PySoneta – Um sistema baseado em Python para visualização de dados do tráfego de rede

Jonathan J. Nefoussi[1], Bruno L. Dalmazo[2], Roben C. Lunardi[1]

[1]Instituto do Rio Grande do Sul (IFRS) – Campus Restinga
Rua 7121, Loteamento Industrial da Restinga, Lote 16, Quadra F, Número 285
CEP: 91791-508 – Porto Alegre – RS – Brasil

[2]CISUC, Department of Informatics Engineering
University of Coimbra, Coimbra, Portugal

{jjnefoussi, roben.lunardi}@restinga.ifrs.edu.br, dalmazo@dei.uc.pt

Abstract. *This paper proposes a prototypical system called PySoneta – currently under development. This system has the objective to allow network administrators to visualize the network packets traffic. The visualizations generated by the system are focused on IPv6 packets to help administrators to understand the network behavior. As preliminary results, it was possible to identify some misleading configurations on the network analyzed through the proposed system.*

Resumo. *Este trabalho propõe um protótipo do sistema PySoneta – atualmente em desenvolvimento. Este sistema tem como objetivo permitir que os administradores de rede possam visualizar o tráfego da rede, tendo como foco o protocolo IPv6 e com isso permitir o entendimento do seu comportamento. Através do sistema proposto obteve-se como resultado a detecção de falhas de configuração na rede do ambiente, utilizada como estudo de caso.*

1. Introdução

Com o fim da alocação da tabela brasileira de endereços IPv4 [NIC.BR, 2014], em junho de 2014, a necessidade da adoção de uma alternativa para endereçamento de *hosts* (dispositivos) na rede, o IPv6, tornou-se inevitável. Paralelamente com a adesão deste novo protocolo, tornou-se necessário estudar o seu comportamento. Entretanto, pouco tem se pesquisado sobre o impacto do uso do IPv6, especialmente quando os equipamentos da rede não são configurados corretamente, e como visualizar de maneira interativa o tráfego do IPv6 como, por exemplo, através de métodos estatísticos e métodos de visualização de dados.

Apesar de o IPv6 possuir os requisitos necessários que o fizeram ser adotado pela indústria e academia na busca da solução para alocação de endereçamento de hosts, ele ainda se encontra em implantação pelas instituições [Qiang *et al.*, 2012]. Desta forma, podem surgir problemas que ainda não foram mensurados devido ao desconhecimento do uso deste novo protocolo, principalmente pelas organizações que ainda estão na fase da implantação do protocolo. Um dos fatores para a causa deste problema é a falta de informação sobre os detalhes do funcionamento do IPv6 em infraestruturas reais. Em contrapartida, muito material é desenvolvido com o propósito de solucionar a dificuldade de implantação e melhorar as questões de mobilidade do IPv6 [Grossetete *et al.*, 2008]. No entanto, existem poucos materiais que exemplifiquem o

impacto que este novo protocolo pode provocar na rede. Além disso, nota-se que muito pouco tem sido feito em questão a ferramentas de apoio que suportem o uso híbrido de endereçamento dos hosts através do uso simultâneo dos protocolos IPv4 e IPv6 [Rafiee, 2012] (fase em que a infraestrutura se encontra durante a implantação).

Com o objetivo de auxiliar no entendimento do uso do IPv6, pretende-se com este trabalho, estudar o comportamento deste protocolo, realizando a coleta de dados, em um cenário real, que disponibiliza o uso do protocolo para os seus usuários finais. Desta forma, será possível realizar uma análise e caracterização dos dados coletados para que consequentemente seja possível proporcionar a criação de formas de visualização com o intuito de facilitar o entendimento do comportamento e impacto do IPv6 na rede.

Para cumprir estes objetivos, este trabalho propõe: (*i*) um protótipo do sistema que está em desenvolvimento que permite apresentar diversos métodos de visualização dos dados correspondentes ao comportamento do protocolo IPv6 na rede; (*ii*) utilização de técnicas de visualização como mecanismo de detecção de falhas na configuração na rede; (*iii*) aplicação do protótipo desenvolvido em cima de dados obtidos através da coleta de dados da rede do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) – Campus Restinga e (*iv*) análise dos dados obtidos através do protótipo do sistema desenvolvido.

O restante do trabalho é apresentado como segue: Na seção 2 são apresentados os trabalhos relacionados que foram utilizados como forma de análise e motivação para este trabalho. Na seção 3 é apresentada a solução conceitual da problematização abordada. Na seção 4 é descrita a implementação da solução proposta. Na seção 5 são apresentadas as formas de avaliação experimental que foram utilizadas e, por fim, na seção 6 são apresentadas as considerações finais.

2. Trabalhos Relacionados

Hui *et al.* [HUI *et al.*, 2011] propõem testes da eficácia do uso isolado do IPSec vinculado ao IPv6, para prevenir ataques DoS/DDoS. No trabalho é proposto que utilizando o IPSec é possível defender-se de ataques DoS/DDoS com endereço de origem falsificado. Também é apresentado que caso a fonte de origem utilize um endereço IPv6 não falsificado um grande número de conexões TCP em um estado semiconectado pode ser encontrado, declarando o IPSec ineficaz quando utilizado como fonte isolada de segurança. Porém, os testes feitos em um ambiente de laboratório, podem não garantir a eficácia do IPSec, quando um grande número de *hosts*, de um ambiente real, for utilizado com endereços falsificados para realizar um ataque DoS/DDoS [HUI *et al.*, 2011].

Barbosa *et al.* [BARBOSA *et al.*, 2010], utilizam o protocolo SNMP como elemento de estudo, baseado em uma metodologia que propõe como devem ser coletados e analisados os dados do tráfego deste protocolo. Ele confronta a falta de uma especificação de um processo de visualização de dados gerados com o uso de técnicas interativas e adaptadas, obtendo quatro protótipos de visualização de dados do protocolo SNMP. No entanto, o trabalho apresenta apenas o SNMP, deixando de explorar o comportamento dos demais protocolos, como por exemplo o IPv6.

HuaYu *et al.* [HUAYU *et al.*, 2010] apresentam um método para melhorar o uso de tabela *hash* como ferramenta de gestão de tráfego do fluxo de pacotes IPv6. O algoritmo proposto, tem uma melhora considerável no desempenho da identificação do fluxo do protocolo IPv6 e diminui a taxa de falsos positivos na tabela *hash*. Contudo, apesar de ter como base o estudo para identificação e análise estatística do protocolo

IPv6 na rede. O trabalho não propõe uma metodologia para a visualização dos dados estudados e coletados.

Como apresentado, ainda que o estudo do protocolo IPv6 esteja acontecendo em diversas linhas de pesquisa, o uso de métodos estatísticos para sua caracterização, juntamente com uma metodologia para visualização interativa dos dados, não foi devidamente investigada. Para tomar ações sobre possíveis impactos que o IPv6 pode causar na rede, as seções seguintes apresentam uma solução conceitual, além do protótipo do sistema desenvolvido para suportá-la e os resultados obtidos.

3. Solução Conceitual

Para auxiliar no entendimento do uso do IPv6, bem como comportamento deste protocolo na rede, foi desenvolvida uma solução para a visualização de dados de redes com foco no protocolo IPv6. Esta solução apresenta diferentes formas de visualização de dados, utilizando técnicas específicas para cada tipo de entrada e de maneira interativa com o usuário. Desta forma, pretende-se facilitar o entendimento do comportamento do protocolo IPv6 na rede, bem como a relação com os demais protocolos, ajudando gerentes de redes nas decisões a respeito de configuração e segurança dos equipamentos. A Figura 1, representação do PySoneta (acrônimo de *Python-based Solution for Network Analysis*), apresenta uma visão geral da solução, destacando os principais componentes.

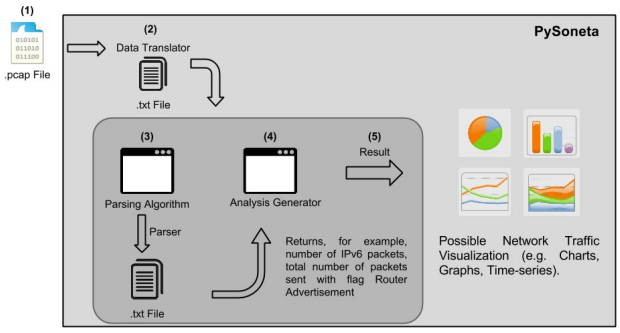

Figura 1: Representação da solução conceitual

Abaixo são descritos os principais componentes do PySoneta, conforme apresentado na Figura 1:

1. Um arquivo. pcap é utilizado como entrada padrão do PySoneta. Este arquivo pode obtido através sniffers de rede já existentes no mercado (Ex.: TCPDump). Uma consideração importante sobre a captura, é de que o host que executa o sniffer deve capturar os dados de toda a rede (promiscuous mode).
2. Depois o módulo denominado *Data Translator* recebe o arquivo. pcap previamente capturando (como foi citado no passo 1) com as informações do tráfego da rede e converte o arquivo .pcap para um arquivo de texto simples (.txt). Esse novo arquivo é então analisado pelo próximo módulo do PySoneta denominado de *Parsing Algorithm*.
3. O módulo *Parsing Algorithm* é composto por diferentes algoritmos. Esses algoritmos foram desenvolvidos baseados nos tipos de informações que o administrador de rede deseja visualizar.
4. Após realizar o processamento dos dados no módulo *Parsing Algorithm*, para cada algoritmo executado são geradas visualizações distintas através do módulo *Analysis Generator*. Como exemplos de visualização podemos citar: (*i*) a soma

da quantidade de pacotes – pode ser utilizado para gerar, por exemplo, um gráfico tipo pizza que permite comparar o uso dos protocolos utilizados na rede (ex.: IPv6 e IPv4); (*ii*) grafos, onde os nós do grafo representam os *hosts* da rede (dispositivos que possuem um endereço IP) e as arestas representam a comunicação entre esses nós através da quantidade que é enviada entre os dispositivos.

5. Os resultados obtidos podem ser tipos de visualizações, tais como gráficos e grafos (ex.: pizza, barras, fluxo de dados).

4. Implementação

A implementação da solução conceitual do sistema PySoneta foi desenvolvida através das seguintes tecnologias:

a. **TCPDump:** Uma poderosa ferramenta de linha de comando para realizar a captura dos dados da rede (*packet sniffer*). O TCPDump foi escolhido como *sniffer* por ser amplamente utilizado e tratar-se de uma ferramenta bem documentada. Além disso, é simples e requer baixo processamento para execução [TCPDump, 2014].
b. **Python:** Linguagem de programação multiparadigma. A linguagem Python foi escolhida por possuir uma sintaxe simples, portabilidade, alta produtividade, comunidade fortes e pela variedade de APIs disponíveis para o desenvolvimento dos métodos de visualizações propostos [Python, 2014].
c. **Matplotlib:** Uma biblioteca de plotagem 2D que produz figuras de alta qualidade, em uma variedade de formatos. A escolha desta biblioteca matplotlib se deu por ser uma API com grande popularidade e por possuir diversas funções implementadas que permitem gerar diversas visualizações de dados, possibilitando diferentes abordagens de estudo antes da versão atual do PySoneta [Matplotlib, 2014].

5. Avaliação experimental

O cenário utilizado para realizar a avaliação experimental do PySoneta foi a infraestrutura de rede do IFRS – Campus Restinga. Deve-se ressaltar que a rede desta instituição provê paralelamente o endereçamento IPv4 e IPv6 para todos os usuários. Devido às limitações de espaço, este artigo apresentará a visualização de uma das sub-redes da infraestrutura onde foi implantando o protótipo do sistema PySoneta. A sub-rede que será discutida é denominada “Rede Admin”, composta pelas configurações quem seguem na tabela 1.

Tabela 1: Configuração da sub-rede “Rede Admin”

Nº Hosts	~200 por dia
Tipo Conexão	*Wireless* (IEEE 802.11) / Cabeada (IEEE 802.3)
Tipo Acesso	Privado
Tipo Usuário	Servidores (funcionários)
Hosts	Heterogêneos (notebook, smarthphone, tablet, PC)

Os dados coletados, a serem apresentados, representam uma fatia de tempo, a qual representa o dia 12 de setembro de 2014 (captura de 24h). Como pode ser visto ao analisar a figura 2, a primeira visualização utilizada é através de um gráfico tipo pizza (a) que apresenta a quantidade de pacotes IPv6 em relação aos pacotes IPv4. Assim, analisando a "Rede Admin", pode-se observar que o IPv6 representa cerca de 27% da quantidade dos pacotes. Além disso, olhando para o gráfico de barras (b) da "Rede

Admin", observa-se cerca de 55 mil pacotes ICMPv6. Isto demonstra a ausência de alguma configuração nos equipamentos que envolvem a infraestrutura da rede. Esta observação pode ser confirmada, por exemplo, se ele teve uma grande quantidade dos pacotes IPv6 enviados para solicitar um roteador na rede (*Router Solicitation*).

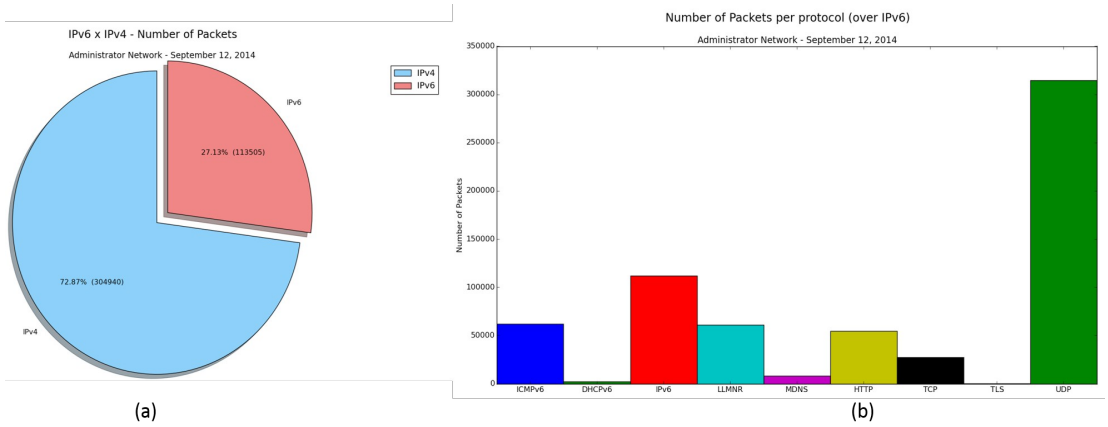

Figura 2: (a) Visualização gerada pelo PySoneta: Quantidade de pacotes IPv6 x IPv4 e (b) Quantidade de pacotes por protocolo (sobre o protocolo IPv6)

Além disso, a figura 3 apresenta a visualização da quantidade de pacotes enviados por cada rede host / IP (a) e o fluxo dos pacotes da rede (b). Através da visualização (a) é possível oferecer aos administradores de rede uma possibilidade de identificar qual host é responsável por gerar grandes quantidades de tráfego, visto que cada nó da visualização está dentro de um dos quatro grupos do filtro de divisão (*Small*, *Medium*, *Large*, *Extra Large*). Caso o equipamento não seja um dos nós centrais da infra-estrutura (servidores, routeadores, *firewalls*), ele pode ser facilmente identificado e configurado corretamente.

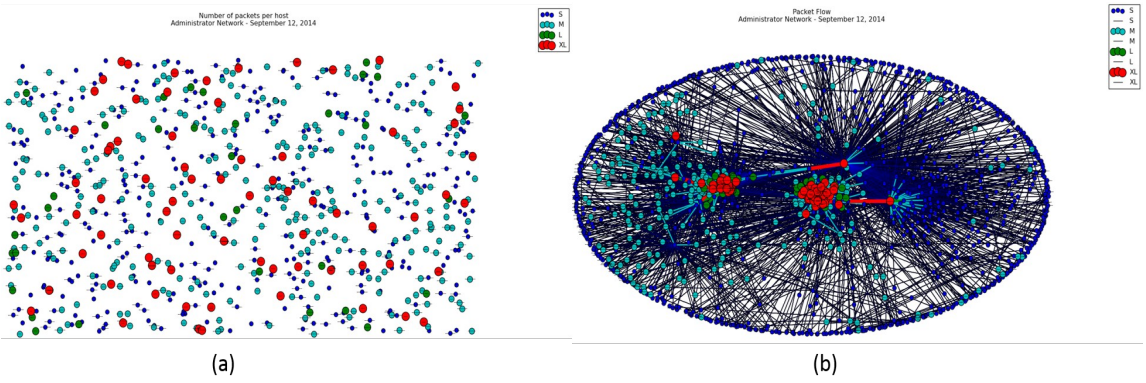

Figura 3: Visualização gerada pelo PySoneta:
(a) Quantidade de pacotes por host e (b) Fluxo de Pacotes

Por outro lado, quando utilizado a visualização do fluxo dos pacotes (b), é possível ajudar os administradores de rede a enteder a comunicação que ocorre entre todos hosts da rede (tanto a comunicação interna quanto externa). Na Figura 3(b) quando a transmissão de pacotes está ativa podemos observar uma alteração na cor do nó correspondente, além disso, ocorre um aumento do seu tamanho para facilitar sua visualização. O mesmo ocorre quando a comunicação entre dois hosts - na mesma rede ou não - ocorre, no entanto é a aresta correspondente a comunicação que realiza as alterações citadas (aumento de tamanho e troca de cor).

6. Considerações finais

Este trabalho propôs o sistema PySoneta, uma solução baseada em técnicas de visualização interativa que ajuda administradores de rede a monitorar o comportamento do tráfego IPv6 em um cenário real. A partir da observação dos resultados obtidos no estudo de caso, nota-se que nosso sistema oferece diversas opções de visualizações para analisar o tráfego da rede, auxiliando os administradores na tomada de decisão. Além disso, nossa solução demostrou-se escalável, uma vez que lidou de forma eficaz com um grande volume de tráfego de rede. Por fim, as visualizações geradas pelo sistema PySoneta mostraram-se de extrema utilidade, permitindo a detecção de falhas e má configuração em equipamentos da rede.

7. Referências

Barbosa, P. E. C. T., Uso de Técnicas de Visualização de Informação para o Estudo de Tráfegos de Gerenciamento de Redes, 2011.

Grossetete, P., Popoviciu, C.P., and Wettling F., “Global IPv6 Strategies: From Business Analysis to Operational Planning”, Cisco Press, 1st Edition. ISBN-13: 978-1587053436, 2008.

HUI, W., SUN, Y., LIU, J. and LU, K., “DDoS/DoS Attacks and Safety Analysis of IPv6 Campus Network”. International Conference on Internet Technology and Applications (iTAP), 2011, pp.1, 4, 16-18 Agosto 2011.

Matplotlib, 2014, User Guide – Introduction, http://matplotlib.org/users/intro.html/ (Outubro, 2014).

NIC.BR, 2014, IPv4 address exhaustion in Latin America, http://www.nic.br/imprensa/releases/2014/rl-2014-19.htm/ (Outubro, 2014).

Python, 2014, What is Python?, https://www.python.org/doc/essays/blurb/ (Outubro, 2014).

Qiang Li; Tao Qin; Xiaohong Guan; Qinghua Zheng, "Empirical analysis and comparison of IPv4-IPv6 traffic: A case study on the campus network," IEEE International Conference on Networks (ICON), 2012 18th, pp.395, 399, 12-14 Dezembro 2012.

Rafiee, H.; von Lowis, M.; Meinel, C., "IPv6 Deployment and Spam Challenges," IEEE Internet Computing, vol.16, no.6, pp.22, 29, Nov.-Dez. 2012.

TCPDump, 2014, http:// http://www.tcpdump.org/ (Outubro, 2014).