Análise Passiva do Tráfego DNS da Internet Brasileira

Kaio Rafael de Souza Barbosa[1], **Eduardo Souto James Pereira**[1]

[1]Departamento de Ciência da Computação
Universidade Federal do Amazonas - UFAM
Av. Gal. Rodrigo Octávio Jordão Ramos, 3000
CEP 69.077-000 – Manaus – AM – Brasil

{kaiorafael, esouto}@dcc.ufam.edu.br

Abstract. *This paper presents the passive monitoring of Brazilian Internet traffic, where it was possible to observe the anomalies that consume computational resources that should answer exclusively valid queries. In an analysis has been found that approximately 43% of most seen resource records are the PTR queries type while others related works indicates that the record of type A is the most common. The observed behavior of Brazilian Internet traffic indicates malicious activities such as network attacks recognition, unauthorized transmission of messages (spam) and domain zone configuration errors.*

Resumo. *Este artigo apresenta o monitoramento passivo do tráfego da Internet Brasileira, onde foi possível observar anomalias que consomem os recursos computacionais que deveriam atender exclusivamente consultas válidas. Em uma análise foi constatado que aproximadamente 43% dos registros de recursos mais vistos são consultas do tipo PTR enquanto outros trabalhos relacionados ao tema indicam que o registro do tipo A é o mais frequente. O comportamento observado no tráfego da Internet Brasileira aponta atividades maliciosas como ataques de reconhecimento de rede, envio de mensagens não autorizadas (spams) e erros de configuração de zona de domínio.*

1. Introdução

O protocolo do DNS (*Domain Name System*) [Mockapetris 1987a] possui um papel crucial para o funcionamento da Internet, a tradução de nomes de máquinas em endereços IP. Tal mecanismo de tradução permite que as aplicações possam obter informações sobre os mais variados tipos serviços disponíveis na rede como, por exemplo, a localização de servidores de correio eletrônico, servidores web, aplicações de comércio eletrônico (*e-commerce*), ou mesmo outros servidores de nomes.

Atualmente, a maioria dos serviços da Internet é baseada em um modelo de funcionamento em que alguma consulta ao sistema DNS é realizada antes de qualquer atividade de comunicação. Através da análise do tráfego DNS é esperado encontrar comportamentos típicos, como solicitações que utilizam a classe padrão de consultas e alguns registros de recursos bem definidos. Por outro lado, por meio da análise da mensagem do protocolo, também é possível identificar tráfego não desejado ou anômalo, ou seja, tráfego que não deveria estar presente nas consultas ou respostas do DNS. Este tipo de tráfego é caracterizado por meio de perguntas com nomes de domínios de primeiro nível (*top-level domains* - TLDs) inválidos, solicitações geradas a partir de servidores DNS

com problemas de configuração de zona de domínio, consultas para o endereçamento privado de rede [Rekhter et al. 1996] e requisições DNS geradas a partir de anomalias de rede como vírus, cavalos de tróia (*trojans*), vermes (*worms*) e spams.

Diversas pesquisas relacionadas à detecção de consultas (ou respostas) DNS que não deveriam ocorrer na Internet têm sido propostas. Em 1992, Danzig et al. [Danzig et al. 1992] observaram certos problemas no tráfego DNS como laços recursivos, consultas repetidas desnecessárias, falhas de configuração e algoritmos de detecção de falhas ineficientes. Muitos desses problemas permanecem até hoje. Através da análise de servidores raiz, Wessels [Wessels 2004] mostra que consultas aos endereços privados (RFC 1918) representam 2.3% do total de tráfego analisado. De acordo com Brownlee et al. [Brownlee et al. 2001], tais tipos de consultas além de não terem utilidade para o tráfego de Internet, consomem grandes recursos computacionais da infra-estrutura de rede.

Este artigo descreve uma análise passiva do tráfego DNS dos servidores de nomes autoritativos que respondem pela zona de domínio *.br*, geridos pelo Registro.br [Registro.br]. A principal motivação para empregar o monitoramento passivo do sistema DNS foi detectar e correlacionar domínios usados em atividades anômalas, independente de terem sido causadas intencionalmente ou não. Neste estudo foi utilizado o tráfego coletado pelo projeto *Day in the Life of the Internet* [DITL 2008] ocorrido durante dois dias do mês de março de 2008, totalizando em média 5.4 bilhões de consultas DNS. Os resultados alcançados sobre o tráfego DNS da zona de domínio *.br* revelam que a base de dados analisada é uma fonte rica de informações de comportamentos anômalos como consultas a TLDs inválidos, que indicam a presença de nomes de domínios como se fossem extensões de arquivos *.gif*, *.css*, *.png* e *.js*, e um grande número de consultas do tipo de registro reverso (PTR), muitas das quais derivadas de atividades de combate ao spam e operações maliciosas a partir de ataques de rede.

O restante do artigo está organizado da seguinte maneira: a Seção 2 discute alguns trabalhos relacionados ao monitoramento e detecção de anomalias em trafego DNS; a Seção 3 descreve a base de dados utilizada para a realização deste trabalho; a seção 4 apresenta uma categorização do tráfego analisado a partir da distribuição de consultas por tipo de registro de recurso, e ainda, apresenta uma análise completa dos resultados dos experimentos e finalmente a Seção 5 apresenta as conclusões e possíveis trabalhos futuros.

2. Trabalhos Relacionados

Este trabalho está relacionado a diversas áreas de estudo sobre o DNS como monitoramento dos recursos de infra-estrutura e detecção de anomalias de rede, algumas das quais são brevemente apresentadas aqui.

Papas et al. [Pappas et al. 2004], apresentam uma análise do tráfego demonstrando que 15% das zonas de domínios encontradas na rede Ásia-Pacífico (APNIC - Asia-*Asia-Pacific Network Information Centre*) são compostas por delações incorretas *lame delegation* [Liu e Albitz 2006], ou seja, zonas cujo registro de servidor de nomes indica um *host* inválido ou não autoritativo pelo domínio. Esse tipo de erro resulta em um grande consumo dos recursos computacionais da infra-estrutura de rede, pois para encontrar o servidor que responde por esse domínio, é necessário consultar o servidor raiz.

Por outro lado, os autores não consideram outros problemas relacionados à configuração da zona de domínio, como consultas direcionadas ao espaço de endereço de rede privado [Rekhter et al. 1996] que além da utilização indevida dos recursos de rede, também podem relevar informações sigilosas das redes que originam esse tipo de tráfego.

Outras abordagens procuram supervisionar o tráfego e a utilização de recursos computacionais dos servidores raiz a fim de identificar e mitigar tráfego não desejado ou anômalo no protocolo DNS, ou seja, tráfego que não deveria estar presente nas consultas ou respostas do sistema de nomes. Wessels e Famenkov [Wessels e Fomenkov 2003] revelam em seus experimentos que consultas do tipo PTR representam 20% de todo o tráfego analisado, das quais 7% foram originadas a partir de um único cliente de Internet banda larga. Entretanto, os autores não descrevem quais são os fatores que contribuíram para o crescimento do registro reverso.

Recentemente, Castro et al. [Castro et al. 2008] apresentaram uma análise da base de dados do projeto DITL (*Day in the Life of the Internet*) [DITL 2008] que visa, através de ações coletivas, monitorar o tráfego DNS de grandes servidores de nomes distribuídos ao redor do mundo. Os resultados da análise mostram que acima de 90% total do tráfego processado pelos servidores raiz nos meses de janeiro de 2006, janeiro de 2007 e março de 2008, são consultas DNS que não deveriam ocorrer na Internet. Os autores observam ainda que consultas do tipo A são os registros mais vistos na rede, representando 60% do total do tráfego. Este comportamento também é constatado por Bojan et al. [Bojan et al. 2007] que atribuem a grande parcela do registro do tipo A no tráfego analisado à ferramentas de combate a mensagens não solicitadas (spam). A detecção de spams depende do DNS para recuperar informações de várias listas negras (RBL - *Realtime blacklists*)[1] em tempo real.

Briodo et al. [Broido et al. 2006] observam as requisições DNS utilizando o espaço de endereçamento privado de rede [Rekhter et al. 1996] em servidores raiz como um fenômeno global ao invés de um problema regional. Por este motivo, os autores argumentam que políticas direcionadas para esse tráfego poderiam ajudar a mitigar esse tipo de problema que não oferece informações relevantes para os servidores raiz.

Em resumo, a poluição do tráfego DNS vem crescendo a cada ano, obrigando os provedores que mantém o funcionamento e o gerenciamento dos recursos computacionais da Internet a desenvolverem novas soluções para atender o grande número de consultas válidas e inválidas, a fim de não interromper o funcionamento das atividades que dependem da operação correta da Internet. Por este motivo, este trabalho investiga a origem de alguns problemas relacionados poluição do tráfego DNS, conforme descrito nas seções seguintes.

3. Metodologia

A Internet brasileira possui seis servidores de nomes autoritativos que respondem pela zona de domínio *.br*, nomeados pelas seis primeiras letras do alfabeto. Tais servidores são mantidos e geridos pelo Registro.br [Registro.br]. O processo de medição deste artigo consiste na análise passiva do tráfego DNS brasileiro coletado durante o projeto DITL

[1]RBLs - Listas atualizadas constantemente que contêm os endereços IPs dos hosts acusados de estarem enviando spams

[DITL 2008], realizado em março de 2008, onde cinco instâncias dos servidores de nomes autoritativos (a-e.dns.br) participaram da coleta do tráfego DNS.

O projeto DITL é uma ação colaborativa entre grandes servidores de nomes distribuídos ao redor do mundo. Cada servidor que participa do evento, coleta durante os dias determinados, de modo passivo, o tráfego DNS, ou seja, o tráfego UDP na porta 53. O tráfego de rede coletado do DILT 2008, durante os dias do evento representa em média 5.4 bilhões de consultas nos servidores brasileiros, correspondendo a um volume de dados de aproximadamente 230 GB. A tabela 1 apresenta um resumo da base de dados utilizada neste trabalho.

DITL 2008		
Dia do Monitoramento	18 de Março de 2008	19 de Março de 2008
Instancias envolvidas	a-e.dns.br	a-e.dns.br
Total de horas	24h	24h
Hora de início	00:00 UTC (+0000)	00:00 UTC (+0000)
Hora de término	23:59:59.999 UTC (+0000)	23:59:59.999 UTC (+0000)
Quantidade de pacotes	2.7 Bilhões	2.6 Bilhões

Tabela 1. Informações sobre base de dados do projeto DITL 2008

Por razões de segurança e privacidade, esses dados são armazenados e mantidos pelo OARC (*Operations, Analysis, and Research Center*) [DNS-OARC].

A arquitetura oferecida pelo OARC dispõe de um sistema de arquivo que mantém o tráfego DNS coletado pelos servidores de cada país em diretórios. Cada diretório é composto por um conjunto de arquivos referente ao tráfego coletado no intervalo de uma hora. Por exemplo, o arquivo *20080319050000.pcap.gz* localizado no diretório da instância *a.dns.br* representa a coleta do tráfego referente ao dia 19 de março de 2008 que tem início às 05h e se estende até às 05h e 59min.

Na análise dos dados, a ferramenta *dnstop* [Wessels 2008] foi utilizada e modificada para visualizar as várias informações do tráfego DNS capturado como tipos de consultas, endereços IP, respostas a TLDs inválidos, etc. Além disso, uma ferramenta foi desenvolvida, em linguagem *Perl*, para obter outras informações como nome de consultas que iniciam com caractere "_"(*underline*).

Nesse artigo, somente consultas e respostas DNS relacionadas ao esquema de endereçamento IP versão 4 (IPv4) serão analisadas. A análise dos dados apresentada a seguir corresponde à soma de todos os registros observados durante os dois dias do evento do projeto DITL 2008.

4. Análise dos Dados

4.1. Distribuição de consultas por tipo de registro.

A figura 1 apresenta a distribuição de consultas por tipo de registro de recurso recebidas pelos servidores autoritativos do domínio *.br*. As consultas do tipo PTR, utilizadas para informar o nome de um domínio a partir de um endereço IP, são os principais tipos de recursos presente no tráfego, correspondendo a 42,91% do total de consultas. Por outro lado, a fração de consultas do tipo A, que corresponde aos registros que mapeiam nomes

de máquinas para endereços IP dos hospedeiros, representa 30,29% do total de tráfego analisado, sendo que sua fração permanece relativamente estável durante dois dias do evento. O terceiro tipo de registro mais observado é o MX com 15,65% do total de registros. Este tipo de registro de recurso indica uma lista de servidores que devem receber e-mails para esse domínio. Os registros de recursos restantes são consultas do tipo AAAA e A6 que indicam consultas DNS buscando domínios utilizando endereços IP na versão 6. Estes recursos representam em média 6% e 2% do total, respectivamente. O tipo SRV, empregado para consultar serviços ou protocolo da rede, apesar da baixa taxa de participação no tráfego, correspondendo em média a 0,03% do total, também contribui para o total de consultas inválidas. Seção 4.2.2 descreve poluição de tráfego DNS a partir de consultas do tipo SRV.

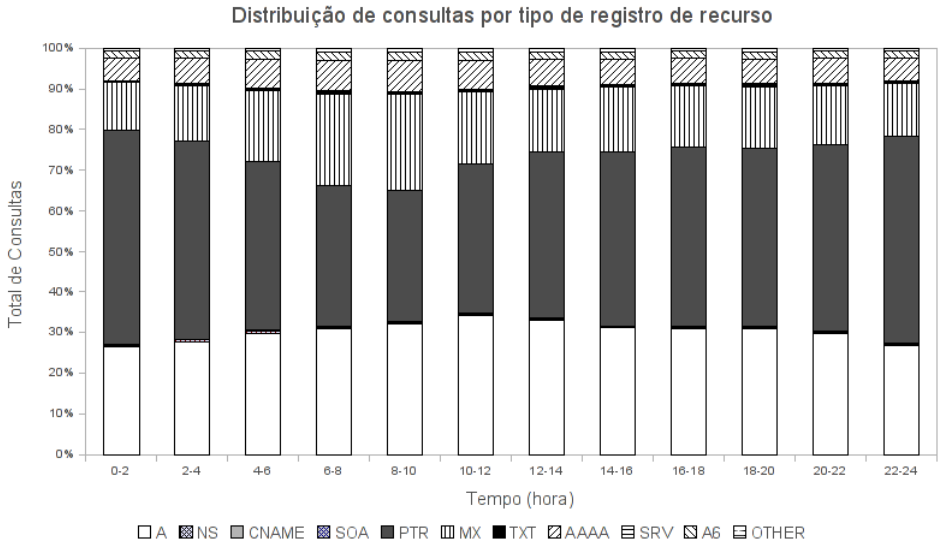

Figura 1. Distribuição de consultas por tipo de registro de recurso.

4.2. Determinando a validade das consultas

A detecção de poluição no tráfego DNS é uma atividade desafiadora porque algumas aplicações apesar de não obterem respostas válidas, ainda continuam funcionando. Wessels e Fomenkov [Wessels e Fomenkov 2003] apresentam em seu trabalho o processo de classificação de consultas inválidas através da utilização de uma lista que indica a categoria da qual uma pergunta pode ser definida. Essa lista também é aplicada em outro trabalhos de classificação e monitoramento do tráfego, como em Castro et al. [Castro et al. 2008] e Brownlee et al. [Brownlee et al. 2001].

Os resultados nas seções seguintes aplicam algumas definições dessa lista, como nomes de domínio compostos de TLD inválido, nomes de consultas com o caractere ”_”e consultas para os endereços definidos na RFC 1918. No entanto, outras métricas são utilizadas para detectar anomalias de rede, como identificação das principais fontes de registros do tipo PTR e classificação das consultas mais freqüentes no tráfego.

4.2.1. Consultas com TLDs inválidos

Como o estudo desse trabalho é restrito as consultas do domínio *.br*, TLDs que não casam com o padrão do sufixo *.br* são considerados como inválidos. Para Wessels [Wessels 2004], interromper o tráfego DNS com o TLD inválido é uma tarefa de grande complexidade, uma vez que o atual sistema DNS não oferece nenhum mecanismo que diferencie entre domínios válidos e inválidos.

Os resultados da análise da base de dados indicam que 0,02% das consultas (média de 300 mil consultas) são perguntas com o domínio de primeiro nível inválido. Entre os TLDs não reconhecidos, destacam-se as consultas que buscam domínios com extensão de nomes de arquivos como *.jpg*, *.css*, *.gif*, *.png*, *.js*, e *.html*. Consultas para domínio *.gif* representam em média 27% do total de consultas com TLD inválido, sendo que tais consultas possuem origem a partir de dois endereços IPs localizados na mesma rede. Comportamento semelhante a este também é observado em [Wessels 2004].

A Tabela 2 lista os dez domínios inválidos mais freqüentes na instância a.dns.br. Foi observado nessa lista uma quantidade surpreendente de números de pedidos gerados por um única fonte, sendo que 10% do tráfego observado nesse servidor são consultas para endereços de servidores que possuem o serviço de *Internet Relay Chat* (IRC) [Kalt 2000]. Esse tipo de tráfego demonstra o comportamento de máquinas infectadas por algum software de código malicioso tentado entrar em contato com o servidor de controle [Chi e Zhao 2007].

IP de Origem	Nome da Consulta
200.XXX.XXX.139	undernet.org:6667
200.XXX.XXX.201	frontal.correo
200.XXX.XXX.139	undernet.org:7000
200.XXX.XXX.78	_msdcs.server1
200.XXX.XXX.139	undernet.org:6669
200.XXX.XXX.201	kolbent.local
200.XXX.XXX.70	mercador.local
200.XXX.XXX.201	alexandriahealthcare.local
200.XXX.XXX.201	joi01.local
201.XXX.XXX.176	dev.null

Tabela 2. Dez domínios mais vistos na instância *a.dns.br*

Para mitigar essa grande quantidade de consultas inválidas, este trabalho sugere estender a lista de domínios marcados como domínios de testes ou inválidos [Eastlake 3rd e Panitz 1999] através da inclusão de nomes de domínios com extensão de nome de arquivos como *.gif*, *.jpg*, *.css* e *.png*. Além disso, como observado em [Wessels 2004], outros nomes domínios também poderiam fazer parte dessa lista como *.local*, *.localdomain*, *.lan*, *.home* e *.bind*. Portanto, esses nomes de domínio seriam considerados como inválidos reduzindo o processamento de carga dos servidores raiz

4.2.2. Consultas que iniciam com '_'

No tráfego analisado foram encontradas diversas consultas referentes ao registro de recurso do tipo SRV, utilizado no funcionamento do *Microsoft Windows Active Directory* [Microsoft 2008] para indicar um protocolo ou serviço de um determinado domínio [Gulbrandsen e Vixie 1996]. A figura 2 mostra um exemplo deste tipo de consulta.

```
1 IP 200.XXX.XXX.XXX.47812 > 200.189.40.10.53:  10937 SRV? _ldap._tcp.pdc._msdcs.EMFLORA.BR. (50)
2 IP 200.XXX.XXX.XXX.33215 > 200.189.40.10.53:  42582 SRV? _ldap._tcp.dc._msdcs.regra.local.br. (64)
3 IP 200.XXX.XXX.XXX.32779 > 200.189.40.10.53:  44264 SRV? _ldap._tcp.dc._msdcs.cvrd.br. (46)
4 IP 200.XXX.XXX.XXX.32806 > 200.189.40.10.53:  18507 SRV? _kerberos._tcp.dc._msdcs.www.ambl.com.br. (58)
5 IP 202.XXX.XXX.XXX.61284 > 200.189.40.10.53:  13590 SRV? _ldap._tcp.au-Kwinana._sites.dc._msdcs.explsbr.com.br. (71)
```

Figura 2. Exemplo de consultas inválidas listadas pelo tcpdump [tcpdump 2008].

Tais consultas correspondem a 9.1 milhões das solicitações analisadas, ou seja, 0.35% do total do tráfego. Requisições do tipo *_ldap._tcp.<DnsDomainName>* alcançam os servidores raiz devido a problemas de zonas integradas no *Active Directory* que podem ser resolvidos através da aplicação de *patches* de atualização do produto, conforme procedimento descrito na base de conhecimento KB62855 [Microsoft 2007b]. Outras consultas como *gc._msdcs.<DnsForestName>* podem estar relacionadas a clientes de rede tentando atualizar dinamicamente a zona de domínio do DNS com suas respectivas informações [Microsoft 2007a].

4.2.3. Utilização de endereços privados

A RFC 1918 [Rekhter et al. 1996] especifica o espaço de endereçamento de rede que deve ser utilizado para fins privados, cujas rotas nunca deverão ser anunciadas para a Internet. Entretanto, requisições utilizando endereços privados na Internet ainda persistem e, geralmente, alcançam os servidores raiz. Esse problema ocorre devido a configurações e utilizações equivocadas das zonas de domínio. O projeto AS112 [AS112 2009] tenta minimizar este problema respondendo às consultas desta natureza indicando que o nome requisitado não existe (*NXDomain*). Além disso, para evitar perguntas repetidas, o bit de autoridade [Mockapetris 1987b] da resposta é habilitado esperando que o resolvedor tenha o *Negative Caching* [Andrews 1998] configurado. Danzig et al. [Danzig et al. 1992] também observam a importância da implementação do *Negative Caching* para reduzir a carga nos servidores de nome devido consultas repetidas.

No Brasil, requisições para o espaço de endereço privado representam 0.01% de perguntas inválidas do total do tráfego observado. A figura 3 ilustra a distribuição de consultas recebidas por cada servidor autoritativo do domínio *.br*. A instância *a.dns.br* responde por 29% desse tipo de tráfego, as instâncias *b.dns.br* e *e.dns.br* respondem em média por 20%, enquanto que as instâncias restantes (*c.dns.br* e *d.dns.br*) respondem por 13% e 16% respectivamente.

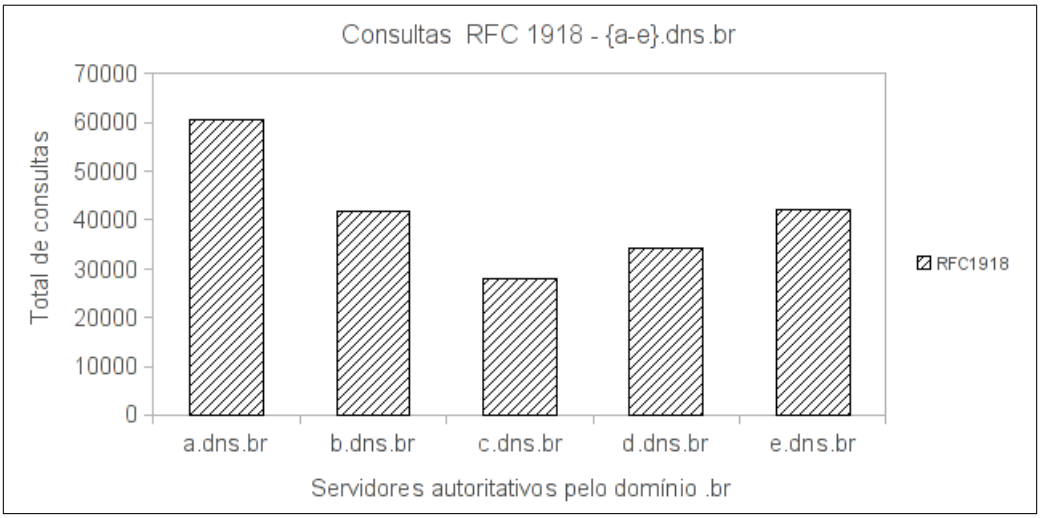

Figura 3. Distribuição de consultas RFC 1918.

Para mitigar consultas para o espaço privado de rede, a figura 4 ilustra os endereços de redes que devem ser configurados pelos administradores de sistema, para que o servidor de nomes, responda de modo autoritário pelas zonas de domínio da RFC 1918. Como resultado, perguntas dessa natureza estão restritas somente ao seu perímetro de rede. A configuração das zonas também é sugerida por diversos trabalhos, como Briodo et al. [Broido et al. 2006], Wessels [Wessels 2004] e Castro et al. [Castro et al. 2008].

- 10.in-addr.arpa
- intervalo entre 16.172.in-addr.arpa até 31.172.in-addr.arpa
- 168.192.in-addr.arpa

Figura 4. Zonas de domínios da RFC 1918 que devem ser configuradas para interromper consultas para endereços privados.

4.3. Ataques de reconhecimento de rede

Várias atividades maliciosas pela Internet utilizam o tráfego DNS para obter informações privilegiadas de rede. Por exemplo, o registro PTR é utilizado como artifício para identificar se um endereço IP está configurado e ativo. Ren et al. [Ren et al. 2006] sugerem que o volume elevado de consultas reversas são resultados de ataques de força bruta contra servidores que oferecem *secure shell* (SSH) [Ylönen 1996]. Esse serviço verifica se o endereço IP do atacante possui uma entrada PTR válida. Oberheide et al. [Oberheide et al. 2007] atribuem a grande incidência do registro PTR à ataques de reconhecimento de rede. Neste tipo de ataque, o atacante percorre uma classe de endereços IPs visando encontrar computadores configurados e ativos.

A figura 5 apresenta uma comparação entre os registros de recurso do tipo A, PTR e MX. Os registros do tipo PTR foram observados com mais freqüência durante todo o evento. O maior pico de consultas é registrado às 23h. Neste instante, as requisições do

tipo reverso apresentaram valores acima de 66,5 milhões de consultas, que correspondem a 2,5% do total de tráfego analisado.

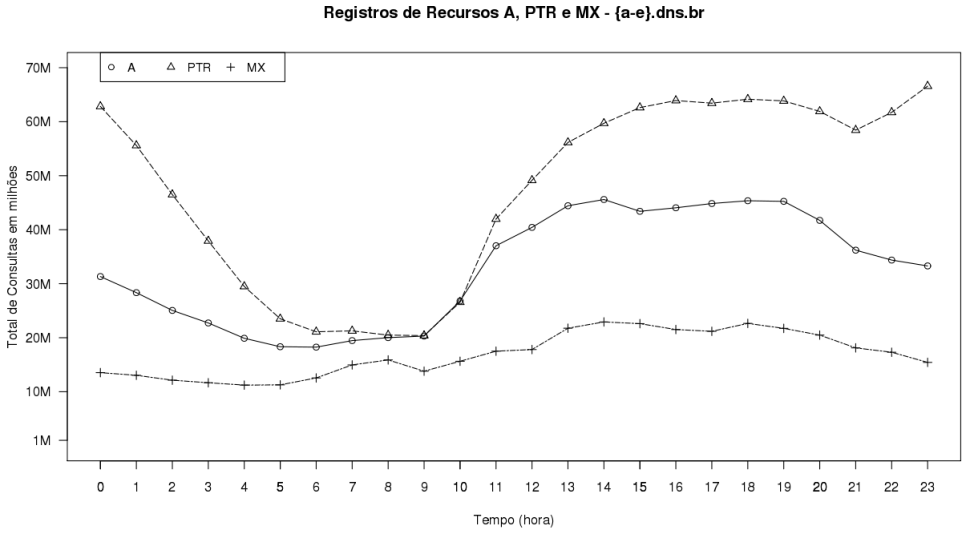

Figura 5. Comparação entre os registros A, PTR e MX.

Em razão do alto volume de consultas do tipo PTR, uma investigação mais detalhada foi realizada para detectar a origem desse tipo de tráfego. A análise mostra que a maioria das consultas foram produzidas por um único cliente de banda larga. A figura 6 mostra o endereço IP deste cliente (*201.XXX.XXX.74*), que é responsável por aproximadamente 1.5 milhões de requisições do total do tráfego observado. Tais consultas apresentam um padrão de ataque de reconhecimento de rede, onde o atacante percorre endereçamentos de redes de países distintos como França, Itália, Estados Unidos, Índia, Alemanha, Canadá, Austrália e Espanha. Devido a questões de sigilo e segurança da informação, alguns dados abaixo foram removidos do tráfego.

```
1 IP 201.XXX.XXX.74.32000 > 200.189.40.10.53:  15098 PTR? 0.XXX.XXX.57.in-addr.arpa.com.br. (62)
2 IP 201.XXX.XXX.74.32000 > 200.189.40.10.53:  58538 PTR? 0.XXX.XXX.151.in-addr.arpa.com.br. (60)
3 IP 201.XXX.XXX.74.32000 > 200.189.40.10.53:  48191 PTR? 1.XXX.XXX.57.in-addr.arpa.com.br. (61)
4 IP 201.XXX.XXX.74.32000 > 200.189.40.10.53:  57046 PTR? 1.XXX.XXX.151.in-addr.arpa.com.br. (61)
5 IP 201.XXX.XXX.74.32000 > 200.189.40.10.53:  11715 PTR? 10.XXX.XXX.57.in-addr.arpa.com.br. (60)
6 IP 201.XXX.XXX.74.32000 > 200.189.40.10.53:  55148 PTR? 10.XXX.XXX.57.in-addr.arpa.com.br. (60)
```

Figura 6. Ataque de reconhecimento de rede.

O ataque detectado percorre primeiramente os endereços que terminam com *.0*, em seguida aqueles que terminam *.10*. Logo após as consultas buscam pelos endereços com final *.1* e os IPs que terminam com *.20* também são percorridos. A estratégia adotada é peculiar, permitindo que o atacante descubra endereços de rede válidos eficientemente. Entretanto, a afirmação de que a predominância de consultas PTR seja evidência de operações maliciosas (por exemplo, de um *malware*) requer ainda uma análise mais detalhada da distribuição da origem dessas consultas.

4.4. Envio de mensagens não solicitadas

Outro fator que explica o grande volume de consultas do tipo PTR na Internet é a utilização de sistemas anti-spam pelos provedores de serviço de Internet e empresas brasileiras. Tais sistemas aplicam mecanismos de filtragem e rejeição de mensagens quando o

servidor de correio eletrônico responsável pela mensagem enviada não possuir o registro de recurso reverso configurado. Neste sentido, as abordagens anti-spam buscam superar o principal problema encontrado no protocolo SMTP (*Simple Mail Transfer Protocol*) [Klensin 2001] que não oferece mecanismos confiáveis de validação e identificação do emissor da mensagem, através de soluções que reconheçam a identidade do remetente, tais como *Sender Policy Framework* [Wong e Schlitt 2006] e [Lyon e Wong 2006]. Estes mecanismos empregam entre outras coisas, a verificação do DNS reverso do IP do servidor que enviou a mensagem.

Para tornar evidente o problema apresentado, considere a análise realizada na instância a.dns.br. A figura 7 mostra a quantidade de consultas do tipo PTR durante o intervalo de uma hora. Os resultados apresentados exibem cinco nomes de consultas (QNAME) mais freqüentes. Por exemplo, a primeira consulta foi observada no tráfego aproximadamente 3 mil vezes, sendo que o total destas consultas representam 0.24% do tráfego monitorado.

As consultas na figura 7 foram produzidas por servidores de nomes e servidores de correio eletrônico para encontrar clientes de Internet banda larga. Através deste tipo de comportamento é razoável inferir que os clientes encontrados no QNAME estão utilizando o tráfego da Internet para exercer atividades maliciosas ou fraudulentas.

Informações sobre a base de dados	Quantidade de Consultas (PTR)	Nome da Consulta (QNAME)
Servidor: a.dns.br Dia do monitoramento: 18/03/2008 Intervalo selecionado: 00:00:00 - 00:59:59 Quantidade de pacotes: 18.8M Total de consultas: 9.9M Total de consultas do tipo PTR: 3.7M	2959	X.X.X.200.in-addr.arpa.
	1885	X.X.X.201.in-addr.arpa.
	1777	X.X.X.200.in-addr.arpa.
	1402	X.X.X.189.in-addr.arpa.
	1181	X.X.X.189.in-addr.arpa.
Percentual das cinco consultas mais frequentes.	0.24%	

Figura 7. Quantidade de consultas do tipo PTR para o intervalo de uma hora do servidor autoritativo *a.dns.br*.

Os resultados apresentados mostram que as atividades de spam executadas nos clientes de banda larga certamente contribuem para enormes prejuízos aos provedores de serviço, devido ao consumo de recursos tais como largura de banda, memória e processamento. Além disso, os spams também consomem inutilmente o tempo dos destinatários e reduzem a credibilidade dos usuários na Internet. Por estas razões, soluções como as propostas pelas RFCs 5068 [Hutzler et al. 2007] e 4409 [Gellens e Klensin 2006] tentam mitigar o envio de mensagens destes clientes através da gerência da porta de saída que é responsável pela submissão de mensagens eletrônicas.

4.5. Uma palavra sobre a RFC 3490

O avanço da Internet permitiu que países que possuem em seu repertório linguístico palavras com acentos gráfico utilizem essa categoria de palavras em nomes de domínios DNS.

No entanto, Castro et al. [Castro et al. 2008] e Brownlee et al. [Brownlee et al. 2001], consideram consultas inválidas aquelas perguntas que não usam os caracteres baseados na tabela ASCII conforme definição da RFC 1035 [Mockapetris 1987b]. Por outro lado, este trabalho discorda da interpretação dada por esses autores quanto à utilização desses caracteres. Iniciativas como internacionalização de nomes de domínios em aplicações (*Internationalizing Domain Names in Applications* - IDNA) [Faltstrom et al. 2003], oferecem um novo mecanismo para utilizar caracteres fora do repertório ASCII sem modificação no protocolo ou comportamento do DNS. Contudo, a única alteração é realizada nos navegadores de Internet que pretendem suportar a funcionalidade sugerida pela IDNA. Por exemplo, o endereço de domínio *previdênciasocial.gov.br* é convertido em código ASCII compatível, que resulta em um novo nome de domínio, *www.xn–previdncia-r7a.gov.br*, transparente para o usuário final.

O processo de adoção da IDNA facilita a criação de novos nomes de domínio, visto que a etapa de codificação de nomes acontece na aplicação do usuário. No entanto, a técnica de mudança pode apresentar problemas nessa fase. A análise realizada neste trabalho não permite concluir se o nome da consulta resultante dessa falha é interpretado como caractere não imprimível. Portanto, é importante que seja realizado uma investigação do impacto dessa solução no tráfego DNS.

5. Conclusões

O monitoramento do tráfego do DNS certamente não é uma atividade inovadora, no entanto, a análise do tráfego permite aos operadores do serviço de DNS entenderem e buscarem soluções para mitigar tráfego não desejado que chegam aos servidores de DNS. No Brasil, este trabalho acredita ter apresentando os primeiros dados concretos da análise de tráfego real do domínio *.br*, apresentando evidências de atividades legítimas e fraudulentas nas consultas DNS.

O tráfego não desejado como consultas inválidas e servidores de nomes com problemas na configuração da zona de domínios influenciam negativamente no tráfego DNS da Internet, pois consomem os recursos computacionais que deveriam atender exclusivamente consultas válidas. Especialmente consultas com o TLD não reconhecidos pela ICANN, como *.local*, *.gif*, *.css* que, não são restringidas de maneira eficiente ou ainda, consultas para o espaço de rede privado encontrado na RFC 1918 que possuem significado inválido para os servidores raiz fora do perímetro de rede. Nesse sentido, alguns problemas mais freqüentes do DNS poderiam ser mitigados com adoções do *negative caching* [Andrews 1998] e da configuração da zona que responderia de forma autoritativa pelos endereços da RFC 1918.

Além disso, a análise dos dados apresentou resultados do tráfego da Internet Brasileira fornecendo indícios de comportamento diferente em relação a outros estudos já feitos sobre medição e caracterização do tráfego DNS. Consultas de resolução reversa são mais freqüentes no tráfego que registros do tipo A, isto é, consultas do tipo PTR corresponderam a 43% do total de consultas em razão de ataques de reconhecimento de rede, envio de SPAM e outras atividades legítimas, como filtros de sistemas de combate de envio de mensagens não solicitadas.

Apesar dos dados fornecerem algumas evidências de atividades fraudulentas como ataques de reconhecimento de rede e presença de programas de código malicioso

(*malware*) no tráfego DNS, como trabalho futuro, os dados referentes aos ataques de rede poderiam ser confrontados e analisados em distribuições de origem destas consultas para validar e confirmar que solicitações desta natureza, são predominante ou não no tráfego de registros do tipo PTR.

Em redes de alta velocidade o monitoramento do tráfego DNS como os servidores raiz é uma atividade de grande complexidade em razão da quantidade de volume de tráfego a ser analisado. Por este motivo, algumas soluções utilizam a teoria da informação a fim de encontrar apenas as variações com base no comportamento da rede como [Xu et al. 2005a] [Xu et al. 2005b] [Lee e Xiang 2001]. Estas alternativas podem ser aplicadas para auxiliar os operadores de rede a detectarem anomalias no tráfego.

5.1. Agradecimentos

Este trabalho agradece ao Centro de Pesquisa, Análise e Operações (OARC) pelos dados e ambientes de teste (*testbed*) fornecidos e, ao Duane Wessels pelas inúmeras discussões que culminaram no desenvolvimento desse artigo.

Referências

Andrews, M. (1998). Negative caching of dns queries (dns ncache).

AS112 (2009). As112 project home page. http://www.as112.net/.

Bojan, Z., Nevil, B., e Duane, W. (2007). Passive monitoring of dns anomalies. In *DIMVA 07: Proceedings of the 4th international conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 129–139, Berlin, Heidelberg. Springer-Verlag.

Broido, A., Hyun, Y., Fomenkov, M., e Claffy, K. (2006). The windows of pivate dns updates. *SIGCOMM Comput. Commun. Rev.*, 36(3):93–98.

Brownlee, N., Claffy, K., e Nemeth, E. (2001). Dns measurements at a root server. In *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, volume 3, pages 1672–1676.

Castro, S., Wessels, D., Fomenkov, M., e Claffy, K. (2008). A day at the root of the internet. *SIGCOMM Comput. Commun. Rev.*, 38(5):41–46.

Chi, Z. e Zhao, Z. (2007). Detecting and blocking malicious traffic caused by irc protocol based botnets. In *NPC '07: Proceedings of the 2007 IFIP International Conference on Network and Parallel Computing Workshops*, pages 485–489, Washington, DC, USA. IEEE Computer Society.

Danzig, P. B., Obraczka, K., e Kumar, A. (1992). An analysis of wide-area name server traffic: a study of the internet domain name system. *SIGCOMM Comput. Commun. Rev.*, 22(4):281–292.

DITL (2008). Day in the life of the internet. March 18-19, 2008 (DITL-2008-03-18) (collection). http://imdc.datcat.org/collection/1-05MM-F=Day+in+the+Life+of+the+Internet%2C+March+18-19%2C+2008+%28DITL-2008-03-18%29 (acesso em 2009/02/13).

DNS-OARC. Domain name system operations, analysis, and research center. https://www.dns-oarc.net/.

Eastlake 3rd, D. e Panitz, A. (1999). Reserved Top Level DNS Names. RFC 2606 (Best Current Practice).

Faltstrom, P., Hoffman, P., e Costello, A. (2003). Internationalizing domain names in applications (idna).

Gellens, R. e Klensin, J. (2006). Message Submission for Mail. RFC 4409 (Draft Standard).

Gulbrandsen, A. e Vixie, P. (1996). A DNS RR for specifying the location of services (DNS SRV). RFC 2052 (Experimental). Obsoleted by RFC 2782.

Hutzler, C., Crocker, D., Resnick, P., Allman, E., e Finch, T. (2007). Email Submission Operations: Access and Accountability Requirements. RFC 5068 (Best Current Practice).

Kalt, C. (2000). Internet Relay Chat: Client Protocol. RFC 2812 (Informational).

Klensin, J. (2001). Simple Mail Transfer Protocol. RFC 2821 (Proposed Standard). Obsoleted by RFC 5321, updated by RFC 5336.

Lee, W. e Xiang, D. (2001). Information-theoretic measures for anomaly detection. In *SP '01: Proceedings of the 2001 IEEE Symposium on Security and Privacy*, page 130, Washington, DC, USA. IEEE Computer Society.

Liu, C. e Albitz, P. (2006). *DNS and BIND (5th Edition)*. O'Reilly Media, Inc.

Lyon, J. e Wong, M. (2006). Sender ID: Authenticating E-Mail. RFC 4406 (Experimental).

Microsoft (2007a). How to enable or disable dns updates in windows 2000 and in windows server 2003. Technical report, Microsoft.

Microsoft (2007b). Problems with many domain controllers with active directory integrated dns zones. Technical report, Microsoft.

Microsoft (2008). Srv resource records.

Mockapetris, P. (1987a). Domain names - concepts and facilities. RFC 1034, Internet Engineering Task Force.

Mockapetris, P. (1987b). Domain names - implementation and specification. RFC 1035, Internet Engineering Task Force.

Oberheide, J., Karir, M., e Mao, Z. M. (2007). Characterizing dark dns behavior. In *DIMVA '07: Proceedings of the 4th international conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 140–156, Berlin, Heidelberg. Springer-Verlag.

Pappas, V., Xu, Z., Lu, S., Massey, D., Terzis, A., e Zhang, L. (2004). Impact of configuration errors on dns robustness. In *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 319–330, New York, NY, USA. ACM.

Registro.br. Registro.br. http://www.registro.br, 2009.

Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G. J., e Lear, E. (1996). Address allocation for private internets.

Ren, P., Kristoff, J., e Gooch, B. (2006). Visualizing dns traffic. In *VizSEC '06: Proceedings of the 3rd international workshop on Visualization for computer security*, pages 23–30, New York, NY, USA. ACM.

tcpdump (2008). Tcpdump - dump traffic on a network. http://www.tcpdump.org/.

Wessels, D. (2004). Is your caching resolver polluting the internet? In *NetT '04: Proceedings of the ACM SIGCOMM workshop on Network troubleshooting*, pages 271–276, New York, NY, USA. ACM.

Wessels, D. (2008). Dnstop. stay on top of you dns traffic. http://dns.measurement-factory.com/tools/dnstop/.

Wessels, D. e Fomenkov, M. (2003). That's a lot of packets. In *in Proc. 2003 Passive and Active Measurements Workshop*.

Wong, M. e Schlitt, W. (2006). Sender Policy Framework (SPF) for Authorizing Use of Domains in E-Mail, Version 1. RFC 4408 (Experimental).

Xu, K., Zhang, Z.-L., e Bhattacharyya, S. (2005a). Profiling internet backbone traffic: behavior models and applications. In *SIGCOMM '05: Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 169–180, New York, NY, USA. ACM.

Xu, K., Zhang, Z.-L., e Bhattacharyya, S. (2005b). Reducing unwanted traffic in a backbone network. In *SRUTI'05: Proceedings of the Steps to Reducing Unwanted Traffic on the Internet on Steps to Reducing Unwanted Traffic on the Internet Workshop*, pages 2–2, Berkeley, CA, USA. USENIX Association.

Ylönen, T. (1996). Ssh: secure login connections over the internet. In *SSYM'96: Proceedings of the 6th conference on USENIX Security Symposium, Focusing on Applications of Cryptography*, pages 4–4, Berkeley, CA, USA. USENIX Association.