Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos e Clusterização K-means

Alexandro M. Zacaron[1], Luiz F. Carvalho[1], Mario H. A. C. Adaniya[1], Taufik Abrão[1], Mario Lemes Proença Jr.[1]

[1]Departamento de Ciência da Computação – Universidade Estadual de Londrina (UEL)
Caixa Postal 6.001 – 86.051-980 – Londrina – PR – Brasil

{zacaron, luizfcarvalhoo, mhadaniya}@gmail.com, {taufik, proenca}@uel.br

Abstract. *This paper presents a model of the Digital Signature of Network Segment Analysis Using Flow and K-means clustering (KM-DSNSF). We used the technique of K-means clustering to generate a profile or baseline of the network through bytes of NetFlow v9 flows, collected during the months of March and April 2012 in the Federal Technological University of Paraná - Campus Toledo, for TCP and UDP protocols, in order to identify the behavior of a given segment after a period of learning thereby establishing thresholds that are considered normal for each managed segment and compare it with the motion generated by NfSen to identify possible anomalies.*

Resumo. *Neste artigo é apresentado um modelo de Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos e clusterização K-means (DSNSF-KM). Foi utilizada a técnica de clusterização K-means para gerar um perfil da rede ou baseline sobre os bytes do fluxos NetFlow v9, coletados durante os meses de março e abril de 2012 na Universidade Tecnológica Federal do Paraná - Câmpus Toledo, para os protocolos TCP e UDP, com objetivo de identificar o comportamento de um determinado segmento após um período de aprendizado estabelecendo assim limiares que serão considerados normais para cada segmento gerenciado e compará-los com o movimento apresentado pelo NfSen visando identificar possíveis anomalias.*

1. Introdução

A caracterização do tráfego de segmento de rede é de vital importância para os administradores de rede, pois possibilita a identificação de comportamentos para cada horário, dia da semana e até mesmo para um serviço específico. Esta tarefa de caracterizar que neste trabalho é chamada por criar uma assinatura digital para o segmento gerenciado, é uma etapa importante e fundamental na detecção de anomalias [Proenca et al. 2006, Fatemipour and Yaghmaee 2007]. Para tanto a coleta e analise de fluxos do tipo IPFIX ou NetFlow se tornaram imprescindíveis nas atuais redes de banda larga [Chang et al. 2010, Muraleedharan et al. 2010].

A detecção de anomalias pode ser classificada com base em assinatura, com a qual o administrador tem um conhecimento prévio sobre o tipo de ataque ou anomalia; em perfis que caracterizam o comportamento normal da rede, nos quais se tem um histórico que representa o comportamento por meio da mineração de dados, de modelos estatísticos, dentre outras técnicas [Denning 1987, Patcha and Park 2007].

Dentre algumas maneiras de se realizar o processo de caracterização do tráfego, a clusterização tem surgido em diversos trabalhos como forma de agrupar dados semelhantes indicando um comportamento para uma determinada aplicação, segmento de rede, serviços, como também para contadores como bytes, pacotes e fluxos [Celenk et al. 2008, Molnar and Moczar 2011, Yingqiu et al. 2007, Singh et al. 2009, Rossi and Valenti 2010].

Este trabalho apresenta um modelo de Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos ou *Digital Signature of Network Segment Using Flow Analysis* (DSNSF). Um fluxo é definido como um conjunto de pacotes passando por um ponto de observação na rede, durante um certo intervalo, compartilhando um conjunto comum de propriedades. NetFlow [Claise 2004] and IP Flow Information eXport (IPFIX) [Claise 2008] são exemplos de protocolos para exportação de fluxos.

Na construção do modelo DSNSF-KM foi utilizada a técnica de clusterização por meio do método K-means [MacQueen 1967], aplicado sobre o fluxos coletados durante o mês de março e abril de 2012, na Universidade Tecnológica Federal do Paraná (UTFPR) - Câmpus Toledo, analisando-se o total de bits a cada 5 minutos para os protocolos Transmission Control Protocol (TCP) and User Datagram Protocol (UDP), comparando-os com o movimento gerado pelo NfSen [Haag 2005], um frond end web para a ferramenta NFDUMP [Haag 2004].

O restante deste artigo se divide da seguinte maneira: na seção 2 se encontram os Trabalhos relacionados; na seção 3, NetFlow e IPFIX; na seção 4, Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos (DSNSF); na seção 5, Análise do método proposto; e na seção 6, a Conclusão.

2. Trabalhos Relacionados

Celenk et al. [Celenk et al. 2008] comentam que a entropia tem sido usada para examinar a rede determinando seu status e detectando anomalias, porém com um tempo alto para essa tarefa. Baseado nisso, os autores propõem uma abordagem para reduzir o tempo de observação das características da rede e seu respectivo tempo médio de entropia aplicando a técnica Fisher Linear Discriminant (FLD). Esse processo visa identificar a hora exata do incidente de segurança com resultados precisos.

Molnar and Moczar [Molnar and Moczar 2011] propõem um framework para caracterização de tráfego para aplicações como P2P, jogos, redes sociais e reprodução de video. Usando clusterização, definem grupos que representam cada tipo de tráfego utilizando o que os autores chamam de caracterização em três dimensões compostas por size, duration and rate. Com esses três tipos de informações, é possível identificar quais aplicações trafegam na rede. Os autores identificam que no comportamento de uma rede social o tamanho dos pacotes varia em 1kB a 350kB, já para o YouTube fica de 320kB a 26MB.

Rossi et al. [Rossi and Valenti 2010] utilizam um algoritmo comportamental que explora os fluxos para classificação do tráfego na rede. Seu classificador é uma extensão do algoritmo de classificação comportamental Abacus e busca identificar uma aplicação utilizando dois campos dos registros de fluxos que são bytes e pacotes. Resultados indicam uma precisão de 90%, no pior caso, para o volume de tráfego.

Singh et al. [Singh et al. 2009], por meio da clusterização, procuram identificar anomalias em dados NetFlow. Apresentam uma abordagem baseada no K-means para analisar os fluxos usando campos como endereço IP, portas, protocolos, entre outros, para detectar anomalias. Os autores apresentam três resultados de seu trabalho: identificação de eventos anômalos na rede, visualização da rede baseada em alguns parâmetros chave e, por fim, visualização dos eventos da rede de uma forma intuitiva.

Yingqiu et al. [Yingqiu et al. 2007] faz a classificação do tráfego em diferentes níveis por máquina de aprendizado, análise de problemas por métodos port-based e payload-based. Seus métodos são avaliados em performance e eficiência pelo algoritmos K-means. Os resultados demonstram que o método pode obter 80% de precisão e 90%, ou mais, se aplicado posteriormente uma transformação log.

Proença [Proenca et al. 2006], através da coleta de objetos SNMP por meio da ferramenta GBA e do algoritmo BLGBA é calculada a moda estatística para determinar um valor para um dado segundo do dia, por meio da análise do mesmo segundo de semanas anteriores, gerando assim a Assinatura Digital de Segmento de Rede ou *Digital Signature of Network Segment* (DSNS). O autor utiliza dois tipos de DSNS, bl-7 que consiste em um para cada dia da semana e o bl-3 que corresponde aos dias úteis, sábados e domingos.

3. NetFlow e IPFIX

NetFlow foi desenvolvido pela Cisco [Claise 2004] como uma opção para a realização da medição/monitoração do tráfego da rede. As informações que esse protocolo exporta inicialmente foram conhecidas com *five-tuples* sendo elas: *sourceIPv4address, destinationIPv4address, sourceTransportPort, destinationTransportPort and protocolIdentifier*, tendo suporte também ao IPv6. O IETF desenvolveu um novo protocolo para exportação chamado IPFIX [Claise 2008], que foi baseado no NetFlow versão 9 com algumas melhorias, como por exemplo o controle de congestionamento e segurança.

Os requisitos do IPFIX foram definidos pela RFC 3917 [Quittek et al. 2004] com objetivo de satisfazer aplicações consideradas importantes hoje e/ou para o futuro das redes IP, sendo elas: *Accounting, Traffic Profiling, Traffic Engineering, Attack/Intrusion Detection and QoS Monitoring*.

Basicamente a terminologia usada para o processo desde a captura até a entrega dos fluxos é dada da seguinte maneira: *Observation Point*, que constitui um ponto na rede no qual os pacotes podem ser observados; *Metering Process*, que gera os registro de fluxos através dos cabeçalhos dos pacotes observados; *Flow Record*, o qual possui informações específicas de um fluxo medido; *Exporting Process*, que envia os registro de fluxos para os coletores; *Collecting Process*, que recebe os registros de fluxos do Exporting Process.

4. Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxo (DSNSF)

Ferramentas que permitem aos administradores de redes caracterizar o tráfego da rede são de vital importância. Elas possibilitam identificar comportamentos para um determinado horário, dia da semana ou até mesmo um serviço específico, tendo um importante papel na detecção de anomalias.

O modelo para criação da Assinatura Digital de Segmento de Rede Utilizando

Análise de Fluxos ou *Digital Signature of Network Segment Using Flow Analysis* (DSNSF) surgiu com esse objetivo, descrevendo um perfil básico do tráfego da rede que possa indicar um comportamento padrão.

O DSNSF constrói para cada dia da semana uma assinatura também conhecida como baseline, esta é baseada no histórico das últimas cinco semanas. Para gerar a assinatura de uma segunda-feira, por exemplo, o DSNSF lê os fluxos das cinco segundas anteriores a ela. Dos fluxos é feita a separação por protocolo, nesse caso dos protocolos TCP e UDP, extraídos os valores do campo bytes de cada um.

Em seguida, com o auxilio da clusterização, que é uma técnica de mineração de dados, pode-se procurar e quantificar dados semelhantes em determinados grupos. Este processo procura minimizar a distância entre os pontos de um determinado grupo e aumentar a distância entre grupos [Fu 2008]. A distância euclidiana (1) geralmente é usada para medir a similaridade entre os dados.

$$J(p) = \sum_{k=1}^{K} \sum_{s=1}^{S} \sqrt{|P_s^k - c^k|^2} \tag{1}$$

Onde K é o número de clusters, S é o número de pontos, P_s^k é o valor dos pontos pertencentes ao cluster k e c^k corresponde ao centro do cluster k. O propósito de usar clusterização é criar um modelo que possa extrair um padrão de informações. Sendo possível identificar dados que tem um comportamento padrão e dados que se distanciam do desse padrão.

4.1. K-means Clustering

K-means (KM) é o processo que divide uma população n-dimensional em K grupos baseado em uma amostra. KM particiona os pontos do vetor ou matrix de dados em k clusters, as linhas da matriz correspondem aos pontos e as colunas as variáveis. Este particionamento procura maximizar a soma das distâncias entre os clusters e diminuir a soma das distâncias dentro de cada cluster. KM sempre retorna um vetor contendo os índices do cluster para cada ponto [MacQueen 1967].

O algoritmo a seguir demonstra o pseudo código para a Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos por meio do KM (DSNSF-KM).

Os parâmetros usados na criação do DSNSF-KM para os protocolos TCP e UDP são diferentes, visto que o comportamento dos dois não é o mesmo dentro da rede analisada. No protocolo TCP os parâmetros usados foram K=4, número de réplicas = 20, número mínimo de pontos pertencentes a um cluster γ=5. Para o protocolo UDP os parâmetros usados foram K=4, número de réplicas = 20, número mínimo de pontos pertencentes a um cluster γ=100.

Como pode ser observado no algoritmo DSNSF-KM, os clusters mais representativos, ou seja, que possuem mais pontos, são usados para compor o baseline. Já os clusters menos representativos não são utilizados na composição. Esse procedimento garante que o baseline não seja formado por pontos que desviam do comportamento normal do segmento (outliers), evitando assim que possíveis anomalias possam influenciar na construção do baseline.

DSNSF-KM algoritmo usado para clusterização.

Entrada: Pontos que representam os bits dos dias anteriores dentro do intervalo, K Número de clusters.
Output: μ: Valor que representa o conjunto de bits no intervalo.
Passo 1 Colocar K pontos no espaço que representa os pontos a serem clusterizados. Estes pontos representam o conjunto de dados inicial de cada centroide.
Passo 2 Atribuir cada ponto para o grupo mais próximo do centroide.
Passo 3 Quando todos os pontos tiverem sido alocados, é recalculada a posição os K centroides.
Passo 4 Repita os passos 2 e 3 até que centroide não se mova mais ou o número de iterações seja excedido.
Se $\gamma <$ **número de pontos do cluster** K
μ = média ponderada entre os clusters mais representativos excluindo-se o que possui pontos inferiores a γ
Fim Se
Retorna μ

A definição da quantidade mínima de pontos para que um cluster faça parte do baseline é dada por γ. Se o cluster não possui a quantidade mínima ele é descartado e não será utilizado no cálculo da média ponderada. O resultado do algoritmo DSNSF-KM para cada instante é a média ponderada (2) dos clusters mais representativos.

$$\mu = \frac{\sum_{j=1}^{K} c_j . p_j}{S} \tag{2}$$

Onde c_j é o centro do cluster j, p_j é o número de pontos pertencentes ao cluster.

5. Análise do Método Proposto

Com o objetivo de avaliar a Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos por meio do KM (DSNSF-KM), foram coletados fluxos durante o mês de março e abril de 2012, na Universidade Tecnológica Federal do Paraná (UTFPR) - Câmpus Toledo, no Gateway principal, conforme pode ser observado na figura (1).

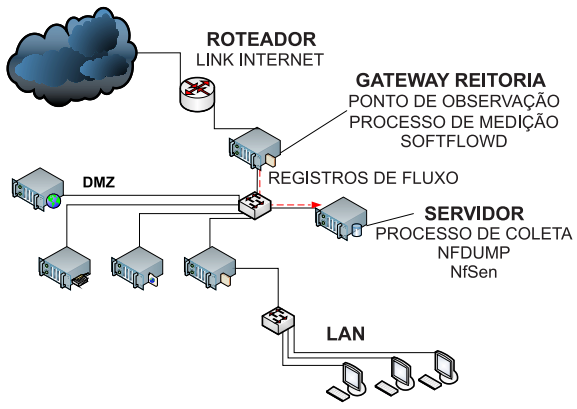

Figura 1. Rede onde os dados foram coletados

A figura (1) identifica o cenário onde foi realizada a coleta dos fluxos, exportados 1:1, ou seja, não foi utilizada nenhuma técnica de amostragem, todos os fluxos foram exportados pelo aplicativo Softflowd [Miller 2010] instalado no gateway. Os fluxos são salvos em arquivos de cinco em cinco minutos para que possam ser analisados posteriormente. O Softflowd é um analisador de rede capaz de exportar dados segundo o padrão NetFlow, através do monitoramento de uma interface de rede ou pela leitura de um arquivo. O Softflowd suporta as versões 1, 5 e 9 do NetFlow.

Os fluxos foram exportados na versão 9 do NetFlow a um servidor CentOS 5.5 rodando o aplicativo NFDUMP, o qual se trata de uma ferramenta para coleta e processamento de dados NetFlow. Juntamente com NFDUMP foi instalado o NfSen, que é um front end para o mesmo, com objetivo de facilitar a visualização, busca, geração de alertas e processamento dos fluxos coletados.

As figuras (2 e 3) a seguir representam a aplicação do DSNSF-KM sobre o movimento gerado pelo NfSen durante o período de 09 a 13 de abril de 2012 para os protocolos TCP e UDP. Como é observado, o modelo DSNSF-KM permite descrever o comportamento da rede. Pode-se verificar que o comportamento do protocolo UDP é diferente do comportamento do TCP, tendo uma menor variação no eixo y que se refere ao volume de tráfego.

Alterações no comportamento do movimento perante os modelos DSNSF são motivo de estudo e servem como motivação a fim de descobrir o que ocasiona essas mudanças, sendo algumas opções como: aumento do número de usuários, alguma atualização de software, um ataque, algum problema nos ativos de rede que pode estar causando por exemplo retransmissões, etc. Com um bom modelo o administrador pode perceber rapidamente que sua rede mudou o comportamento e investigar as possíveis causas.

Para o protocolo TCP na segunda-feira, o movimento apresenta um comportamento bem superior ao DSNSF-KM, porém pode-se observar que é um movimento uniforme, pricipalmente no período entre 8h e 23h.

Para uma análise mais detalhada do modelo proposto, realizamos o cálculo da correlação para indicar como o modelo esta relacionado com o movimento de cada dia tanto para o protocolo TCP quanto para UDP. Os resultados são apresentados na tabela (1).

Tabela 1. Correlação entre DSNSF-KM e o movimento do NfSen para os protocolos TCP e UDP.

protocolo TCP					
	Mon	Tue	Wed	Thu	Fri
DSNSF-KM	0,8266	0,7203	0,8255	0,8047	0,6561
Protocol UDP					
	Mon	Tue	Wed	Thu	Fri
DSNSF-KM	-0,0327	0,2562	0,0231	0,2706	0,7930

Para correlação, se o valor apresentado for próximo de 1 significa um excelente resultado, quer dizer que se o movimento subir o DSNSF-KM também terá que subir na mesma proporção, e vice-versa. Caso o valor seja próximo de 0, significa que existe

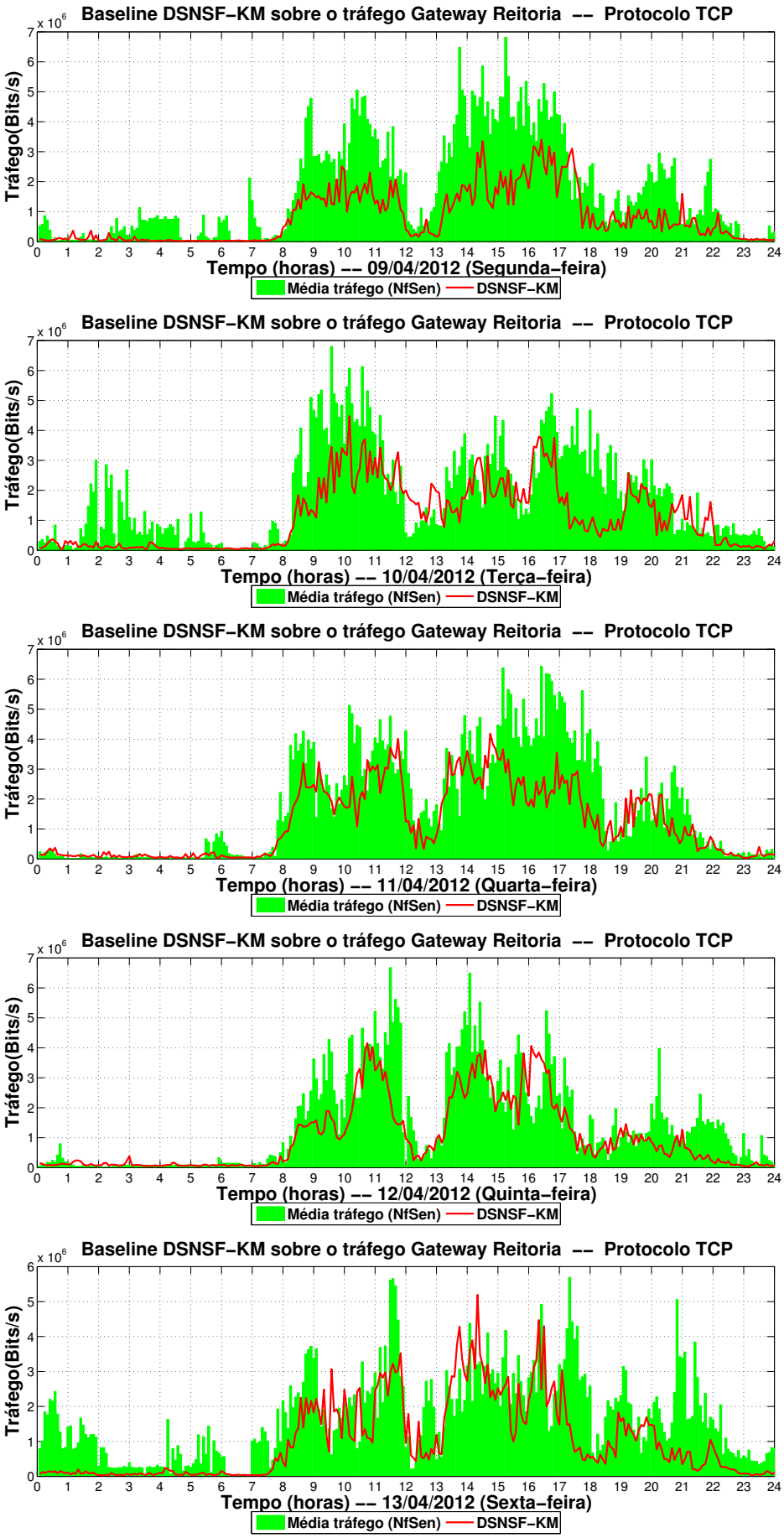

Figura 2. DSNSF-KM sobre o movimento registrado pelo NfSen para o protocolo TCP.

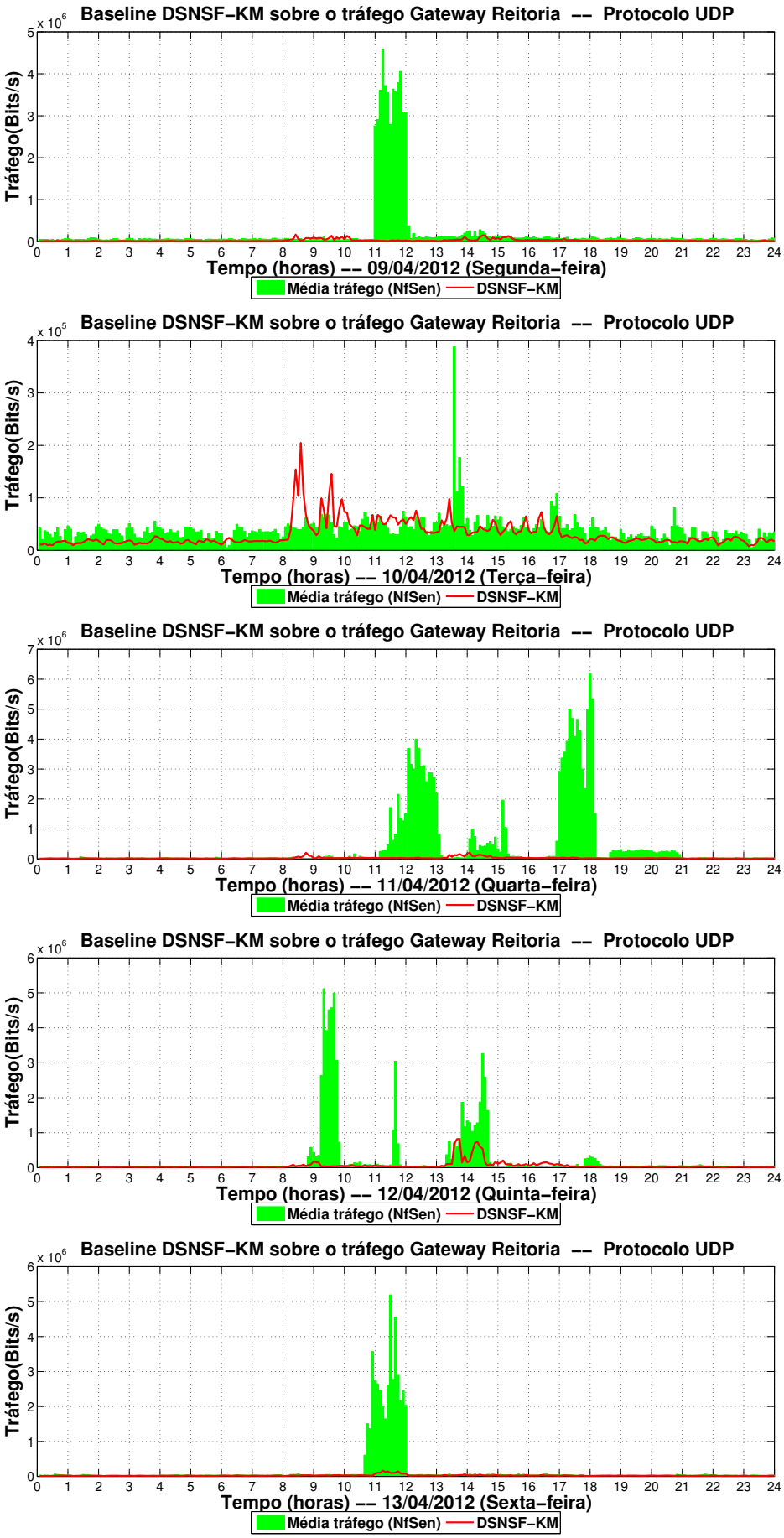

Figura 3. DSNSF-KM sobre o movimento registrado pelo NfSen para o protocolo UDP.

uma pequena correção ou seja o movimento não é mais proporcional. Caso o valor seja próximo de -1 indica que o DSNSF-KM está descorrelacionado do movimento, ou seja se o movimento subir, o DSNSF-KM descerá e vice-versa.

Na avaliação do DSNSF-KM se tratando da correlação para o protocolo TCP os resultados foram bons ficando com o valor acima de 0,65 para correlação. Indicando que o modelo consegue perceber a tendência do movimento para este protocolo.

Para o protocolo UDP houve uma queda no nível de correlação do modelo em relação ao TCP, o DSNSF-KM ficou o melhor valor 0,79, porém apresentou para segunda-feira um valor negativo de -0,0327, indicando uma pequena descorrelação. Considera-se que o protocolo UDP possui um comportamento diferente do TCP, sendo necessário avaliar os parâmetros dos modelos afim de melhorar os resultados.

Para definir a proximidade do modelo apresentado com o movimento, foi utilizado o erro quadrático médio normalizado (NMSE), para os protocolos TCP e UDP, os resultados são apresentados na tabela (2). Para esse quesito, pode-se interpretar que quanto menor o valor obtido mais próximo o modelo estará do movimento.

Tabela 2. Erro quadrático médio normalizado entre DSNSF-KM e o movimento do NFSEN para os protocolos TCP e UDP.

Protocol TCP					
	Mon	Tue	Wed	Thu	Fri
DSNSF-KM	0,8242	0,9580	1,6050	1,5396	0,7401
Protocolo UDP					
	Mon	Tue	Wed	Thu	Fri
DSNSF-KM	0,6214	0,3668	2,1514	4,2354	0,6567

Como é observado na tabela (2), o protocolo TCP apresentou para sexta-feira melhor resultado com 0,7401. Para o protocolo UDP, a terça-feira foi melhor com 0,3668. Observa-se que para o protocolo UDP os valores dos modelos ficaram bem mais próximos que para o protocolo TCP, caracterizando um melhor ajuste nesse quesito para o UDP.

Com objetivo de analisar o comportamento do DSNSF-KM frente ao movimento e também as anomalias que possam existir realizamos alguns experimentos controlados que podem ser observados na figura (2), tendo o conhecimento que eles influenciariam nos resultados para correlação e NMSE.

Como pode ser observado na figura (2) o gráfico que corresponde a segunda-feira dia 09/04/2012 há uma diferença entre o movimento e o DSNSF-KM no período em torno das sete horas da manhã, onde foi realizado, por meio da ferramenta LOIC [Technologies 2006], um DoS (Denial of Service) entre 6h45min e 7h05min totalizando, 9494 fluxos, 159803 pacotes, e 7,2MB de tráfego gerado.

No gráfico que corresponde a terça-feira 10/04/2012, na figura (2), durante o período entre 1h30min e 3h, onde foi realizado, por meio da ferramenta LOIC, um DoS (Denial of Service) em três etapas: primeiro entre 1h32min e 1h52min, segundo entre 2h02min e 2h20min, e por fim entre 2h30min e 2h53min totalizando, respectivamente: 9514, 7735, 11392 fluxos, 250301, 185895, 207400 pacotes, e 12,8MB, 8,3MB, 9,4MB de tráfego gerado.

No gráfico que corresponde a sexta-feira 13/04/2012, na figura (2), durante o período entre 0h e 8h, onde foi realizado, por meio da ferramenta LOIC, um DoS (Denial of Service) em três etapas: primeiro entre 7min e 39min, sengundo entre 1h18min e 1h49min, e por fim entre 6h51min e 7h21min totalizando, respectivamente: 2625, 1003, 617 fluxos, 590087, 467602, 606220 pacotes, e 211,2MB, 231,5MB, 217,2MB de tráfego gerado.

Durante os experimentos de criação do DSNSF-KM para o protocolo UDP, observou-se que ocorreram picos que se diferenciavam enormemente do comportamento apresentado. Ao aplicar o DSNSF-KM sobre o movimento referente ao protocolo UDP para todos os dias úteis da semana, identificaram-se esses picos de horários e duração variada.

Em paralelo, procurou-se identificar por meio das informações disponíveis nestes fluxos, sua origem, destino e motivo destes *outliers*. Após analise foi encontrado um único host que estava executando um cliente torrent que solicitava diversas conexões com destino a porta 8080 através protocolo UDP.

6. Conclusão

O modelo apresentado neste trabalho para construção de Assinatura Digital de Segmento de Rede, utilizando Análise de Fluxos (DSNSF-KM) apresentou bons resultados, possibilitando que o comportamento do segmento analisado seja descrito de forma automática e com isto permitindo que seja automatizada a tarefa de monitoramento dos segmentos por parte do administrador na medida em que se utiliza o DSNSF-KM como limiar base para anomalias ocorridas.

Conforme foi apresentado na figura (2) os ataques DoS gerados por meio da ferramenta LOIC podem ser facilmente identificados pois destoam do modelo DSNSF-KM. Para um sistema de detecção de anomalias eficiente a caracterização do tráfego é um passo fundamental que deve ser realizada com o objetivo de se conhecer o padrão e estabelecer limiares que serão considerados normais para cada segmento gerenciado.

Como trabalho futuro, o modelo deve ser aprimorado no sentido de aumentar a variação do número de semanas analisadas objetivando aproximá-lo ainda mais do movimento, assim permitirá identificar pequenas variações no tráfego com mais precisão. Também realizar a combinação de mais elementos dos fluxos como pacotes, número de fluxos, além de bytes.

Agradecimento

Este trabalho tem suporte financeiro apoiado pela SETI/Fundação Araucária e MCT/CNPq para o Projeto Rigel. Agradecemos também à Universidade Tecnológica Federal do Paraná - Câmpus Toledo.

Referências

Celenk, M., Conley, T., Willis, J., and Graham, J. (2008). Anomaly detection and visualization using fisher discriminant clustering of network entropy. In *Digital Information Management, 2008. ICDIM 2008. Third International Conference on*, pages 216 –220.

Chang, S., Qiu, X., Gao, Z., Liu, K., and Qi, F. (2010). A flow-based anomaly detection method using sketch and combinations of traffic features. In *Network and Service Management (CNSM), 2010 International Conference on*, pages 302 –305.

Claise, B. (2004). Cisco Systems NetFlow Services Export Version 9. RFC 3954 (Informational).

Claise, B. (2008). Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information. RFC 5101 (Proposed Standard).

Denning, D. (1987). An intrusion-detection model. *Software Engineering, IEEE Transactions on*, SE-13(2):222 – 232.

Fatemipour, F. and Yaghmaee, M. (2007). Design and implementation of a monitoring system based on ipfix protocol. In *Telecommunications, 2007. AICT 2007. The Third Advanced International Conference on*, page 22.

Fu, H. (2008). A novel clustering algorithm with ant colony optimization. In *Computational Intelligence and Industrial Application, 2008. PACIIA '08. Pacific-Asia Workshop on*, volume 2, pages 66 –69.

Haag, P. (2004). NFDUMP - NetFlow processing tools.

Haag, P. (2005). NetFlow visualisation and investigation tool.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In Cam, L. M. L. and Neyman, J., editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press.

Miller, D. (2010). Softflowd - traffic flow monitoring. [Online; accessed 28-May-2011].

Molnar, S. and Moczar, Z. (2011). Three-dimensional characterization of internet flows. In *Communications (ICC), 2011 IEEE International Conference on*, pages 1 –6.

Muraleedharan, N., Parmar, A., and Kumar, M. (2010). A flow based anomaly detection system using chi-square technique. In *Advance Computing Conference (IACC), 2010 IEEE 2nd International*, pages 285 –289.

Patcha, A. and Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, 51(12):3448 – 3470.

Proenca, M., Coppelmans, C., Bottoli, M., and Souza Mendes, L. (2006). Baseline to help with network management. In *e-Business and Telecommunication Networks*, pages 158–166. Springer Netherlands.

Quittek, J., Zseby, T., Claise, B., and Zander, S. (2004). Requirements for IP Flow Information Export (IPFIX). RFC 3917 (Informational).

Rossi, D. and Valenti, S. (2010). Fine-grained traffic classification with netflow data. In *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*, IWCMC '10, pages 479–483, New York, NY, USA. ACM.

Singh, M., Subramanian, N., and Rajamenakshi (2009). Visualization of flow data based on clustering technique for identifying network anomalies. In *Industrial Electronics Applications, 2009. ISIEA 2009. IEEE Symposium on*, volume 2, pages 973 –978.

Technologies, P. (2006). Low orbit ion cannon. [Online; accessed 20-Jun-2011].

Yingqiu, L., Wei, L., and Yunchun, L. (2007). Network traffic classification using k-means clustering. In *Computer and Computational Sciences, 2007. IMSCCS 2007. Second International Multi-Symposiums on*, pages 360 –365.