TRÁFEGO DE REDES SEM FIO: UMA CLASSIFICAÇÃO BASEADA EM INSPEÇÃO PROFUNDA DE PACOTES - DPI

TRAFFIC IN WIRELESS NETWORKS: A CLASSIFICATION BASED ON DEEP PACKET INSPECTION - DPI

Marcos Iran da Silva Waquim[1]
Harilton da Silva Araújo[2]

Resumo

A inspeção profunda de pacotes é um recurso usado para ajudar na manutenção de redes de computadores, monitorando parte do tráfego de dados de entrada e saída dos equipamentos conectados à rede. Dessa forma, é possível realizar uma filtragem dos dados classificando os tipos de serviços utilizados a fim de aumentar a gerência da rede. Com o avanço dos equipamentos e dispositivos móveis e sem fio cada vez mais portáteis, faz-se necessário o controle dos canais de comunicação e do fluxo entre tais ativos. Com esse avanço o numero de equipamentos conectados pode ser um fator que necessite de uma análise mais apurada dos fluxos de dados da rede. Este trabalho propõe uma análise de fluxos de dados a fim de realizar a classificação de tráfego nesse tipo de infraestrutura de comunicação.

Palavras-chave: rede sem fio, fluxo de dados, gerenciamento de rede.

Abstract

The deep packet inspection is a feature used to assist in the maintenance of computer networks, monitoring of the traffic data input and output devices connected to the network. Thus, it is possible to filter the data classifying the types of services used to increase the network management. With the advancement of mobile devices and wireless and increasingly mobile, it is necessary to control the channels of communication and flow between these assets. With this advance the number of connected equipment can be a factor that requires a more thorough analysis of the data streams of the network. This paper proposes an analysis of data streams in order to perform traffic classification in this type of communication infrastructure.

Keywords: wireless, flow, network management.

1 INTRODUÇÃO

[1] Bacharelando em Sistemas de Informação na Faculdade das Atividades Empresariais de Teresina - FAETE; Técnico Judiciário do Tribunal de Justiça do Estado do Piauí; E-mail: marcosiran@gmail.com

[2] Doutorando em Informática Aplicada – UNIFOR
Professor da Faculdade das Atividades Empresariais de Teresina – FAETE; E-mail: hariltonaraujo@faete.edu.br

O volume de dados e o número de dispositivos conectados nas redes são fatores que aumentam a complexidade do gerenciamento. O monitoramento é necessário para evitar mudanças bruscas no tráfego e identificar anomalias que podem vir a prejudicar o desempenho e a qualidade dos fluxos de dados da rede.

A Classificação de tráfego em uma rede através do fluxo caracteriza-se por coletar pacotes de rede de forma a gerar um banco de dados para análises de padrões, serviços mais solicitados, quantidade de *hosts* trocando pacotes e fluxo realizado entre todos os dispositivos detectáveis na rede. As tarefas de gestão de rede, tais como caracterização de carga de trabalho (*workload*), planejamento de capacidade, provisão de rotas, modelagem e policiamento de tráfego dependem da identificação e classificação do tráfego de rede. Os operadores de rede necessitam obter informações sobre o que está fluindo através das suas redes em tempo real. Tais informações possibilitam a tomada de decisão a fim de evitar problemas e alcançar os objetivos das regras de negócio. Portanto a precisa classificação de tráfego de rede é fundamental para diversas atividades relacionadas às redes, tais como: monitoramento de segurança, auditorias, provisão de qualidade de serviço e previsões de fornecimento a longo prazo.

A fim de obter a caracterização de tráfego, são descritas na literatura, três abordagens para classificação, são elas: baseada em portas, baseada em fluxos e inspeção profunda de pacotes (Deep Packet Inspection, DPI). A identificação de tráfego utiliza as portas da camada de transporte, não sendo esta uma técnica eficiente. Isso ocorre porque essa técnica depende de aplicações que façam uso de portas definidas pela IANA[3][2] - (tabela de classificação das portas dos serviços de rede). Para classificação baseada em fluxos, uma importante abordagem é a utilização da mineração de dados e aprendizagem de máquina. O processamento desses fluxos provoca o surgimento de outros pontos importantes no projeto de um classificador de tráfego, como a capacidade de armazenamento de fluxos de dados e a velocidade dessa classificação. Deep Packet Inspection (DPI) é uma tecnologia de vigilância de rede que permite os operadores analisar o tráfego da rede em tempo real e tomar decisões automatizadas sobre o que fazer, filtrando e inspecionando os pacotes coletados.

[3] http://www.iana.org/. Acessado em dezembro de 2012

2 TRABALHOS RELACIONADOS

Essa seção apresenta alguns trabalhos que investigam diferentes maneiras de classificação de tráfego em redes de computadores. Alguns desses trabalhos adotam as metodologias de análise em fluxo de dados utilizando algoritmos de redes neurais abordando a capacidade de armazenamento desses fluxos e a velocidade de classificação implementando a mineração nos bancos de dados como o algoritmo GPU based Streaming Decision Tree (GSDT).

Em Carela et al., (2011) foi abordada a classificação de tráfego com informações de fluxos IP (protocolo de internet) de roteadores e *switches* exportados com o protocolo *Netflow* (Cisco 2008) desenvolvido pelo Cisco. Neste trabalho foi aplicada a técnica de arvore de decisão C4.5. Foi feita a análise da classificação usando a teoria de amostragem (*sampling*), a qual usa uma porcentagem do total de amostras coletadas, onde variando a taxa de amostras foi possível mostrar que ela tem um grave impacto sobre o desempenho do método de classificação.

Petrônio e Fernandes (2012) propuseram um algoritmo voltado para classificação de tráfego em redes baseado na mineração de fluxos de dados, paralelizado através de uma GPU (Graphics Processing Unit). Para o processo de classificação realizado na GPU, foram efetuadas duas etapas. Na primeira etapa a árvore construída é copiada para a memória da unidade de processamento gráfico, a fim de ser utilizada durante toda a classificação. O segundo passo consiste em acumular um número configurável de registros de fluxos, que são transferidos para a GPU e classificados em paralelo. Neste caso, um grande número de registros é classificado de uma só vez, reduzindo o tempo de classificação.

Em Yulios et al., (2011) foi proposta uma abordagem que trata da classificação de tráfego baseado em multifractais, precisamente o conceito de Cascata Multiplicativa Binomial, onde cada nível da cascata é calculado uma variância. Todos esses valores de variâncias obtidos são armazenados em um vetor, denominado vetor de características. A técnica mostrou uma otimização na tarefa de classificar os registros dos tráfegos avaliados obtendo taxas de detecção acima de 90%.

Michael et al., (2012) apresentaram uma avaliação de desempenho de quatro classificadores de tráfego IP baseados em aprendizagem de máquina, são elas: Árvores de Decisão, Naive Bayes, Redes Bayseanas e Redes Neurais. As métricas avaliadas forma acurácia, qualidade na classificação e complexidade do ponto de vista da arquitetura de Serviços Diferenciados da Internet e no contexto de Qualidade de Serviço.

Os resultados obtidos mostraram que uma rede Neural não é adequada ao problema, pois precisa de tempo de treinamento longo. Porém todos os classificadores, após a conclusão do modelo de classificação, estão aptos para atuarem no modelo *DiffServ* em se tratando do tempo de processamento de fluxos para a classificação do tráfego.

Vilela (2006) apresentou uma classificação através dos fluxos de comunicação que abstrai-se analisando um conjunto de pacotes que possuem características em comum, tais como número da porta, endereço de origem e de destino, verificando três principais variáveis: tamanho, duração e taxa. O método proposto divide os fluxos em N classes, indicando seus comportamentos e impactos no tráfego da rede. Por essa divisão é possível conhecer melhor as características do tráfego da rede. Pode-se verificar o comportamento do tráfego que se aproxima da Distribuição de Pareto (Joseph Moses Juran, 1904), onde uma das principais características dessa distribuição é a existência de uma grande quantidade de fluxos de pequeno porte que é responsável por grande maioria do tráfego em bytes.

Este trabalho, diferentemente das pesquisas disponíveis na literatura, propõe a utilização da DPI para análise das redes, considerando a necessidade de precisão na classificação. A inspeção de pacotes, neste trabalho, surge como uma importante alternativa. Entretanto, a alta complexidade computacional e a dificuldade de inspecionar a carga útil do pacote em trânsito são problemas que podem inviabilizar o uso do DPI.

3 CLASSIFICAÇÃO DE TRÁFEGO USANDO DPI (*DEEP PACKET INSPECTION*)

O DPI (*Deep Packet Inspection*) monitora parte do tráfego de entrada e saída dos equipamentos conectados à rede. Dessa forma, é possível realizar uma filtragem desses dados ao identificar desvios de protocolo de rede, conteúdo que indique um ataque ou violação da política de segurança e, assim, encaminhar para um destino diferente ou armazenar *logs* para futura análise.

O DPI pode operar no modo detecção ou prevenção para proteger as redes ou sistemas. O detalhamento dos eventos fornece informações valiosas, incluindo o responsável pelo ataque, a data do evento e o alvo da tentativa de exploração. Os administradores podem ser notificados automaticamente por meio de alertas na ocorrência de um incidente.

Além de usar o DPI para proteger suas redes internas, os ISP (*Internet Service Provider*)

aplicam essa tecnologia nas redes de seus clientes. Os usos mais comuns do DPI pelos ISPs são interceptação legal de dados, definição e aplicação de políticas de acesso e qualidade, publicidade.

A tecnologia tem o potencial para fornecer aos *ISPs* e a outras organizações, amplo acesso a grandes quantidades de informações pessoais enviadas através da Internet para:

- Publicidade segmentada com base no comportamento dos usuários durante a navegação na Internet;
- Monitoramento do tráfego da rede de conteúdo indesejável ou ilegal, como: a distribuição não autorizada de material protegido ou disseminação de materiais proibidos ou obscenos;
- Captura e gravação de pacotes como parte da vigilância para a segurança nacional e outros fins de investigação de crimes, e;
- Monitoramento de tráfego para medir o desempenho da rede e planejar investimentos futuros para instalações.

Será discutido com maior ênfase o último ponto, com a utilização de ferramentas para capturar e armazenar os pacotes de rede a fim de realizar a inspeção mais detalhada do tráfego de uma rede. Os pacotes serão armazenados em banco de dados conforme a captura seja feita através de um dispositivo conectado a rede. Filtros de pacotes serão aplicados para detectar serviços e dados que passam por toda rede, tais como: *FTP* (File Transfer Protocol), *HTTP* (Hypertext Transfer Protocol), *p2p* (Peer to Peer), *VPN* (Virtual Private Network), etc. Com isso a inspeção dos pacotes será criteriosa para identificar, com base em médias, os recursos mais utilizados para gerar relatórios de desempenho e medição do tráfego para auxiliar um administrador de rede a tomar decisões planejando e inferindo configurações de forma a deixar a rede em pleno funcionamento, fator esse essencial em empresas que utilizam diariamente recursos e serviços em sua infraestrutura de tecnologia da informação.

O fluxo analisado pela DPI é filtrado e processado de forma a gerar informações relevantes aos tipos de necessidades encontradas. Conforme mostra a figura 1, o fluxo de pacotes é submetido às inspeções nos tipos de serviços utilizados na rede, onde são processados, a fim de buscar características, identificar padrões e até mesmo revelar conteúdos presentes no tráfego analisado.

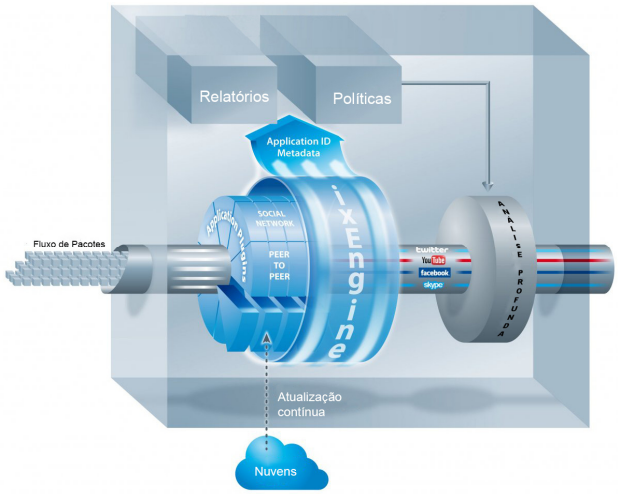

Figura 1. apresenta um exemplo da análise de um fluxo utilizando DPI

Fonte: http://www.qosmos.com/products/deep-packet-inspection-engine/

Os relatórios mostram os resultados obtidos ao longo das análises que servirão para tomada de decisões na criação de políticas de rede e definições de regras dentro do ambiente de infraestrutura de rede.

4 RESULTADOS

As pesquisas realizadas até o momento registraram que o uso de serviços da *WEB* são os mais utilizados, bem como os protocolos de comunicação entre os dispositivos, que são de extrema importância para o funcionamento da rede. A figura 2 mostra os serviços mais utilizados durante a coleta de dados feita em uma rede de uma determinada instituição do Estado do Piauí.

Os serviços mostrados na figura 2 são classificados por portas, onde cada porta é caracterizada de acordo com a tabela IANA. Na referida empresa os serviços e os sistemas são utilizados através da web utilizando criptografia *SSL* (Secure Sockets Layer).

O serviço mais solicitado refere-se ao serviço de firewall do sistema operacional Windows. Para esse cenário 99% dos computadores utilizam esse sistema.

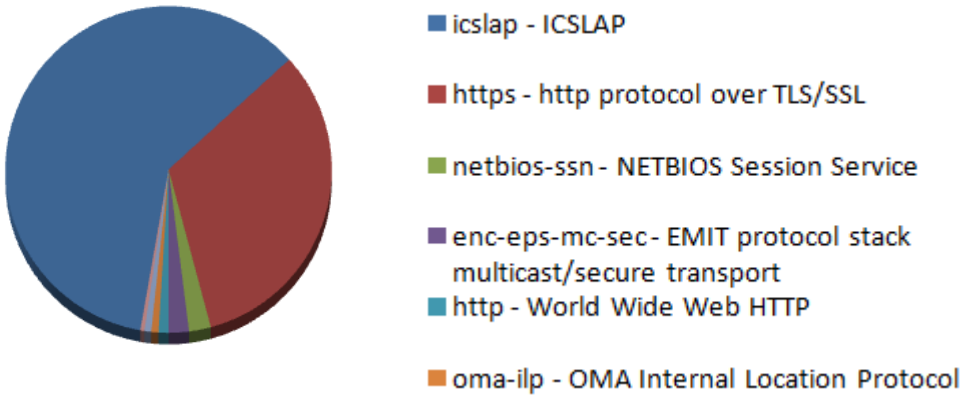

Figura 2 – Principais serviços mais requisitados na rede

A figura 3 classifica os *IPs* mais requisitados durante a coleta dos pacotes. Neste caso, é possível perceber que um host ultrapassa os demais hosts. Esse cenário mostra um serviço de descoberta de rede caracterizado por um protocolo baseado no IP para a propagação e descoberta de serviços de rede. Esse procedimento é feito sem a colaboração de mecanismos baseados no servidor de configuração, tais como: o Protocolo de Configuração Dinâmica (DHCP) ou o Servidor de nomes e Domínios (DNS). Para este caso, não foi utilizada a configuração estática especial de um host da rede.

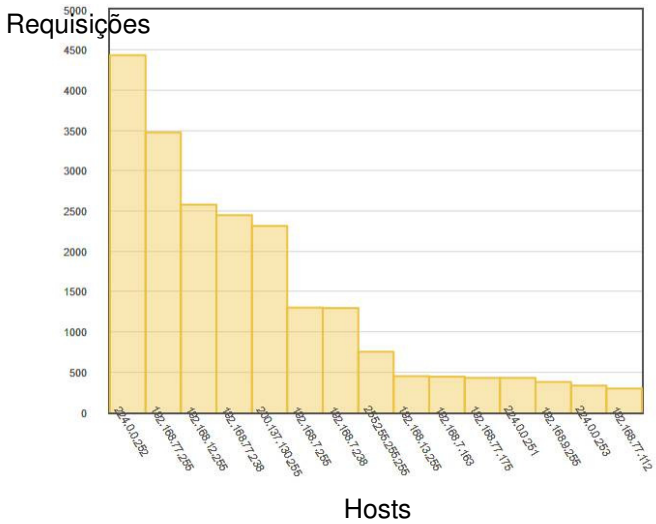

Figura 3 – Número de requisições por IP

A figura 4 mostra os maiores fluxos de tráfego identificados na classificação realizada. Tal classificação foi feita através do calculo das médias dos tamanhos dos pacotes enviados e recebidos em um determinado período do tempo realizado durante a coleta. Ainda na figura 4 é

possível observar que o IP 192.168.77.238 possui o maior valor, pois durante a execução do *sniffer*, esse host encontrava-se transferindo arquivos na rede.

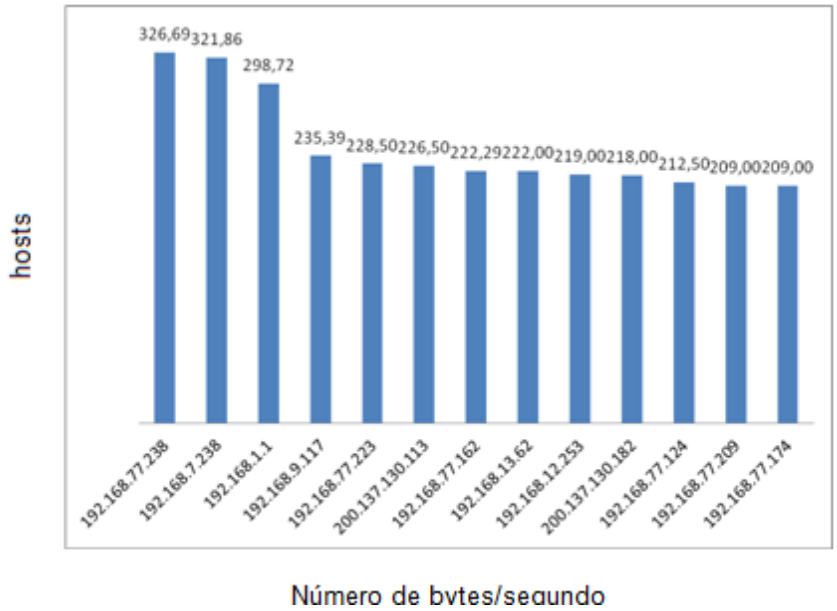

Figura 4 – Média dos tamanhos dos pacotes por host

Através de uma análise mais profunda nos atributos dos pacotes coletados, especificamente o tamanho (*length*) de cada um, foi feita uma média, em intervalos de 15 minutos a fim de obter uma média parcial da velocidade de tráfego na rede estudada, conforme mostra a figura 5.

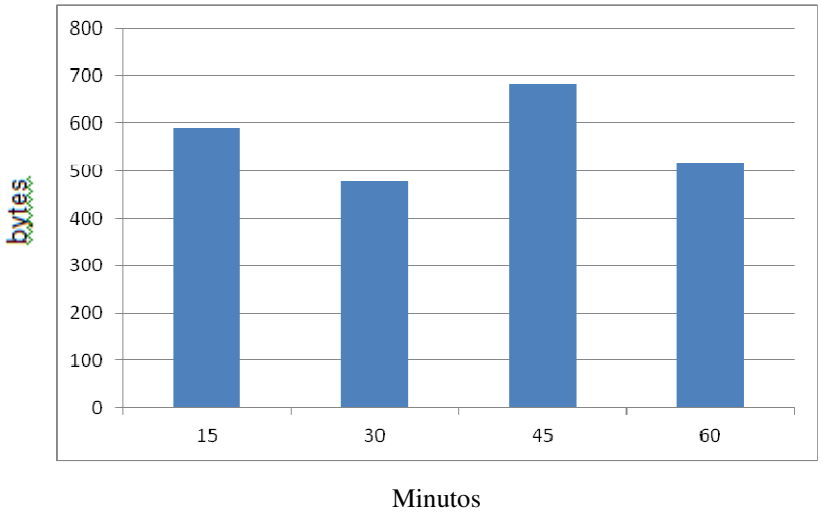

Figura 5 - Média de velocidade de Tráfego em bytes durante uma hora

A figura 5 mostra uma variância nos intervalos representada nos 45 minutos do experimento realizado. Essa característica é semelhante à apresentada na figura 4, onde um dos hosts da rede realiza transferência de arquivos.

As figuras 1, 2 e 5 mostram que foi possível detectar diversidades de serviços convergindo para os protocolos web. Essa característica apresenta um dos motivos mais críticos nas instituições atualmente. Nesta análise, foi possível observar uma variedade de informações e conteúdos que podem prejudicar o pleno funcionamento da infraestrutura de comunicação dos equipamentos de rede. Nos resultados apresentados foi possível observar que não foram identificadas anomalias na infraestrutura ou funcionamento inadequado devido o excesso de tráfego, indisponibilidade de recursos ou uso de *softwares spam*.

CONCLUSÃO E TRABALHOS FUTUROS

A classificação de tráfego é uma estratégia necessária para monitorar alterações no tráfego e identificar anomalias que podem vir a prejudicar o desempenho e a qualidade dos fluxos de dados da rede. A classificação de tráfego em uma rede utilizando o DPI (*Deep Packet Inspection*) é de grande importância para monitorar parte do tráfego de entrada e saída dos equipamentos conectados à rede. O uso dessa estratégia realiza uma filtragem de dados ao identificar desvios de protocolo de rede. Os resultados apresentados mostram que a inspeção de pacotes surge como uma importante alternativa, mostrando, principalmente, os atributos do pacote e a troca de fluxo entre os dispositivos. Como trabalhos futuros pretende-se propor um classificador de tráfego que utilize o DPI para identificar o fluxo de rede p2p.

REFERÊNCIAS

TANENBAUM, Andrew S.; WETHERALL, David. **Redes de Computadores**. 5ª Edição. São Paulo: Campus, Ano 2011.

KUROSE, James F. **Redes de computadores e a Internet: uma abordagem top-down**. São Paulo: Pearson Addison Wesley, 2006.

COMER, Douglas Earl. **Redes de computadores e internet**. 4ª Edição. Porto Alegre: Bookman, 2007.

VIDELA,Guilherme Silva. **Caracterização de tráfego utilizando classificação de fluxo de comunicação**. 88 f. Dissertação (Mestrado em Ciências em Engenharia de Sistemas e computação) — Coordenação de Pós-Graduação de Engenharia da Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

BARROS, Michael Taynnan, GOMES, Reinaldo Cézar de Morais. **Avaliação de Classificação de Tráfego IP baseado em Aprendizagem de Máquina Restrita à Arquitetura de Serviços Diferenciados**. Revista de tecnologia da informação e comunicação, Campina Grande/PB, v. 1, n. 2, p. 10-19, maio 2012.

Deep Packet Inspection (DPI): o que é e para que serve. <https://www.ibm.com/developerworks/community/blogs/fd26864d-cb41-49cf-b719-d89c6b072893/entry/deep_packet_inspection_dpi_o_que__C3_A9_e_para_que_serve?lang=en>. Acesso em 11 mar. 2013.

JPCAP - a network Packet Capture Library for Applications Written in Java. Disponível em: <http://jpcap.sourceforge.net/javadoc/index.html>. Acesso em: 05 dez. 2012.

Service Name and Transport Protocol Port Number Registry. Disponível em: <http://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xml>. Acesso em: 04 mar. 2013.

Cisco IOS Flexible NetFlow. Disponível em: <http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6555/ps6601/ps6965/product_data_sheet0900aecd804b590b.html>. Acesso em: 04 mar. 2013.

Apresentado em: 30.05.2013
Aprovado em: 28.06.2013