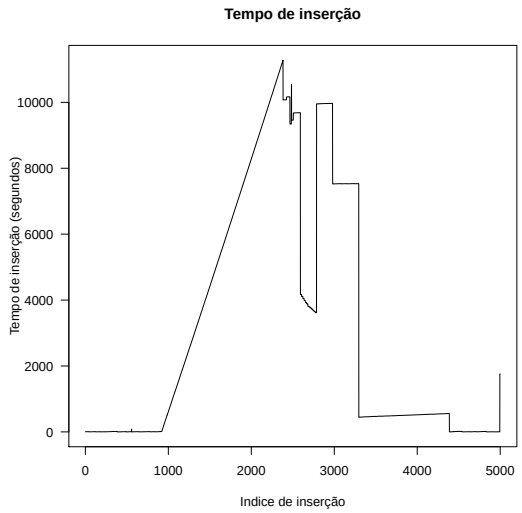
A Figura 5.6 ilustra o tempo de inserção no banco de dados em períodos distintos. A Figura 5.6(a) apresenta o tempo de inserção para 9.000 requisições, em um período em que ocorreram diversas chegadas de requisições (em rajadas). Observa-se que, durante um ataque, o tempo de inserção aumenta até que todas as *threads* pendentes tenham conseguido escrever no banco de dados, quando o tempo então se torna próximo a zero. No período mostrado, o tempo máximo de inserção chegou a 453,3 s (mais de 7,5 min), com uma média de 125,4 s (mais de 2 min). O comportamento apresentado é normal em momentos em que estão sendo realizados ataques. A Figura 5.6(b) apresenta o tempo de inserção para 5.000 requisições durante um ataque maciço. Esse ataque atinge um tempo de inserção máximo de 3 h 8 min, com média de 52,9 min, quando o desejável é que esse tempo de inserção fosse na ordem de segundos, no pior caso. Em períodos como o mostrado no gráfico, o DNSpot se torna não responsivo, deixando de processar novas requisições; a solução é reiniciar a aplicação, com a consequente perda dos dados associados às *threads* pendentes.

Cabe ressaltar que o tempo de inserção no BD não é considerado no cômputo da duração dos ataques DoS (Tabela 4.21, página 59), que se baseia em *timestamps* atribuídos antes dos dados serem gravados. Com isso, a análise da duração dos ataques não é prejudicada.



(a) Período com vários ataques

Fonte: O próprio autor



(b) Período com ataque maciço

Fonte: O próprio autor

Figura 5.6: Tempo de inserção no banco de dados.

5.3 Discussão

Os resultados discutidos acima e especialmente os ataques descritos conseguem demonstrar a limitação do DNSpot para tratar todas as transações que estão sendo recebidas pelo sistema. Esta quantidade de perdas pode ser atribuída ao tempo que o sistema acaba levando para inserir novas informações no banco de dados e ao rápido crescimento na quantidade de requisições que o sistema acaba recebendo durante um ataque.

É importante compreender de que maneira esse índice de perda de requisições afeta a análise dos resultados. Como nem todo o tráfego recebido é processado pelo DNSpot, os dados estatísticos apresentados neste trabalho são na verdade uma subestimativa do tráfego que atinge o DNSpot. As diferenças tendem a ser maiores em períodos de ataques intensos, quando o DNSpot recebe grandes quantidades de requisições em curtos espaços de tempo e não consegue dar vazão a todas essas requisições. Em outras palavras, o cenário de ataques envolvendo o DNSpot é ainda pior do que o apresentado aqui. Por outro lado, cabe lembrar que o DNSpot oferece um único ponto de observação sobre o tráfego DNS malicioso, e que, por isso mesmo, sua perspectiva é limitada por natureza. Assim, ainda que os números estejam subestimados, a representatividade dos dados não pode ser questionada além do que se poderia questionar caso o DNSpot tivesse a capacidade de processar todas as requisições recebidas.

Conforme apresentado na Seção 4.1 (página 37), o DNSpot executa em uma má-

quina com uma configuração de *hardware* que pode ser considerada modesta. Um incremento na capacidade do *hardware*, especialmente da memória e disco, provavelmente melhoraria o desempenho da ferramenta, mas dificilmente seria capaz de resolver completamente as limitações apontadas neste capítulo. Está em andamento um TCC em que alternativas para melhorar o desempenho do subsistema de armazenamento estão sendo investigadas (DIATEL, 2016). Uma outra possibilidade seria rearquitetar o DNSpot de modo que ele fosse capaz de lidar com intensidades de tráfego iguais ou superiores às mostradas na Seção 5.1; essa possiblidade seria mais complexa do que substituir o subsistema de armazenamento, mas teria o potencial de oferecer ganhos mais significativos de escalabilidade.

5.4 Considerações do Capítulo

A utilização do DNSpot durante um período ampliado de monitoramento de 250 dias permitiu apontar limitações de escalabilidade que não tinham sido identificadas no trabalho inicial, no qual a ferramenta foi usada durante apenas 49 dias (LONGO, 2015). Essas limitações estão associadas à interação entre a intensidade de tráfego DNS, especialmente durante ataques DoS, e o banco de dados usado para o registro de informações sobre as requisições processadas.

Por um lado, é possível afirmar que as limitações de escalabilidade prejudicaram a análise dos dados, sem no entanto comprometê-la. Por outro lado, o conhecimento mais aprofundado sobre os limites do DNSpot abre novos rumos de investigação no sentido de ampliar esses limites.

6 Considerações

O DNS é um elemento vital na infraestrutura da Internet. Em vista disso, a sua segurança é crítica para o bom funcionamento da rede. Essa segurança depende não apenas de mecanismos de prevenção contra ameaças e ataques, mas também de um trabalho de monitoração e análise que permita conhecer melhor essas ameaças e ataques, e acompanhar sua evolução. Esse conhecimento é importante justamente para desenvolver novos mecanismos de proteção e refinar os mecanismos existentes.

O DNSpot é um *honeypot* específico para DNS que tem por objetivo propiciar essa monitoração sob o prisma de servidores DNS recursivos. A análise de dados conduzida na validação inicial da ferramenta revelou a rapidez com que servidores recursivos abertos são recrutados como refletores em ataques de negação de serviço distribuída, e algumas características do DNS que são exploradas para potencializar esses ataques.

Este trabalho apresentou uma análise de dados coletados do DNSpot durante um período de 250 dias, no qual foram processadas quase 32,3 milhões de requisições. Em relação ao estudo original (LONGO, 2015), foram percebidas algumas mudanças relevantes no tráfego, que incluem:

- aumento significativo na intensidade do tráfego direcionado ao DNSpot;
- aumento na duração e no volume de tráfego em ataques DDoS que usam o DNSpot como refletor;
- evidências de uso do DNSpot como servidor DNS recursivo regular por parte de usuários finais ou em nome destes.

No decorrer do estudo foi possível observar algumas limitações de escalabilidade do DNSpot. Essas limitações dificultaram a operação da ferramenta, prejudicando a coleta e análise de dados, sem contudo comprometer a validade dos resultados.

Para a continuidade deste trabalho, algumas perspectivas podem ser apontadas:

- Manter a coleta de dados do DNSpot por períodos ainda maiores, conjugando análises de longo prazo, que a evolução do tráfego, com análises de curto prazo, que permitem

capturar e explicar dados efêmeros, como os discutidos na Seção 4.5.2 para o domínio `test003.com`;

- Desenvolver melhorias para o DNSpot, incrementando sua escalabilidade e ampliando a observabilidade de parâmetros relevantes para o seu desempenho, como o número de *threads* ativas no sistema e o tempo de inserção de requisições;
- Investigar como múltiplas instâncias do DNSpot, em diferentes pontos da Internet, podem ser usadas em um sistema distribuído de monitoramento e detecção de ataques ao DNS.

Referências Bibliográficas

AEGINS. *aegins.com*. jun 2017. https://www.aegins.com/.

ALBITZ, C.; LIU, P. *DNS and BIND (5th Edition)*. : O'Reilly Media, Inc., 2006. ISBN 0596100574.

ANDREWS, M. *Negative Caching of DNS Queries (DNS NCACHE)*. mar 1998. RFC 2308. https://tools.ietf.org/html/rfc2308.

ATKINS, D.; AUSTEIN, R. *Threat Analysis of the Domain Name System (DNS)*. agosto de 2004. RFC 3833. Disponível em: <http://tools.ietf.org/html/rfc3833>.

ATLASIS, A. An attack-in-depth analysis of multicast DNS and DNS service discovery. In: *Hack in the Box*. Amsterdam: , 2017. https://www.secfu.net/app/download/10962123597/AAtlasis+-+An+Attack-in-Depth+Analysis+of++multicast+DNS+and+DNS+Service+Discovery_wp.pdf?t=1493477029.

BARBOSA, K. R.; PEREIRA, E. S. J. Análise passiva do tráfego DNS da Internet brasileira. In: *IX Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSEG 2009)*. 2009. p. 203–216.

BORTZMEYER, S. *DNS Privacy Considerations*. agosto de 2015. RFC 7626. Disponível em: <http://tools.ietf.org/html/rfc7626>.

BROWNLEE, N.; CLAFFY, k.; NEMETH, E. DNS Measurements at a Root Server. In: *IEEE Global Telecommunications Conference*. San Antonio, TX: IEEE Global Telecommunications Conference (GLOBECOM), 2001.

CASTRO, S. et al. A day at the root of the Internet. *SIGCOMM Comput. Commun. Rev.*, ACM, New York, NY, USA, v. 38, n. 5, p. 41–46, set. 2008. ISSN 0146-4833. Disponível em: <http://doi.acm.org/10.1145/1452335.1452341>.

CASTRO, S. et al. Understanding and preparing for DNS evolution. In: *Traffic Monitoring and Analysis Workshop (TMA)*. Zurich, Switzerland: TMA 2010, 2010. p. 1–6.

CERT.BR. *Servidores DNS recursivos abertos*. dec 2014. Disponível em http://www.cert.br/docs/whitepapers/dns-recursivo-aberto/. Acessado em 16/11/2016.

CERT.BR. *Recomendações para Melhorar o Cenário de Ataques Distribuídos de Negação de Serviço (DDoS)*. apr 2016. Disponível em http://www.cert.br/docs/whitepapers/ddos/.

CHESHIRE, S.; KROCHMAL, M. *DNS-Based Service Discovery*. fev. 2013. RFC 6763. https://tools.ietf.org/html/rfc6763.

CONRAD, D. *Towards Improving DNS Security, Stability, and Resiliency*. 2012. Disponível em http://tinyurl.com/h5z49y5.

DANZIG, P. B.; OBRACZKA, K.; KUMAR, A. An analysis of wide-area name server traffic: a study of the Internet Domain Name System. *ACM SIGCOMM Computer Communication Review*, ACM, v. 22, n. 4, p. 281–292, 1992.

DIATEL, M. *Análise comparativa de ferramentas de armazenamento para honeypots DNS*. Dissertação (Trabalho de conclusão de curso) — Bacharelado em Ciência da Computação, Universidade do Estado de Santa Catarina, Joinville, dezembro de 2016.

GAO, H. et al. An empirical reexamination of global DNS behavior. *SIGCOMM Comput. Commun. Rev.*, ACM, New York, NY, USA, v. 43, n. 4, p. 267–278, ago. 2013. ISSN 0146-4833. Disponível em: <http://doi.acm.org/10.1145/2534169.2486018>.

GROUP, Z. W. *Zero Configuration Networking (Zeroconf)*. jun 2017. http://www.zeroconf.org/.

HOEPERS, C.; STEDING-JESSEN, K.; CHAVES, M. *Honeypots e Honeynets: Definições e Aplicações*. Oct 2007. Disponível em: <http://www.cert.br/docs/whitepapers/honeypots-honeynets/>.

IANA. *Root Servers*. nov 2016. Iana. https://www.iana.org/domains/root/servers;.

ICANNWIKI. *Domain Monetization*. jun 2017. https://icannwiki.org/Domain_Monetization.

JUNG, J. et al. DNS performance and the effectiveness of caching. *IEEE/ACM Trans. Netw.*, IEEE Press, Piscataway, NJ, USA, v. 10, n. 5, p. 589–603, out. 2002. ISSN 1063-6692. Disponível em: <http://dx.doi.org/10.1109/TNET.2002.803905>.

KARRENBERG, D. *Root Servers*. nov 2016. http://www.root-servers.org/; The 13 root name servers.

KHALIMONENKO, A. *DDOS attacks in Q1 2017*. may 2017. Disponível em: <https://securelist.com/analysis/quarterly-malware-reports/78285/ddos-attacks-in-q1-2017/>.

LONGO, F. S. *Honeypot para Servidores DNS Recursivos: Adaptação, Coleta e Análise de Resultados*. Monografia (Trabalho de conclusão de curso) — Bacharelado em Ciência da Computação, Universidade do Estado de Santa Catarina, Joinville, dezembro de 2015.

MAXMIND. *geolocation of IP addresses*. may 2017. Disponível em: <http://freegeoip.net/?q=187.107.120.249>.

MCAFEE. *McAfee products using Global Threat Intelligence*. may 2017. Disponível em: <http://tinyurl.com/yamqh2oq>.

MEDEIROS, H. de. *Análise de Desempenho do Uso de TCP para Consultas DNS*. Monografia (Trabalho de conclusão de curso) — Bacharelado em Ciência da Computação, Universidade do Estado de Santa Catarina, Joinville, junho de 2011.

MOCKAPETRIS, P. *Domain Names – Concepts and Facilities*. nov. 1987. RFC 1034. Disponível em: <http://tools.ietf.org/html/rfc1035>.

MOCKAPETRIS, P. *Domain Names – Implementation and Specification*. nov 1987. RFC 1035. Disponível em: <https://tools.ietf.org/html/rfc1035>.

MOCKAPETRIS, P.; DUNLAP, K. J. Development of the Domain Name System. *SIGCOMM Comput. Commun. Rev.*, ACM, New York, NY, USA, v. 18, n. 4, p. 123–133, ago. 1988. ISSN 0146-4833. Disponível em: <http://doi.acm.org/10.1145/52325.52338>.

NAUGLE, M. G. *Network Protocol Handbook*. 1st ed. New York, NY, USA: McGraw-Hill, Inc., 1998. ISBN 0070466033.

OPENBSD. *OpenBSD PF: User's Guide*. jun. 2017. Disponível em: <http://www.openbsd.org/faq/pf/index.html>.

PCWORLD. *Malware increasingly uses DNS as command and control channel to avoid detection, experts say*. jun 2017. https://www.pcworld.idg.com.au/article/417011/malware_increasingly_uses_dns_command_control_channel_avoid_detection_experts_say/.

PERDISCI, R. et al. Detecting malicious flux service networks through passive analysis of recursive DNS traces. In: *Twenty-Fifth Annual Computer Security Applications Conference, ACSAC 2009, Honolulu, Hawaii, 7-11 December 2009*. 2009. p. 311–320. Disponível em: <http://dx.doi.org/10.1109/ACSAC.2009.36>.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2008. ISBN 3-900051-07-0. Disponível em: <http://www.R-project.org>.

REKHTER, Y. et al. *Address Allocation for Private Internets*. fev. 1996. RFC 1918. https://tools.ietf.org/html/rfc1918.

RIEDEL, W. *Domain Name System (DNS) Parameters*. dec 2016. http://www.iana.org/assignments/dns-parameters;.

RUHR UNIVERSITY BOCHUM, G. *Amplification DDoS Tracker Project*. jun 2017. http://scanresearch1.syssec.ruhr-uni-bochum.de/.

SALUSKY, W.; DANFORD, R. *Know Your Enemy: Fast-Flux Service Networks*. jul. 2007. Disponível em http://www.honeynet.org/papers/ff.

SCHNEIER, B. *Lessons From the Dyn DDoS Attack*. novembro de 2016. Security Intelligence. Disponível em: <http://tinyurl.com/jjet78r>.

SPITZNER, L. Honeypots: Catching the insider threat. In: *Proceedings of the 19th Annual Computer Security Applications Conference*. Washington, DC, USA: IEEE Computer Society, 2003. (ACSAC '03), p. 170–. ISBN 0-7695-2041-3. Disponível em: <http://dl.acm.org/citation.cfm?id=956415.956438>.

SQLITE. *Database File Format*. jun 2017. Disponível em: <https://sqlite.org/fileformat.html>.

SQLITE. *File Locking And Concurrency In SQLite Version 3*. jun. 2017. Disponível em: <https://sqlite.org/lockingv3.html>.

SQLITE. *SQLite Query Language: VACUUM*. jun. 2017. Disponível em: <https://sqlite.org/lang_vacuum.html>.

SQLITE. *The sqlite3_analyzer.exe Utility Program*. jun 2017. Disponível em: <https://sqlite.org/sqlanalyze.html>.

STEDING-JESSEN, K.; VIJAYKUMAR, N. L.; MONTES FILHO, A. Using low-interaction honeypots to study the abuse of open proxies to send Spam. *InfoComp*, v. 7, n. 1, p. 44–52, 2008. ISSN 1807-4545. Acesso em: 19 ago. 2016.

TENTLER, D. *phobos*. may 2017. Disponível em: <https://phobos.io/>.

THIELMAN, S.; JOHNSTON, C. Major cyber attack disrupts Internet service across Europe and US. *The Guardian*, 2016. Publicado em 21/10/2016. Disponível em: <http://tinyurl.com/zvdbrym>.

THOMAS, M.; WESSELS, D. An analysis of TCP traffic in root server DITL data. In: *DNS-OARC 2014 Fall Workshop*. 2014. Disponível em: <http://tinyurl.com/zzw4kud>.

VAKALI, A.; PALLIS, G. Content delivery networks: status and trends. *IEEE Internet Computing*, v. 7, n. 6, p. 68–74, Nov 2003. ISSN 1089-7801.

WHOIS. *WHOIS – test003.com*. jun 2017. https://www.whois.com/whois/test003.com.

WHOIS. *Whois.com – Domain Names & Identity for Everyone*. may 2017. Disponível em: <https://www.whois.com>.

ZDRNJA, B.; BROWNLEE, N.; WESSELS, D. Passive monitoring of DNS anomalies. In: *Proceedings of the 4th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Berlin, Heidelberg: Springer-Verlag, 2007. (DIMVA '07), p. 129–139. ISBN 978-3-540-73613-4.

ZHAO, G. et al. Detecting APT malware infections based on malicious DNS and traffic analysis. *IEEE Access*, v. 3, p. 1132–1142, 2015.

A Apêndice: Cronograma

No início deste período de estudo, que durou um total de três meses, foi realizado a formulação de um plano para o desenvolvimento deste trabalho, algumas das características abordadas neste plano podem ser observadas como:

(1)Formulação do plano do TCC, especificação de algumas características para o início do trabalho;

(2)Revisão sobre DNS, *honeypots* e DNSpot, para um melhor entendimento das principais características e funcionalidade da ferramenta que está sendo utilizada para a captura de informação;

(3)Revisão de trabalhos correlatos, buscando identificar questões importantes para a análise dos dados do DNSpot;

(4)Adaptação no DNSpot para coletas de longo prazo: algumas modificações foram realizadas para que o DNSpot consiga lidar com o período de coleta, removendo algumas características que não eram essenciais para este estudo;

(5)Coleta de dados: o início da coleta de dados foi no dia 17/09/2016;

(6)Definição das análises a serem realizadas, com um melhor entendimento e caracterização de alguns trabalhos na área, foi possível definir algumas características para a análise a ser realizada;

(7)Escrita da monografia da primeira parte (TCC-I).

A.1 Atividades

(8)Análise dos resultados obtidos, com uma análise a longo prazo será possível observar algumas características e comportamentos não vistos em uma análise realizada em um período pequeno (um mes ou até mesmo dias);

(9)Escrita da monografia da segunda parte (TCC-II).

A.2 Cronograma

O cronograma proposto para o TCC pode ser observado na Tabela A.1.

Etapas	2016						2017					
	J	A	S	O	N	D	J	F	M	A	M	J
1	X	X										
2		X	X									
3			X	X								
4			X									
5				X	X	X	X	X	X	X	X	X
6			X	X	X							
7				X	X							
8								X	X	X	X	X
9										X	X	X

Tabela A.1: Cronograma