Identificação e Caracterização de Comportamentos Suspeitos Através da Análise do Tráfego DNS

Kaio R. S. Barbosa, Eduardo Souto, Eduardo Feitosa, Gilbert B. Martins

ALUNO: WEVERTON BUENO DA SILVA

Motivação

- Acesso livre e distribuído do protocolo DNS.
- É um problema em aberto, especialmente quando o tráfego é analisado em servidores de Domínio de Primeiro Nível.
- É possível assumir que hosts maliciosos apresentem padrões de comportamentos similares na rede.

Metodologia

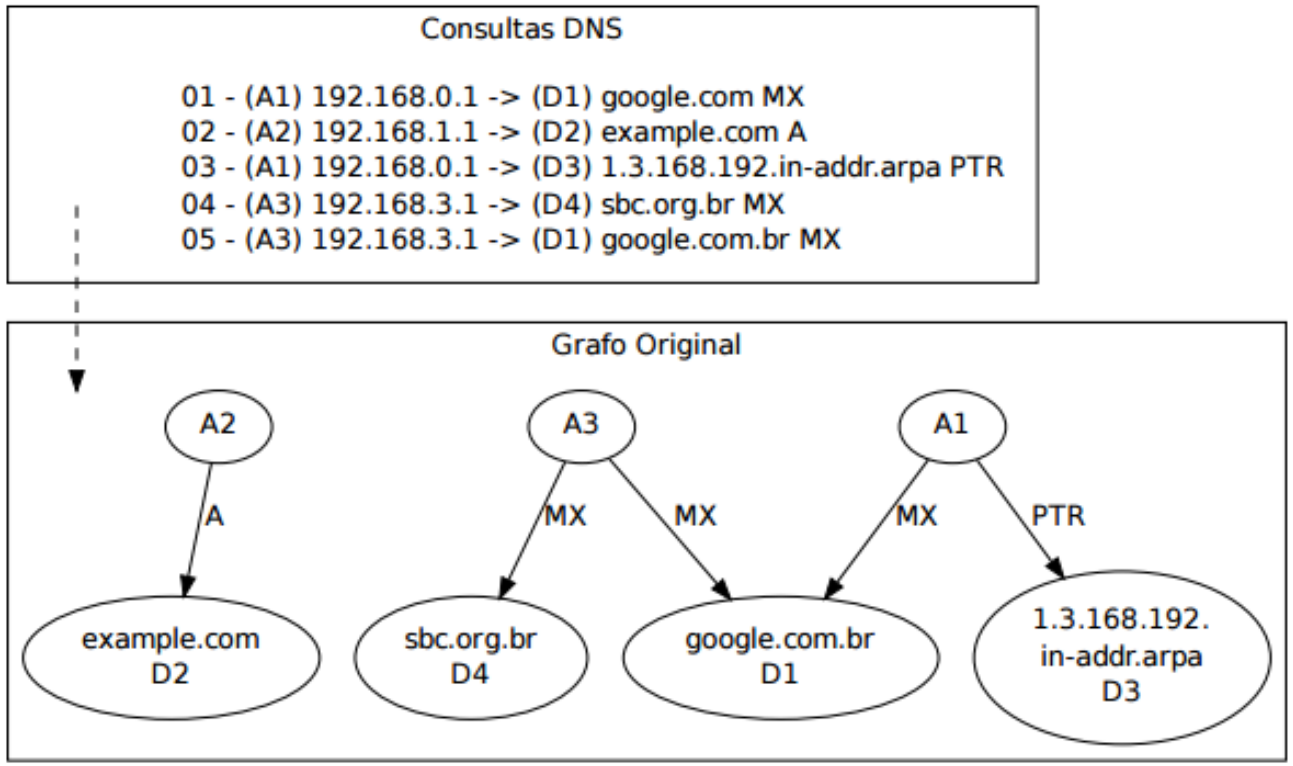
1. Construção do grafo original, onde o tráfego DNS é modelado em grafo direcionado. Os vértices são formados por hosts e nomes de domínios e as arestas, as comunicações entre os vértices.

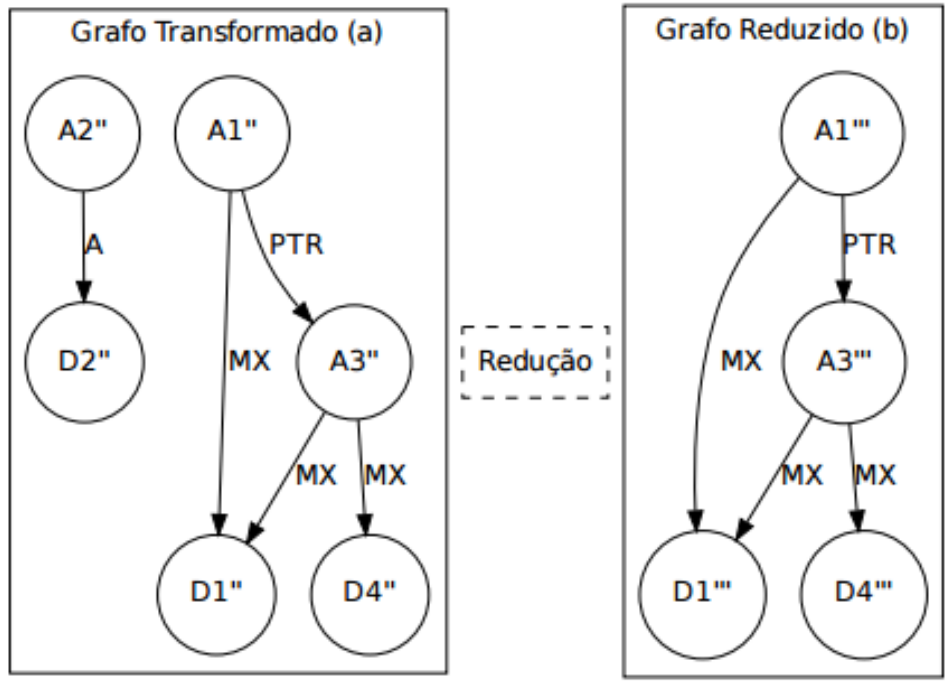
2. Transformação do grafo, tem objetivo reforçar as conexões do grafo para encontrar padrões de comunicações que não foram modelados inicialmente.

3. Redução do grafo, onde os componentes conexos irrelevantes do grafo transformado são eliminados.

4. Classificação das consultas, onde um conjunto de métricas, definidas para descrever as propriedades estruturais do grafo, e usado para classificar os nós e identificar possíveis comportamentos maliciosos nos hosts associados.

Exemplo


Consultas DNS
01 - (A1) 192.168.0.1 -> (D1) google.com MX
02 - (A2) 192.168.1.1 -> (D2) example.com A
03 - (A1) 192.168.0.1 -> (D3) 1.3.168.192.in-addr.arpa PTR
04 - (A3) 192.168.3.1 -> (D4) sbc.org.br MX
05 - (A3) 192.168.3.1 -> (D1) google.com.br MX
Grafo Original
A2
A
example.com
D2
A3
MX
MX
sbc.org.br
D4
A1
MX
PTR
google.com.br
D1
1.3.168.192.
in-addr.arpa
D3


Grafo Transformado (a)
A2"
A
D2"
A1"
PTR
MX
A3"
MX
MX
D1"
D4"
Redução
Grafo Reduzido (b)
A1'''
PTR
MX
A3'''
MX
MX
D1'''
D4'''

Métricas

i) alto grau de entrada do nó são observados principalmente nos que possuem alta incidência de consultas dos registros de recurso do tipo PTR e MX.

ii) alto grau de saída de nó são observados nos que realizam grandes quantidades de consultas empregando principalmente os registros PTR, MX e NS.

iii) razão entre o grau saída e entrada do vértice a partir dessa relação e possível identificar padrões de consultas semelhantes no tráfego.

iv) frequência relativa dos registros de consultas enviadas;

v) frequência relativa dos registros de consultas recebidas. Tais frequências permitem agrupar hosts que abusam dos registros (e.g: PTR e MX) independente da quantidade de consultas enviadas.

Classe Primária Q	**X**	**Y**	**Z**	**W**
Grau de Entrada (deg^-)	deg^+	Razão	Freq. RR Enviados	Freq. RR Recebidos
Grau de Saída (deg^+)	Razão	deg^-	Freq. RR Enviados	Freq. RR Recebidos
Razão Saída / Entrada	deg^-	deg^+	Freq. RR Enviados	Freq. RR Recebidos
Frequência Registros Recursos Enviados	deg^-	deg^+	Razão	Freq. RR Recebidos
Frequência Registros Recursos Recebidos	deg^-	deg^+	Razão	Freq. RR Enviados

Base de Dados

Foi utilizado o tráfego DNS real coletado durante o projeto DITL [DITL 2014], uma cortesia da DNS-OARC (Centro de Pesquisa, Operações e Análise DNS). O tráfego obtido é composto por 15 servidores DNS autoritativos que respondem pelo domínio *.br*, sendo cinco em 2008, quatro em 2009 e seis em 2010.

	DITL 2008	DITL 2009	DITL 2010
Dias de Coleta	[18-19]/03	[30-31]/03, 01/04	[14-15]/04
Servidores DNS	{a-e}.dns.br	{a,b,e,f}.dns.br	{a-f}.dns.br
Total de Pacotes	5.3 bilhões	6.4 bilhões	6.9 bilhões
Tamanho da base	229GB	236GB	282GB

Servidor *a.dns.br* durante o período de 00:00 até 00:59 dos dias 18/03, 30/03 e 14/04 de 2008, 2009 e 2010, respectivamente.

Resultados

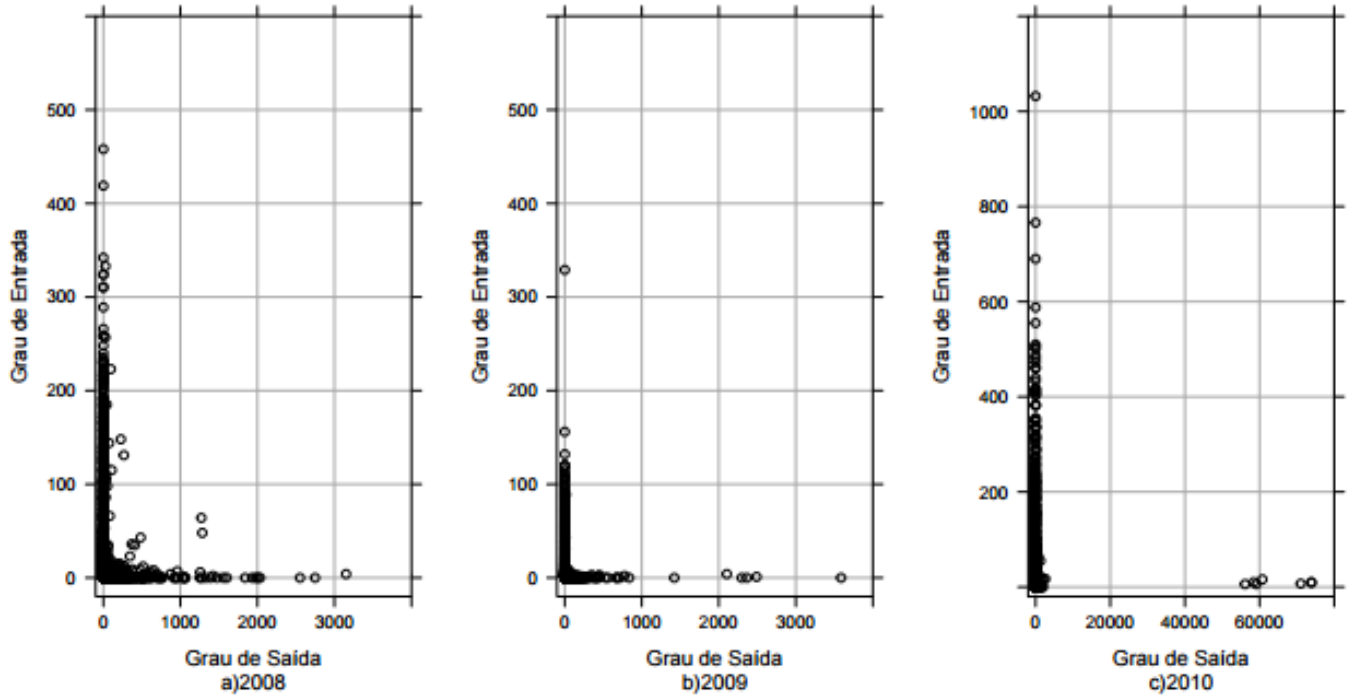
Tabela 3. Resumo dos hosts relevantes encontrados para análise.

	DITL 2008	DITL 2009	DITL 2010
Total de Consultas	9.985.841	23.190.993	25.193.658
Total de Hosts Únicos	263.036	351.353	342.045
Total de Hosts Relevantes	138.558	178.519	306.124
Dist. Freq. - A|PTR|MX	32% | 53% | 12%	20% | 74% | 4%	39% | 43% | 16%

Através da metodologia proposta foi possível reduzir em média 35% do total de hosts a serem analisados.

A figura apresenta o grau de saída e o grau de entrada dos vértices analisados em 2008, 2009 e 2010.

Por razões de espaço, 20 mil hosts de cada ano foram escolhidos aleatoriamente e plotados.



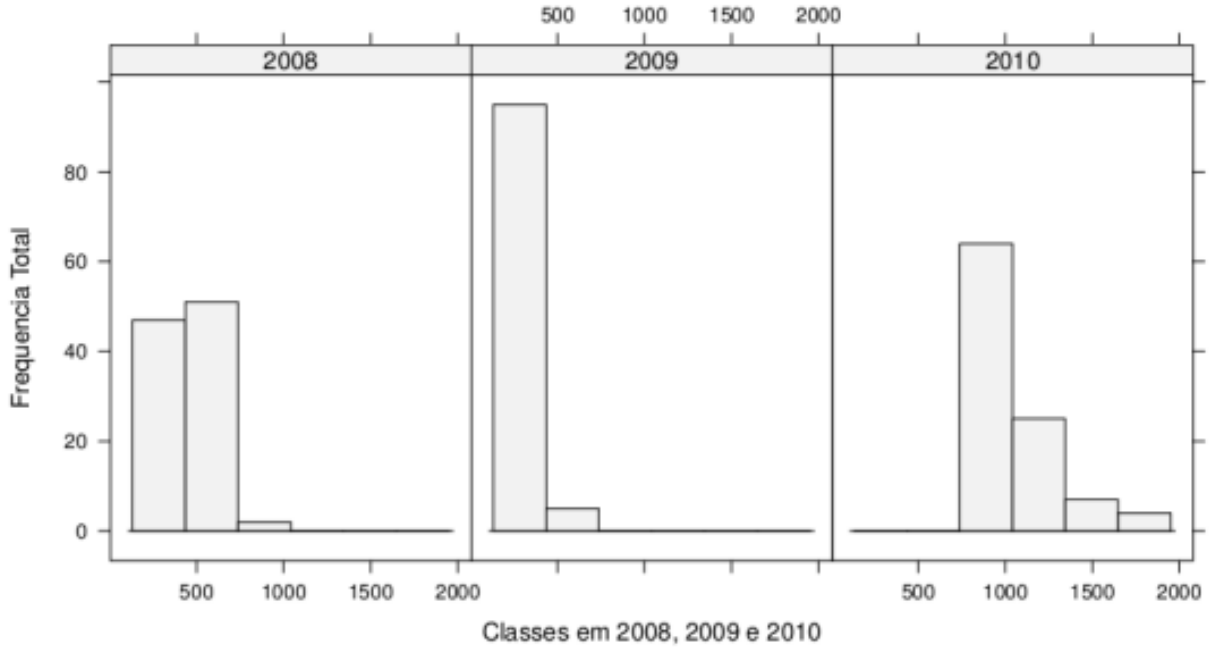
Os valores identificados mostram que 22 hosts enviaram mais de 60 mil requisições e, por isso, foram investigados.

Tabela 4. Hosts relevantes em 2010 que enviaram mais de 60 mil consultas DNS.

Host	deg^-	deg^+	Razão	Freq. RR Entrada	Freq. RR Saída
1	50	204.616	4092.32	100%	75.2% | 14.7% | 14.4%
2	45	177.259	3930.08	100%	73.6% | 10.8% | 15.5%
3	37	130.205	3519.05	100%	70.3% | 9.4% | 19.8%
4	13	105.763	8135.61	92.4% | 7.6%	74.4% | 6.6% | 18.8%
5	37	97.858	2644.81	100%	81.1% | 6.9% | 11.8%

A frequência dos registros de recursos de entrada denotam 100% de consultas do tipo *PTR*, com exceção do host de número 4, o qual recebeu consultas *PTR* (92.4%) e do tipo * (7.6%). As frequências dos registros de saída mostram os valores dos registros do tipo *A*, *PTR* e *MX*, entre os quais o tipo *A* e o mais frequente.

Para ilustrar como os hosts são observados nas classes de consultas, a classe primária do grau de entrada e demonstrada a seguir. Os 100 primeiros hosts que tiveram maior incidência de consultas foram identificados e atribuídos em intervalos distintos das classes de comportamentos.



Em 2010, 60% dos hosts que possuem maior incidência de consultas estão dentro do intervalo entre 500 e 1000 consultas.

Vale destacar dois grupos:

i) o conjunto de hosts que apenas recebem consultas.

Representa 70% do total de hosts analisados. Tais hosts são consultados através dos registros de recurso *PTR*, * e *CNAME*.

ii) conjunto de hosts que enviam e recebem consultas.

Representa 30% do intervalo citado e a frequência dos registros de recursos de saída denota comportamento suspeito. Os hosts identificados nesse grupo abusam do registro de recurso do tipo *MX*. Em média, esse registro representa 79% do total de consultas enviadas, enquanto os registros do tipo *A* e *PTR* representam 20% e 1%, respectivamente.

Análise Passiva dos Hosts Relevantes

Assume como comportamento suspeito um endereço de banda larga abusando dos registros PTR ou MX, ou ainda, quando múltiplos servidores de email, dentro do mesmo espaço de tempo, consultam por um endereço de banda larga.

Considerando os hosts relevantes que apenas receberam consultas, 88.5% são endereços de banda larga e os demais denotam nome de domínio inexistente.

Uma verificação superficial mostrou que 85.8% dos hosts estavam cadastrados em listas negras por não seguirem as definições da RFC 2142.

O grau de entrada durante 10s do tempo total analisado, denota que esses hosts receberam em média 6 consultas a partir de endereços IPs distintos.

A maioria dos endereços IPs de origem que consultaram esses hosts são servidores de email e DNS.

Um conjunto de hosts relevantes são consultados por um ou mais endereços IP de origem.

Diante desse contexto, e possível assumir que os hosts relevantes estão sendo consultados e validados após o envio em massa de spam.

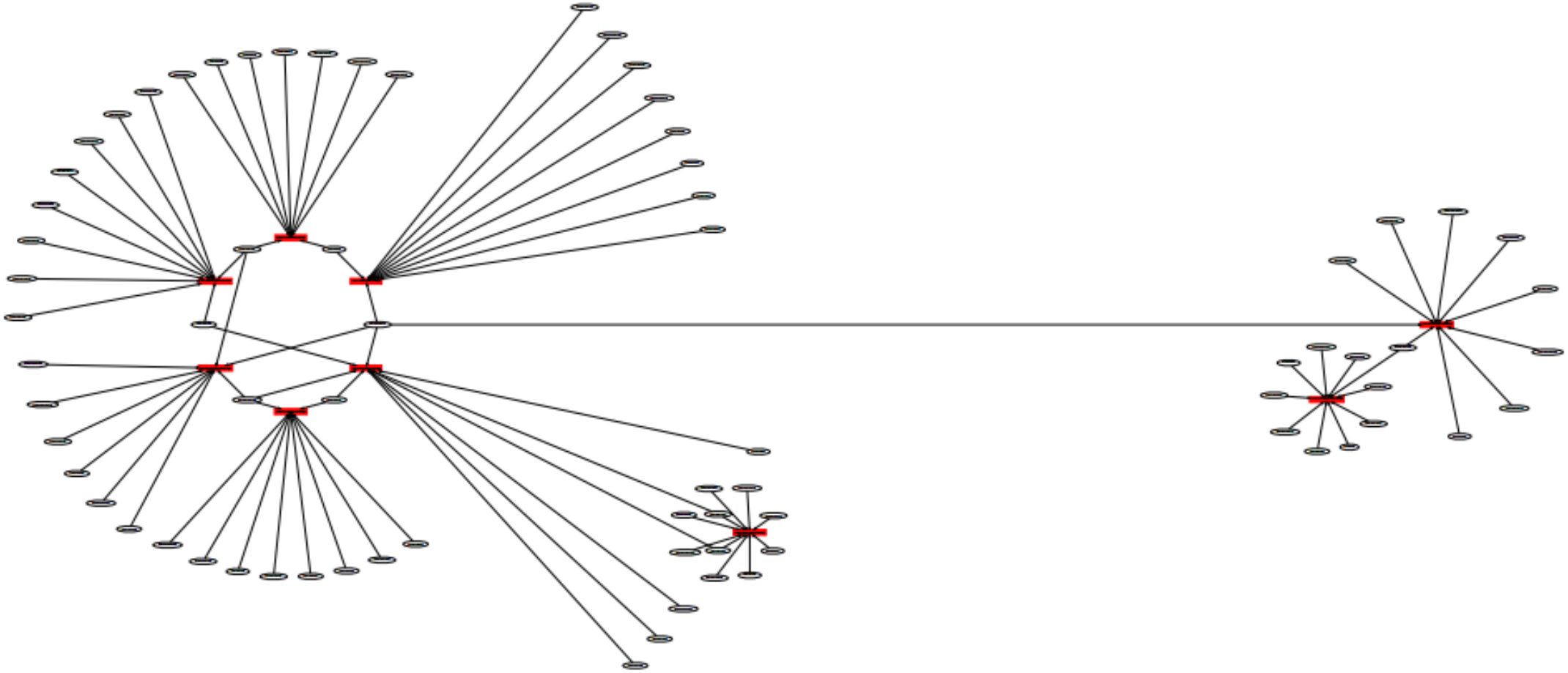


Figura 5. Padrão de consultas recebidas por *bots* para envio de spam. Quadrados vermelhos denotam nome da consulta, enquanto as elipses representam o IP de origem.

```
01 - 201.b.c.x2z <domínio>.com.br MX
02 - 201.b.c.x2z ns1.<domínio>.com.br A
03 - 201.b.c.x2z itajuba.com.br MX
04 - 201.b.c.x2z <domínio>.com.br MX
05 - 201.b.c.x2z <domínio>.com.br MX
...
06 - 200.202.252.2 d.c.b.a.in-addr.arpa PTR
07 - 200.202.252.2 d.c.b.a.in-addr.arpa PTR
08 - 200.202.252.2 x2z.c.b.201.in-addr.arpa PTR
09 - 200.202.252.2 d.c.b.a.in-addr.arpa PTR
10 - 200.202.252.2 d.c.b.a.in-addr.arpa PTR
```

Figura 6. Exemplo de consultas realizadas por um *bot* para envio de spam.

Tal comportamento pode ser identificado caso exista uma aresta entre o endereço IP do servidor de email e o endereço IP de origem inicial da consulta DNS. No entanto, para obter o endereço IP do servidor de email é necessário realizar uma nova consulta, a qual aumenta o tempo de análise dos hosts relevantes.

Trabalhos Futuros

- Outros registros do tráfego DNS podem ser utilizados para identificar comportamentos suspeitos.

- Utilizar algoritmos de aprendizagem de máquina para detectar e classificar comportamentos suspeitos em tempo real a partir do emprego das características extraídas dos hosts relevantes.

Referência

Kaio R. S. Barbosa, Eduardo Souto, Eduardo Feitosa, Gilbert B. Martins. Identificação e Caracterização de Comportamentos Suspeitos Através da Análise do Tráfego DNS. Presente em XIV SBSEG, pág. 167–180.