Plataforma para Análise de Dados de Honeypots no Curto Prazo

Gustavo José Neves da Silva
gustavo.neves@yandex.com
Universidade do Estado de Santa Catarina – UDESC
Joinville, Santa Catarina

Rafael R. Obelheiro
rafael.obelheiro@udesc.br
Universidade do Estado de Santa Catarina – UDESC
Joinville, Santa Catarina

ABSTRACT

Honeypots are computing resources whose value lies in being probed, attacked, or compromised. They are security tools instrumental in observing attackers' behavior and understanding attack dynamics. Honeypots generate a large amount of monitoring data, which are often processed and analyzed manually. This introduces a delay in the perception of atypical occurrences and changes in trends, often rendering a deeper analysis of such phenomena unfeasible. We introduce a platform for the periodic automated processing of data collected by honeypots. The platform generates statistics and graphs from a set of traffic capture files in the PCAP format, and its modular design allows it to be easily extended and adapted to produce different outputs.

KEYWORDS

Honeypots, Segurança computacional, Redes de computadores

1 INTRODUÇÃO

Um *honeypot* é um recurso computacional de segurança dedicado a ser sondado, atacado ou comprometido [1]. O monitoramento do tráfego de um *honeypot* permite a observação de diversos tipos de comportamento malicioso, tais como ataques. Como resultado do monitoramento, *honeypots* podem produzir grandes volumes de dados, que muitas vezes são processados e analisados manualmente. Isso pode fazer com que anomalias e mudanças de tendências no tráfego sejam percebidas muito tempo após terem acontecido, impossibilitando uma investigação mais aprofundada sobre essas ocorrências e suas causas.

Diversas descrições de arquiteturas de *honeypots* podem ser encontradas na literatura [1–8]. Em [6, 8] são mencionadas algumas ferramentas de análise de dados produzidos por *honeypots*, mas as ferramentas citadas não estão disponíveis publicamente. Ainda, em [6] é proposto um formato para representação e troca de dados coletados por *honeypots*. No entanto, as referências encontradas não detalham o processo de geração de estatísticas a partir dos dados coletados pelos *honeypots*.

2 SOLUÇÃO PROPOSTA

O presente trabalho tem como objetivo propor uma plataforma para o processamento automatizado diário dos dados coletados por *honeypots*. Essa plataforma será usada no processamento de dados do HReflector [9], um *honeypot* para a observação de ataques distribuídos de negação de serviço por reflexão (DRDoS, *distributed reflection denial of service*). Ataques DRDoS exploram protocolos requisição-resposta com características de amplificação (que podem gerar respostas muito maiores que requisições), e utilizam endereços IP de origem forjados para que as respostas sejam redirecionados para o alvo do ataque [10]. O HReflector usa o Tcpdump [11] para captura de tráfego, e armazena os dados no formato PCAP, o qual é comumente usado em *honeypots* [6]. Para evitar a manipulação de arquivos muito grandes, o próprio Tcpdump separa os arquivos em *chunks* de 100 MB; com isso, o tráfego de um dia pode estar disperso por vários arquivos dentro de um mesmo diretório.

O fluxo de processamento da plataforma proposta é mostrado na Figura 1. No passo 1, determina-se quais arquivos PCAP contêm o tráfego do dia de interesse, usando a ferramenta capinfos [12]. No passo 2, esses arquivos são combinados em um único arquivo usando a ferramenta mergecap [12]. No passo 3, os dados de pacotes contidos no arquivo PCAP são agregados em fluxos (*flows*) [13], ou seja, dados de sessões em intervalos de 5 minutos, usando a ferramenta argus [14]. No passo 4, o arquivo de fluxos é convertido para um arquivo de texto puro usando a ferramenta ra [14]. No passo 5, o arquivo gerado no passo 4 é processado para obter tabelas e gráficos com estatísticas de tráfego por ASN (5a), protocolo (5b) e país (5c), usando o pacote estatístico R [15]. Todo o fluxo de processamento é orquestrado por um script R.

A Figura 2 mostra uma tabela e um gráfico gerados pela ferramenta, com estatísticas do tráfego diário por país. Como os dados referem-se a ataques DRDoS, as vítimas de ataques são caracterizadas pelos endereços IP de origem dos pacotes, devido ao uso de IP *spoofing*. Pode-se observar nesse exemplo que, no dia 29/09/2018, a maior parte das vítimas de ataques DRDoS estava situada nos Estados Unidos (US), com o total de 32,61 MB, e que o pico de tráfego para esse país foi de 1,72 KB/s. O ataque mais intenso (pico mais à direita no gráfico) ocorreu às 22:35, e teve vítimas situadas na China, com intensidade de 22,29 KB/s. De acordo com a Figura 3, essas vítimas estavam concentradas no sistema autônomo identificado pelo ASN 4134.

3 CONSIDERAÇÕES FINAIS

O processamento automatizado de dados gerados por *honeypots* é essencial para permitir o acompanhamento contínuo dos ataques observados por essas ferramentas. Com esse acompanhamento é possível identificar anomalias e mudanças de tendências no tráfego pouco tempo depois do sua ocorrência, e consequentemente aprofundar a investigação sobre suas causas, tanto pela ampliação da instrumentação dos *honeypots* quanto pelo acesso a dados de monitoramento mais voláteis que ajudem a elucidar os fenômenos. Este trabalho apresenta uma plataforma que facilita o processamento periódico de dados armazenados no formato PCAP. Embora desenvolvida com foco no HReflector, a plataforma pode ser facilmente adaptada a outros *honeypots* que armazenem dados nos formatos PCAP ou de fluxos. Na continuidade desta pesquisa, pretende-se colocar a plataforma em produção e disponibilizar seu código fonte sob uma licença livre.

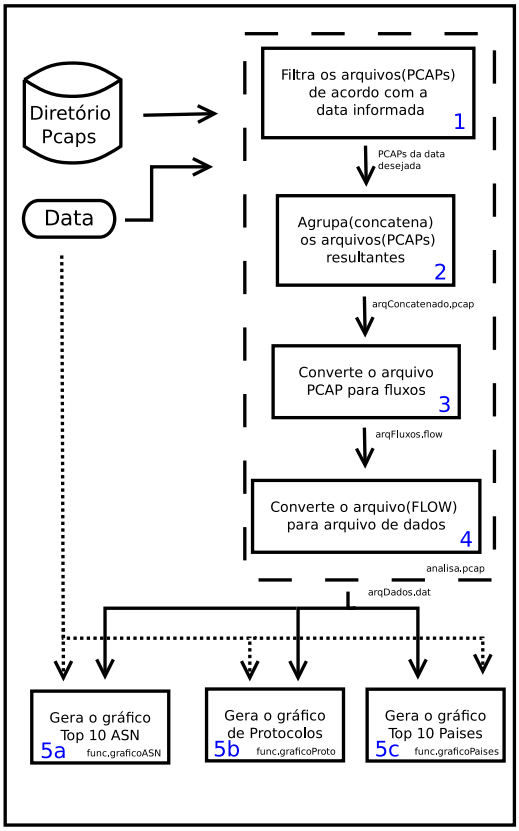


Figura 1: Fluxo de processamento da plataforma proposta.

AGRADECIMENTOS

Os autores agradecem o apoio da UDESC e da FAPESC para a realização desta pesquisa. Gustavo Neves da Silva foi bolsista PROBIC/UDESC.

REFERÊNCIAS

[1] Lance Spitzner. *Honeypots: Tracking Hackers*. Addison-Wesley Professional, 2002. ISBN 9780321108951,0321108957.

[2] Eric Alata, Vincent Nicomette, Mohamed Kaâniche, Marc Dacier, and Matthieu Herrb. Lessons learned from the deployment of a high-interaction honeypot. In *2006 Sixth European Dependable Computing Conference*, pages 39–46. IEEE, 2006.

[3] Cristine Hoepers, Klaus Steding-Jessen, Luiz ER Cordeiro, and Marcelo HPC Chaves. A national early warning capability based on a network of distributed honeypots. In *17th Annual FIRST Conference on Computer Security Incident Handling, Singapore*, pages 2–5, 2005.

[4] Eric Alata, Marc Dacier, Yves Deswarte, M Kaaâniche, Kostya Kortchinsky, Vincent Nicomette, Van-Hau Pham, and Fabien Pouget. Collection and analysis of attack data based on honeypots deployed on the internet. In *Quality of Protection*, pages 79–91. Springer, 2006.

[5] David Watson and Jamie Riden. The honeynet project: Data collection tools, infrastructure, archives and analysis. In *2008 WOMBAT Workshop on Information Security Threats Data Collection and Sharing*, pages 24–30. IEEE, 2008.

[6] Cristine Hoepers, Nandamudi L Vijaykumar, and Antonio Montes. Hidef: a data exchange format for information collected in honeypots and honeynets. *INFOCOMP*, 7(1):86–95, 2008.

[7] Peter Pisarčík and Pavol Sokol. Framework for distributed virtual honeynets. In *Proceedings of the 7th International Conference on Security of Information and Networks*, page 324. ACM, 2014.

[8] Marcin Nawrocki, Matthias Wählisch, Thomas C. Schmidt, Christian Keil, and Jochen Schönfelder. A survey on honeypot software and data analysis. *CoRR*, abs/1608.06249, 2016. URL http://arxiv.org/abs/1608.06249.

[9] Tiago Heinrich and Rafael R. Obelheiro. Brasil vs mundo: Uma análise comparativa de ataques DDoS por reflexão. In *XIX Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg)*, São Paulo, 2019.

[10] Cert. Recomendações para melhorar o cenário de ataques distribuídos de negação de serviço (ddos). URL https://www.cert.br/docs/whitepapers/ddos/.

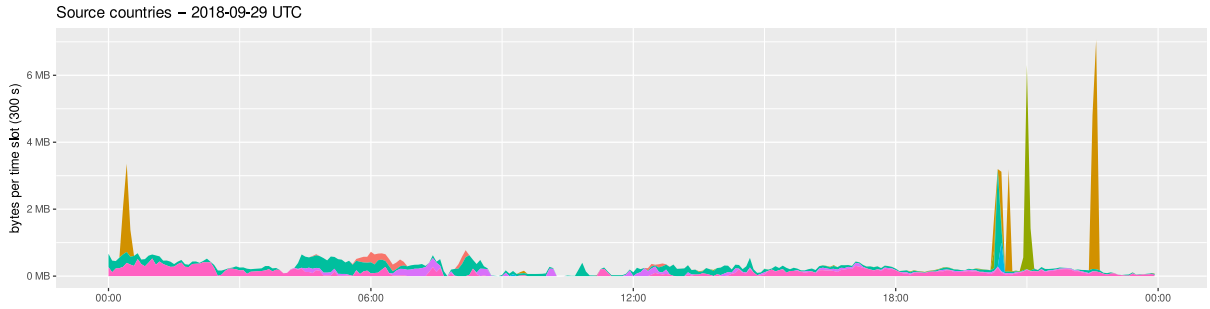
[11] The Tcpdump team. Tcpdump/libpcap public repository. URL http://www.tcpdump.org/.

[12] Wireshark Foundation. Wireshark. URL https://www.wireshark.org/.

[13] Richard Bejtlich. *The Practice of Network Security Monitoring*. No Starch Press, 2013.

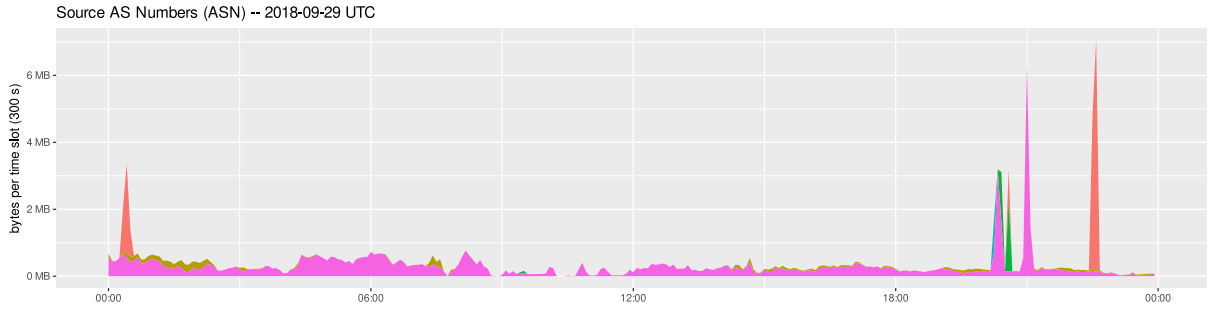
[14] Qosient. Argus - auditing network activiy. URL https://www.qosient.com/argus/.

[15] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.



Country_code	Country_code	Total	Max	Avg
US	United States	32.61 MB 29.89 %	1.72 KB/s	0.39 KB/s
CN	China	20.98 MB 19.23 %	22.30 KB/s	0.25 KB/s
GB	United Kingdom	9.18 MB 8.42 %	0.91 KB/s	0.11 KB/s
FR	France	8.56 MB 7.84 %	19.86 KB/s	0.10 KB/s
AU	Australia	3.50 MB 3.21 %	1.02 KB/s	0.04 KB/s
RU	Russia	0.88 MB 0.81 %	2.80 KB/s	0.01 KB/s
IE	Ireland	0.03 MB 0.02 %	0.04 KB/s	0.00 KB/s
TW	Taiwan	0.01 MB 0.01 %	0.03 KB/s	0.00 KB/s
-	Others	33.35 MB 30.57 %	9.41 KB/s	0.40 KB/s

Figura 2: Estatísticas diárias de tráfego por país.



ASNs	Name	Country_code	Total	Max	Avg
4134	No.31,Jin-rong Street	CN	16.51 MB 15.13 %	22.30 KB/s	0.20 KB/s
7922	Comcast Cable Communications, LLC	US	9.45 MB 8.66 %	0.88 KB/s	0.11 KB/s
4837	CHINA UNICOM China169 Backbone	CN	3.52 MB 3.23 %	6.63 KB/s	0.04 KB/s
133774	Fuzhou	CN	0.95 MB 0.87 %	3.23 KB/s	0.01 KB/s
8075	Microsoft Corporation	US	0.01 MB 0.01 %	0.05 KB/s	0.00 KB/s
0	Others	-	78.67 MB 72.10 %	20.35 KB/s	0.93 KB/s

Figura 3: Estatísticas diárias de tráfego por ASN.