Características do Ruído de Fundo da Internet Brasileira a partir de dados do Consórcio Brasileiro de Honeypots

Eduardo G Barros[1], Antonio Montes[2], Stephan Stephany[1]

[1]Laboratório Associado de Computação Aplicada – Instituto Nacional de Pesquisas Espaciais (INPE)
Av dos Astronautas, 1758 – Jd Granja – CEP 12227-010 – São José dos Campos – SP – Brasil

[2]Centro de Pesquisas Renato Archer (CENPRA)
Rodovia Dom Pedro I, km 143,6 – Amarais – CEP 13069-901 – Campinas – SP – Brasil
edugdb@cea.inpe.br, antonio.montes@cenpra.gov.br, stephan@lac.inpe.br

Abstract. *During the last years it has been noticed the presence of a constant Internet traffic: the background radiation. It is composed of malicious traffic, like the one due to worms and bots, and/or benign traffic due to misconfigurations. Once this radiation is not fully characterized, malicious agents can use it to hide their activities. The Brazilian Honeypots Consortium (CBH) uses IP addresses assigned to known institutions on machines intended to capture malicious traffic. This research employs CBH traffic data to characterize the background radiation to help security administrators to detect early-stage suspicious activities.*

Resumo. *Nos últimos anos tem-se percebido a existência de um tráfego constante na Internet: o ruído de fundo. Este é composto por tráfego malicioso, como o originado por worms e bots, e/ou benigno devido a má configuração. Como esse ruído ainda não foi plenamente caracterizado, agentes maliciosos podem utilizá-lo para encobrir suas atividades. O Consórcio Brasileiro de Honeypots (CBH) usa endereços IP atribuídos a instituições em máquinas destinadas a capturar tráfego malicioso. Este trabalho emprega dados de tráfego do CBH para a caracterização do ruído de fundo, de forma a auxiliar administradores a detetar atividades maliciosas ainda nos estágios iniciais.*

1. Introdução

Todo tráfego que possa ser caracterizado como complexo, altamente automatizado, malicioso e que sofre mutações num curto espaço de tempo constitui o que se chama de ruído ou de radiação de fundo.

Uma ferramenta que fornece a representatividade e a distribuição necessária para a aquisição de dados para caracterização do ruído de fundo são as redes telescópio. Elas monitoram tráfego enviado para porções não alocadas do endereçamento IP. Segundo Pang et all (2004), a maioria dos estudos que envolvem a coleta de dados maliciosos ou de ruído de fundo da Internet têm sido realizadas em redes telescópio.

Segundo Savage (2006), a combinação da capacidade de detecção em larga escala da rede telescópio com a capacidade de resposta dos honeypots permite obter

mais e melhores informações. Honeypots individuais não são as ferramentas mais adequadas para monitoramento do ruído de fundo por não apresentarem a distribuição geográfica necessária. Esta deficiência é superada com o uso de um grande número de honeypots.

Esta é a proposta do Consórcio Brasileiro de Honeypots (CBH). Uma aliança de mais de 40 instituições distribuídas por todo o Brasil, coordenadas pelo CenPRA e pelo CERT.br, que usam honeypots de baixa interatividade[1] como ferramentas para geração de avisos precoces e análise de tendências.

2. Metodologia

Dos vários sensores componentes do CBH levantou-se quais apresentam uma "característica de normalidade" que pode ser extrapolada para a internet brasileira. Por "característica de normalidade" entende-se que os dados do sensor foram examinados e avaliados segundo os seguintes critérios:

- número de dias que cada sensor forneceu dados: número de dias com dados para os anos de 2005 e de 2006 separadamente, por sensor, entre 00:00 de 01/01/2005 até 23:00 de 30/06/2006;
- relação entre número de fluxos TCP, UDP e ICMP: verifica a relação percentual dos fluxos TCP, UDP e ICMP em relação ao total de fluxos. A "característica de normalidade" esperada para este critério é de aproximadamente 90% de fluxos TCP, 7% de fluxos UDP e 3% de fluxos ICMP. Percentual médio obtido consolidando-se os dados de todos os sensores; e,
- relação entre número de fluxos com menos de três pacotes e fluxos com três ou mais pacotes: verifica a relação percentual ente estes fluxos e o total de fluxos TCP. A "característica de normalidade" esperada é de aproximadamente 75% de fluxos com menos de 3 pacotes e 25% de fluxos com 3 ou mais pacotes. Percentual médio obtido após consolidação dos dados de todos os sensores.

Através de uma métrica que associou pontos a cada uma dos critérios anteriores avaliou-se quais dos sensores melhor representam a "característica de normalidade" desejada. O subconjunto dos sensores com menor número de pontos é o que melhor representa o tráfego de dados na parcela brasileira da internet.

3. Resultados Iniciais

O fluxo TCP é muito maior do que a soma dos demais (≈ 90%); os fluxos UDP e ICMP pequenos com o UDP (≈ 7%) geralmente, superior ao ICMP (≈ 3%). Ainda é possível identificar, dentro do fluxo TCP dois fluxos distintos: fluxos com menos de 3 pacotes (≈ 80%, e geralmente associados a varreduras) e fluxos com 3 ou mais pacotes (≈ 20%, geralmente associados a conexões).

Para caracterizar o fluxo TCP e as portas de destino acessadas é usada uma representação visual como descrita em Grizzard et all (2005). É gerada uma imagem em que cada posição representa uma intensidade de acordo com a formulação proposta.

Analisou-se as portas de destino e verificou-se que 90% do fluxo com menos de 3 pacotes foi direcionado para somente 12 portas sendo que a porta 445 foi a mais varrida com quase duas vezes mais fluxos do que a segunda colocada, a porta 139.

1 Atacantes interagem com ferramentas que emulam sistemas operacionais e serviços.

Realizou-se a associação entre endereços IP de origem e fluxos, parâmetro usado por Grizzard J. et all. (2005) e Pang et all (2004), usando, para tal, o conceito de endereço IP único: endereço IP único é a quantidade de endereços mas não necessariamente a quantidade de máquinas.

Levantou-se, também, a quantidade de fluxos por endereço agrupando-se os endereços IP no primeiro byte de sua representação decimal independentemente da classe a qual pertence. Verificou-se que o endereço de origem 200 seguido do 201 são os que têm mais fluxos.

Foi estudada a associação entre endereços IP de origem e portas de destino usando a visualização já apresentada anteriormente. Ficou claro a existência de blocos de endereços que realizam varreduras na parcela brasileira da internet.

4. Conclusões

Estão sendo empregados os dados do CBH para levantar características que permitam caracterizar o ruído de fundo na parcela brasileira da internet.

As métricas sugeridas refletem o que se chamou de "característica de normalidade" e que se espera encontrar como ruído de fundo nas máquinas conectadas à Internet na sua parcela brasileira.

Verifica-se que a grande maioria das varreduras concentram-se em um número pequeno de portas que possuem tráfego continuado. As portas que apresentam conexão não devem ser extrapoladas do CBH para a internet porque nem todos os sensores possuem recursos para completar as conexões. A grande maioria das varreduras e, consequentemente, do ruído de fundo, tem origem em endereços IP iniciados por 200. No tráfego originado por varreduras nem todos os endereços são usados porém, nas conexões, todos os endereços, válidos ou não, são usados, caracterizando tentativas de ataques usando IPs forjados.

É necessário aprofundar alguns tópicos: (i) procurar correlações ou tendências entre as portas acessadas nas varreduras e aquelas acessadas nas conexões, (ii) buscar correlações ou tendências entre os endereços IP de origem e as portas, (iii) procurar quantificar/visualizar acelerações na velocidade das varreduras que caracterizem a presença de ataques por worms. Finalmente, empregar o CBH como gerador de alertas precoces.

Referências

Grizzard J. et all. (2005) "Flow Based Observations from NETI@home and Honeynet Data", In: Proceedings of the 2005 IEEE Workshop on Information Assurance and Security. United States Military Academy, West Point, NY, USA.

Pang, R. et all. (2004) "Characteristics of Internet Background Radiation", In: Proceedings of the IMC'04, Taomina, Italy.

Savage, S. et all. (2006) "Center for Internet Epidemiology and Defenses", http://www.cs.ucsd.edu/~savage/papers/CIEDProposal.pdf, October.