UNIVERSIDADE ESTADUAL DE LONDRINA

ALEXANDRO MARCELO ZACARON

ASSINATURA DIGITAL DE SEGMENTO DE REDE UTILIZANDO ANÁLISE DE FLUXOS E CLUSTERIZAÇÃO K-MEANS

Londrina
2013

ALEXANDRO MARCELO ZACARON

ASSINATURA DIGITAL DE SEGMENTO DE REDE UTILIZANDO ANÁLISE DE FLUXOS E CLUSTERIZAÇÃO K-MEANS

Dissertação apresentada à Universidade Estadual de Londrina como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Mario Lemes Proença Jr.

Londrina
2013

Catalogação elaborada pela Divisão de Processos Técnicos da Biblioteca Central da Universidade Estadual de Londrina.

Dados Internacionais de Catalogação-na-Publicação (CIP)

Z13a Zacaron, Alexandro Marcelo.
Assinatura digital de segmento de rede utilizando análise de fluxos e clusterização K-means / Alexandro Marcelo Zacaron. – Londrina, 2013.
109 f. : il.

Orientador: Mário Lemes Proença Júnior.
Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Ciência da Computação, 2013.
Inclui bibliografia.

1. Redes de computadores. 2. Telecomunicações – Tráfego. 3. Redes de computadores – Protocolos. I. Proença Júnior, Mário Lemes. II. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação. III. Título.

CDU 519.68.04

ALEXANDRO MARCELO ZACARON

ASSINATURA DIGITAL DE SEGMENTO DE REDE UTILIZANDO ANÁLISE DE FLUXOS E CLUSTERIZAÇÃO K-MEANS

Dissertação apresentada à Universidade Estadual de Londrina como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

BANCA EXAMINADORA

Prof. orientador Dr. Mario Lemes Proença Jr
UEL – Londrina – PR

Prof. Dr. Lourival Aparecido de Góis
UTFPR – Ponta Grossa – PR

Prof. Dr. Bruno Bogaz Zarpelão
UEL – Londrina – PR

Prof. Dr. Sylvio Barbon Júnior
UEL Londrina – PR

Londrina, 18 de março de 2013.

A Deus pela oportunidade e à família pelo apoio

AGRADECIMENTOS

À Deus, por tudo.

Ao meu orientador, Professor Dr. Mario Lemes Proença Jr., pela orientação, pelos conselhos, pela oportunidade e pela motivação que se iniciou em 2006 ainda no curso de Pós-Graduação atuando como um grande incentivador.

Ao Professor Dr. Taufik Abrão, que, além de um grande professor, foi mais um incentivador nesse período.

Ao Professor Dr. Rodolfo Miranda de Barros, à Professora Dra. Jandira Guenka Palma, pela contribuição no meu aprendizado.

Aos funcionários do Departamento de Computação.

Ao Moisés pelos n “galhos” quebrados, principalmente pelas inscrições feitas como aluno especial...

Aos colegas de Processos Estocásticos: Fábio Engel, Álvaro Souza, Gilson e ao Héber. Grande fase essa!

Aos colegas Everton, Ricardo, Tiago, Marc.

Aos amigos e colegas do grupo Rigel: Gilberto, Marcos, Eduardo, principalmente ao Luiz F., Carvalho e Mario H. A. C. Adaniya (Pajeh), que me ajudaram muito nessa fase de minha vida.

Aos servidores da UTFPR Câmpus Toledo: Frigori, Daniela, Araceli, Simone, Gerson, ao diretor Beto e principalmente aos colegas da COGETI: Vin´ıcius, Dalila e Sidgley que “seguraram as pontas” em todo esse período.

Aos meus pais e familiares e principalmente à Dani, que me suportou durante todo esse período.

“Embora ninguém possa voltar atrás e fazer um novo começo
qualquer um pode começar agora e fazer um novo fim.”

(Chico Xavier)

ZACARON, Alexandro M. **Assinatura digital de segmento de rede utilizando análise de fluxos e clusterização K-means**. 2013. 109f. Dissertação (Mestrado) – Universidade Estadual de Londrina, Londrina. 2013.

RESUMO

A caracterização do tráfego de rede é de grande 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. Este trabalho propõe um método para construção de uma assinatura utilizando análise de fluxos do tipo IPFIX, sFlow ou NetFlow a fim de caracterizar o tráfego de um determinado segmento de rede analisando-se o histórico dos fluxos por meio da clusterização. A tarefa de caracterizar, que neste trabalho é nominada por criar uma Assinatura Digital de Segmento de Rede utilizando Análise de Fluxos e Clusterização K-means ou *Digital Signature of Network Segment using Flow analysis and K-means Clustering* (DSNSF-KM), pode ser utilizada como limiar base na detecção de anomalias auxiliando aos administradores na gerência de redes. Para avaliar o método proposto foram coletados fluxos IP na Universidade Tecnológica Federal do Paraná - Câmpus Toledo e na Universidade Estatual de Londrina. Os fluxos coletados foram utilizados na criação de cinco cenários diferentes com objetivo de avaliar as assinaturas geradas pelo DSNSF-KM frente cada cenário. Os resultados alcançados demonstram que o DSNSF-KM possibilita caracterizar o tráfego de um segmento de rede e pode auxiliar os administradores na detecção de problemas.

Palavras-chave: DSNSF. DSNSF-KM. Caracterização de tráfego. Fluxos IP.

ZACARON, Alexandro M. **Digital signature of network segment using flow analysis and K-means clustering**. 2013. 109f. Dissertation (Master) – State University of Londrina, Londrina. 2013.

ABSTRACT

The characterization of the traffic network segment is vitally important for network administrators, as it enables the identification of behaviours for each time, day of week or even for a specific service. This work proposes a method for building a signature using flow analysis of type IPFIX, sFlow or NetFlow, in order to characterize the traffic from a specified network segment analyzing the historical flows through clustering. This task of characterizing which is in this work is called for creating a Digital Signature of Network Segment using Flow analysis and K-means Clustering (DSNSF-KM), can be use as a threshold based on anomaly detection helping administrators in network management. To evaluate the proposed method flows were collected at the Federal University of Technology Paraná - Toledo Campus and at the State University of Londrina. The collected flows were used to create five different scenarios to evaluate the signatures generated by DSNSF-KM front each. Results show that the DSNSFKM allows to characterize traffic from a network segment and can assist administrators to detect problems.

Keywords: DSNSF-KM. Traffic characterization. Clustering. K-means.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

BLGBA	*Baseline para gerenciamento de backbone automático*
CPU	*Central Processing Unit*
CRAN	*Common Reliable Accounting for Network Element Protocol*
DNS	*Domain Name System*
DPI	*Deep Packet Inspection*
DSNS	*Digital Signature of Network Segment*
DSNSF	*Digital Signature of Network Segment Using Flow Analysis*
DSNSF-KM	*Digital Signature of Network Segment Using Flow Analysis and K-means Clustering*
DoS	*Denial-of-service*
DDoS	*Distributed denial-of-service*
FLD	*Fisher Linear Discriminant*
FNF	*Flexible NetFlow*
FTP	*File Transfer Protocol*
GBA	*Gerenciamento de Backbone Automático*
GMM	*Gaussian Mixture Modeling*
HTTP	*Hypertext Transfer Protocol*
IETF	*Internet Engineering Task Force*
IEEE	*Institute of Electrical and Electronic Engineers*
IP	*Internet Protocol*
IPFIX IP	*Flow Information eXport*
ISP	*Internet Service Provider*
KM	*K-means*
MIB	*Management Information Base*
NMSE	*Normalized Mean Square Error*
NBAR	*Cisco Network Based Application Recognition*
OSI	*International Organization for Standardization*
P2P	*Peer-to-peer*
QoS	*Quality of service*
RFC	*Request for Comments*
RTP	*Real-time Transport Protocol*
SCTP	*Stream Control Transmission Protocol*

SIP	*Session Initiation Protocol*
SOHO	*Small Office/Home Office*
SNMP	*Simple Network Management Protocol*
TCP	*Transmission Control Protocol*
UDP	User Datagram Protocol
UEL	*Universidade Estatual de Londrina*
UTFPR	*Universidade Tecnológica Federal do Paraná*
VoIP	*Voice Over Internet Protocol*

CONVENÇÕES E LISTA DE SÍMBOLOS

Na notação das equações, as seguintes convenções foram utilizadas:

- letras maiúsculas expressam vetores, exemplo: X;

Os seguintes símbolos foram utilizados:

Símbolo	*Descrição*
S	Número de amostras de um conjunto.
K	Número de clusters.
P_s^k	Valor do ponto pertencente ao cluster k.
c^k	Centro do cluster k.
$a(i)$	Dissimilaridade média de i com todos os objetos do mesmo cluster.
$b(i)$	Dissimilaridade média miníma de i em relação ao objeto i com todos os objetos dos outros clusters.
$\overline{X}_w$	Média ponderada.
p^k	Número de objetos pertencentes ao cluster j.
γ	Número de pontos mínimos pertencentes a um cluster.
D	Número de dimensões.
I	Número máximo de iterações.
X	Representa a assinatura gerada pelo DSNSF-KM.
Y	Representa o movimento do dia/período.
ρ	Correlação.
σ	Desvio padrão.
σ_{XY}	Desvio padrão da multiplicação de X e Y.
$NMSE$	Erro Quadrático Médio Normalizado.
FSD	Desvio Fracional Padrão.

SUMÁRIO

1 INTRODUÇÃO

O rápido crescimento dos serviços disponibilizados por meio da Internet representam um aumento constante da quantidade de tráfego encaminhado para as redes. À medida que mais serviços críticos dependem de estarem conectados, especialmente os que são oferecidos por meio da Internet, é necessário maior gerenciamento com um nível detalhado de informações, tonando-se imprescindível que os *Internet Service Providers* (ISP) e seus administradores tenham uma visão de tudo o que acontece em sua rede [1, 2, 3].

Nesse sentido, evitar quaisquer falhas ou indisponibilidades na rede exige uma gerência eficiente, por isso detectar anomalias é um dos grandes desafios para os administradores. Essa detecção pode ser classificada com base em assinatura, a partir da qual o administrador tem um conhecimento prévio sobre o tipo de ataque ou anomalia, entende-se como conhecimento prévio, o comportamento esperado para um determinado ataque ou anomalia, por exemplo a sequência de pacotes e/ou *flags* utilizados; em perfis que caracterizam o comportamento normal da rede, em que ao se utilizar um histórico diário, mensal, semanal ou anual da rede gera-se um perfil por meio da mineração de dados, de modelos estatísticos, dentre outras técnicas; e finalmente por técnicas híbridas que se caracterizam pela união das duas anteriores [4, 5].

Uma das etapas importantes nesse processo de detecção é a caracterização do tráfego de um segmento. Tal etapa torna-se de grande importância, pois possibilita a identificação de comportamentos para cada horário, dia da semana e até mesmo para um serviço específico, facilitando a busca e identificação de anomalias. Para tanto, a coleta e análise de fluxos do tipo IPFIX, sFlow ou NetFlow se tornaram ferramentas de grande utilidade para os administradores nas redes atuais [6, 7].

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, serviço, como também para contadores *bytes*, pacotes e fluxos [8, 9, 10, 11, 12]. Entende-se

por clusterização, que é uma técnica de mineração de dados, o processo de classificação de uma amostra de dados em grupos, onde a partir de cada grupo é extraído um valor que melhor o representa.

Diante da necessidade emergente de novas tecnologias de monitoramento capazes de suportar o rápido crescimento das redes e de fornecer informações detalhadas de sua utilização, os fluxos - definidos como um conjunto de pacotes passando por um ponto de observação na rede, durante um certo intervalo, compartilhando um conjunto comum de propriedades - surgiram a fim de superar essas dificuldades e de fornecer aos administradores um nova visão de gerência de rede. IP Flow Information eXport (IPFIX)[13], NetFlow [14] e sFlow [15], são exemplos de protocolos para exportação de fluxos.

Dentre as vantagens de se utilizar fluxos IP, destaca-se a possibilidade de extração de diversas informações de um único registro de fluxo, como por exemplo: contadores para bytes, pacotes e fluxos; identificar endereços IP de origem e destino; identificar portas de origem e destino; protocolos como TCP, UDP, ICMP; verificar as *flags* utilizadas na comunicação; período e a duração dos fluxos.

Neste trabalho, propõe-se a construção de um modelo para Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos e Clusterização K-means (ou *Digital Signature of Network Segment Using Flow Analysis and K-means Clustering* - (DSNSF-KM)), modelo esse baseado na análise do histórico do comportamento da rede por meio da clusterização. Com isso, espera-se auxiliar os administradores na detecção de problemas que possam ser identificados como anomalias em um segmento de rede.

Na construção do modelo DSNSF-KM, utilizou-se a técnica de clusterização por meio do método K-means (KM), processo que divide uma população n-dimensional em K grupos (ou clusters) baseado em uma amostra. KM particiona os objetos do vetor ou matriz de dados em K clusters, sendo que as linhas da matriz correspondem aos objetos, e as colunas, às dimensões. Esse particionamento procura maximizar a soma das distâncias entre os clusters e diminuir a soma das distâncias dentro de cada um deles. Dessa forma, o KM analisa uma amostra de dados dividindo-a em grupos, permitindo que padrões possam ser identificados e classificados [16].

O modelo proposto, DSNSF-KM, tem como objetivo construir uma assinatura de segmento de rede baseada na análise dos fluxos coletados a partir de uma base histórica. Para o desenvolvimento, aplicação, testes e validação do modelo proposto, foram utilizados fluxos coletados da rede da Universidade Tecnológica Federal do Paraná (UTFPR) - Câmpus Toledo e da Universidade Estadual de Londrina (UEL), sendo utilizados fluxos do

tipo Netflow versão 9 e sFlow versão 5, respectivamente para os cenários de testes.

Para construção da assinatura, aplicou-se o modelo a cinco cenários reais diferentes com objetivo de avaliar parâmetros como número de centros para clusterização, número mínimo de pontos pertencentes a um cluster, número de semanas, dias da semana utilizados, horários, tráfego total de bits/s, pacotes/s e fluxos/s, além da geração de assinaturas para protocolos específicos.

Após a geração das assinaturas para os diferentes cenários, pôde-se, por meio de testes analíticos, compará-las com o movimento real gerado pelo NfSen de cada cenário, com o objetivo de avaliar se as assinaturas permitem descrever o comportamento, e assim, serem utilizadas para auxiliar o gerencimento das redes. A aplicação da Assinatura Digital nos cenários reais das redes da UTFPR e UEL foi de grande valia, pois possibilitou validar o modelo de uma forma prática.

O restante deste trabalho esta organizado da seguinte seguinte forma: O capítulo 2 apresenta trabalhos relacionados à caracterização de tráfego e detecção de anomalias com a utilização de fluxos. O capítulo 3 aborda alguns padrões e recursos na gerência de redes com a utilização de fluxos. O capítulo 4 explica a construção da Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxo e Clusterização K-means (DSNSF-KM). O capítulo 5 demonstra a aplicação da assinatura sobre um segmento real de rede. Por fim no capítulo 6 apresentam-se as considerações do trabalho realizado.

2 TRABALHOS RELACIONADOS

Neste capítulo, relacionam-se trabalhos que utilizam fluxos IP para: caracterização de tráfego, clusterização, criação de *templates* específicos com objetivo de extrair informações adicionais visando suprir necessidades dos administradores e detecção de anomalias.

Shawky *et al.* [3] procuram modelar e caracterizar o tráfego de um *backbone* coletando alguns parâmetros como: *inter-arrival time*, endereços IP, número de portas, protocolos de transporte e tamanho dos *frames*. A modelagem se realiza em dois processos: o de análise e o de geração. No primeiro coletam-se e exportam-se dados com a ferramenta *Wireshark*, e com esses dados faz-se uma análise estatística que entrega sub-modelos ao segundo processo, no qual se gera o tráfego estimado que é comparado com o tráfego real do *backbone*. Os autores concluem que cinco parâmetros são suficientes para caracterizar e modelar o tráfego do *backbone*, e a pesquisa provou que a utilização de fluxos para modelar aplicativos como por exemplo HTTP, FTP e DNS se mostra potencial.

Celenk *et al.* [8] comentam que a entropia tem sido usada para examinar a rede determinando seu *status* e detectando anomalias. Baseados nisso, os autores propõem uma abordagem para reduzir o tempo de observação das características da rede aplicando a técnica Fisher Linear Discriminant (FLD). Esse processo realiza o agrupamento ou *clusterização* para identificar a hora exata do incidente ou ataque na rede. O método proposto é testado em tempo real, sobre os dados coletados na conexão principal de Internet da Universidade de Ohio, apresentando resultados precisos na identificação de anomalias.

Molnar and Moczar [9] propõem um framework para caracterização de tráfego para aplicações como P2P, jogos, redes sociais e reprodução de vídeo. Usando clusterização, definem grupos que representam cada tipo de tráfego utilizando o que os autores chamam de caracterização em três dimensões compostas por tamanho dos pacotes, duração e taxa. Com esses três tipos de informações, é possível identificar quais aplicações utilizam a rede.

Rossi *et al.* [12] utilizam o algoritmo comportamental Abacus, que explora

os fluxos para classificação do tráfego na rede. Seu classificador - extensão desse algoritmo - busca identificar uma aplicação utilizando dois campos dos registros de fluxos consistentes em bytes e pacotes. Testes foram realizados sobre um vasto conjunto de dados contendo uma mistura heterogênea de aplicações que vão de compartilhamentos P2P até *streaming* P2P. Os resultados indicam uma precisão de 90%, no pior caso, na classificação de aplicações utilizando-se o volume do tráfego.

Proença Jr. [1], na coleta de objetos SNMP por meio da ferramenta GBA e do algoritmo BLGBA, calcula a moda estatística para determinar um valor esperado para um dado segundo do dia, analisando o 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 uma assinatura para cada dia da semana, e o bl-3 que corresponde a uma assinatura para os dias úteis, sábados e domingos.

O NetFlow e o IPFIX suportam a criação de templates, permitindo que os fluxos transmitam informações específicas que sejam do interesse do administrador da rede. Hyeongu et al. [17] estabelecem dois templates no IPFIX para *Session Initiation Protocol* (SIP) e *Real-time Transport Protocol* (RTP), adicionando quatro campos no padrão *Institute of Electrical and Electronic Engineers* (IEEE) 802.11 com objetivo de monitorar por meio dos fluxos roteadores *wireless*. Os autores propõem um algoritmo de detecção para ataques do tipo flooding e ataques do tipo DoS por meio de mensagens do tipo CANCEL/BYE.

Tin and Chang [18] procuram medir *burstiness* (rajadas) de fluxos por meio da técnica Fractal. Usando autosimilaridade de fluxos de rede, transformam o tráfego em domínio de tempo para domínio do espaço, adotando variações de dimensões fractal. Os resultados das variações de rajadas indicam uma situação irregular da rede. Resultados experimentais mostram que a variação da dimensão fractal é proporcional a variação das rajadas. Assim a variação que excede um limite preestabelecido indica um tráfego irregular e dispara um alarme.

Pescape *et al.* [19] investigam o impacto de se usar amostragem no monitoramento do tráfego de uma rede. Para isso, utilizam três diferentes técnicas de amostragem, *systematic, uniform* e *stratified*, e diferentes taxas de amostragem. Para verificar a quantidade de degradação e tendência que a amostragem introduz em relação ao tráfego não amostrado, os autores utilizam medidas estatísticas como *Hellinger Distance* e *Fleiss Chi-Square* e um vasto conjunto de características de fluxos. Aplicam sua metodologia em um grande conjunto de dados heterogêneos e concluem que por um lado a amostragem causa uma certa degradação, mostrando que pode ser um desafio para avaliar corretamente seu impacto.

Tchrakian *et al.* [20] apresentam um algoritmo para implementação em tempo real baseado em análise espectral. A previsão consiste na caracterização de fluxos, com base em funções modais associadas a uma matriz de covariância que é construída a partir do histórico de fluxos, em que o algoritmo é capaz de fazer a previsão para o que os autores chamam de *forecast horizon* (previsão para vários passos ou para um passo), a cada 15 minutos, faz-se a previsão para 1 hora e 15 minutos seguintes. Os autores utilizam uma técnica de média em sucessivas previsões.

Chih-Ming e Feng-Li [21] tratam a análise de tráfego de fluxos dinâmicos, processo que ocorre no tempo e espaço, como uma importante questão no gerenciamento e controle de sistemas. Eles propõem um estimador dinâmico de tráfego de fluxos chamado *Gaussian Mixture Modeling* (GMM). Os dados de tráfego de fluxos podem ser representados por uma combinação linear de multiplas funções Gaussianas para reduzir a grande quantidade de dados. Os resultados experimentais mostram cinco padrões de tráfego de fluxo baseados em um conjunto de dados reais. Cada padrão de tráfego pode indicar uma interpretação diferente de um comportamento de tráfego dinâmico.

Marques *et al.* [22] procuram caracterizar o comportamento do usuário de banda larga de um provedor de serviços, no qual os usuários se dividem em dois grandes grupos: residencial e *Small-Office/Home-Office* (SOHO). Para cada grupo, a caracterização se baseia em quatro critérios adquiridos por meio da coleta de fluxos, sendo eles: processo de chegada de sessão, duração de sessão, volume de tráfego dentro da sessão e padrão de solicitação de usuário. Os autores concluem que o grupo residencial possui taxas elevadas durante o dia; já o grupo SOHO possui uma maior variação, porém as sessões são de maior duração e maior tráfego. Para caracterização, utilizou-se a clusterização K-means.

Moghaddam *et al.* [23] apresentam um projeto para redes móveis impulsionado por modelos realísticos de comportamento de usuários on-line com base na mineração de registros de redes sem fio. Introduzem um método sistemático para grande escala multidimensional *co-clustering* de atividades na web por milhares de usuários móveis em 79 locais. As informações provêm de fluxos NetFlow e são clusterizadas para geração dos grupos. Os autores identificam que os usuários podem ser modelados utilizando 10 grupos.

Lakhina *et al.* [24] propõem a entropia como ferramenta de sumarização verificando que a análise das distribuições de características leva a avanços significativos. O primeiro consiste na possibilidade de detecção altamente sensível de uma vasta gama de anomalias. Outro avanço reside na classificação automática de anomalias por meio de técnicas de aprendizado não supervisionado, utilizando-se dois tipos de algoritmos de clusterização: a)

particional, em que se explora a estrutura global dividindo os dados em K centros, e b) algoritmos hierárquicos trabalhando com uma estrutura de baixo para cima ou cima para baixo. Na pesquisa coletaram-se fluxos de dois *backbones*: Abilene - 3 semanas de fluxos com amostragem 1:100; Géant - 3 semanas de fluxos com amostragem 1:1000. Os autores concluem que a entropia é uma métrica efetiva para capturar mudanças de tráfegos incomuns causados por anomalias.

Com o uso de esboços de tráfego, Chang *et al.* [6] propõem um novo método para caracterizar e detectar de anomalias utilizando análise de fluxos. Estabelecem-se esboços para endereços de origem, destino e porta de destino, sendo os registros de fluxo gravados utilizando funções *hash*. Para cada esboço, utiliza-se a técnica *Holt-Winters* para alcançar a previsão e predição criando um certo limite. Ao se exceder o limite, subalarmes fazem a comparação do tráfego com características de vários ataques. Em seguida, geram-se alarmes finais. Enquanto se constroem os esboços, gravam-se os endereços de destino em listas ligadas para serem utilizados posteriormente na identificação das vítimas.

Singh *et al.* [11], por meio da clusterização, procuram identificar anomalias em fluxos NetFlow. Apresentam uma abordagem baseada no K-means para analisar os fluxos usando campos como endereços IP, portas, protocolos, entre outros, para detectar anomalias. Os autores apresentam três resultados no 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.* [10], introduzem diferentes níveis na análise de tráfego de rede e relevante conhecimento em domínio de máquina de aprendizado. São analisados problemas por métodos *port-based* e *payload-based*. Avaliam-se seus métodos em performance e eficiência pelo algoritmos K-means e os resultados demonstram que o método pode obter 80% de precisão.

Ao se analisar os trabalhos relacionados, verifica-se a tendência em se utilizar informações de fluxos IP. Em relação às técnicas de classificação e previsão, são relacionadas diversas como entropia, *Fisher Linear Discriminant*, clusterização, Abacus, BLGBA, Fractal, matrizes de covariância, GMM, *K-means* e *Holt Winters*. Algumas técnicas utilizam o histórico dos fluxos coletados, outras realizam uma análise do tráfego em tempo real. Nesse trabalho é proposto um modelo para criação de uma Assinatura Digital analisando-se o histórico dos fluxos por meio da clusterização especificamente do método *K-means*, dessa forma o modelo procura identificar padrões que retratam o comportamento ou perfil do segmento analisado.

3 GERÊNCIA UTILIZANDO ANÁLISE DE FLUXO

O rápido crescimento das redes IP tem criado novos negócios e serviços. Estes novos serviços como VoIP, multimídia e jogos tem gerado necessidade de aumento da largura de banda, de melhora de performance e de qualidade dos serviços. Em paralelo surge a necessidade emergente de tecnologias para medição/controle que suportem esse rápido crescimento e forneçam informações detalhadas da rede e sua utilização. Os fluxos IP surgiram a fim de superar essas necessidades e fornecer aos administradores uma nova visão de gerência de rede.

A definição para fluxo inicialmente ficou conhecida como *five-tuples* onde sua composição é feita pela união dos seguintes campos: endereço IP de origem e destino, porta de origem e destino, e protocolo de transporte. Tramell *et al.* [25] trouxe uma definição menos restritiva de fluxo em relação a apresentada como *five-tubles*. Ele sugere a definição de fluxo como sendo um conjunto de pacotes passando por um ponto de observação na rede, durante um certo período, que compartilham um conjunto comum de propriedades.

Alguns exemplos dessa definição podem ser: todos os pacotes que contenham como destino a porta 80; Todos os pacotes que tenham como origem o endereço IP x.x.y.y; Pode-se ainda fazer combinações como todos os pacotes com destino ao endereço IP x.y.y.y e destino a porta 443. Como pode ser observado nos exemplos, existem inúmeras combinações possíveis, elas geralmente são realizadas conforme a necessidade dos administradores de rede.

A Figura 3.1, extraída do RFC3917 [26], ilustra o processo básico de um cenário clássico em que é utilizada a gerência de fluxos e são identificados seus principais componentes. Primeiro, define-se o local/segmento na rede onde será feita a coleta dos fluxos, em sequência ativa-se a exportação de fluxos no dispositivo correspondente, que pode ser um *switch*, roteador ou um equipamento que possua alguma funcionalidade para exportação de fluxos, configurando-o para enviar os fluxos a um ou mais coletores onde os fluxos serão processados e armazenados.

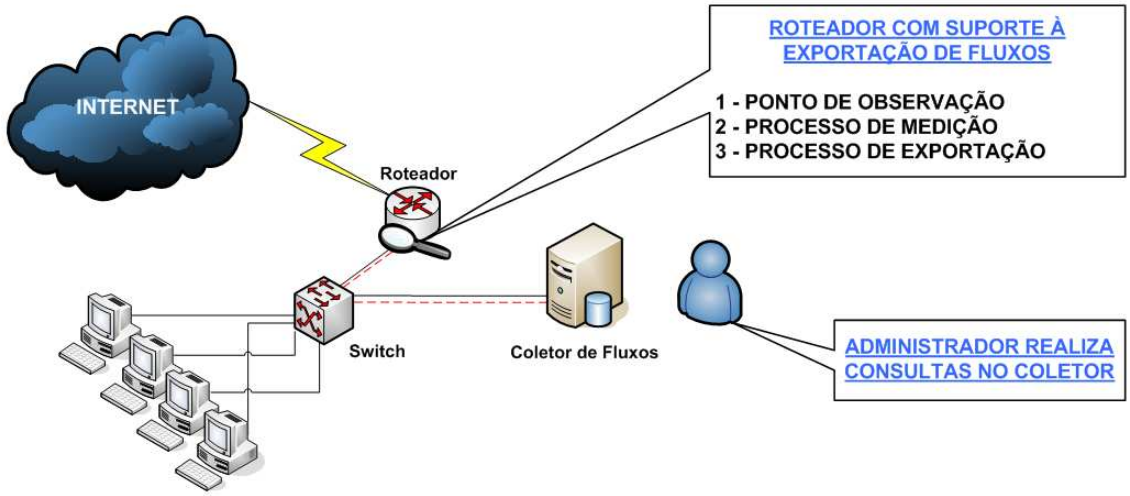

Figura 3.1: Modelo de envio e coleta de fluxos.

Existem outras opções de coleta de fluxo. Há casos em que os dispositivos, roteadores e *switches* não têm suporte aos protocolos de fluxos. Sendo assim uma alternativa possível é o espelhamento de uma porta do *switch* para um servidor executando uma aplicação para tratar os pacotes espelhados e transformá-los em fluxo, em seguida enviá-los a uma aplicação para coleta dos mesmos [27].

A Figura 3.2 apresenta seis amostras de fluxo que foram coletadas no padrão NetFlow versão 9, porém exibe somente alguns campos, dentre os fornecidos pelo NetFlow. Nessa figura pode-se observar os seguintes campos: Data, em que temos a data/hora/minuto, Protocolo utilizado na comunicação, endereços de origem e destino, versão do IP, além dos contadores para bytes, pacotes e fluxos.

«Data"	"Protocolo"	"IP Porta Origem"		"IP e Porta Destino"	"Versão IP"	"Bytes"	"Pacotes"	"Fluxos"
2011-04-28 15:06	TCP	200.234.196.185:80	->	10.1.1.102:53958	4	19655	28649	3
2011-04-28 15:06	TCP	10.1.1.102:53958	->	200.234.196.185:80	4	13677	8649	7
2011-04-28 15:06	UDP	200.147.255.105:53	->	10.1.1.102:50680	4	7719655	7728649	16
2011-04-28 15:06	UDP	10.1.1.102:50680	->	200.147.255.105:53	4	13677196	13677286	10
2011-04-28 15:06	TCP	67.174.72.86:53680	->	10.1.1.102:41762	4	456283	638233	6
2011-04-28 15:06	UDP	200.147.255.105:53	->	10.1.1.102:2923	4	1371027691	1371027711	1

Figura 3.2: Exemplo de fluxos listados no coletor.

Neste capítulo, são apresentados os conceitos relacionados à gerência de rede utilizando fluxos IP, baseado nos protocolos NetFlow [14], Flexível NetFlow [28], sFlow [15] e IPFIX [13], e uma visão geral sobre cada um dos protocolos citados.

3.1 Netflow

O NetFlow foi desenvolvido pela Cisco [14] como uma solução para os desafios emergentes na 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 e protocolIdentifier*, também com suporte ao IPv6. Segundo o RFC 3954 [14], o NetFlow possibilita várias aplicações como:

- Monitoramento da rede: pode ser utilizado para visualizar padrões de tráfego de roteadores, *switches*, detectar problemas de forma proativa e facilitar a resolução de problemas.
- Criação de perfis e monitoramento de aplicações e usuários: possibilita aos administradores da rede obterem uma análise detalhada com base no tempo e no uso das aplicações. Essa informação é usada para planejar e compreender novas aplicações e também alocar recursos de rede nas demandas que surgem.
- Planejamento da rede: pode ser usado para capturar um longo período de tempo produzindo um perfil da rede e antecipando o crescimento da mesma no sentido de aumentar o *backbone*, aumentar o número de roteadores e planejar novas políticas.
- Análise de segurança: identificar e classificar ataques DDoS, vírus e *worms* em tempo real. Mudanças no comportamento da rede podem indicar anomalias.
- Contabilidade: provê medição detalhada com fina granularidade. Por exemplo permite checar um IP específico ou porta tanto de origem quanto destino. Pode ser usado para bilhetagem por tempo, bytes ou serviços.
- Armazenamento e mineração de dados: pode ser usado para armazenamento dos fluxos para uso posterior de departamentos da empresa como marketing e programas de serviço ao consumidor.

O formato do registro NetFlow versão 9 consiste de um cabeçalho seguido por pelo menos um ou mais *templates* ou *Data FlowSet*. Um *template* fornece uma descrição dos campos que estão presentes em dados futuros. A Tabela 3.1 ilustra a estrutura de um pacote NetFlow na versão 9.

Tabela 3.1: Pacote NetFlow v9.

Packet Header	Template FlowSet	Data FlowSet	Data FlowSet	....	Data FlowSet

O NetFlow trabalha criando entradas *cache NetFlow* que contêm as informações de todos os fluxos ativos. Elas são construídas pelo processamento do primeiro pacote de um fluxo através do percurso de comutação padrão. Cada registro de fluxo na *cache NetFlow* contém campos chaves que podem ser depois usados para explorar os dados dentro de um dispositivo coletor. Cada registro de fluxo é criado pela identificação dos pacotes com características de fluxo similares, e contagem ou controle, de pacotes e bytes por fluxo. As informações da *cache* são exportadas a um servidor coletor de fluxos periodicamente, baseadas em temporizadores de fluxo.

O NetFlow trabalha de uma forma muito eficiente, a quantidade de dados exportados corresponde a 1,5% do total do tráfego no dispositivo. Ele faz a contagem de todos os pacotes, e não trabalha no modo de amostragem, promovendo uma visão condensada e detalhada de todo o tráfego da rede do dispositivo.

Segundo o *NetFlow Services Solutions Guide v4* [29], a chave para o NetFlow versão 9 conseguir escalabilidade e performance são algoritmos inteligentes de gerenciamento de *cache* de fluxos, especialmente para populações densas e roteadores de borda, que tratam um grande número de concorrentes e possuem fluxos de curta duração. O software de gerenciamento do *cache NetFlow* contém um sofisticado conjunto de algoritmos para determinar com eficiência se um pacote é parte de um fluxo existente ou se deve ser gerada uma nova entrada *cache NetFlow*. Os algoritmos são capazes de atualizar dinamicamente medições por fluxo que residem na *cache NetFlow* e de determinar a validade/expiração dos fluxos.

As regras para determinar quando um fluxo NetFlow expira são as seguintes:

- Os fluxos que estiverem ociosos/inativos por um certo tempo são expirados e removidos da *cache*, o tempo padrão é 15 segundos;
- Fluxos ativos são expirados por padrão. Eles são removidos ao alcançarem 30 minutos;
- Quando o *cache* fica cheio, uma série de heurísticas são aplicadas aos fluxos simultaneamente a fim de expirar os fluxos mais antigos;
- Conexões TCP com *flags* FIN ou RST são expiradas.

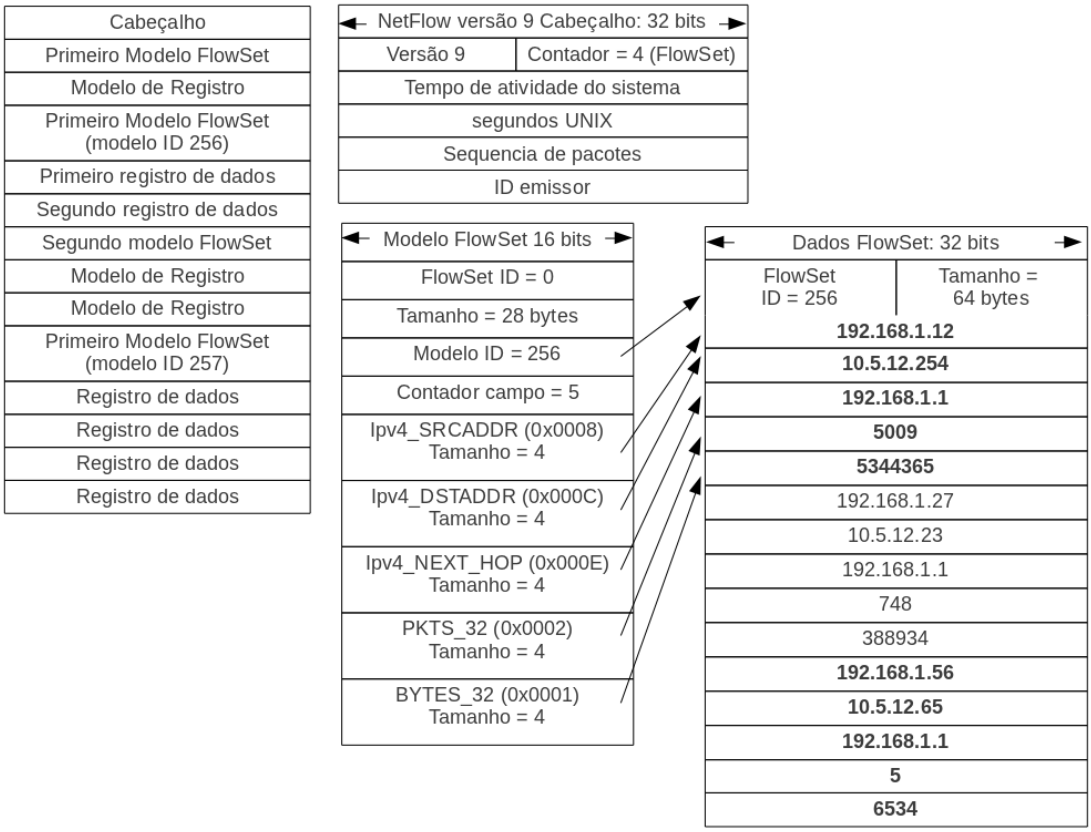

Figura 3.3: Exemplo pacote NetFlow v9.

Os fluxos expirados são agrupados juntos no datagrama *NetFlow Export* para serem enviados/exportados. O *NetFlow Export* pode consistir de até 30 registros de fluxo para as versões 5 e 9 do NetFlow. Para configurar o NetFlow basta o usuário especificar o endereço IP e o número da porta do aplicativo coletor.

A Figura 3.3, adaptada de [29], exibe um típico registro de fluxo NetFlow na versão 9. Esta versão se difere das anteriores por apresentar um modelo / *template* que descreve o conjunto de dados NetFlow contido nos fluxos, sendo os fluxos associados a seu *template*. Isto permite uma flexibilidade na exportação, por exemplo, pode-se criar um *template* específico para enviar mensagens do protocolo SIP, ou informações de endereços IPv6. Isso significa que o administrador não fica limitado as informações padrão que o exportador envia, e sim ele pode criar um *template* contendo as informações que ele considera importantes. O pacote NetFlow pode conter mais de um *template*.

Na Figura 3.3, pode-se identificar alguns campos como cabeçalho, o qual traz informações como versão do NetFlow, o número de registros de fluxos que este pacote possui, temporizadores e sequência do pacote. Em seguida primeiro modelo *FlowSet*, traz informações como tamanho do *FlowSet*, identificação do modelo (ID=256) e na sequência os

registros de fluxos contendo as informações como endereço IP de origem e destino, *Next Hop*, número de pacotes e Bytes.

3.1.1 Flexible NetFlow (FNF)

Desenvolvido também pela Cisco, o *Flexible NetFlow*, segundo o *Cisco IOS XE Flexible NetFlow Overview* [28], é a próxima geração em tecnologia de fluxo que permite a otimização da infraestrutura da rede, reduzindo os custos de operação, melhorando o planejamento, capacidade e detecção de incidentes de segurança, com maior flexibilidade e escalabilidade.

Entre as diferenças, o FNF utiliza amostragem com objetivo de reduzir a carga que ele pode gerar na rede, limitando o número de pacotes analisados. A amostragem pode ser determinística ou randômica. A taxa de amostragem pode ser de 1 pacote amostrado entre o intervalo de 2 até 32.768 pacotes.

A principal diferença entre o NetFlow v9 e o FNF é a habilidade para exportar seletivamente os dados. Basicamente, define-se a informação que se deseja exportar, daí a origem ao nome de flexível. O FNF tem suporte a *Deep Packet Inspection* (DPI), em que é possível por exemplo, examinar os pacotes e procurar desconformidades de protocolo, vírus, spam, ou até mesmo para definir critérios para tomada de decisão. Também oferece suporte a *Network Based Application Recognition* (NBAR), monitoramento de VoIP, monitoramento de tráfego de vídeo e coletores de fluxo ilimitados.

O FNF tem por objetivo melhorar o NetFlow original adicionando a capacidade de customizar os parâmetros de análise de tráfego para necessidades específicas dos administradores de rede.

3.2 sFlow

O nome sFlow pode ser definido como um acrônimo de *sampling flow* ou seja, amostragem de fluxos. Ele usa amostragem para alcançar escalabilidade. Por essa razão geralmente é aplicado em redes de alta velocidade. Segundo a RFC 3176 [15] o sFlow é uma tecnologia para monitoramento de tráfego de redes inserida em *switches* e roteadores. Define implementações de mecanismos de amostragem como: sFlow Agent para monitorar o tráfego; sFlow MIB para controlar sFlow Agent; Define também o formato dos dados amostrados usados pelo sFlow Agent quando os dados são encaminhados para o coletor.

A arquitetura e as técnicas de amostragem usadas no sistema de monitoramento sFlow são desenvolvidas para prover o contínuo monitoramento em toda rede inclusive para altas taxas de transmissão em *switches* e roteadores.

O projeto do sFlow aborda questões associadas com:

- Monitoramento preciso de redes Gigabit e superiores;
- Permite gerenciar dezenas de milhares de agentes em um único coletor sFlow;
- Fornecer uma visão do uso e das rotas de toda a rede. É uma técnica escalável para medição, coleta, armazenamento e análise do tráfego;
- Com custo baixo na execução do agente, tem sido implementado por fabricantes sem o uso de requisitos adicionais como memória e CPU.

O sistema de monitoramento sFlow consiste em um agente sFlow (incorporado em um *switch*, roteador ou como uma sonda em um dispositivo) e um coletor central de dados ou um analisador sFlow. Os agentes sFlow utilizam tecnologia de amostragem para capturar estatísticas de tráfego do dispositivo que estão monitorando e enviá-las ao(s) coletor(es).

O sFlow utiliza duas formas de amostragem:

Statical packet-based sampling of switched flows em que um fluxo é definido como todos os pacotes que são recebidos em uma interface e encaminhados para outra interface. No caso de um roteador, pode ser considerada a mesma interface. O mecanismo de amostragem deve garantir que qualquer pacote envolvido em um fluxo tenha chances iguais de ser amostrado independente do fluxo que ele pertença. A amostragem de fluxos é realizada da seguinte maneira: quando um pacote chega a uma interface, uma decisão de filtragem é tomada para determinar se o pacote deve ser descartado ou encaminhado a uma interface. Neste ponto é tomada a decisão se o pacote será ou não amostrado. O mecanismo envolve um contador que é decrementado a cada pacote, quando o contador chega a zero a amostragem é realizada ou seja, é retirada uma amostra.

Nessa forma de amostragem, para se ter a taxa de amostragem, é utilizado um contador *Total_Packets* que é incrementado independentemente se um pacote é amostrado ou não, já o *Total_Samples* é incrementado a cada pacote amostrado. A taxa de amostragem pode ser obtida por meio do total de pacotes dividido pelo total de pacotes amostrados.

Na *Time-based sampling of network interface statistics* o objetivo do contador é realizar de forma eficiente e periódica a sondagem de cada fonte de dados nos dispositivos e extrair os principais dados estatísticos. Por razões de eficiência e escalabilidade, o sistema sFlow implementa contadores *polling* (sondagem) nos agentes sFlow. Um intervalo de sondagem máximo é atribuído para o agente, mas ele é livre para definir esse parâmetro com o intuito de maximizar a eficiência interna. Amostragem de fluxo e amostragem de contadores são desejáveis como parte de um sistema integrado. Ambos os tipos são combinados nos datagramas sFlow.

A Figura 3.4, exibe por meio da ferramenta Wireshark [30], um pacote sFlow capturado. Pode-se identificar as duas formas de amostragem citadas anteriormente exibidas como *Flow sample* com uma entrada e *Counters sample* com sete entradas, também a versão do sFlow, IP do dispositivo exportador, número de sequência e número total de amostras.

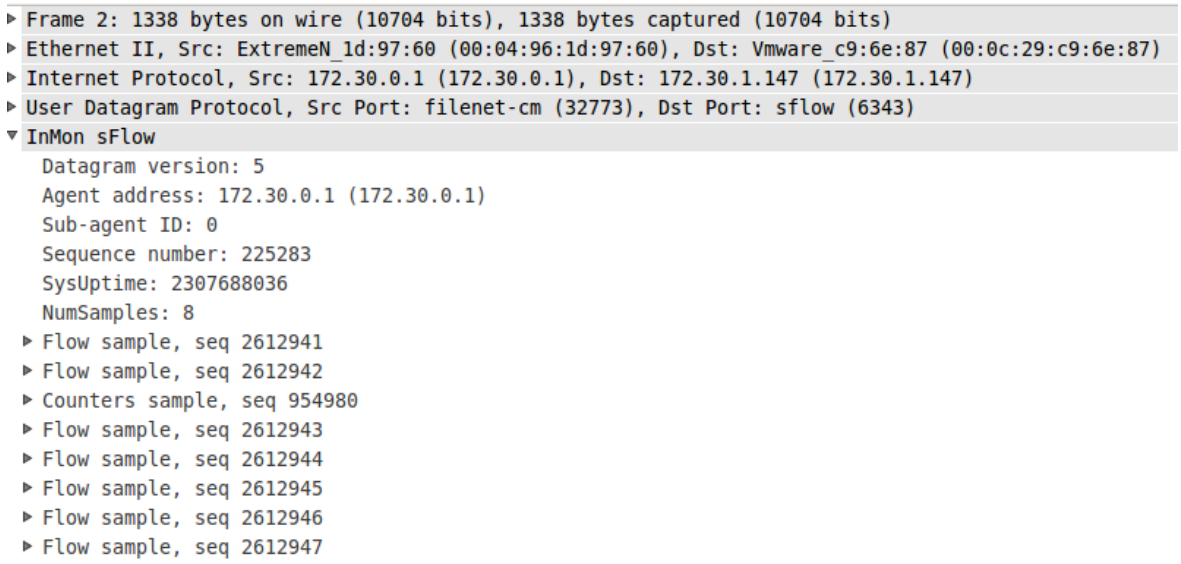

Figura 3.4: Amostra de pacote sFlow capturado.

Na Figura 3.5 é exibido o mesmo pacote porém expandido, sendo possível identificar e diferenciar os dois tipos de amostras. Em *Counters sample* o sFlow está exportando fluxos referentes à interface *3001*, também um subgrupo para os contadores genéricos dessa interface. Para *Flow sample*, também é possível identificar a interface *3011* que está exportando os fluxos com taxa de amostragem *1 para 1024* dentre outras informações.

Alguns exemplos de aplicações com o uso de dados sFlow:

- Detecção, diagnóstico e resolução de problemas de rede;
- Gerenciamento de congestionamento em tempo real;

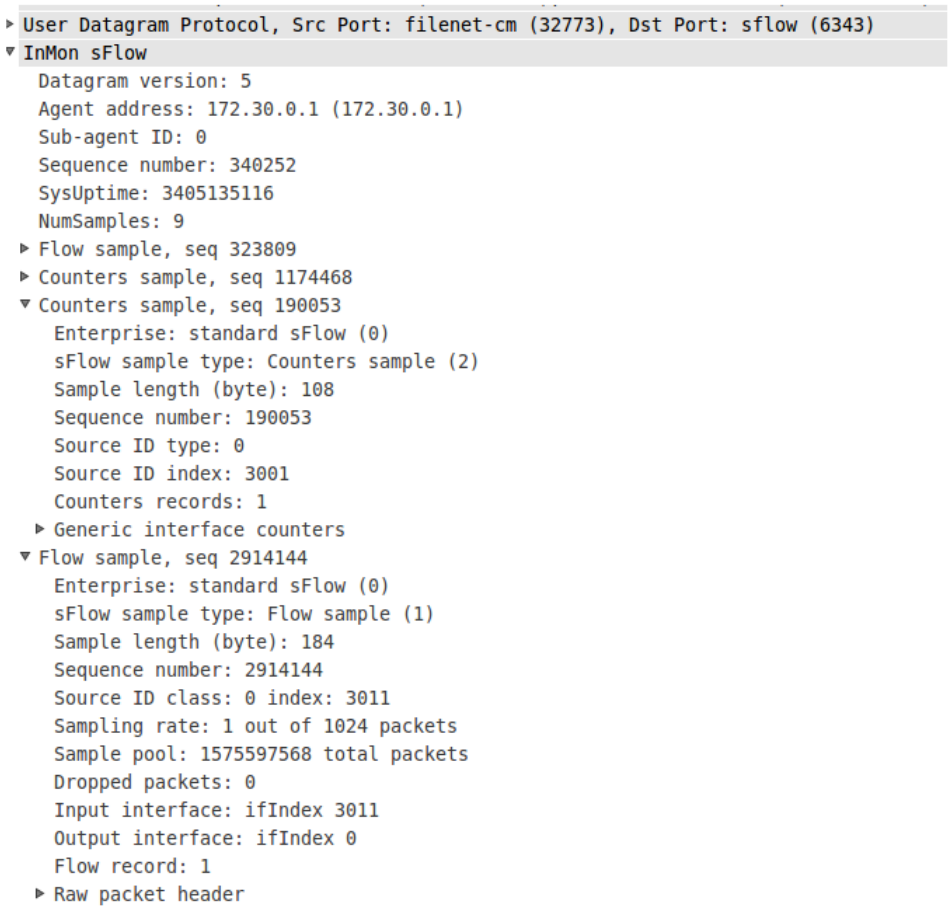

Figura 3.5: Amostra de pacote sFlow capturado expandido.

- Compreensão de diversas aplicações como: P2P, HTTP, DNS;
- Auditoria, análises para identificar atividades não autorizadas;
- Perfis de rotas;
- Tendências e planejamento de capacidade.

A Figura 3.6, adaptada de [31], ilustra a história da amostragem de pacotes. A Hewlett Packard (HP) foi a primeira a demonstrar o monitoramento de uma rede com o uso de amostragem de pacotes na Universidade de Genebra e no CERN. No entanto, a aceitação desta técnica ocorreu com o surgimento das altas taxas de transmissão e a transição de redes não comutadas (HUB) para comutadas (*switched*) [31].

Na Figura 3.6, pode-se observar que somente em meados de 2001 o sFlow

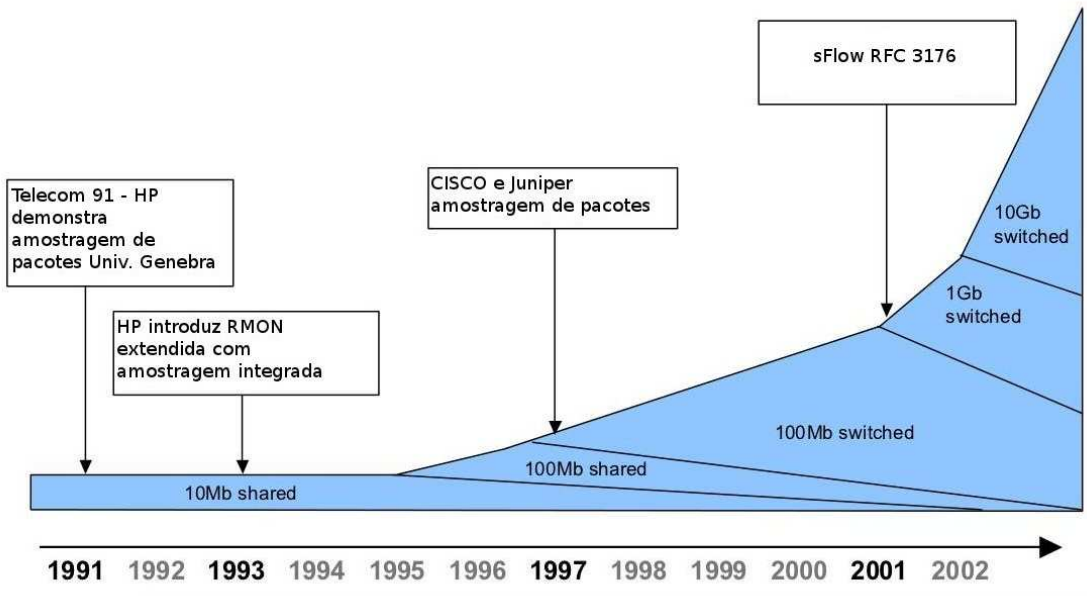

Figura 3.6: Histórico amostragem de pacotes.

se tornou um padrão IETF com a RFC 3176 aliado à tendência do aumento das taxas de transmissão que passaram de 1Gb/s, e a partir dessa data, *switches* e roteadores de alguns fabricantes tiveram a tecnologia sFlow incorporada.

O monitoramento sFlow de redes de alta velocidade, roteadores e redes comutadas possui as seguintes propriedades:

- Acurácia: amostragem é simples o suficiente para ser realizada via hardware. Em adicional, o sistema sFlow foi projetado para que a acurácia possa ser alcançada em qualquer medição;
- Detalhamento: cabeçalhos de pacotes completos e informações de comutação e roteamento permitem análises de fluxo de tráfego desde a camada 2 até a 7 do modelo OSI;
- Escalável: o sistema sFlow é escalável em tamanho e velocidade, pode monitorar redes superiores a 10Gbps e 100Gbps, permite monitorar milhares de dispositivos através de um único coletor sFlow;
- Baixo custo: o agente sFlow é muito simples de ser implementado e introduz um custo insignificante a um *switch* ou roteador;
- Conveniente: o coletor sFlow sempre fornece a visualização do tráfego minuto a minuto. Essa informação é importante, pois possibilita prover controles em tempo real.

A funcionalidade do sFlow é determinada pela escolha do analisador sFlow, pois o mesmo tem a função de fazer toda a análise dos dados. Com FNF, grande parte da análise é executada no próprio dispositivo, o que pode limitar a funcionalidade de coletores NetFlow, pois os mesmos simplesmente gravam os dados e geram relatórios. Em contraste, o sFlow passa a função de análise para um analisador sFlow central que determina como tratar os dados e apresentar os resultados.

Enquanto muitos coletores NetFlow têm suporte ao sFlow, eles tendem a ser limitados, ignorando características específicas do sFlow, tratando o mesmo como se fosse uma versão básica do NetFlow v5.

Atualmente, o sFlow encontra-se na versão 5 de 2004, e uma das maiores diferenças dessa versão quando comparada às versões anteriores é em relação ao seu datagrama, nos quais existem especificações que envolvem extensibilidade dos fornecedores, permitindo assim que os mesmos tenham a liberdade de adicionar seus próprios tipos de registros [32].

3.3 Internet Protocol Flow Information Export (IPFIX)

Quando o IETF iniciou os trabalhos para criar um protocolo de troca de informações de tráfego ou protocolo de exportação de fluxos chamado *Internet Protocol Flow Information Export* (IPFIX), definiu-se por meio da RFC3917 [26] os requisitos e objetivos desse protocolo, que tem como objetivo satisfazer as aplicações consideradas de importância significativa hoje e/ou para o futuro das redes IP. As aplicações podem ter um grande número de diferentes implementações e por isso os requisitos derivam de forma geral das seguintes aplicações selecionadas:

- Contabilidade: pode ser baseada no tempo, volume, por usuário ou grupo e um serviço IP;
- Perfil de Tráfego: processo de caracterização dos fluxos para gerar um modelo que representa os parâmetros chave da rede; é um pré-requisito para que seja feito, por exemplo, um planejamento da rede;
- Engenharia de Tráfego: compreende os métodos para medir, modelar, caracterizar e controlar a rede; o objetivo é a otimização dos recursos da rede e a performance do tráfego;

- Detecção de Ataques/Intrusões: capturar informações de fluxo tem um importante papel para segurança da rede, tanto no sentido de detectar alguma violação de segurança quanto na defesa subsequente. Em caso de ataque de negação de serviço, o monitoramento de fluxo pode permitir a detecção por meio de tráfegos incomuns ou suspeitos. Um segundo passo seria analisar as informações de fluxo do ataque para fazer uma estratégia de defesa;

- Monitoramento da Qualidade de Serviço (QoS): é uma medição passiva dos parâmetros de qualidade para fluxos IP. Em contraste com medição ativa, a passiva utiliza o tráfego existente na rede para fazer a análise; não é gerado nenhum tráfego adicional.

Após a definição dos requisitos do IPFIX, a RFC3955 [33] procurou avaliar cinco protocolos candidatos para o IPFIX, baseando-se nos requisitos. Os candidatos foram os seguintes: *Common Reliable Accounting for Network Element Protocol* (CRANE); Diameter; LFAP; NetFlow v9 e Streaming IPDR. Os protocolos foram caracterizados e agrupados em amplas categorias e após esse processo foram avaliados em relação aos requisitos. Como resultado, foi recomendado o protocolo NetFlow v9 como base para a especificação do IPFIX.

Baseado no NetFlow v9 [14], o IETF desenvolveu o IPFIX que é descrito pela RFC5101 [13], trazendo algumas melhorias em diferentes domínios, como controle de congestionamento e segurança. O IPFIX suporta os protocolos de transporte *Stream Control Transmission Protocol* (SCTP), UDP e TCP. Preferencialmente é utilizado o protocolo SCTP, pois ele apresenta entrega sequencial de pacotes com confiabilidade. O UDP não é recomendado, pois não possui mecanismos que forneçam confiabilidade, a não ser que a comunicação seja feita por meio de links dedicados, os quais não são suscetíveis a congestionamentos. O IPFIX opera com TCP da mesma maneira com que o SCTP.

As terminologias usadas no IPFIX, segundo as RFC3917 e RFC5101, são:

- Ponto de observação: constitui um local na rede em que os pacotes podem ser observados. Podem ser únicos ou formarem um superconjunto de vários pontos de observação;

- Processo de medição: gera os registros de fluxo, por meio dos cabeçalhos dos pacotes observados no ponto de observação e o seu tratamento. Suas funções incluem: captura dos pacotes, *timestamps*, amostragem, classificação e manutenção dos registros de fluxo;

- Registro de fluxo: o qual possui informações sobre fluxo específico que foi medido no ponto de observação. Contém propriedades de medida como número total de bytes

de todos os pacotes do fluxo, características como: IP de origem/destino, porta de origem/destino, entre outras, conforme pode ser verificado na RFC5102 [34], que define as informações do IPFIX;

- Processo de exportação: envia os registros de fluxos para um ou mais coletores. Tais registros são gerados por um ou mais processos de medição;
- Processo de coleta: em que se recebem os registros de fluxo de um ou mais processos de exportação e em que se pode processar e/ou armazenar os fluxos.

A Tabela 3.2 é um exemplo ou descreve o modelo de uma mensagem IPFIX. Uma mensagem IPFIX consiste basicamente de conjunto de dados (*Data Set*), onde as informações são carregadas depois que o *Template Set* é definido e transmitido ao coletor, sendo esse responsável pela identificação das informações que o *Data Set* carrega. A maioria das mensagens IPFIX consistem apenas de *Data Set*.

Tabela 3.2: Exemplo mensagem IPFIX.

Message Header	Template Set	Data Set	....	Options Template Set	Data Set

Na Tabela 3.2 observa-se ainda o campo *Options Template Set*, nesse caso, existia um espaço na mensagem, por isso o campo pode ser adicionado, senão ele seria transmitido junto com a próxima mensagem. Cada *Data Set* possui um identificador referente ao *Template Set* que ele pertence, no entanto o *Template Set* é enviado somente a primeira vez.

Para ilustrar as terminologias citadas na RFC3917 e RFC5101 , a Figura 3.7 é um exemplo de um roteador, com três pontos de observação, um processo de medição e um de exportação enviando fluxos IPFIX para o coletor. Quando é configurado um dispositivo para exportação de fluxos, são definidas quais interfaces serão observadas, ou seja a coleta de informações será feita na(s) interface(s) definida(s).

A Figura 3.8 procura explicar a sequência desde o ponto de obvervação até o envio dos fluxos para o coletor. Primeiramente, assim que os pacotes entram no ponto de observação, eles entram na fase do processo de medição (*Metering Process*). Nessa fase é realizada a captura dos pacotes para geração dos fluxos, em sequência faz-se a marcação do tempo relativo a cada fluxo. O próximo passo é a classificação que é composta pela amostragem, onde pode-se aplicar alguma técnica para diminuir o volume de fluxo gerado, e pela aplicação de algum filtro como por exemplo tráfego em uma porta específica. Ainda

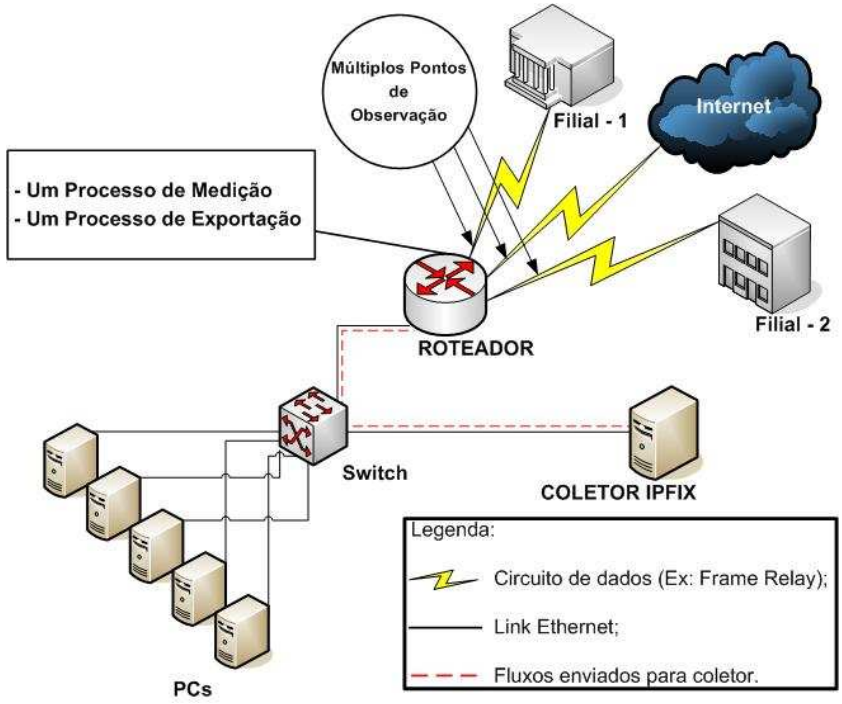

Figura 3.7: Roteador com vários pontos de observação.

nessa fase é feita a manutenção dos fluxos, em que é verificado por exemplo se um fluxo deve ser expirado e enviado ao processo de exportação.

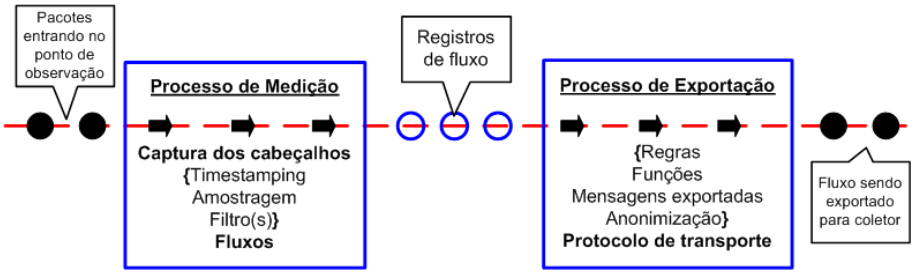

Figura 3.8: Descrição do processo de exportação de fluxos desde o ponto de observação até a exportação.

Ainda na Figura 3.8, no processo de exportação (*Exporting Process*), inicialmente são aplicadas regras e funções que definem os atributos para cada fluxo medido, como por exemplo versão do protocolo IP, contadores, etc. Também são tratadas questões de congestionamento e segurança. Ainda no processo de exportação, pode ser realizado o ato de anonimizar endereços IP, que consiste em ocultar informações de usuários para manter seu anonimato. Por exemplo, o administrador quer manter o monitoramento sobre o tráfego da rede porém sem identificar o tráfego por usuário, para isso ele faz uso dos contadores dos fluxos e ignora os endereços de origem e destino. O último passo é o envio dos fluxos para o coletor onde serão processados e armazenados.

A Tabela 3.3 resume as RFCs que descrevem o protocolo IPFIX, tendo como período inicial outubro de 2004 até outubro de 2012. Pelo número de RFCs e pelas datas percebe-se que esse protocolo está em constante atualização.

Tabela 3.3: RFCs do IPFIX.

Número	Descrição	Data
RFC 3917	Requisitos para IPFIX	Outubro 2004
RFC 3955	Avaliação dos protocolos candidatos para IPFIX	Outubro 2004
RFC 5101	Especificações do IPFIX, protocolo para troca de informações de tráfego de fluxo	Janeiro 2008
RFC 5102	Modelo de informações para IPFIX	Janeiro 2008
RFC 5103	Exportação de fluxo bidirecional usando IPFIX	Janeiro 2008
RFC 5153	Orientações para implementação	Abril 2008
RFC 5470	Arquitetura para monitoramento seletivo de fluxos IP e exportação de um dispositivo IPFIX para um coletor	Março 2009
RFC 5471	Orientações/diretrizes para IPFIX (Lista de testes)	Março 2009
RFC 5472	Aplicabilidade IPFIX para uma variedade de aplicações	Março 2009
RFC 5473	Método para economizar largura de banda, as considerações aplicam-se para PSAMP	Março 2009
RFC 5610	Descreve um extensão do IPFIX para exportação de tipos de informações extendidas	Julho 2009
RFC 5655	Descreve o formato do arquivo para armazenar os fluxos	Outubro 2009
RFC 5815	Definição de objetos gerenciados para IPFIX	Abril 2010
RFC 5982	Descreve alguns problemas de medição baseada em fluxo	Agosto 2010
RFC 6183	Framework estende o modelo de referência especificado na RFC 5470	Abril 2011
RFC 6235	Técnicas para anonimizar endereços IP, esquemas comuns e definição de parâmetros	Maio 2011
RFC 6313	Extensão do IPFIX, permite estrutura de dados complexas como listas de tamanho variavél e especificações referente a hierarquias na relação entre *template*	Julho 2011
RFC 6526	Utilização do protocolo SCTP, uma das vantagens é reduzir a probabilidade na perca de dados	Março 2012
RFC 6615	Definição de objeto gerenciados para IPFIX	Junho 2012
RFC 6727	Definição de objetos gerenciados para pacotes amostrados (PSAMP) módulo IPFIX-SELECTOR-MIB-MODULE	Outubro 2012
RFC 6728	Especifica o modelo de dados para IPFIX e PSAMP, na configuração e monitoramento	Outubro 2012

3.4 Comparação entre Fluxos IP e SNMP

O modelo de gerenciamento do *Simple Network Management Protocol* (SNMP) [35], é constituído basicamente por três componentes básicos: dispositivo, agente e um sistema de gerenciamento. No dispositivo é executado o agente SNMP que fornece as informações da MIB para a estação de gerenciamento. A estação de gerenciamento faz consultas utilizando comandos como, por exemplo, GET e GETNEXT. O comando GET retorna uma informação de gerenciamento como um contador de uma interface. O comando GETNEXT retorna uma sequência de informação. A estação de gerenciamento também pode fazer alguma configuração no dispositivo por meio do comando SET.

O dispositivo também pode enviar informações não solicitadas para a estação de gerenciamento por meio do comando TRAP, utilizado para reportar alguma notificação.

Ao se comparar o modelo de gerenciamento proposto pelo SNMP e pela análise de fluxos IP (IPFIX, sFlow e NetFlow), pode-se concluir que com o SNMP o administrador disponibiliza uma série de informações individuais, fornecidas por objetos contidos na MIB, normalmente em forma de contadores. No modelo de gerência utilizando a análise de fluxos, tem-se um conjunto de informações mais rico que podem fornecer o período, duração, endereço e porta de origem e destino, seus contadores, conforme já citado nas seções anteriores.

Ao se monitorar, individualmente, o uso de Internet dos dispositivos de um segmento de rede utilizando SNMP, é necessário que em todos os dispositivos do segmento tenham um agente sendo executado. Já com fluxos IP, pode-se instalar no *gateway* desse segmento um exportador de fluxos, dessa forma o administrador terá informações de tráfego de Internet de todos os dispositivos desse segmento sem a necessidade de instalar qualquer software nos mesmos.

4 ASSINATURA DIGITAL DE SEGMENTO DE REDE UTILIZANDO ANÁLISE DE FLUXO

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

Com esse objetivo, nesse capítulo é proposto e implementado um modelo para geração da Assinatura Digital, chamado de DSNSF-KM (Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos e Clusterização K-means ou *Digital Signature of Network Segment Using Flow Analysis and K-means Clustering*), que descreve um perfil básico do tráfego de um segmento de rede e permite indicar um comportamento padrão baseado na análise de fluxos.

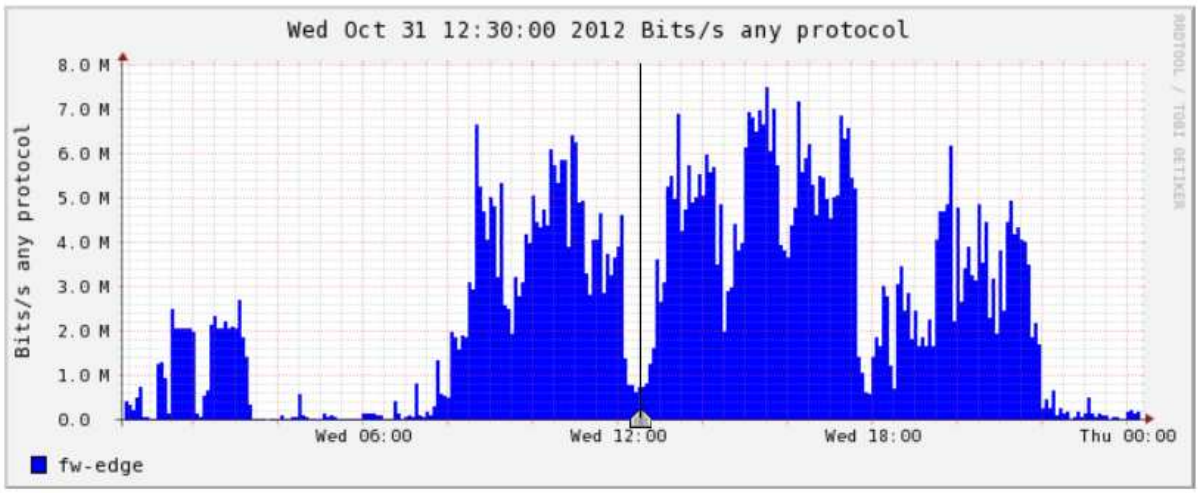

Figura 4.1: Movimento em bits/s de um segmento da UTFPR - Câmpus Toledo.

A Figura 4.1 representa o movimento em bits/s do segmento *fw-edge*, pertencente a rede da UTFPR - Câmpus Toledo, para o dia 31 de outubro de 2012, gerado

por meio da ferramenta NfSen [36], que é um *front end* para o aplicativo NFDUMP [37] que faz a coleta e processamento de fluxos.

A Figura 4.2 representa o movimento em pacotes/s do mesmo segmento e período apresentado na Figura 4.1, também gerado por meio da ferramenta NfSen e NFDUMP.

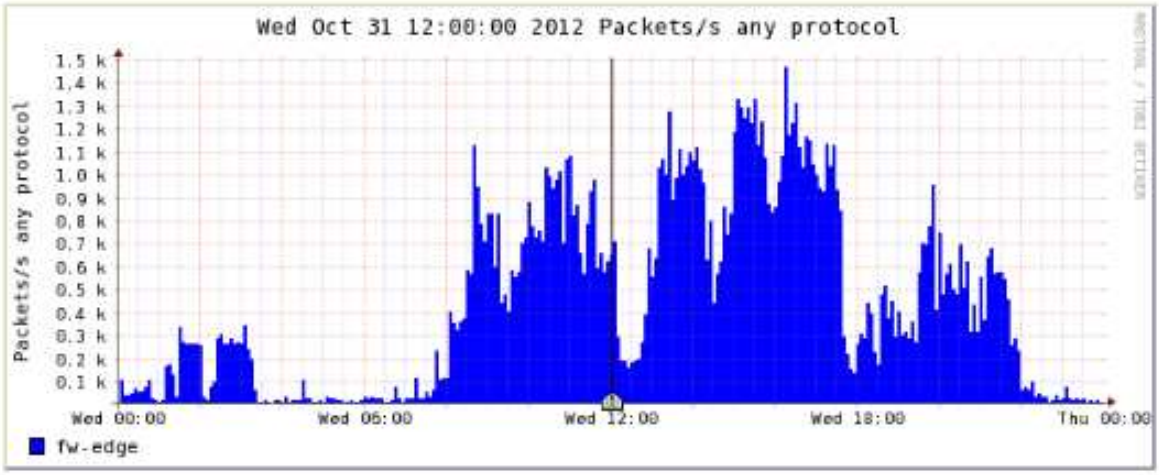

Figura 4.2: Movimento em pacotes/s de um determinado segmento.

Em seguida, na Figura 4.3 tem-se a representação por volume de fluxos/s para o mesmo segmento e período como apresentado nas figuras 4.1 e 4.2, também gerado por meio da ferramenta NfSen e NFDUMP.

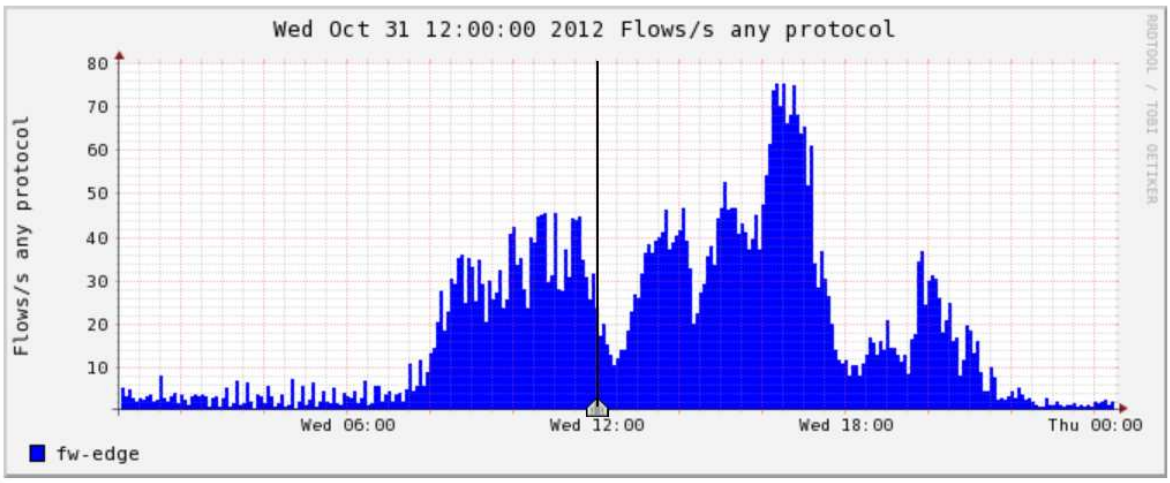

Figura 4.3: Movimento em fluxos/s de um determinado segmento.

Existem diversos softwares que permitem o monitoramento do tráfego de uma rede, dentre eles o NfSen como citado anteriormente nas figuras 4.1, 4.2 e 4.3. Esses softwares utilizam dados como por exemplo: objetos SNMP, espelhamento de portas e coleta de fluxos. Para o administrador da rede, se torna uma tarefa difícil monitorar gráfico por

gráfico a cada instante de todos os seus segmentos, pois ele perderá muito tempo observando cada gráfico que possui.

Grande parte desses softwares permitem definir limites (*threshold*) para o tráfego, volume, ou um determinado contador, para que a partir do momento em que esses limites sejam ultrapassados, alarmes sejam gerados. Porém, como o administrador poderá definir um limite para cada segmento? Como pode saber se um comportamento é normal ou anômalo? Será baseado simplesmente na capacidade de seu enlace? Essas respostas não são simples e são uma das motivações deste trabalho.

Por isso, além do DSNSF-KM descrever um perfil do segmento ele possibilita ainda a automatização do processo de detecção de anomalias por meio de um sistema para geração de alarmes, baseado no(s) limite(s) definido(s) pelo próprio DSNSF-KM para o segmento. Pode-se ter como resultado final, uma ferramenta a partir da qual o administrador não precisará monitorar visualmente todos os seus segmentos, tarefa essa que se torna difícil em casos onde existam muitos. Com o DSNSF-KM e a automatização de alarmes, o administrador focará somente nos segmentos que apresentam alguma anomalia, facilitando seu dia-a-dia e a automatizando suas tarefas, dispensando a necessidade de que um operador fique monitorando cada segmento.

O DSNSF-KM constrói uma assinatura por meio da leitura de fluxos, esta é baseada no histórico das últimas semanas. Uma das metas desse trabalho é definir a quantidade de semanas mínimas para que a assinatura possa identificar corretamente o comportamento do segmento analisado.

Com a leitura dos fluxos, pode-se extrair diversas informações, como já citado no capítulo 3. A partir delas, o administrador define quais serão utilizadas pelo DSNSF-KM para construção da assinatura, como por exemplo, contadores de bytes, pacotes, fluxos, entre outros. Além disso, pode-se dividir esses contadores por protocolo, endereço IP de origem e destino, porta de origem e destino, entre outros. As combinações possíveis limitam-se ao conteúdo dos fluxos coletados, por exemplo, NetFlow v9 possui mais campos de informações e a possibilidade de serem adicionados *templates* quando comparado ao NetFlow v5. Dessa forma a assinatura fica condicionada as informações dos fluxos coletados e a necessidade do administrador.

Basicamente, para geração da assinatura, os fluxos foram utilizados da seguinte maneira: extraiu-se do histórico de fluxos o campo desejado, como por exemplo bytes em seguida convertido em bits. Com essas informações, o DSNSF-KM faz a clusterização, aplica a regra de pontos mínimos pertencentes a cada centro, e em seguida calcula a média

ponderada para cada instante de cinco minutos, resultando no valor que será considerado como comportamento normal ou característico para aquele instante no segmento.

Para construir a assinatura, o DSNSF-KM utilizará a técnica de clusterização K-means. Diferentemente da classificação, a clusterização trabalha com dados não rotulados sendo um processo de aprendizagem não supervisionada. A clusterização foi definida como ferramenta para criação da assinatura por ser uma técnica de mineração de dados, a qual identifica e mensura semelhanças de um conjunto de informações. Esse processo procura minimizar a distância entre os objetos de um determinado grupo e aumentar a distância entre grupos [38]. A distância Euclidiana (4.1) geralmente é usada para medir a similaridade entre os dados.

$$J(p) = \sum_{k=1}^{K} \sum_{s=1}^{S} \sqrt{|P_s^k - c^k|^2} \tag{4.1}$$

em que $\boldsymbol{K}$ é o número de clusters, $\boldsymbol{S}$ é o número de objetos a serem clusterizados, $\boldsymbol{P_s^k}$ é o valor do ponto pertencente ao cluster k e $\boldsymbol{c^k}$ corresponde ao centro do cluster k. O propósito de usar clusterização é criar um modelo que possa extrair padrões a fim de construir um padrão normal, sendo possível identificar dados com um comportamento padrão e dados que se diferem desse padrão.

4.1 Clusterização K-means

K-means (KM) é o processo que divide uma população n-dimensional em K grupos (clusters) baseado em uma amostra. O KM particiona os objetos do vetor ou matriz de dados em k clusters, sendo que as linhas da matriz correspondem aos objetos e as colunas as dimensões. Esse particionamento procura maximizar a soma das distâncias entre os clusters e diminuir a soma das distâncias dentro de cada cluster. O KM sempre retorna os centroides de cada cluster e um vetor contendo os índices do cluster para cada objeto [16].

Os índices servem para representar em qual cluster cada objeto está associado, e em cada cluster existe um objeto para representá-lo que é chamado de centroide. Quando clusterizamos um conjunto de objetos, o algoritmo primeiramente inicializa os K pontos, podendo ser de forma aleatória, objetivando encontrar o centro que melhor represente o conjunto de objetos. O algoritmo realiza o cálculo da distância entre os objetos diversas vezes por meio das iterações, sendo as iterações um parâmetro de entrada para o algoritmo KM. Após a conclusão do algoritmo, são retornados os índices dos objetos e os centroides de

cada cluster.

Um problema que pode ocorrer no KM é o de cluster vazio ou *empty cluster*, que ocorre quando não é associado nenhum objeto a um cluster durante o processo de atribuição dos objetos aos clusters. Aumentando-se o número de repetições pode-se ajudar a evitar um cluster vazio, pois se em uma repetição/iteração esse problema acontece, na seguinte pode não acontecer, assim o algoritmo KM retornará o melhor resultado encontrado durante sua execução [39].

O Algoritmo 4.1 demonstra o pseudo código do KM.

Algoritmo 4.1 K-means
KM algoritmo usado para clusterização.

Entrada: S objetos que representam os dados a serem clusterizados, K Número de clusters, I Número de iterações.
Saída: Índices dos objetos para cada cluster, Cluster centroide.

Passo 1 Colocar k pontos no espaço que representa os objetos 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 dos k centroides.
Passo 4 Repita os passos 2 e 3 até que o centroide não se mova mais ou o número de repetições seja excedido.

A Figura 4.4 ilustra uma amostra de dados (objetos), gerada de forma randômica, distribuída em um plano bidimensional. Após o processo de clusterização, os dados serão agrupados em um número de clusters preestabelecido.

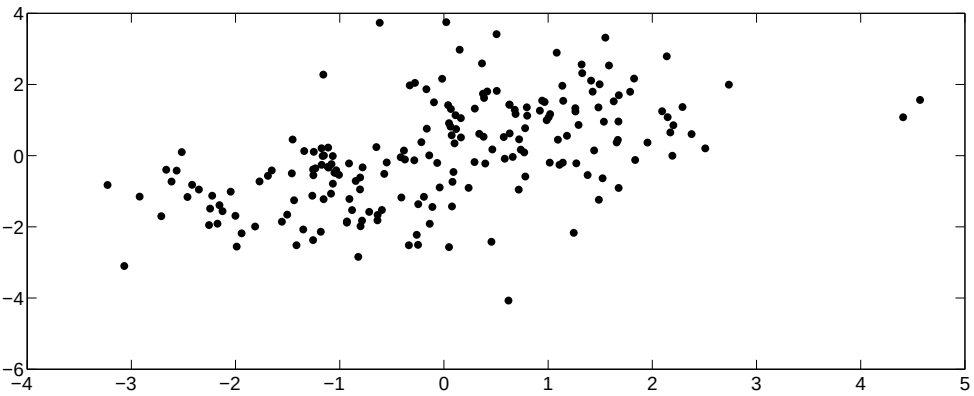

Figura 4.4: Conjunto de objetos antes da clusterização.

A Figura 4.5 exibe cada centro com seu respectivo centroide. O centroide é o ponto em que o algoritmo definiu como mais representativo para o grupo que ele pertence,

sendo o melhor valor para representá-lo.

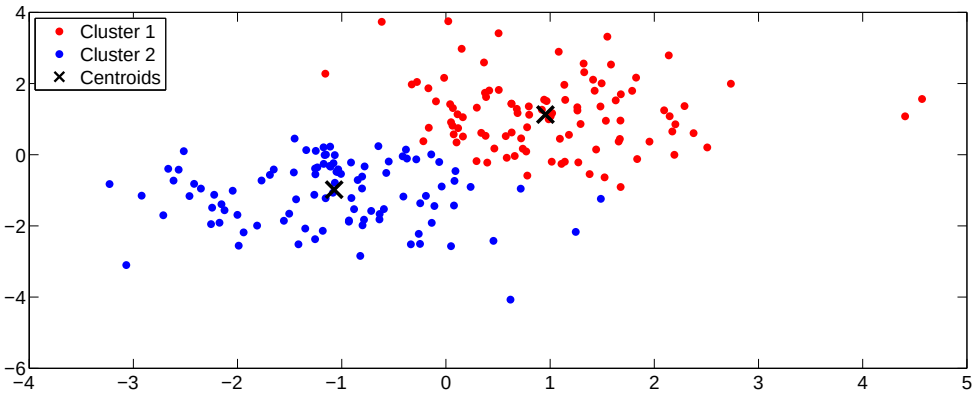

Figura 4.5: Objetos distribuídos em dois centros, com seus respectivos centroides.

Para determinar quantos centros seriam usados no processo de clusterização, utilizou-se o método Silhouette [40] para interpretação e validação dos centros. A técnica fornece uma representação gráfica sucinta de como cada objeto se encontra dentro de seu cluster e para cada objeto é atribuído um valor que está entre -1 e 1. Quando $S(i)$ está próximo de 1, significa que o dado está clusterizado adequadamente. Se $S(i)$ estiver próximo de -1, significa que deveria ser agrupado em outro cluster. Se $S(i)$ estiver próximo de zero, significa que ele está entre dois clusters. A função de *Silhouette*, $S(i)$, é descrita pela Equação (4.2).

$$S(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]} \tag{4.2}$$

onde, para cada objeto $S(i)$, $a(i)$ é a dissimilaridade média de i com todos os objetos do mesmo cluster, $b(i)$ é a dissimilaridade média miníma de i em relação ao objeto i com todos os objetos dos outros clusters. O método Silhouette assume que os dados tenham sido clusterizados com qualquer técnica, como por exemplo K-means. A Figura 4.6 ilustra a aplicação do método sobre um mesmo conjunto de dados, variando-se o número de centros, K, entre 2 e 7, sendo os mesmos identificados pelo eixo y de cada um dos seis gráficos apresentados.

Na Figura 4.6 observa-se que para essa amostra de dados referente a Figura 4.4, segundo o método Silhouette, o melhor número de centros a serem usados é 5, em que nenhum centro apresentou valor negativo. Todas as demais opções de centros apresentaram pelo menos um centro com valor negativo.

Com base no resultado do método Silhouette, foi realizada a clusterição da amostra de dados utilizando KM com 5 centros. Também foi realizada clusterização com 6

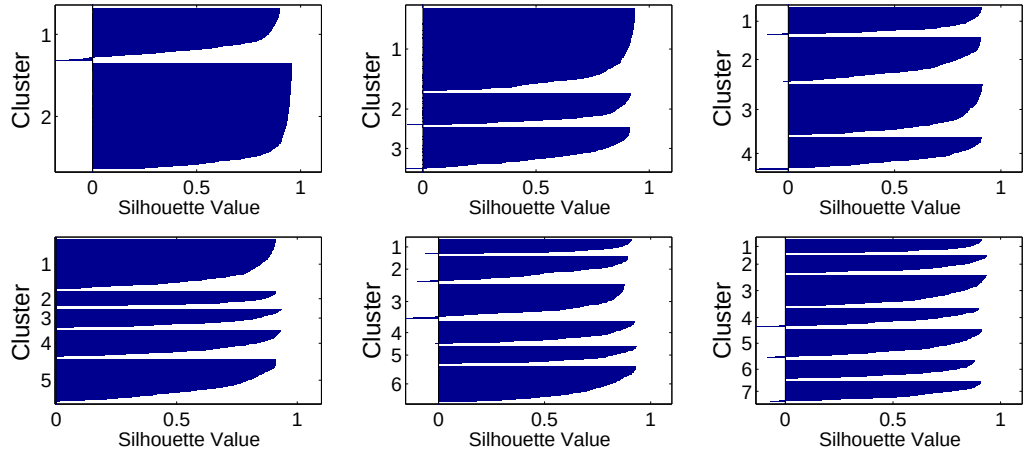

Figura 4.6: Silhouette aplicado a uma amostra de dados.

centros, objetivando verificar as possíveis diferenças na utilização de 5 e 6 centros para essa amostra. Conforme Figura 4.7, observa-se que utilizando 5 centros existe uma distribuição mais uniforme ao se comparar com 6 centros, em que o cluster 5 ficou desproporcional, em altura e comprimento, em relação aos outros clusters.

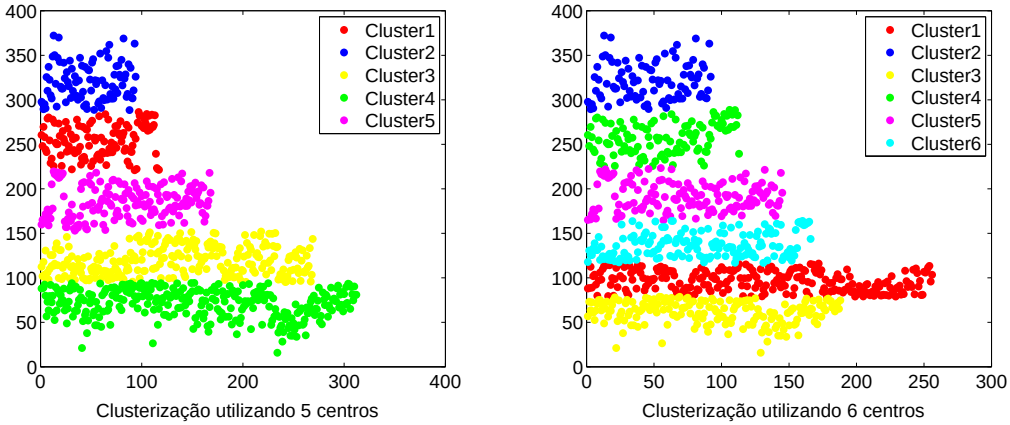

Figura 4.7: Clusterização dos objetos com 5 e 6 centros.

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

Nesta seção é apresentado o modelo para criação da Assinatura Digital, chamado de Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos e Clusterização K-means (DSNSF-KM) ou *Digital Signature of Network Segment Using Flow Analysis and K-means Clustering* (DSNSF-KM). Após a definição dos parâmetros de entrada do algoritmo DSNSF-KM, o algoritmo realiza a clusterização dos fluxos e faz o cálculo da

média ponderada dos clusters mais representativos.

Com a finalidade de classificar um cluster como mais ou menos representativo, definiu-se uma quantidade mínima de objetos para que o mesmo faça parte da assinatura, o valor é representado por γ. Essa definição é importante para evitar que possíveis anomalias ou *outliers* façam parte do processo de criação da assinatura. Por meio de clusterizações realizadas, concluiu-se que o número mínimo de pontos por cluster, ou seja o valor de γ fica em 2% em uma amostra com 300 pontos. Se o cluster não possui a quantidade mínima, é descartado e não será utilizado no cálculo da média ponderada. A média ponderada (4.3) dos clusters mais representativos é o resultado do algoritmo DSNSF-KM para cada instante.

$$\overline{X}_w = \frac{\sum_{k=1}^{K} c^k . p^k}{S} \tag{4.3}$$

em que c_k é o centro do cluster k, p_k é o número de objetos pertencentes ao cluster k, S é o número total de objetos clusterizados naquele intervalo.

Para se chegar ao valor de γ, foram realizados testes analíticos na rede da UTFPR - Câmpus Toledo, aplicou-se dois testes estatísticos para se obter uma avaliação numérica na definição do valor de γ. Os testes utilizados foram a Correlação e o Erro Quadrático Médio Normalizado (NMSE), sendo os mesmos explicados na seção 4.3. Como resultado da correlação e NMSE temos a Figura 4.8.

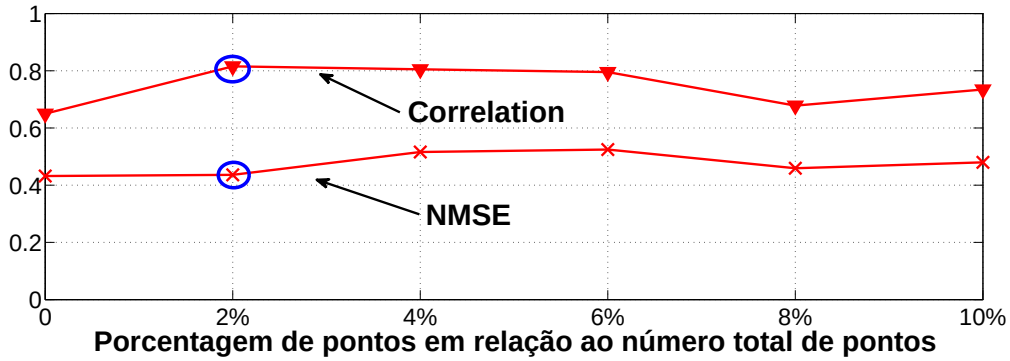

Figura 4.8: Definição do número mínimo de pontos pertencentes a um cluster.

Na Figura 4.8, com γ variando entre dois e seis por cento do total de pontos, obteve-se os melhores índices para correlação. Para o NMSE, γ obteve bons índices até dois por cento e de oito a dez por cento. Com essas informações definiu-se o valor de γ em 2%, pois representa os melhores índices entre os testes analíticos utilizados, desta forma definiu-se o número mínimo de pontos pertencentes a um cluster.

A Figura 4.9 procura ilustrar os passos do DSNSF-KM a partir do momento da entrada dos dados, passando pela clusterização, teste de pontos mínimos, e a conclusão com o próprio DSNSF-KM.

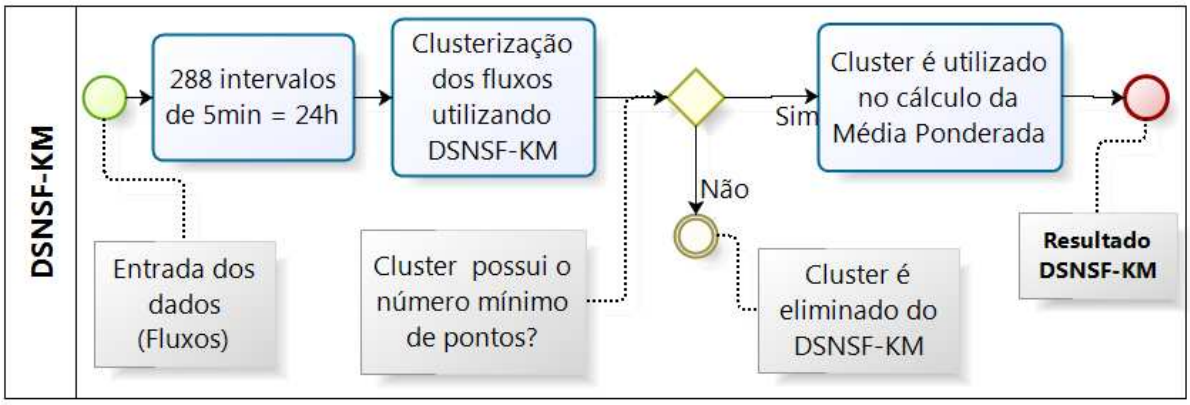

Figura 4.9: Execução do DSNSF-KM.

Conforme a Figura 4.9, o funcionamento do DSNSF-KM é o seguinte:

- Inicialmente é definido que tipo de informação será utilizada para criação da assinatura, como bits, pacotes, fluxos, etc;
- Faz-se a leitura dos fluxos utilizando-se o histórico dos mesmos, por exemplo, para geração da assinatura de uma segunda-feira é realizado a leitura de n segundas anteriores, conforme será definido no subcapítulo 4.3.5. Esse processo é feito para cada intervalo de cinco minutos, totalizando para um dia 288 intervalos;
- O próximo passo é a clusterização de cada intervalo de cinco minutos, que contém as informações referente ao históricos dos fluxos das n semanas;
- Em seguida é realizado um teste para verificar a quantidade mínima de pontos pertencentes a cada cluster. Se um cluster possui a quantidade ele é utilizado no cálculo da média ponderada. Caso contrário ele é excluído;
- Cálculo da média ponderada dos clusters mais representativos, ou seja, somente são utilizados os que possuem a quantidade mínima de pontos definidos pela variável γ;
- DSNSF-KM retorna o valor calculado para cada intervalo de cinco minutos, resultando na assinatura digital para o segmento;

O Algoritmo 4.2 demonstra o pseudocódigo para a Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos por meio do KM (DSNSF-KM).

Algoritmo 4.2 DSNSF-KM

DSNSF-KM algoritmo usado para criação da assinatura.

Entrada: S informação dos n dias ou semanas a serem clusterizados, K Número de clusters, I Número de repetições e γ número mínimo de objetos pertencentes a um cluster.
Saída: DSNSF-KM

Passo 1 Colocar k pontos no espaço que representa os objetos 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 dos k centroides.
Passo 4 Repita os passos 2 e 3 até que centroide não se mova mais ou o número de repetições seja excedido.
Se $\gamma <$ número de objetos do cluster k
$\overline{X}_w =$ média ponderada entre os clusters mais representativos
Fim Se
Retorna DSNSF-KM

Como pôde ser observado no algoritmo DSNSF-KM, os clusters mais representativos, ou seja que possuem mais objetos, são usados para compor a assinatura. Já os clusters menos representativos não são utilizados na composição. Esse procedimento garante que a assinatura não seja formada por objetos que desviam do comportamento normal do segmento como (*outliers*), evitando assim que possíveis anomalias possam influenciar na construção da assinatura.

Para demonstrar o motivo na definição de objetos mínimos, a Figura 4.10 lista os fluxos pertencentes a um determinado período, classificados em ordem decrescente por Bytes. Pode-se observar que o primeiro fluxo possui o valor 40,7MB, o segundo 12,8MB e a partir da terceira entrada os valores possuem uma discrepância menor e mais uniforme. Os contornos simbolizam a clusterização realizada com KM, utilizando-se 4 centros. Na figura somente são exibidas 3 cores que representam três centros, a primeira linha ficou isolada em um cluster 1 (verde escuro), da segunda até a sexta no cluster 2 (azul) e o terceiro cluster a pare em vermelho. Os dados do último centro possuem valores inferiores e estão ao final da lista, por isso não são exibidos.

A Figura 4.11 demostra a criação do DSNSF-KM para o mesmo intervalo da Figura 4.10. Pode-se observar que a utilização da variável γ é de extrema importância, pois é o mecanismo que evitará que o cluster 3 faça parta da assinatura, pois o mesmo possui apenas um ponto e está totalmente isolado dos outros clusters. Possivelmente esse ponto pode ser uma anomalia ou um *outlier* para esse instante nesse segmento de rede. Sua causa e consequência podem ser estudadas em trabalhos futuros.

```
** nfdump -M /usr/local/nfsen/profiles-data/live/fw-edge  -T  -r 2012/09/26/nfcapd.201209261100 -n 50 -s record/bytes
nfdump filter:
any
Aggregated flows 14924
Top 50 flows ordered by bytes:
```

Date flow start	Duration	Proto	Src IP Addr:Port		Dst IP Addr:Port	Packets	Bytes	Flows
2012-08-07 17:57:23.686	4294667.215	TCP	89.238.229.75:1935	->	10.1.1.10:21132	27670	40.7 M	1
2012-08-07 18:02:03.691	4294666.760	UDP	200.19.73.231:5247	->	10.1.1.10:35638	19775	12.8 M	1
2012-08-07 17:55:27.839	4294897.645	TCP	173.194.27.168:80	->	10.1.1.10:56179	3690	5.5 M	1
2012-08-07 17:57:39.676	4294666.205	TCP	173.192.94.176:554	->	10.1.1.10:51386	3542	5.0 M	1
2012-08-07 17:57:36.010	4294740.115	TCP	89.238.227.28:1935	->	10.1.1.10:41403	2760	3.6 M	1
2012-08-07 17:54:54.229	4294847.332	TCP	173.194.27.88:80	->	10.1.1.11:20657	2315	3.4 M	1
2012-08-07 17:58:27.043	4294673.079	TCP	74.125.234.65:80	->	10.1.1.11:20867	2051	2.9 M	1
2012-08-07 18:02:03.691	4294666.760	UDP	10.1.1.10:35638	->	200.19.73.231:5247	11513	2.7 M	1
2012-08-07 17:56:55.272	4294884.141	TCP	173.194.27.168:80	->	10.1.1.10:58447	1804	2.7 M	1
2012-08-07 17:59:57.733	4294705.207	TCP	74.125.234.68:443	->	10.1.1.11:54833	1703	2.3 M	1
2012-08-07 18:01:04.700	4294664.908	TCP	74.125.234.71:443	->	10.1.1.11:3364	1720	2.2 M	1
2012-08-07 17:55:12.598	4294831.355	TCP	10.1.1.10:5127	->	79.45.47.163:30685	1159	1.5 M	1
2012-08-07 17:56:42.355	4294932.451	TCP	200.19.73.116:80	->	10.1.1.10:10461	836	1.2 M	1
2012-08-07 17:54:34.092	4294830.893	TCP	184.169.73.33:80	->	10.1.1.10:50996	590	863961	1
2012-08-07 17:54:36.679	4294832.731	TCP	74.125.234.100:80	->	10.1.1.11:11112	600	842701	1
2012-08-07 17:59:22.608	4294667.120	TCP	69.127.231.235:23456	->	10.1.1.10:54861	595	719626	1
2012-08-07 17:57:23.686	4294667.215	TCP	10.1.1.10:21132	->	89.238.229.75:1935	15039	712560	1
2012-08-07 17:58:57.675	4294666.872	TCP	85.51.173.15:25167	->	10.1.1.10:14219	852	710010	1
2012-08-07 17:57:18.699	4294666.287	TCP	74.125.234.96:443	->	10.1.1.10:27457	538	700342	1
2012-08-07 17:58:04.641	4294666.625	TCP	189.25.192.231:43932	->	10.1.1.10:16498	852	675190	1
2012-08-07 17:57:52.449	4294667.055	TCP	151.55.95.204:31429	->	10.1.1.10:41484	835	673870	1
2012-08-07 18:00:55.521	4294667.026	TCP	86.67.7.242:62273	->	10.1.1.10:58010	678	670685	1
2012-08-07 17:58:33.588	4294666.185	TCP	89.84.18.20:36292	->	10.1.1.10:52163	893	653514	1
2012-08-07 17:57:38.574	4294665.768	TCP	190.44.230.59:37101	->	10.1.1.10:17555	790	652128	1
2012-08-07 17:59:14.230	4294863.002	TCP	200.19.73.171:443	->	10.1.1.10:25942	472	651689	1
2012-08-07 17:59:02.651	4294664.232	TCP	89.168.91.235:45935	->	10.1.1.10:48756	733	651612	1
2012-08-07 17:54:36.014	4294910.469	TCP	200.123.194.139:80	->	10.1.1.10:20356	476	647185	1
2012-08-07 17:59:20.670	4294666.504	TCP	212.122.69.120:55196	->	10.1.1.10:42565	904	643412	1
2012-08-07 17:59:33.605	4294665.880	TCP	37.11.19.174:6891	->	10.1.1.10:31402	563	635965	1
2012-08-07 17:59:18.581	4294667.644	TCP	10.1.1.10:61912	->	82.60.165.157:59317	738	634134	1
2012-08-07 17:58:46.245	4294669.908	TCP	62.147.128.1:6881	->	10.1.1.10:26502	690	631772	1
2012-08-07 17:59:04.531	4294667.176	TCP	177.100.63.20:56530	->	10.1.1.10:1799	662	631468	1
2012-08-07 17:57:22.311	4294667.134	TCP	181.14.188.220:20645	->	10.1.1.10:36531	475	631389	1
2012-08-07 17:58:57.614	4294666.131	TCP	82.161.138.100:60405	->	10.1.1.10:35103	530	628488	1
2012-08-07 17:59:04.697	4294666.286	TCP	77.70.121.24:49165	->	10.1.1.10:59746	521	623644	1
2012-08-07 17:58:43.200	4294670.365	TCP	24.62.7.29:63045	->	10.1.1.10:30718	530	622705	1
2012-08-07 18:00:25.582	4294665.984	TCP	78.129.52.126:44283	->	10.1.1.10:16263	609	621445	1

Figura 4.10: Intervalo de fluxos não uniforme.

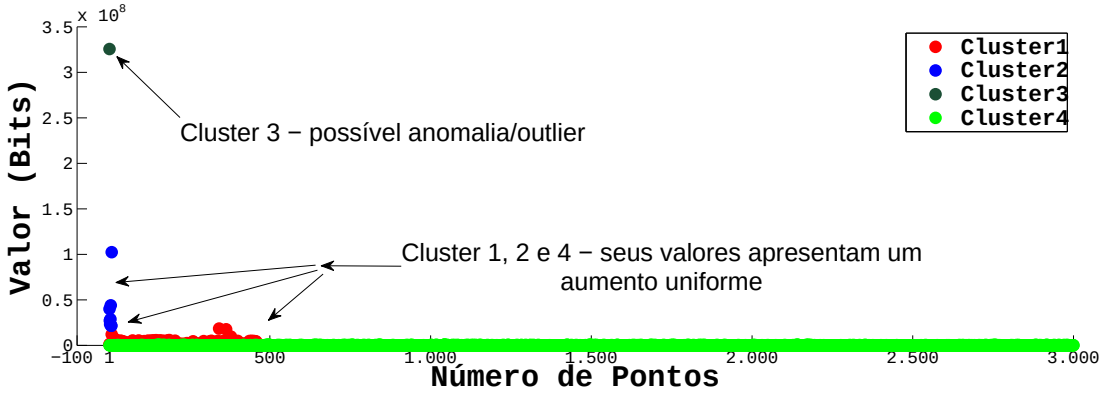

Figura 4.11: Clusterização dos fluxos com movimento não uniforme.

Ao se observar a Figura 4.12 que representa outro instante do mesmo segmento, verifica-se que ela apresenta valores mais uniformes e nenhum valor que destoe enormemente dos outros, como apresentado na Figura 4.10. Os contornos simbolizam a clusterização realizada com KM, utilizando-se 4 centros. Na figura somente são exibidas 3 cores que representam três centros, os dados do último centro possuem valores inferiores e estão ao final da lista, por isso não são exibidos.

```
** nfdump -M /usr/local/nfsen/profiles-data/live/fw-edge  -T  -r 2012/09/26/nfcapd.201209261235 -n 50 -s record/bytes
nfdump filter:
any
Aggregated flows 9360
Top 50 flows ordered by bytes:
Date flow start          Duration Proto      Src IP Addr:Port          Dst IP Addr:Port   Packets    Bytes Flows
2012-08-07 19:36:22.932 4294669.675 TCP     74.125.137.189:443   ->       10.1.1.10:16863     2277   2.4 M     1
2012-08-07 19:36:08.646 4294666.330 TCP      62.75.158.77:51413  ->       10.1.1.10:46353     1556   2.1 M     1
2012-08-07 19:35:52.607 4294666.410 TCP      82.141.83.9:46952   ->       10.1.1.10:21464     1421   1.8 M     1
2012-08-07 19:35:45.688 4294666.656 TCP     46.105.124.214:6911  ->       10.1.1.10:3464      1353   1.8 M     1
2012-08-07 19:35:51.531 4294666.779 TCP       86.4.159.6:6883    ->       10.1.1.10:45588     1345   1.8 M     1
2012-08-07 19:36:11.450 4294666.426 TCP    213.251.186.147:51413 ->       10.1.1.10:46103     1353   1.8 M     1
2012-08-07 19:35:45.634 4294666.710 TCP     86.139.228.116:6929  ->       10.1.1.10:31015     1960   1.8 M     1
2012-08-07 19:35:45.468 4294666.876 TCP      88.181.46.83:51413  ->       10.1.1.10:46281     1351   1.8 M     1
2012-08-07 19:35:46.383 4294666.879 TCP      94.251.77.181:2005  ->       10.1.1.10:30079     1322   1.8 M     1
2012-08-07 19:35:56.628 4294666.010 TCP     37.221.162.108:49010 ->       10.1.1.10:50190     1311   1.8 M     1
2012-08-07 19:36:06.461 4294665.346 TCP      89.22.97.149:7995   ->       10.1.1.10:36897     1314   1.8 M     1
2012-08-07 19:35:45.683 4294666.661 TCP     82.246.196.182:51413 ->       10.1.1.10:41553     1290   1.8 M     1
2012-08-07 19:35:59.249 4294666.839 TCP      37.144.52.254:51413 ->       10.1.1.10:13133     1323   1.8 M     1
2012-08-07 19:35:44.698 4294668.506 TCP      94.242.218.48:61078 ->       10.1.1.10:40320     1311   1.8 M     1
2012-08-07 19:35:44.571 4294668.767 TCP      94.23.211.202:61046 ->       10.1.1.10:5784      1290   1.7 M     1
2012-08-07 19:35:48.611 4294666.194 TCP     188.165.219.174:17501 ->       10.1.1.10:8863      1281   1.7 M     1
2012-08-07 19:35:48.699 4294666.125 TCP     217.85.143.39:35622  ->       10.1.1.10:9895      1233   1.6 M     1
2012-08-07 19:36:21.932 4294667.459 TCP       10.1.1.10:57761    ->    91.39.244.41:53413     2671   1.5 M     1
2012-08-07 19:35:50.672 4294667.016 TCP      71.63.133.94:9505   ->       10.1.1.10:19215     1480   1.1 M     1
2012-08-07 19:28:30.036 4294951.635 TCP     173.214.196.101:80   ->       10.1.1.10:57264      757   1.1 M     1
2012-08-07 19:28:05.732 4294932.565 TCP      200.201.88.35:80    ->       10.1.1.10:22556      593   847595    1
2012-08-07 19:32:20.501 4294666.517 TCP       10.1.1.10:35012    ->    81.97.168.221:63276     602   755696    1
2012-08-07 19:28:43.235 4294937.917 TCP     173.214.196.101:80   ->       10.1.1.10:35914      487   692917    1
2012-08-07 19:31:38.036 4294894.011 TCP      139.82.30.201:80    ->       10.1.1.10:14619      469   688755    1
2012-08-07 19:36:08.182 4294666.941 TCP      66.43.207.226:51413 ->       10.1.1.10:34568      742   653391    1
2012-08-07 19:31:37.954 4294893.775 TCP      139.82.30.201:80    ->       10.1.1.10:48283      434   633342    1
2012-08-07 19:31:04.887 4294935.105 TCP       10.1.1.15:8080     ->   177.133.206.40:49901     417   610254    1
2012-08-07 19:35:45.322 4294667.022 TCP      62.113.182.248:6987 ->       10.1.1.10:28574      668   538536    1
2012-08-07 19:28:04.963 4294934.434 TCP      200.201.88.35:80    ->       10.1.1.10:41157      368   528758    1
2012-08-07 19:35:50.650 4294666.547 TCP     220.233.65.10:13376  ->       10.1.1.10:13245      600   485607    1
2012-08-07 19:28:35.074 4294946.070 TCP     173.214.196.101:80   ->       10.1.1.10:52786      334   475382    1
2012-08-07 19:28:35.446 4294946.039 TCP     173.214.196.101:80   ->       10.1.1.10:62865      318   448614    1
2012-08-07 19:31:52.854 4294955.001 TCP      139.82.193.19:80    ->       10.1.1.10:49889      325   441028    1
2012-08-07 19:33:22.417 4294858.251 TCP     74.125.234.151:80    ->       10.1.1.10:36679      320   408849    1
2012-08-07 19:35:55.128 4294668.172 TCP     62.147.192.231:61421 ->       10.1.1.10:23298      537   403740    1
2012-08-07 19:36:12.651 4294666.356 UDP      64.94.18.78:17771   ->       10.1.1.10:36523      696   378818    1
2012-08-07 19:36:12.651 4294666.356 UDP       10.1.1.10:36523    ->    64.94.18.78:17771     1469   370180    1
```

Figura 4.12: Intervalo de fluxos uniforme.

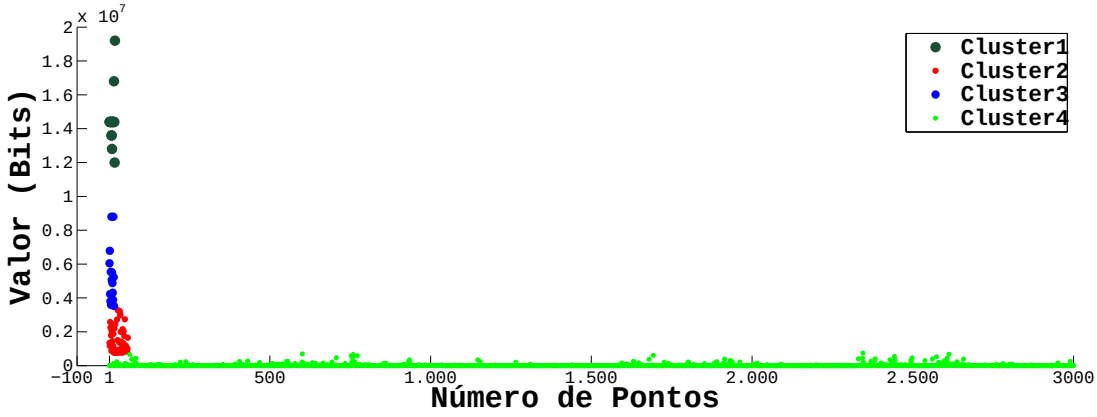

Figura 4.13: Clusterização dos fluxos com movimento uniforme.

A Figura 4.13 demostra a criação do DSNSF-KM para o mesmo intervalo da

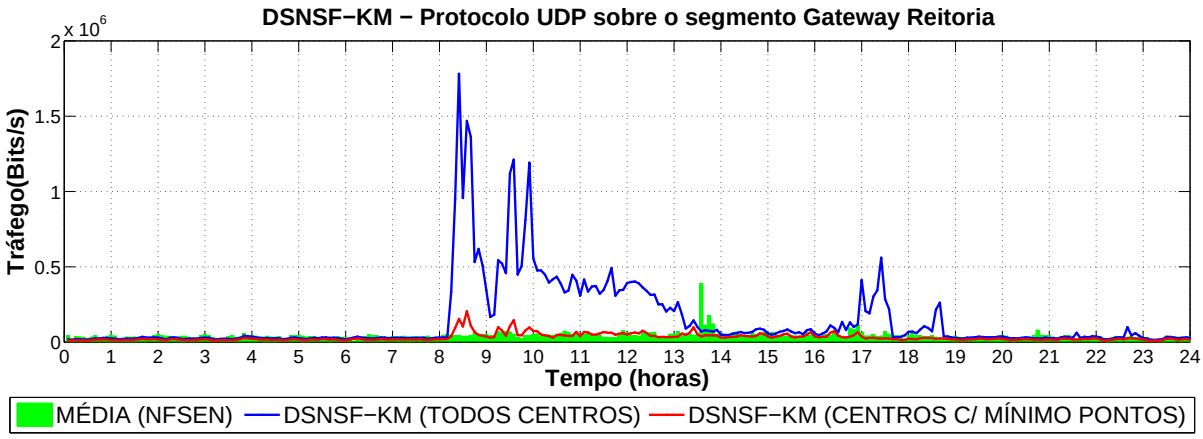

Figura 4.14: DSNSF-KM com e sem o parâmetro de pontos mínimos por centro.

Figura 4.12, observa-se que ela apresenta um comportamento mais uniforme dos seus clusters. A maior parte dos objetos estão agrupados ao cluster 4, o qual abriga os objetos de valores mais baixos do conjunto de dados. Os objetos com maiores valores estão distribuídos nos outros três clusters. Eles estão em menor número e seguem um comportamento uniforme, pois não existem discrepâncias entre os clusters ao se comparar com a Figura 4.11. Logo, não existe um único objeto nesse conjunto de dados que possa representar um *outlier*, segundo a definição de γ para os pontos mínimos pertencentes a um cluster. Portanto todos os centros seriam usados na criação da assinatura.

Para demostrar na prática a utilização ou não do parâmetro γ, a figura 4.14 aplica o DSNSF-KM de duas maneiras e o compara com o movimento gerado pelo NfSen. Na primeira é gerada a assinatura sem o parâmetro γ, nesse caso o DSNSF-KM gera a assinatura utilizando todos os centros independentemente do número de pontos e calcula a média ponderada. Para segunda maneira é gerada a assinatura com o parâmetro γ, nesse caso somente os clusters mais representativos foram usados para criação da assinatura. Conclui-se observando a Figura 4.14, que a utilização dos centros mais representativos evita que a assinatura seja influenciada por *outliers* e por anomalias, obtendo uma melhor previsão do movimento.

Os parâmetros de entrada usados na criação do DSNSF-KM podem ser alterados, pois o comportamento de um determinado campo do fluxo pode ser diferente de outro campo. Por exemplo, se for feita uma análise para os protocolos TCP, UDP, ICMP pode-se chegar a conclusão que os valores usados para o número de centros K, número de iterações e número mínimo de objetos pertencentes a um cluster γ tendem ser alterados. No entanto com os padrões adotados obteve-se resultados satisfatórios.

4.2.1 Complexidade do DSNSF-KM

Com relação a complexidade, DSNSF-KM possui ordem inicial de $O(X)$, tendo X como conjunto de objetos a ser agrupado. Para cada objeto X é feito o cálculo de associação. Como DSNSF-KM é baseado em centros, a cada atualização de centros é gerado um custo K, tornando a complexidade $O(XK)$. Outro parâmetro utilizado é o número de repetições/iterações I e a dimensão D dos dados coletados, ficando a complexidade da seguinte forma $O(XKDI)$.

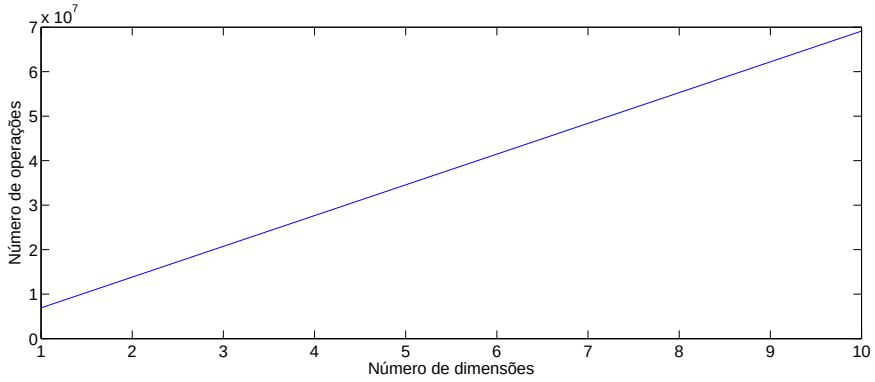

Figura 4.15: Simulação da complexidade do DSNSF-KM em número de operações efetuadas.

A Figura 4.15 é uma simulação para estimar o número de operações realizadas pelo DSNSF-KM. Os parâmetros utilizados foram com base em fluxos coletados em um segmento de rede num intervalo de vinte quatro horas, que corresponde a 86400 segundos. Dessa forma $X = 86400$, $K = 4$, $I = 20$, $D = [1, 10]$, onde D representa o número de campos extraídos dos fluxos coletados como por exemplo: IP e porta de origem e destino, contadores de entrada e saída (bytes, pacotes e fluxos) entre outros.

4.3 Testes Utilizados

Nessa seção, são definidas algumas métricas para avaliação das assinaturas geradas pelo DSNSF-KM quando comparadas ao movimento do tráfego de um segmento de rede.

4.3.1 Silhouette

Como já mencionado, para se determinar quantos centros seriam usados no processo de clusterização, utilizou-se do método Silhouette [40] para interpretação e validação dos centros. Foram realizados testes partindo de 3 até 6 centros. Após os testes, realizados

em amostras de cinco minutos de fluxos do segmento da UTFPR - Câmpus Toledo, o melhor resultado obtido foi utilizando 4 centros, conforme pode ser observado na Figura 4.16.

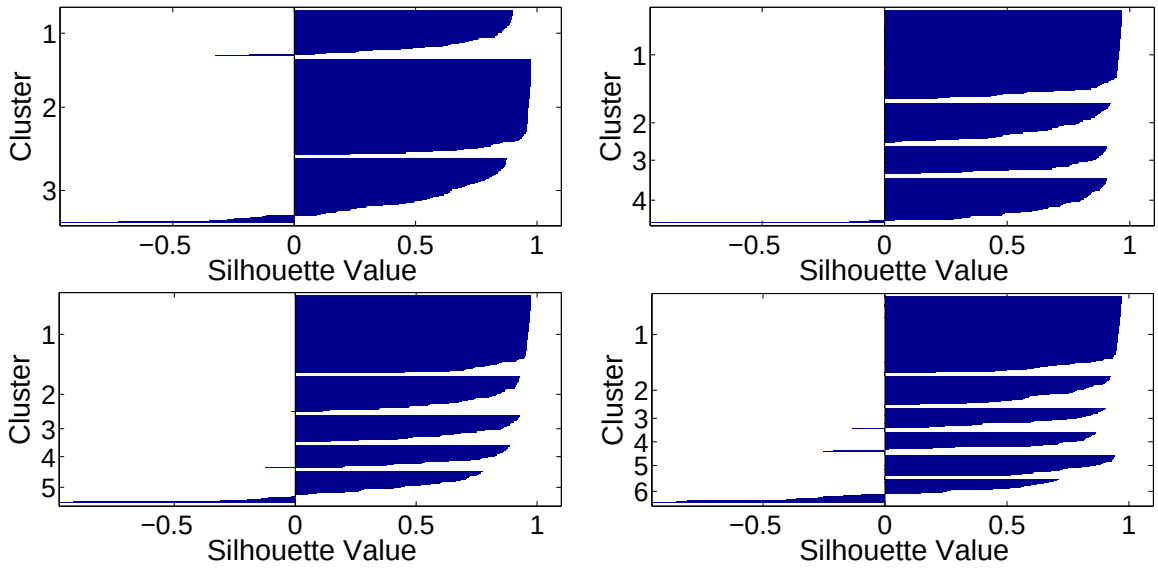

Figura 4.16: Silhouette - Método de interpretação e validação dos centros dos dados.

O resultado do método Silhoutte, na Figura 4.16, apresenta para 3 centros, uma pequena porção com valor negativo para o cluster 1 e para o cluster 3 um valor superior ultrapassando -0,5, o cluster 2 obteve um bom resultado. Ao se utilizar 4 centros, tem-se do primeiro ao terceiro bons resultados, já no 4 cluster uma pequena porção negativa. Para 5 centros, temos no quarto e quinto cluster um pequeno valor negativo e nos restantes um bom índice. Com 6 centros, observa-se que o número de centros com valor negativo aumenta apresentando 3 centros com valor negativo sendo o sexto com uma porção mais representativa. Analisando-se os resultados conclui-se que o método de Silhouette utilizando 4 centros obteve melhor distribuição dos dados sobre os centros e apresentou apenas um centro com uma pequena porção negativa.

4.3.2 Correlação

Dentre as ferramentas estatísticas utilizadas na avaliação do DSNSF-KM, a correlação indica a força e a direção do relacionamento entre duas variáveis, utilizou-se dela para avaliar se a assinatura consegue prever a tendência do movimento, no sentido de que para cada instante analisado, ela indica se o DSNSF-KM acompanha movimento de subida ou descida no segmento analisado. Como resultado, a correlação apresenta valores entre $(-1$ e $1)$, caso 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 uma pequena correção ou seja, o movimento do DSNSF-KM e o tráfego não são mais proporcionais. Caso o valor seja próximo de -1 indica que o DSNSF-KM está descorrelacionado ao movimento, ou seja se o movimento subir, o DSNSF-KM descerá e vice-versa.

A equação da correlação (4.4) realiza a correlação entre os vetores X e Y, em que X representa a assinatura gerada pelo DSNSF-KM e Y representa o movimento do dia/período em questão, retornando um valor entre -1 e $+1$ conforme explicado anteriormente.

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\mathbb{E}\{X.Y\} - \mathbb{E}\{X\}.\mathbb{E}\{Y\}}{\sqrt{\mathbb{E}\{(X - \overline{X})^2\}.\mathbb{E}\{(Y - \overline{Y})^2\}}} \tag{4.4}$$

onde σ_{XY} é o desvio padrão da multiplicação de X e Y, σ_X desvio padrão de X e σ_Y desvio padrão de Y, $\mathbb{E}\{X.Y\}$ esperança da multiplicação de X e Y, $\mathbb{E}\{X\}$ esperança de X e $\mathbb{E}\{Y\}$ esperança de Y.

4.3.3 Erro Quadrático Médio Normalizado (NMSE)

Outra ferramenta utilizada foi o Erro Quadrático Médio Normalizado (*Normalize Mean Squared Error* - NMSE). O NMSE é uma maneira de se quantificar a diferença entre valores estimados e o valor real obtido. Dessa forma utilizou-se dele para definir a proximidade do modelo DSNSF-KM com o movimento gerado pelo NfSen. Para esse quesito, pode-se interpretar que quanto menor o valor obtido mais próximo o modelo está do movimento real.

A Equação (4.5) realiza o cálculo do NMSE entre os vetores X e Y.

$$MNSE = \frac{\left(\sqrt{\sum_{s=1}^{S}(X - Y)^2}\right)^2}{\left(\sqrt{\sum_{s=1}^{S} Y^2}\right)^2} \tag{4.5}$$

onde S tamanho da amostra, X sendo DSNSF-KM e Y o movimento para o período em questão.

4.3.4 Desvio Fracional Padrão

Para checar a tendência do DSNSF-KM de subestimar ou sobreestimar o movimento, utilizou-se o Desvio Fracional Padrão (ou *Fractional Standard Deviation* - FSD). Seus índices variam entre (-2 e 2), ao tender-se para -2 significa que o DSNSF-KM está subestimado em relação ao movimento e ao tender-se para 2 significa que está sobreestimado em relação ao movimento. Para o FSD tem-se como medida ideal o valor zero.

A Equação (4.6) realiza o cálculo do desvio fracional padrão entre os vetores X e Y, nesse caso X sendo DSNSF-KM e Y, o movimento para o período em questão.

$$FSD = 2\frac{\sigma_X - \sigma_B}{\sigma_X + \sigma_Y} \tag{4.6}$$

onde σ_A é o desvio padrão do DSNSF-KM e σ_B é o desvio padrão do movimento.

4.3.5 Período de Aprendizagem

Nesta seção, é apresentada a forma como foi definido o número de semanas a serem utilizadas pelo DSNSF-KM para geração das assinaturas. Definiu-se que seriam analisadas de uma a dez semanas e os critérios de escolha foram correlação e NMSE, medidas essas que podem indicar se a assinatura possui uma relação com o movimento e verificar a distância entre a assinatura e o movimento, dessa forma pode-se avaliar a assinatura perante o movimento real.

Conforme já citado, para geração da assinatura os fluxos foram utilizados da seguinte maneira: extraiu-se o valor do campo bytes em seguida convertido em bits. Com essas informações, o DSNSF-KM faz a clusterização, aplica a regra de pontos mínimos pertencentes a cada centro, e em seguida calcula a média ponderada para cada instante de cinco minutos, resultando no valor que será considerado como comportamento normal ou característico para aquele instante no segmento. Foram realizadas análises de dez semanas de fluxos, durante 24 horas por dia, nas redes onde os testes foram realizados.

A Figura 4.17, representa a avaliação das dez semanas utilizadas na geração do DSNSF-KM, perante o movimento real dos segmentos utilizados para o teste de período de amostragem. Nessa figura, pode-se observar em relação a correlação que a partir da terceira semana já temos um resultado significativo, pois o valor fica em quase 0,8 aproximadamente e se mantém até a décima semana, contudo a sétima semana apresenta o melhor resultado, pois é a que está mais próxima de 1. Em relação ao erro quadrático, verifica-se da mesma

forma que a correlação, em que a partir da terceira semana os índices se estabilizam até a décima semana, porém a sétima semana representa novamente o melhor resultado chegando mais próximo de zero.

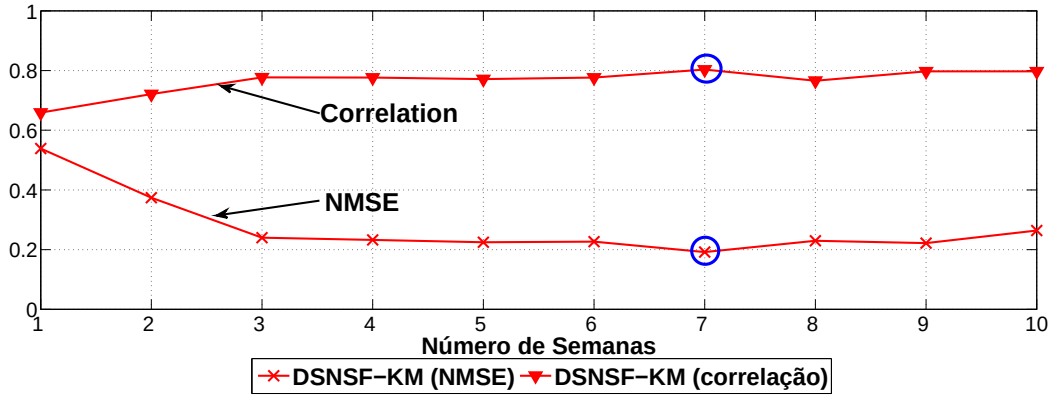

Figura 4.17: Avaliação do número de semanas a serem usadas no DSNSF-KM.

A diferença entre a terceira até a décima semana quando comparadas com a sétima é pequena, para correlação é entre 2% e 3%, para o NMSE é entre 3% e 4%. Entretanto observou-se que ao utilizar um número menor do que sete semanas para criação da assinatura, faz com que a assinatura gerada fique muito suscetível à influência em períodos que tenham, por exemplo: feriados ou recessos acadêmicos. Nesse caso optou-se por utilizar sete semanas.

Baseando-se nos resultados apresentados pela Figura 4.17, definiu-se que o número de semanas utilizadas pelo DSNSF-KM para criação das assinaturas seria sete semanas. Vale salientar também que novos estudos deverão ser realizados com outras redes e com fluxos coletados com base em amostragem, que poderão apresentar um resultado diferente, estes não foram coletados e estudados aqui por não terem sido escopo deste trabalho.

5 RESULTADOS

Este capítulo tem o objetivo de apresentar testes e resultados que visam avaliar a Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos e clusterização K-means (DSNSF-KM) em uma rede real. Para tanto, foram utilizados 5 cenários diferentes:

1) No primeiro criou-se uma assinatura para cada dia útil da semana, utilizando-se fluxos NetFlow e extraindo-se somente o tráfego do protocolo TCP para Bytes - convertido em bits -, pacotes e fluxos, baseado no histórico de sete semanas para cada dia útil;

2) No segundo cenário, o procedimento foi mesmo porém o protocolo utilizado foi UDP;

3) No terceiro cenário, seguiu-se o mesmo procedimento dos cenários anteriores, porém utilizou-se o tráfego total do segmento.

4) No quarto cenário, também utilizaram-se fluxos NetFlow e a assinatura foi criada para os protocolos TCP e UDP utilizando-se os campos Bytes - convertido em bits -, pacotes e fluxos. As assinaturas desse cenário foram aplicadas ao movimento de uma semana, em que ataques de negação de serviço foram realizados no segmento com objetivo de avaliar se os ataques poderiam ser identificados pela assinatura.

5) No quinto cenário, a assinatura foi gerada para cada dia útil da semana, utilizando-se fluxos sFlow e extraindo-se o tráfego total para Bytes - convertido em bits -, pacotes e fluxos, baseado no histórico de sete semanas para cada dia útil;

Os protocolos TCP e UDP foram definidos para os cenários 1, 2 e 4, pois representam mais que 90% do tráfego em uma rede conforme a pesquisa de Shawky *et al.* [3], além disso, realizou-se uma verificação nos cenários 1, 2, 3 e 4 utilizados e concluiu-se que a utilização dos dois protocolos citados passa de 95%.

Em relação à criação da assinatura somente para os dias úteis, pode-se justificar da seguinte maneira: os dias de sábado e domingo não foram considerados pois os

mesmos mantêm um tráfego diferente dos dias úteis. Pode-se observar esse comportamento na Figura 5.1, extraída do software NfSen e que expressa o movimento iniciando de domingo a sábado. O objetivo desse trabalho inicialmente é a caracterização do tráfego dos dias úteis.

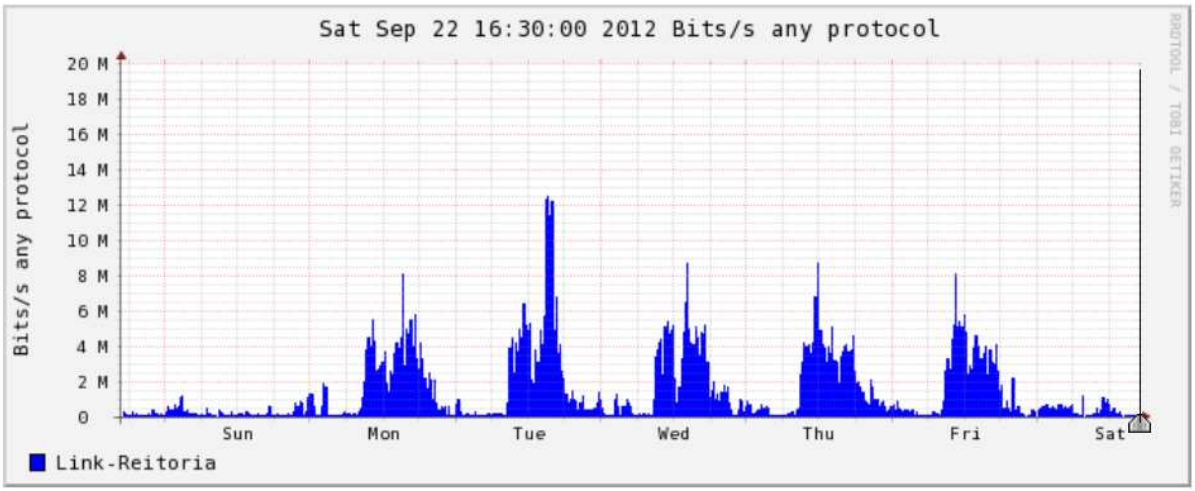

Figura 5.1: NfSen - Exibindo uma semana de tráfego entre Domingo e Sábado.

A Tabela 5.1 resume perante cada cenário: o período usado para criação da assinatura, padrão de fluxos utilizados, os testes realizados como: correlação, NMSE, sendo NMSE - Erro Quadrático Médio Normalizado (ou *Normalize Mean Squared Error*) e FSD - Desvio Fracional Padrão (ou *Fractional Standard Deviation*), para avaliar o DSNSF-KM em relação ao movimento gerado pelo NfSen e por último o tipo de tráfego utilizado para construção da assinatura no segmento de rede, separadas por protocolo TCP e UDP, e utilizando o tráfego total.

Tabela 5.1: Cenários utilizados para avaliar o DSNSF-KM.

			Testes			Tráfego		
	Período	Fluxos	Correlação	NMSE	FSD	TCP	UDP	Total
Cenário 1	7 sem	NetFlow	X	X	X	X		
Cenário 2	7 sem	NetFlow	X	X	X		X	
Cenário 3	7 Sem	NetFlow	X	X	X			X
Cenário 4	7 Sem	NetFlow	X	X	X	X	X	
Cenário 5	7 Sem	sFlow	X	X	X			X

Os fluxos foram coletados para os cenários 1, 2, 3 e 4 na Universidade Tecnológica Federal do Paraná (UTFPR) - Câmpus Toledo, no Gateway principal, conforme pode ser observado na Figura 5.2, e o período de coleta iniciou-se a partir de novembro de 2011. A rede da UTFPR em Toledo, possui aproximadamente 400 *hosts*, três links de internet totalizando 12 Mbps e aproximadamente 900 usuários entre alunos, técnicos administrativos e professores, seu horário de funcionamento é das sete até as vinte e três horas.

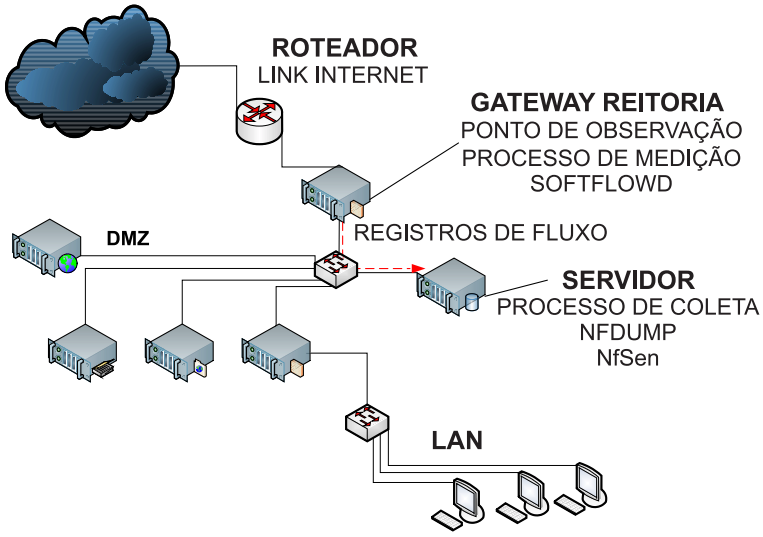

Figura 5.2: Rede em que os dados foram coletados

A Figura 5.2 identifica o cenário onde foi realizada a coleta dos fluxos, exportados 1:1, ou seja, não se utilizou de nenhuma técnica de amostragem. Todos os fluxos foram exportados pelo aplicativo Softflowd [41] instalado no gateway. O Softflowd é um analisador de rede capaz de exportar dados segundo o padrão NetFlow, por meio 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 são exportados na versão 9 do NetFlow e salvos em arquivos de cinco em cinco minutos em um servidor com sistema operacional CentOS 5.5 executando o aplicativo NFDUMP [37], o qual se trata de uma ferramenta para coleta e processamento de dados NetFlow. Juntamente com NFDUMP, foi instalado o NfSen [36], que é um *front end* para o NFDUMP, com objetivo de facilitar a visualização, busca e processamento dos fluxos coletados.

Para o cenário 5, os fluxos foram coletados na Universidade Estadual de Londrina (UEL). Foram exportados no padrão sFlow v5, com amostragem 1:256 por meio de um *switch* marca Extreme. Os fluxos sFlow são salvos da mesma forma que os fluxos NetFlow em um servidor com sistema operacional SUSE SLES 11.1 executando os aplicativos NFDUMP e NfSen.

Para cada cenário, o DSNSF-KM analisou sete segundas, terças, quartas, quintas e sextas-feiras de fluxos, com objetivo de extrair uma assinatura para cada dia útil da semana. Com essas informações o DSNSF-KM retorna para cada intervalo de cinco minutos, o valor que será considerado como comportamento normal ou característico para aquele intervalo no segmento.

5.1 CENÁRIO 1: DSNSF-KM - Bits, Pacotes e Fluxos - NetFlow - Protocolo TCP

Para esse cenário procurou-se avaliar o DSNSF-KM na construção das assinaturas especificamente para o protocolo TCP. Para isso, separou-se o tráfego desse protocolo e extraíram-se as seguintes informações: Bytes/s - convertido em bits/s -, pacotes/s e fluxos/s para geração das assinaturas.

A Figura 5.3 representa o movimento previsto pelo DSNSF-KM para o segmento em bits/s para o protocolo TCP, sobre o movimento gerado pelo NfSen para duas semanas subsequentes, correspondente aos dias 30 de abril a 04 de maio, organizados na primeira coluna, e 07 a 11 de maio de 2012 na segunda coluna. Observa-se que o DSNSF-KM permite descrever um comportamento padrão do segmento. A Figura 5.3 possui um gráfico para cada dia útil da semana, sendo cada um identificado em sua base com a data e o dia por extenso, também há no canto superior direito as letras (A, B, C, D e E) correspondendo a cada dia da semana.

Na Figura 5.3 observa-se ainda que durante o período da madrugada para alguns dias obteve-se um leve aumento do tráfego, enquanto os dias restantes mantiveram-se com o tráfego dentro do esperado. Em relação aos outros horários, pode-se observar que o DSNSF-KM identifica os horários de picos onde temos por exemplo: a chegada dos servidores e alunos a partir das sete horas e trinta minutos; os intervalos às dez, doze, quinze, dezessete e vinte horas; e por fim a retomada do movimento as dezenove até vinte duas quando o movimento começa a baixar com a saída dos alunos e servidores.

Observa-se no dia 01 de maio, em que o movimento ficou bem abaixo do tráfego esperado, esse comportamento se deve ao fato desse dia ser feriado nacional correspondente ao Dia do Trabalhador.

A Figura 5.4 representa o movimento previsto pelo DSNSF-KM para o segmento em pacotes/s para o protocolo TCP, sobre o movimento gerado pelo NfSen, referente ao mesmo período citado na Figura 5.3. Observa-se que o DSNSF-KM também permite descrever um comportamento padrão do segmento para os pacotes. Apesar dos valores serem diferentes ao se comparar bits e pacotes em seu eixo y, tendo em vista que os bits ficam em uma escala de Mb enquanto os pacotes não ultrapassam o valor de 1200 em sua escala, percebe-se que o comportamento dos mesmos é muito semelhante. Conforme pode-se identificar na figura 5.4 em alguns dias obteve-se um movimento superior ao DSNSF-KM durante a madrugada, como também foi constatado no DSNSF-KM para bits/s.

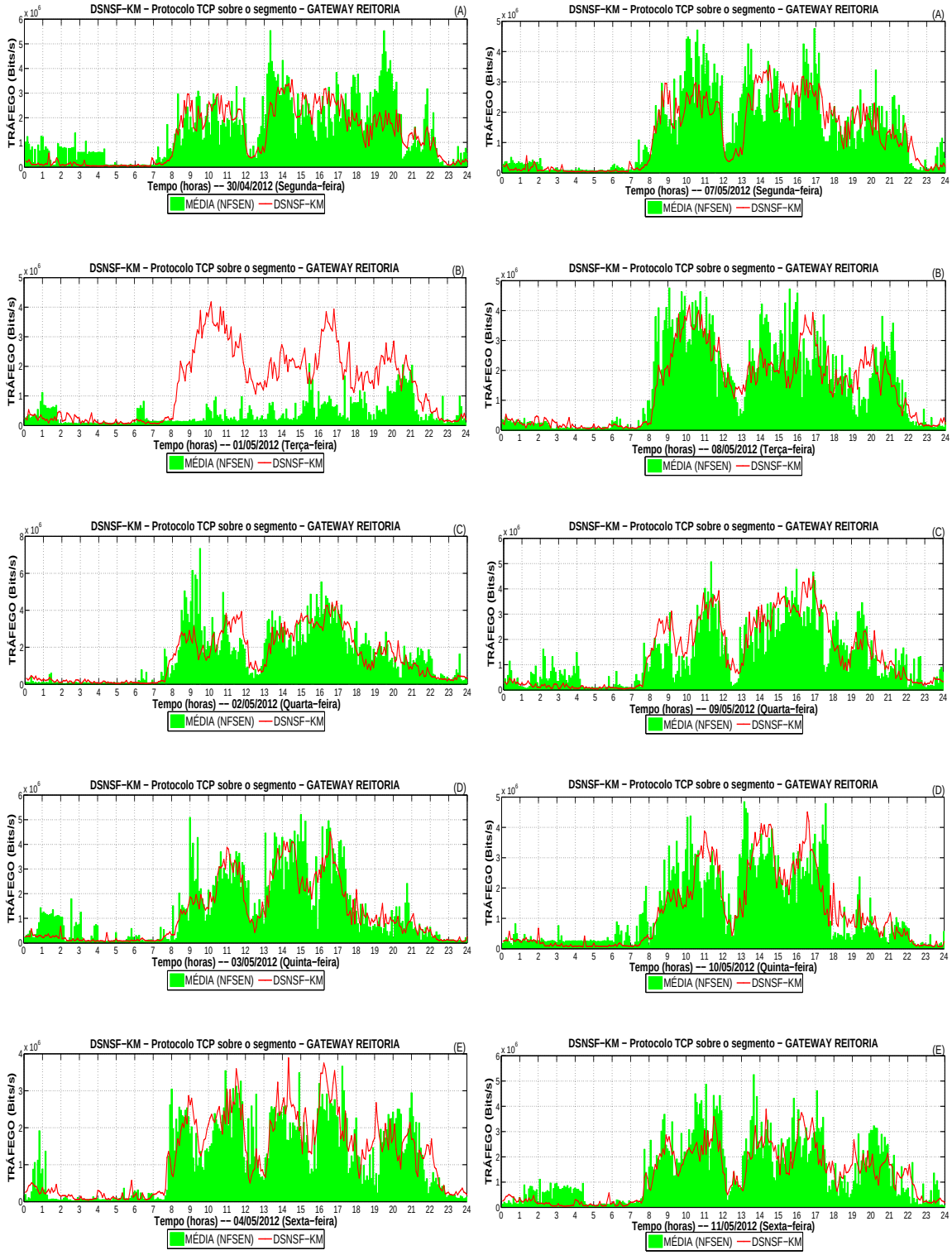

Figura 5.3: DSNSF-KM sobre o movimento registrado pelo NfSen para o protocolo TCP.

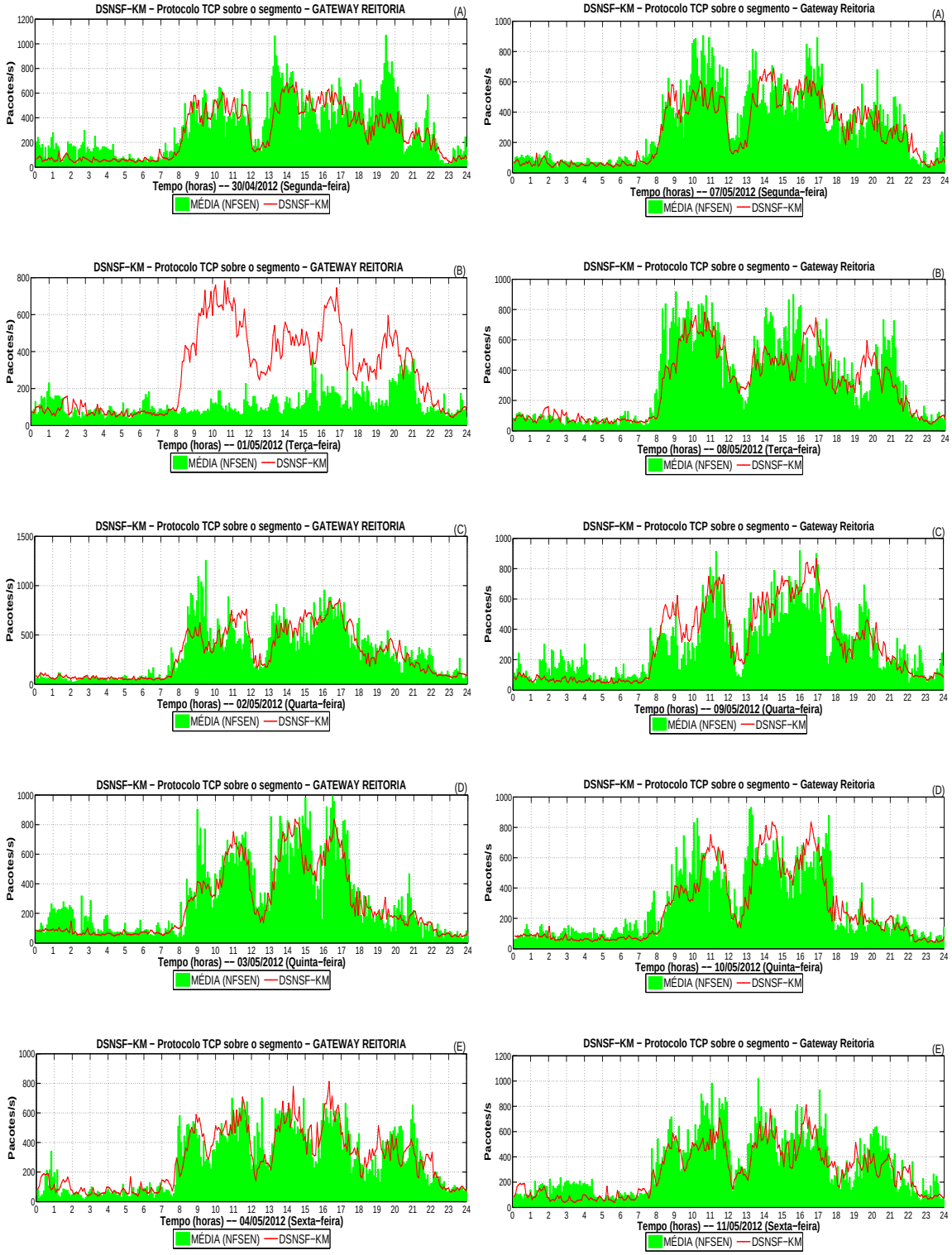

Figura 5.4: DSNSF-KM sobre os pacotes registrados pelo NfSen para o protocolo TCP.

Como se observa na Figura 5.4, que o movimento apresentou tanto na primeira coluna quanto na segunda pequenos picos que ultrapassam a assinatura, sendo que esse comportamento também pode ser verificado na Figura 5.3 com relação aos bits. Além dos picos, existem ocasiões em que o movimento fica um pouco abaixo da assinatura, essas poderiam ser consequência de uma queda em algum dos links da universidade, pois o GATEWAY REITORIA é composto por de 3 links, ou pelo baixo uso da rede mesmo. A queda no movimento devido ao feriado de 1 de maio pode ser identifica para os pacotes também.

A Figura 5.5 representa o movimento previsto pelo DSNSF-KM em relação ao número de fluxos TCP, sobre o movimento gerado pelo NfSen, referente ao mesmo período citado na Figura 5.3. Constatou-se que o comportamento dos fluxos é mais uniforme, seu eixo y possui uma variação menor do que apresentado por bits e por pacotes, ficando na faixa de 40 fluxos/s. Apesar de ser possível perceber o aumento dos fluxos a partir das oito até as dezessete horas, percebe-se que a variações em seu eixo y são discretas e diferentes das apresentadas para bits e pacotes. Pode-se concluir que bits e pacotes possuem uma correlação maior entre si quando comparados individualmente com fluxos.

No gráfico de quarta-feira dia 09, apresentado na Figura 5.5 pode-se observar um pico entre uma e duas horas que chega a quase 35 fluxos/s quando pelo DSNSF-KM não poderia passar de 15. Em relação aos outros dias e com exceção ao período citado na quarta-feira, a assinatura para os fluxos obteve bons resultados. Para terça-feira dia 01, correspondente ao feriado, observa-se que diferença entre o movimento e assinatura é visível no intervalo das oito horas até dezessete horas.

Para uma análise mais detalhada das assinaturas construídas pelo DSNSF-KM perante o movimento real gerado pelo NfSen, as Tabelas 5.2, 5.3 e 5.4, apresentam os resultados da correlação, erro quadrático médio normalizado e o desvio fracional padrão. As linhas das tabelas correspondem aos valores para bits, pacotes e fluxos. As colunas aos dias da semana de segunta até sexta-feira. Em cada intersecção entre as linhas e colunas são apesentados dois valores separados por dois pontos ":" representando os dois dias pertencentes as Figuras 5.3, 5.4 e 5.5. Por exemplo na intersecção da coluna Seg com a linha Bits/s, apresenta o valor para segunda dia 30 de abril e segunda dia 07 de maio.

Em se tratando da correlação entre o DSNSF-KM e o movimento, a Tabela 5.2 indica como a assinatura está relacionada com o movimento para cada dia útil da semana para bits, pacotes e fluxos.

Conforme a Tabela 5.2, tem-se para quarta-feira dia 09 de maio, o melhor índice chegando a 0,8776 para pacotes, e o pior índice de 0,0917 para fluxos na terça-feira dia

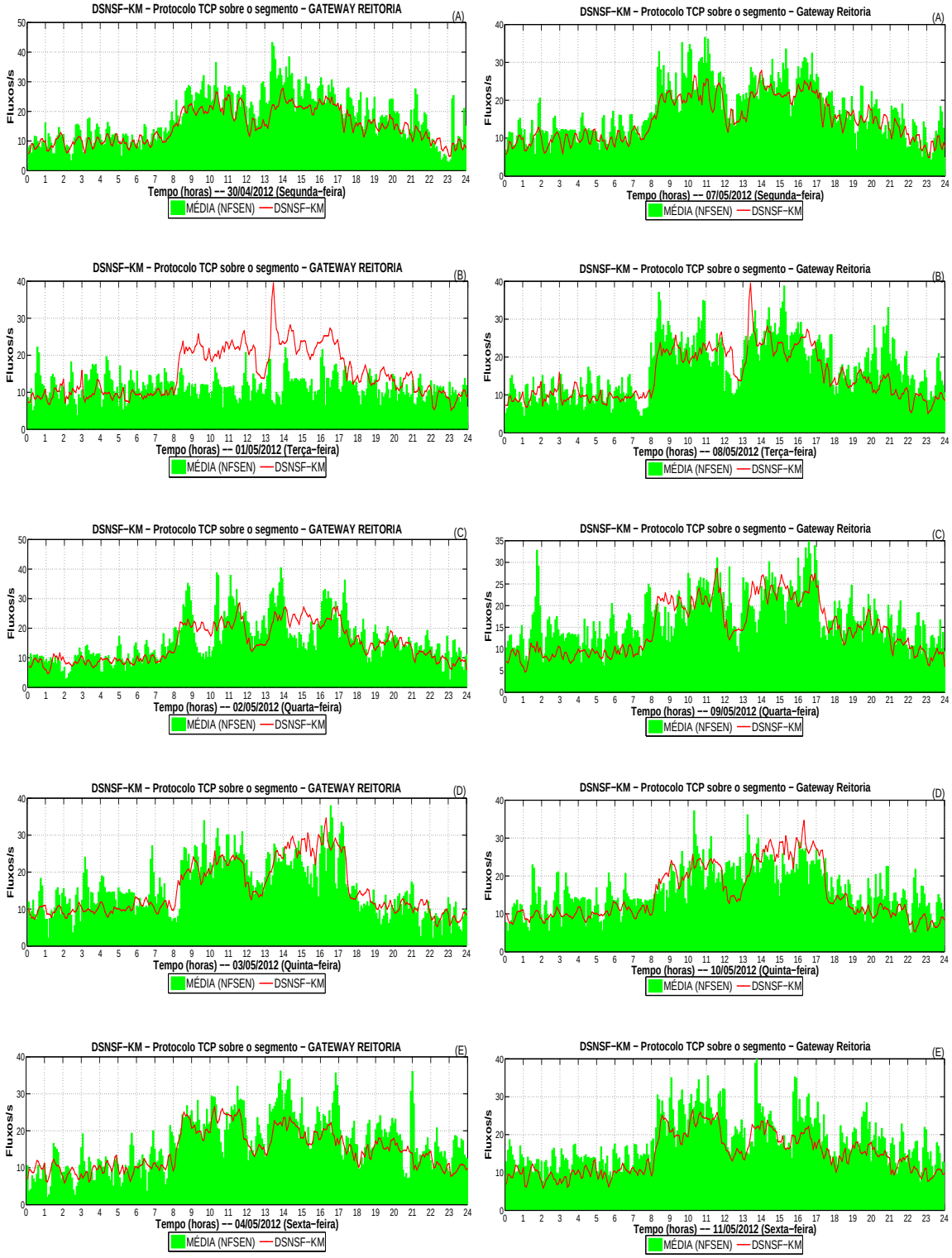

Figura 5.5: DSNSF-KM sobre os fluxos registrados pelo NfSen para o protocolo TCP.

Tabela 5.2: Correlação entre DSNSF-KM e o movimento do NFSEN para bits, pacotes e fluxos para o protocolo TCP.

	Seg	Ter	Qua	Qui	Sex
Bits/s	0,7589 : 0,8144	0,3156 : 0,8201	0,7922 : 0,8154	0,8439 : 0,7991	0,8089 : 0,7802
Pacotes/s	0,7963 : 0,8407	0,2936 : 0,8436	0,8269 : 0,8313	0,8776 : 0,8289	0,8460 : 0,8184
Fluxos/s	0,8644 : 0,8335	0,0917 : 0,7243	0,6892 : 0,6792	0,7624 : 0,7207	0,7510 : 0,7791

01 de maio. Observa-se que essa terça-feira obteve os piores índices devido ao feriado.

Para definir a proximidade do modelo apresentado com o movimento, o erro quadrático médio normalizado foi utilizados e os resultados são apresentados na tabela 5.3.

Tabela 5.3: Erro quadrático médio normalizado entre DSNSF-KM e o movimento do NFSEN para bits, pacotes e fluxos para o protocolo TCP.

	Seg	Ter	Qua	Qui	Sex
Bits/s	0,5053 : 0,5785	39,8111 : 1,1160	0,8676 : 1,1866	0,6903 : 0,6393	2,2207 : 0,8933
Pacotes/s	0,2134 : 0,1696	7,9111 : 1,0693	0,2197 : 0,3335	0,2769 : 0,1763	0,5364 : 0,2062
Fluxos/s	0,0928 : 0,0767	0,6820 : 0,1200	0,1585 : 0,0954	0,2467 : 0,0902	0,1795 : 0,0772

Em se tratando do NMSE na Tabela 5.3, a segunda-feira dia 07 de maio obteve o melhor índice chegando a 0,0767 para fluxos, e a terça-feira dia 01 de maio o pior índice com 39,8111 para bits, concluindo-se que para esse quesito obtiveram-se mais expressivos para fluxos, seguido de pacotes e bits.

A Tabela 5.4 apresenta os resultados, Para o desvio fracional padrão (FSD), a fim de indicar se o DSNSF-KM está subestimando ou sobre-estimando o movimento para cada dia da semana. Para o FSD, tem-se como medida ideal o valor zero.

Tabela 5.4: Desvio Fracional Padrão entre DSNSF-KM e o movimento do NFSEN para bits, pacotes e fluxos para o protocolo TCP.

	Seg	Ter	Qua	Qui	Sex
Bits/s	-0,1406 : -0,1602	0,9522 : -0,1965	-0,1405 : 0,1008	-0,1353 : -0,0288	-0,0089 : -0,2113
Pacotes/s	-0,1388 : -0,1623	1,0589 : -0,2091	-0,0898 : 0,1435	-0,1048 : 0,0114	-0,0564 : -0,1988
Fluxos/s	-0,3744 : -0,2421	0,6548 : -0,1611	-0,2339 : 0,0381	-0,0430 : 0,1605	-0,3309 : -0,2154

Ao se observar a tabela 5.4, verifica-se que a sexta-feira 04 de maio, obteve o melhor índice de -0,0089 para bits, e a terça-feira o pior índice com 1,0589 para pacotes, concluindo-se que para esse quesito obtiveram-se bons resultados tendo em vista que esse índice varia de -2 a 2.

Conclui-se ao fim desse cenário que o DSNSF-KM obteve uma boa previsão em relação as duas semanas analisadas, conforme pôde-se observar os gráficos que representam os dias para bits, pacotes e fluxos. Também por meio dos testes estatísticos apresentou um

bom desempenho. A criação de uma assinatura para um protocolo específico como apresentado nesse cenário, permite ao administrador da rede filtrar o tráfego de interesse a fim de facilitar a busca e resoluções de problemas.

5.2 CENÁRIO 2: DSNSF-KM - Bits, Pacotes e Fluxos - NetFlow - Protocolo UDP

Para esse cenário procurou-se avaliar o DSNSF-KM na construção das assinaturas especificamente para o protocolo UDP, para isso, separou-se o tráfego desse protocolo e extraiu-se as seguintes informações: Bytes/s - convertido em bits/s -, pacotes/s e fluxos/s para geração das assinaturas.

A Figura 5.6 representa o movimento previsto pelo DSNSF-KM para um segmento em bits/s para o protocolo UDP, sobre o movimento gerado pelo NfSen para duas semanas subsequentes, correspondente aos dias 30 de abril a 04 de maio, organizados na primeira coluna, e 07 a 11 de maio de 2012 na segunda coluna. Observa-se que o DSNSF-KM permite descrever um comportamento padrão do segmento para segunda, terça e quarta-feira na primeira coluna e para segunda, terça e sexta na segunda coluna.

Para quinta e sexta-feira da primeira coluna e quarta e quinta-feira da segunda coluna, verifica-se na Figura 5.6 que o comportamento do movimento real se destoa enormemente do previsto pelo DSNSF-KM. Ao se observar os outros dias da semana constata-se que o tráfego fica em torno de 50Kb/s e algumas vezes passa de 100Kb/s. O eixo y para esses dias fica com valor máximo de 200Kb/s a 250Kb/s, e para os dias com comportamento diferente da assinatura os valores desse eixo variam de 3Mb/s a 10Mb/s com duração aleatória. Entende-se que nesses períodos possa ter ocorrido alguma anomalia nesse segmento.

A Figura 5.7 representa o movimento previsto pelo DSNSF-KM para um segmento em pacotes/s para o protocolo UDP, sobre o movimento gerado pelo NfSen, referente ao mesmo período citado na Figura 5.6. Para esse quesito, o DSNSF-KM consegue prever o comportamento do segmento para segunda, terça e quarta-feira da primeira coluna, na quinta e sexta-feira existe um pico do movimento que destoa da assinatura durante um curto intervalo de tempo.

Nesta Figura 5.7, observa-se para quarta e quinta-feira da segunda coluna o movimento também destoa da assinatura, por um período superior a duas horas, da mesma forma que do que apresentado na Figura 5.6 relacionada ao movimento em bits/s. Observa-se ainda que o eixo y em geral não ultrapassaria em média 200 pacotes/s porém na quarta esse

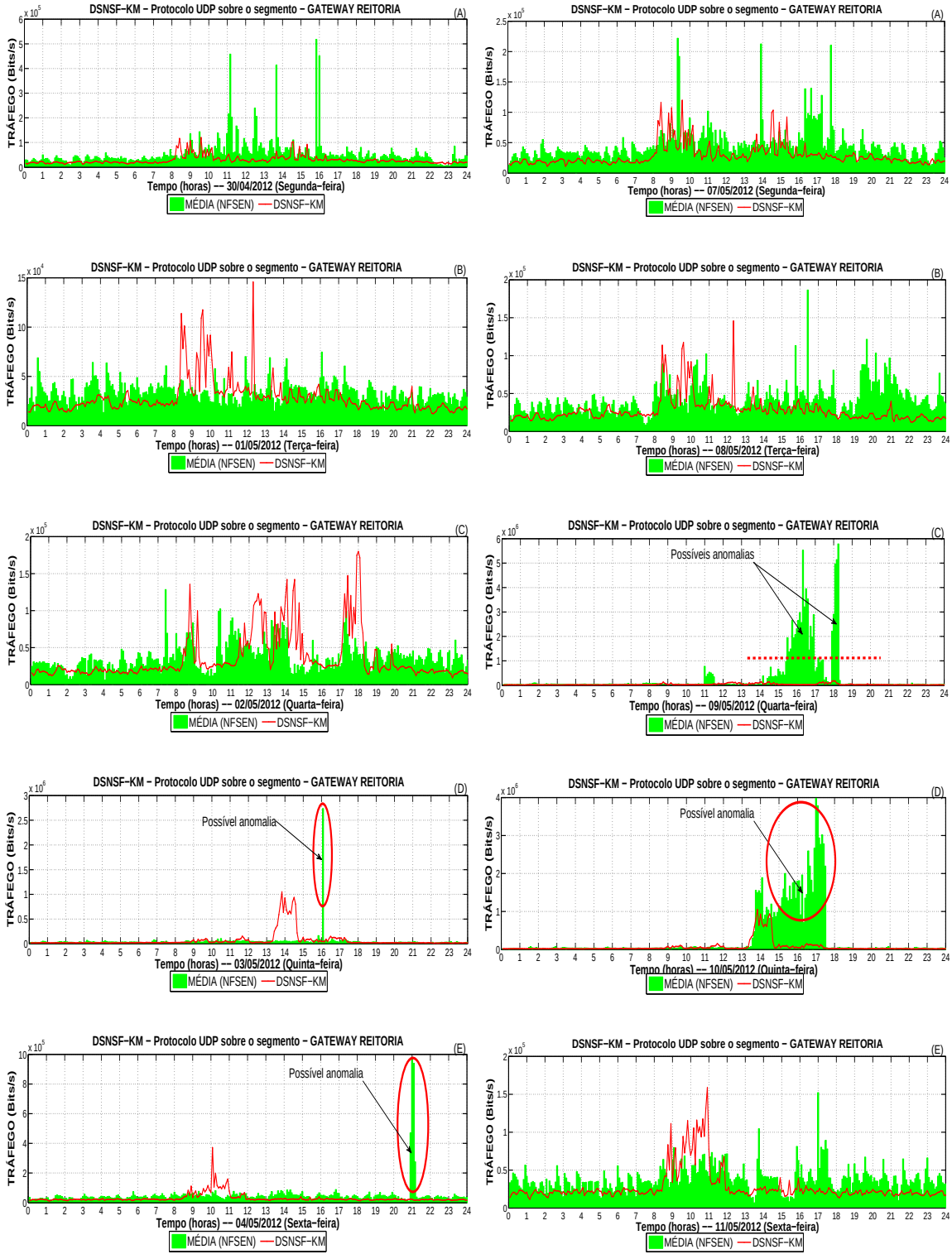

Figura 5.6: DSNSF-KM sobre o movimento registrado pelo NfSen para o protocolo UDP.

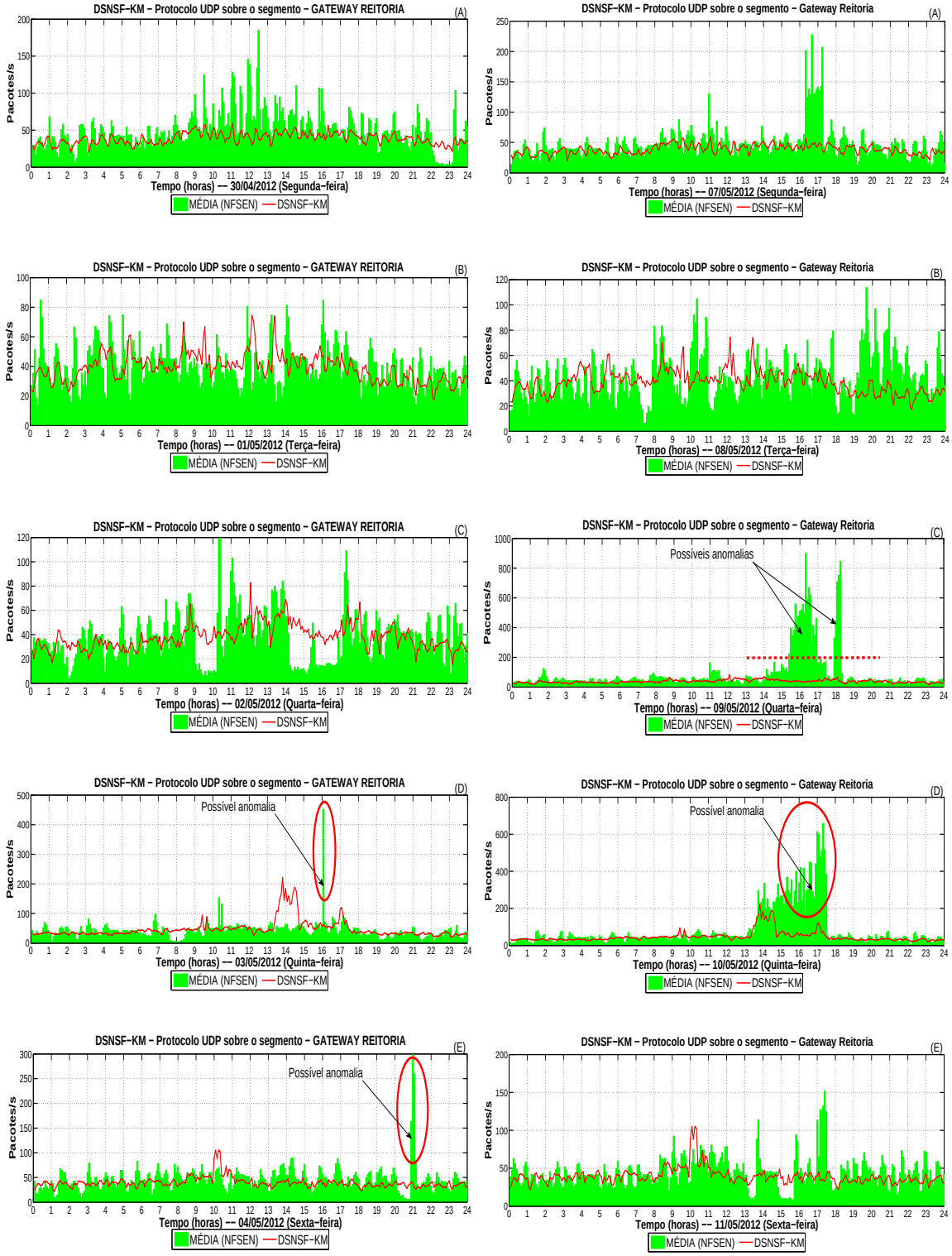

Figura 5.7: DSNSF-KM sobre os pacotes registrados pelo NfSen para o protocolo UDP.

valor chega a 1000 e na quinta a 700. Mais uma vez, entende-se que alguma anomalia poderia estar acontecendo nesse intervalo.

A Figura 5.8 representa o movimento previsto pelo DSNSF-KM para um segmento em fluxos/s para o protocolo UDP, sobre o movimento gerado pelo NfSen conforme já apresentado anteriormente. Verifica-se que o DSNSF-KM permite descrever um comportamento padrão do segmento para todos os dias da semana.

Ainda referente a Figura 5.8, verifica-se na primeira coluna para quarta, quinta e sexta-feira algumas quedas do movimento em relação a assinatura. Esse comportamento também é observado na segunda coluna para quarta-feira as dezessete horas e para sexta as treze e quinze horas. Essas quedas do movimento podem ser uma indicação de problema em algum dos links de Internet da universidade. Observa-se ainda que ao comparar a Figura 5.8 com as Figuras 5.6 e 5.7 não existe correlação entre os fluxos e bits ou fluxos e pacotes além de um pequeno aumento no número de fluxos no horário de funcionamento da universidade.

Observa-se que os picos/anomalias identificados nas Figuras 5.6 e 5.7, são resultados de clientes torrent realizando download na rede. Isso pode ser constatado verificando as informações nos fluxos e constatando-se que na maioria dos casos um único *host* foi responsável pela maior parte do tráfego gerado tanto para bits quanto para pacotes.

Novamente, para uma análise mais detalhada das assinaturas construídas pelo DSNSF-KM perante o movimento gerado pelo NfSen, as Tabelas 5.5, 5.6 e 5.7, apresentam os resultados da correlação, erro quadrático médio normalizado e o desvio fracional padrão.

Em se tratando da correlação entre o DSNSF-KM e o movimento, a Tabela 5.5 indica como a assinatura está relacionada com o movimento para cada dia útil da semana para bits, pacotes e fluxos.

Tabela 5.5: Correlação entre DSNSF-KM e o movimento do NFSEN para bits, pacotes e fluxos para o protocolo UDP.

	Seg	Ter	Qua	Qui	Sex
Bits/s	0,3613 : 0,3036	0,0224 : 0,1157	0,3239 : 0,2285	0,2000 : 0,3730	0,0073 : 0,2373
Pacotes/s	0,5850 : 0,3461	0,1819 : -0,0515	0,2185 : 0,1332	0,1869 : 0,5276	0,0862 : 0,2016
Fluxos/s	0,6320 : 0,2396	0,2395 : 0,7209	0,1401 : 0,2152	0,2793 : 0,3264	0,2026 : 0,2367

Conforme a Tabela 5.5, tem-se para terça-feira dia 08 de maio, o melhor índice chegando a 0,7209 para fluxos e o pior índice de -0,0515 para pacotes, nesse caso apresentando uma pequena descorrelação entre o DSNSF-KM e o movimento.

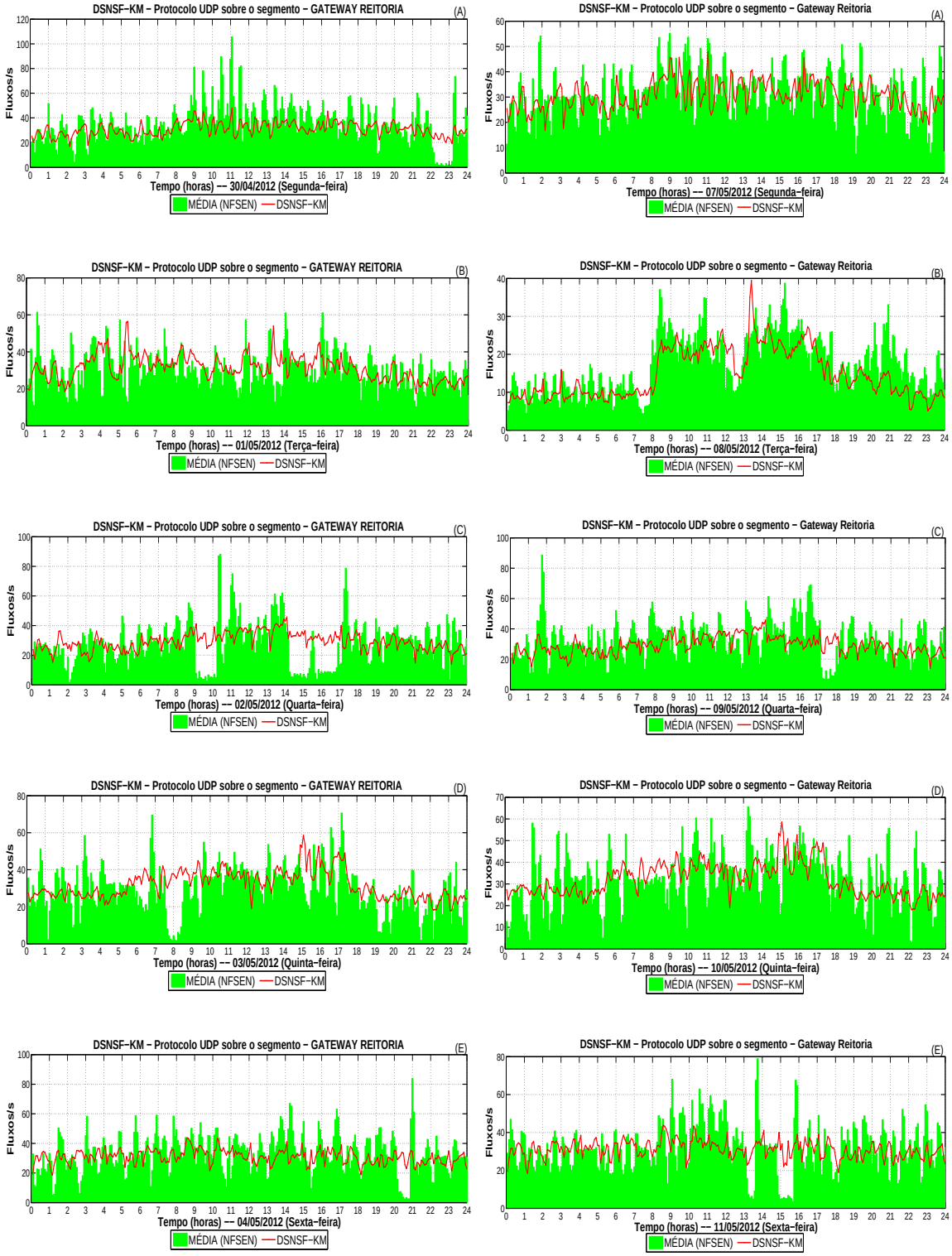

Figura 5.8: DSNSF-KM sobre os fluxos registrados pelo NfSen para o protocolo UDP.

Para definir a proximidade do modelo apresentado com o movimento, o erro quadrático médio normalizado foi utilizado e os resultados são apresentados na tabela 5.6.

Tabela 5.6: Erro quadrático médio normalizado entre o DSNSF-KM e o movimento do NFSEN para bits, pacotes e fluxos para o protocolo UDP.

	Seg	Ter	Qua	Qui	Sex
Bits/s	0,3613 : 0,2311	0,3611 : 0,3240	0,9713 : 0,6783	9,2228 : 0,4562	0,5069 : 0,4791
Pacotes/s	2,6244 : 0,1584	0,2216 : 0,4727	1,7265 : 0,2503	2,3689 : 0,4038	0,6801 : 0,5111
Fluxos/s	11,4877 : 0,1949	0,2327 : 0,1187	4,3963 : 0,3771	7,2226 : 0,6415	1,9907 : 1,1287

Em se tratando do NMSE na Tabela 5.6, a terça-feira obteve o melhor índice chegando a 0,1187 para fluxos, e a segunda-feira o pior índice com 11,4877, concluindo-se que para esse quesito obtivemos bons resultados, pois a maioria dos índices ficaram abaixo de 0,5.

A Tabela 5.7 apresenta os resultados para o desvio fracional padrão (FSD), a fim de indicar se o DSNSF-KM esta subestimando ou sobre-estimando o movimento para cada dia da semana.

Tabela 5.7: Desvio Fracional Padrão entre DSNSF-KM e o movimento do NFSEN para bits, pacotes e fluxos para o protocolo UDP.

	Seg	Ter	Qua	Qui	Sex
Bits/s	-1,0830 : -0,4493	0,5100 : -0,1580	0,5217 : -1,8591	-0,0389 : -1,2589	-0,9099 : 0,5991
Pacotes/s	-1,0915 : -1,1329	-0,3693 : -0,6692	-0,6865 : -1,7321	-0,0470 : -1,1900	-0,8629 : -0,6513
Fluxos/s	-0,9736 : -0,5435	-0,3747 : -0,1658	-0,8915 : -0,7338	-0,5002 : -0,4633	-0,9472 : -0,9381

Para o desvio fracional padrão conforme a Tabela 5.7, a quinta-feira dia 03 de maio, obteve o melhor índice chegando a-0,0389 para bits, e a quarta-feira dia 09 de maio, o pior índice com -1,8591. Os resultados indicam que o DSNSF-KM tendeu a subestimar o movimento, pois obteve-se somente três valores positivos enquanto os outros foram negativos.

Conclui-se, ao fim desse cenário, que o DSNSF-KM obteve uma boa previsão na comparação das duas semanas utilizadas, conforme pôde-se observar os gráficos que representam os dias das semanas para bits, pacotes e fluxos. Também por meio dos testes estatísticos, seu desempenho é considerado satisfatório, tendo em vista que o comportamento do protocolo UDP é mais linear/uniforme quando comparado ao comportamento do protocolo TCP. No entanto para o administrador é importante conhecer o comportamento dos protocolos pois, pode ser mais simples identificar algum comportamento anômalo por meio da separação dos mesmos.

5.3 CENÁRIO 3: DSNSF-KM - Total de Bits, Pacotes e Fluxos - NetFlow

Para esse cenário, procurou-se avaliar o DSNSF-KM na construção das assinaturas utilizando todo o tráfego do segmento, diferentemente dos cenários anteriores onde foram criadas assinaturas para os protocolos TCP e UDP. Nesse cenário extraíram-se as seguintes informações: Bytes/s - convertido em bits/s -, pacotes/s e fluxos/s para geração das assinaturas.

Mais uma vez, o DSNSF-KM analisou sete segundas, terças, quartas, quintas e sextas-feiras de fluxos, com objetivo de extrair uma assinatura para cada dia útil da semana. Com essas informações o DSNSF-KM retorna para cada intervalo de cinco minutos, o valor que será considerado como comportamento normal ou característico para aquele intervalo no segmento.

A Figura 5.9 representa o movimento previsto pelo DSNSF-KM para o segmento em bits/s, sobre o movimento gerado pelo NfSen para duas semanas subsequentes, correspondente aos dias 30 de abril a 04 de maio, organizados na primeira coluna, e 07 a 11 de maio de 2012 na segunda coluna. Na Figura 5.9 para segunda-feira dia 07 de maio, o movimento se manteve como esperado segundo o DSNSF-KM, tendo pequenas variações ou picos de curta duração. Para terça-feira dia 01 de maio, mais uma vez observamos que o movimento está distante do DSNSF-KM devido ao fato de ser feriado.

Na quarta-feira dia 09 de maio, tem-se pequeno aumento durante a madrugada, porém na faixa de 1Mb/s. Em seguida, o movimento diminui durante o início da manhã e se normaliza em seguida perante o DSNSF-KM. Nesse dia tem-se dois picos entre dezesseis e dezoito horas, que tendem a indicar uma possível anomalia principalmente na faixa de horário das dezoito horas onde tem-se um pico com duração média.

Quinta-feira dia 10 de maio, inicialmente tem-se uma excelente previsão. Na sequência a partir das oito horas um leve aumento do tráfego em relação ao DSNSF-KM que em seguida se normaliza, entre dezessete e dezoito horas temos um pico e no restante o movimento fica como previsto. Para sexta-feira dia 11 de maio, o comportamento do movimento frente ao DSNSF-KM, se iniciou um pouco superior durante a madrugada e depois seguiu como previsto pelo DSNSF-KM. Para sexta-feira pode-se perceber pequenos picos, porém eles não destoam enormemente do comportamento previsto pelo DSNSF-KM. Fazendo uma análise visual dos dias analisados na Figura 5.9, percebemos que o DSNSF-KM consegue descrever o comportamento da rede para as duas semanas com exceção do feriado de 01 de maio.

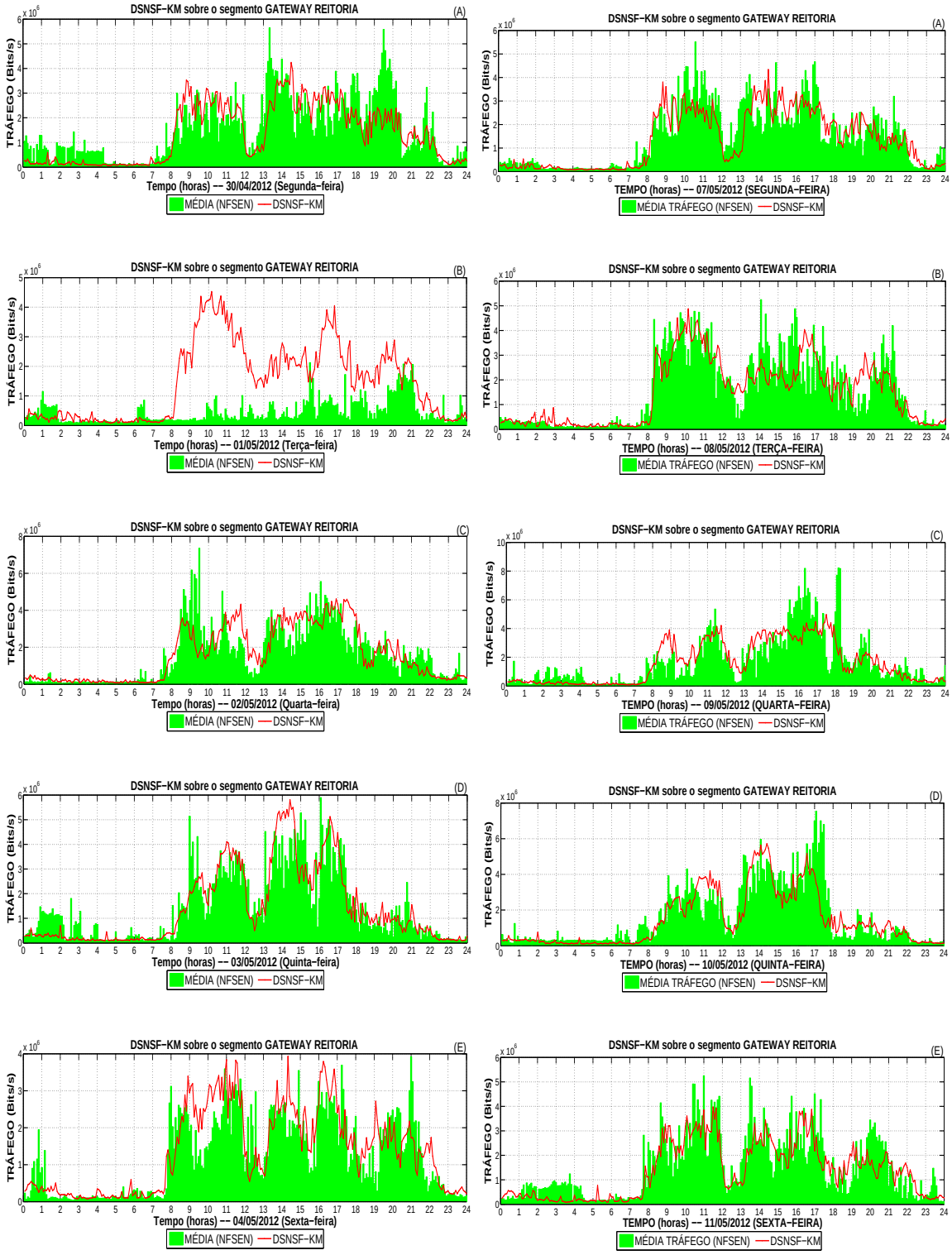

Figura 5.9: DSNSF-KM sobre o movimento registrado pelo NfSen.

A Figura 5.10 apresenta o movimento previsto para os pacotes em relação ao movimento gerado pelo NfSen, no mesmo período apresentado pela Figura 5.9. Analisando-se a Figura 5.10 percebe-se que o DSNSF-KM permite descrever o comportamento do tráfego de pacotes. Para ambas segundas-feiras, ele obteve um bom ajuste, com exceção de períodos com alguns picos, porém seguiram a tendência do DSNSF-KM.

Para os outros dias da semana na Figura 5.10 o DSNSF-KM também consegue descrever o comportamento da rede. Na quarta-feira dia 09 de maio, temos um pico às dezessete horas onde para o DSNSF-KM o movimento deveria cair, porém logo próximo as dezoito horas o movimento volta a seguir o comportamento previsto pelo DSNSF-KM. Novamente, observa-se que existe um comportamento semelhante entre bits e pacotes neste cenário.

A Figura 5.11 apresenta o movimento previsto referente aos fluxos em relação ao tráfego total, no mesmo período apresentado pela Figura 5.9. Em se tratando de fluxos, o DSNSF-KM consegue prever seu comportamento para todos os dias. Nas segundas-feiras, a assinatura parece estar visualmente mais ajustada com o movimento em relação aos outros dias.

Para quarta-feira, dia 09 de maio, existe um pico que destoa enormemente da assinatura entre uma e duas horas da manhã, e sexta dia 04 de maio, às vinte e uma horas, o DSNSF-KM prevê quarenta fluxos, porém o movimento ficou em cento e vinte fluxos. Esse pico poderia representar alguma anomalia. Nos outros dias não há nenhum pico nessa proporção.

Como nos cenários anteriores, para uma análise mais detalhada das assinaturas construídas pelo DSNSF-KM perante o movimento gerado pelo NfSen, as tabelas 5.8, 5.9 e 5.10, apresentam os resultados da correlação, erro quadrático médio normalizado e o desvio fracional padrão. Conforme explicado no cenário 1.

Em se tratando da correlação entre o DSNSF-KM e o movimento, a Tabela 5.8 indica como a assinatura está relacionada com o movimento para cada dia útil da semana para bits, pacotes e fluxos. Observa-se que, quanto mais próximo de um, melhor.

Tabela 5.8: Correlação entre DSNSF-KM e o movimento do NfSen.

	Seg	Ter	Qua	Qui	Sex
Bits/s	0,7600 : 0,8039	0,2753 : 0,2924	0,7710 : 0,7383	0,8443 : 0,8512	0,7944 : 0,7901
Pacotes/s	0,8001 : 0,8312	0,2636 : 0,8526	0,8233 : 0,7703	0,8700 : 0,8500	0,8204 : 0,8248
Fluxos/s	0,7626 : 0,5488	0,1904 : 0,2958	0,3087 : 0,4211	0,5203 : 0,5070	0,4311 : 0,4684

Conforme a Tabela 5.8, para quinta-feira dia 03 de maio, o melhor índice

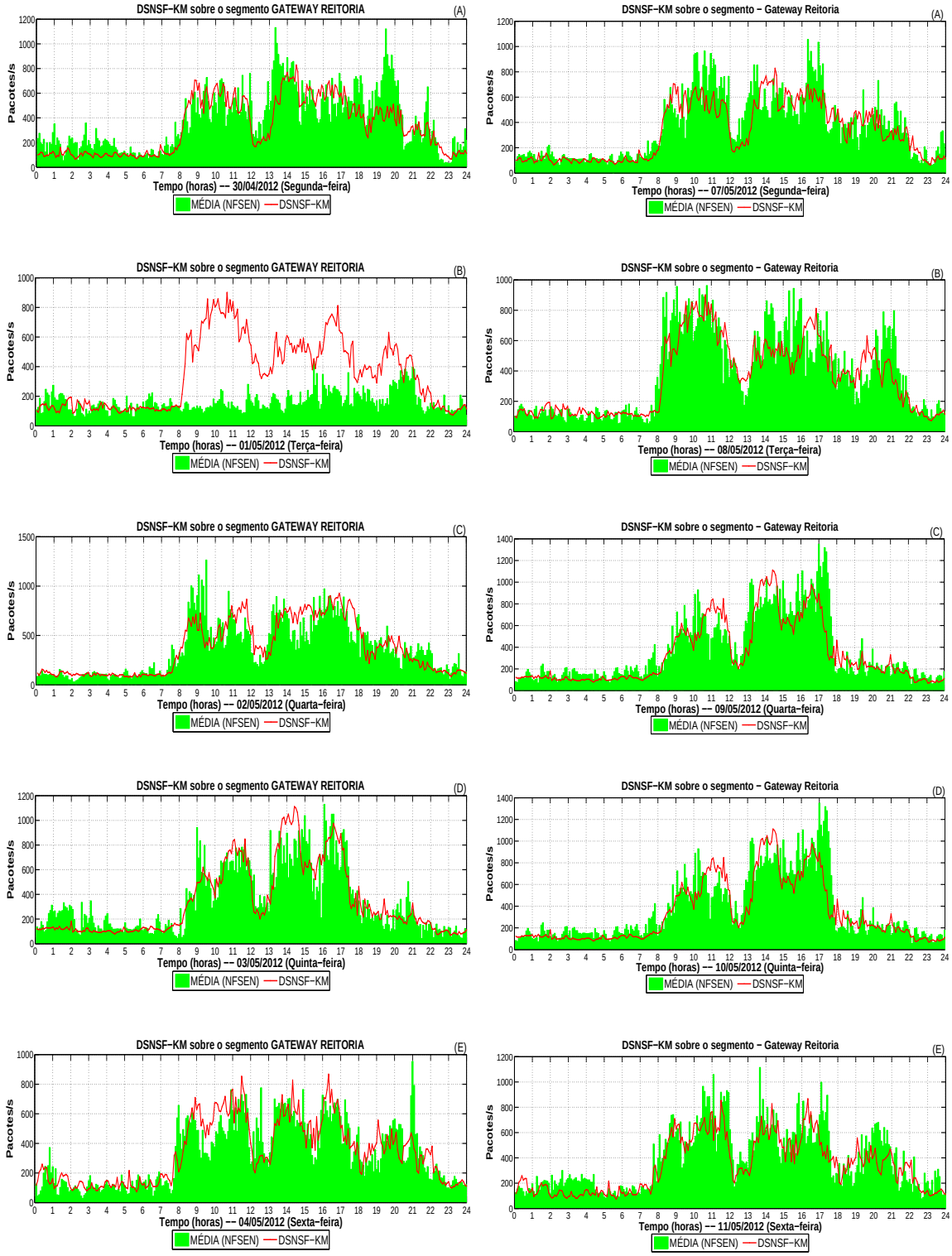

Figura 5.10: DSNSF-KM sobre os pacotes registrados pelo NfSen.

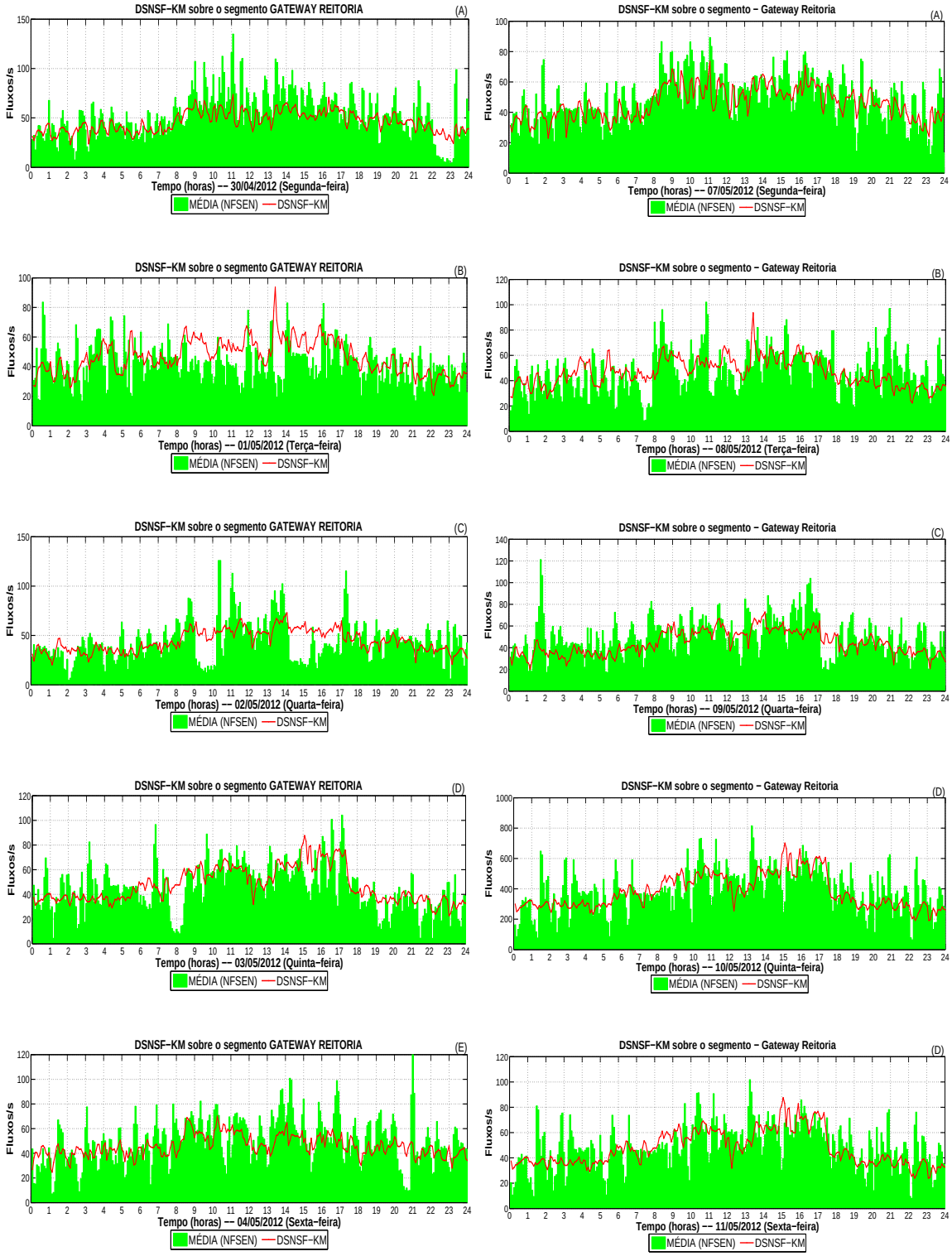

Figura 5.11: DSNSF-KM sobre os fluxos registrados pelo NfSen.

chegou a 0,8700. O pior índice conforme previsto, ficou para terça-feira dia 01 de maio com o valor de 0,1904 para fluxos. Pode-se concluir que para bits e pacotes todos os dias obtiveram uma boa correlação com exceção da terça-feira. Para fluxos as terças-feiras obtiveram os piores índices também.

Para definir a proximidade do modelo apresentado com o movimento, o erro quadrático médio normalizado foi utilizado e os resultados são apresentados na Tabela 5.9. Para esse quesito quanto mais próximo de zero melhor.

Tabela 5.9: Erro quadrático médio normalizado entre DSNSF-KM e o movimento do NFSEN.

	Seg	Ter	Qua	Qui	Sex
Bits/s	0,4289 : 0,4411	35,3100 : 1,1160	0,7977 : 0,9517	0,5648 : 0,6121	1,1713 : 1,4289
Pacotes/s	0,2047 : 0,1329	4,6346 : 0,1612	0,2320 : 0,1224	0,2929 : 0,1422	0,3314 : 0,1433
Fluxos/s	0,8059 : 0,1243	0,2568 : 0,3113	0,7789 : 0,1415	1,0771 : 0,2368	0,4213 : 0,1460

Em se tratando do NMSE, na Tabela 5.9, a quarta-feira dia 09 de maio, obteve o melhor índice chegando a 0,1224 para pacotes, e a terça-feira dia 01 de maio o pior índice com 35,3100 devido ao feriado. Concluindo-se que para esse quesito obteve-se bons resultados, pois a maioria dos índices estão abaixo de 0,45.

Para o desvio fracional padrão (FSD), a fim de indicar se o DSNSF-KM esta subestimando ou sobre-estimando o movimento para cada dia da semana. A Tabela 5.10 apresenta os resultados. Para o FSD tem-se como medida ideal o valor zero e variação de -2 a 2.

Tabela 5.10: Desvio Fracional Padrão entre DSNSF-KM e o movimento do NFSEN.

	Seg	Ter	Qua	Qui	Sex
Bits/s	-0,0676 : -0,0955	1,0148 : -0,1361	-0,0319 : -0,0379	0,0760 : -0,1162	0,0580 : 0,0484
Pacotes/s	-0,1013 : -0,1369	1,0917 : -0,1447	-0,0045 : -0,0856	0,0556 : -0,0842	-0,0194 : -0,1546
Fluxos/s	-0,7642 : -0,4110	-0,1638 : -0,4568	-0,6227 : -0,4705	-0,3062 : -0,2278	-0,7780 : -0,7391

Conforme os resultados apresentados na tabela 5.10, conclui-se que na quarta-feira dia 02 de maio obtivemos o melhor índice de -0,0045 e o pior de 1,0148 para terça-feira dia 01 de maio. Os valores apresentados pelo FSD indicam que o DSNSF-KM permite ter uma boa previsão do volume para bits, pacotes e fluxos, tendo em vista que a maioria dos índices ficaram muitos próximos de zero.

A aplicação do DSNSF-KM para criação da assinatura neste cenário, apresentou bons resultados para as duas semanas analisadas, possibilitando que o comportamento do segmento seja previsto tanto para bits, pacotes e fluxos. Ele pode permitir ainda a automatização da tarefa de monitoramento dos segmentos, tornando-se um parâmetro chave na detecção de possíveis anomalias.

5.4 CENÁRIO 4: DSNSF-KM - Bits, Pacotes e Fluxos - NetFlow - Protocolo TCP e UDP - Injeção de ataques DoS.

Para esse cenário, procurou-se avaliar o DSNSF-KM na construção das assinaturas separando o tráfego dos protocolos TCP e UDP. Nesse cenário extraiu-se as seguintes informações: Bytes/s - convertido em bits/s -, pacotes/s e fluxos/s para geração das assinaturas. O período de coleta foi durante o mês de fevereiro, março e abril de 2012, para geração das assinaturas que em seguida foram comparadas com o movimento durante o período de 9 e 13 de abril de 2012, na Universidade Tecnológica Federal do Paraná – Câmpus Toledo.

Neste cenário, da mesma forma que os cenários anteriores, utilizou-se o período de um dia inteiro (24 horas). Permitindo que ataques experimentais do tipo DoS (*Denial of Service*) pudessem ser realizados em horários onde a rede não estivesse sendo utilizada a fim de não interferir nos trabalhos da universidade. Apesar dos ataques influenciarem nos resultados de correlação, NMSE e FSD, eles foram importantes no sentido de se verificar o comportamento da rede perante a assinatura. Além dos ataques durante esse cenário, foi observado um comportamento que se diferenciava do normal durante alguns instantes para o protocolo UDP, esse será explicado e observado no decorrer desta seção, sendo o mesmo comportamento identificado para o cenário 2.

Partiu-se para criação das assinaturas exibidas nas Figuras 5.12, 5.13, 5.14, 5.15, 5.16 e 5.17, que 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. As figuras possuem um gráfico para cada dia útil da semana sendo, cada um identificado no canto superior direito como (A, B, C, D ou E) e também por extenso na base dos gráficos correspondente a cada dia. Como pode ser observado nas figuras, o modelo DSNSF-KM permite descrever o comportamento da rede. Mais uma vez, pode-se verificar que o comportamento do protocolo UDP é diferente do TCP, tendo uma menor variação no eixo y no que se refere ao volume de tráfego.

Observando a Figura 5.12 para quarta-feira (C), observa-se que o DSNSF-KM está bem correlacionado com o movimento até as dez horas. Entre dez e onze há uma descorrelação pois o movimento sobe e o DSNSF-KM desce. No período da tarde a partir das quinze horas há um aumento do movimento até as dezoito e trinta aproximadamente. Na quinta-feira (D), pode-se considerar que o movimento manteve-se dentro do esperado pelo

DSNSF-KM.

Ainda em relação a Figura 5.12 para os dias de segunda, terça e sexta-feira, objetivando-se analisar o comportamento do DSNSF-KM frente ao movimento e também as anomalias que possam existir, realizou-se alguns experimentos controlados que podem ser observados, tendo-se o conhecimento que eles influenciariam nos resultados para correlação, NMSE e FSD.

Como pode ser observado nas Figuras 5.12, 5.13 e 5.14. Nos gráficos que correspondem à segunda-feira (A) dia 09/04/2012, há uma diferença entre o movimento e o DSNSF-KM no período em torno das sete horas da manhã, quando foi realizado, por meio da ferramenta LOIC [42], um DoS (*Denial of Service*) entre 6h45min e 7h05min totalizando, 9494 fluxos, 159803 pacotes, e 7,2MB de tráfego gerado.

Nos gráficos que correspondem à terça-feira (B) 10/04/2012, nas Figuras 5.14, 5.13 e 5.12, durante o período entre 1h30min e 3h, onde foi realizado, por meio da ferramenta LOIC, um DoS 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.

Nos gráficos que correspondem à sexta-feira (E) 13/04/2012, nas Figuras 5.14, 5.13 e 5.12, durante o período entre 0h e 8h, quando foi realizado, por meio da ferramenta LOIC, um DoS em três etapas: primeiro entre 7min e 39min, segundo 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.

Conforme foi apresentado nas Figuras 5.12, 5.13 e 5.14 os ataques DoS gerados por meio da ferramenta LOIC podem ser facilmente identificados pois destoam da assinatura gerada pelo DSNSF-KM. Para um sistema de detecção de anomalias eficiente, a caracterização do tráfego é um passo fundamental que deve ser realizado com o objetivo de identificar o padrão e estabelecer limiares que serão considerados normais para cada segmento gerenciado.

A Figura 5.15 representa o DSNSF-KM aplicado sobre o tráfego do protocolo UDP conforme já citado. Durante os experimentos de criação do DSNSF-KM para o protocolo UDP, observou-se que ocorreram picos que se diferenciavam enormemente do comportamento apresentado para segunda, quarta, quinta e sexta-feira. Sendo esses identificados como possíveis anomalias nos gráficos correspondentes.

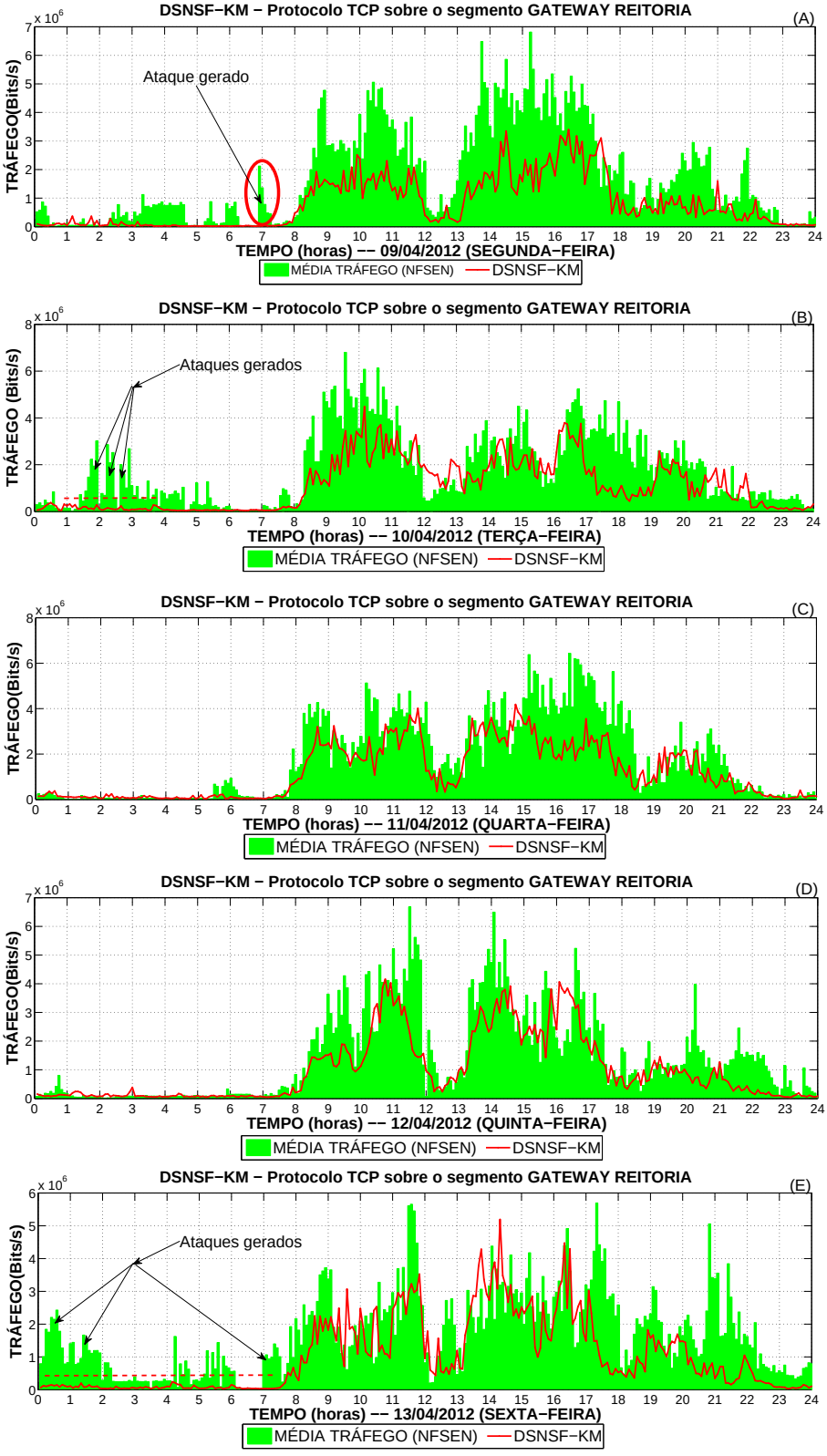

Figura 5.12: DSNSF-KM sobre o movimento registrado pelo NfSen para o protocolo TCP.

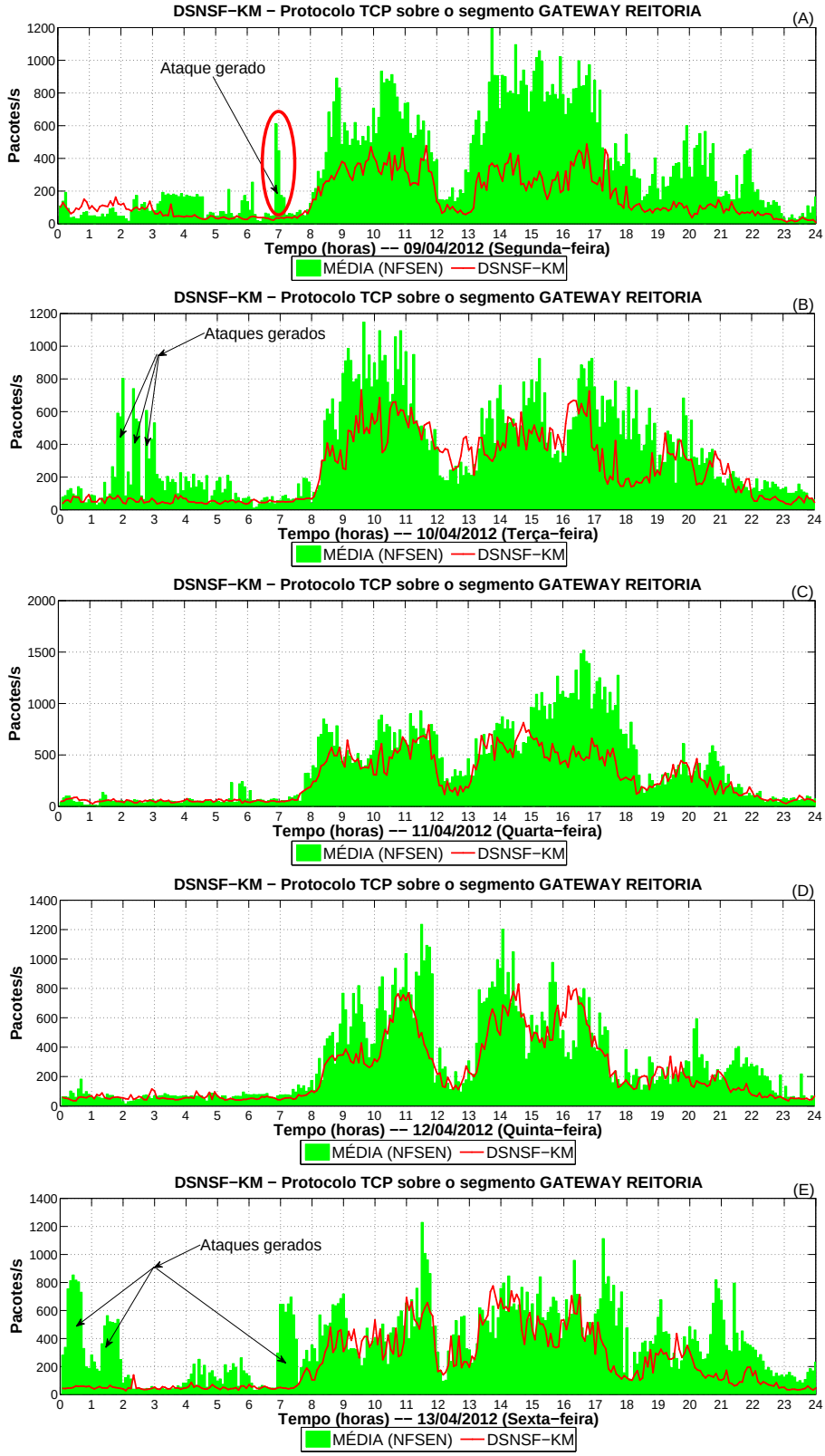

Figura 5.13: DSNSF-KM sobre os pacotes registrados pelo NfSen para o protocolo TCP.

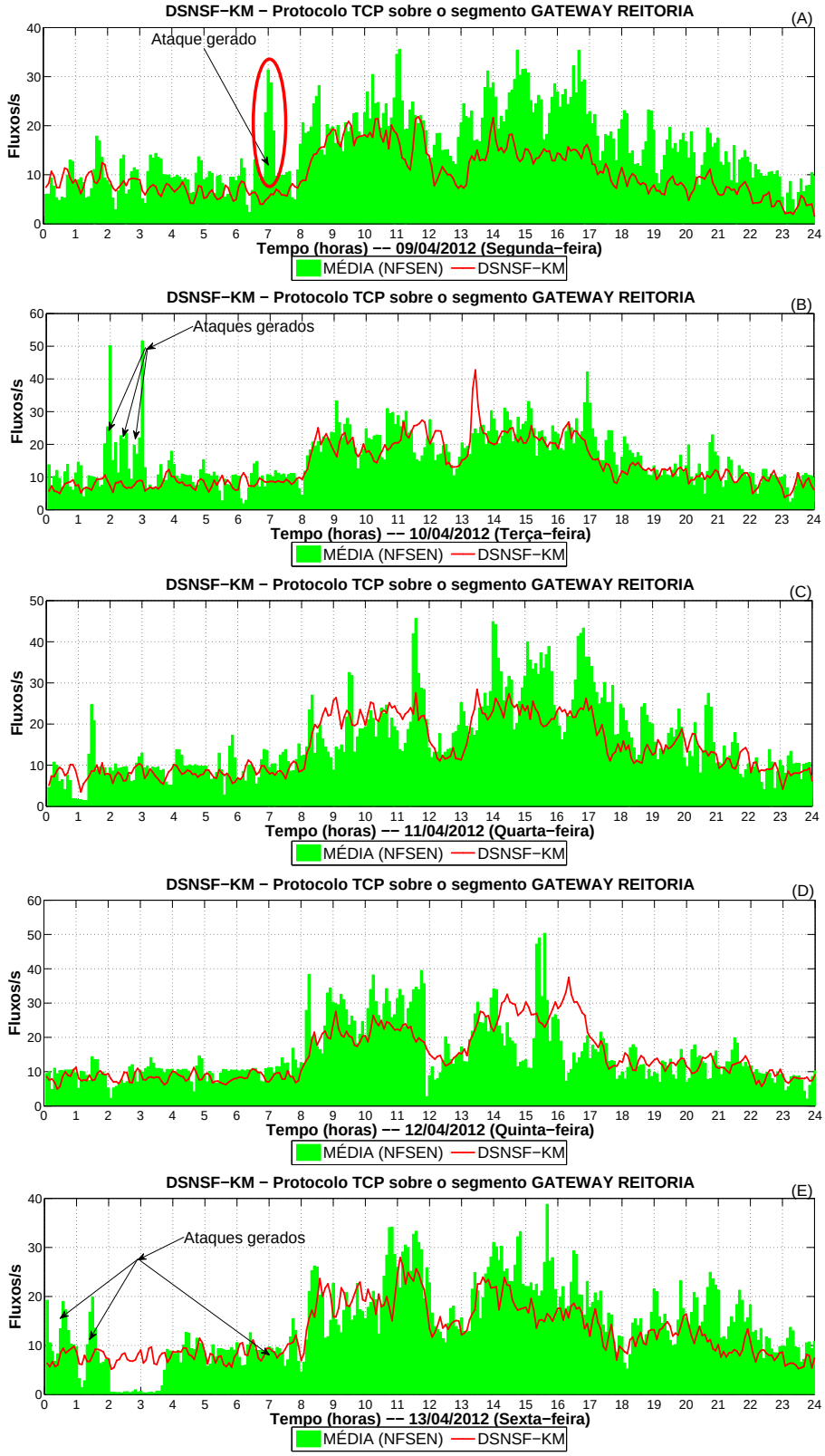

Figura 5.14: DSNSF-KM sobre os fluxos registrados pelo NfSen para o protocolo TCP.

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 conforme pode ser observado na Figura 5.15. Esses picos ou *outliers* foram destacados com objetivo de tentar rastrear sua origem e/ou motivo.

Na segunda-feira (A), houve somente um pico, porém contínuo de aproximadamente uma hora, chegando a quase 5Mbps conforme foi destacado como possível anomalia. Durante a terça-feira (B), verificou-se um pequeno pico de curta duração próximo a 4Mbps, no entanto o movimento procura seguir a tendência do DSNSF-KM. Para quarta-feira (C), tem-se dois picos com maior duração sendo que um quase alcança a taxa de 7Mbps, ambos estão identificados como possíveis anomalias também. A quinta-feira (D), se assemelha à quarta-feira, porém os picos são em horários diferentes. Para sexta-feira (E), tem-se um pico semelhante ao de segunda-feira.

Nota-se que o comportamento do protocolo UDP para esse segmento em geral não ultrapassa os 500Kbps, conforme pode ser observado na terça-feira e também pode ser verificado no cenário 2. Observa-se ainda que os ataques DoS com a ferramenta LOIC foram realizados utilizando somente com protocolo TCP, conforme já descrito.

A Figura 5.16 representa o movimento previsto referente aos pacotes do protocolo UDP conforme já citado. Nela também pode-se identificar as alterações no comportamento para quarta, quinta e sexta-feira como na Figura 5.15, sendo as mesmas identificadas.

Tem-se na Figura 5.17 o movimento previsto referente aos fluxos do protocolo UDP para o mesmo período. Ao contrário das Figuras 5.15 e 5.16, nela não podemos identificar as alterações no comportamento para os dias citados anteriormente. Uma das possíveis causas dessas alterações não aparecerem nos gráficos de fluxo, poderia ser um único usuário gerando esse tráfego, cada fluxo do mesmo estaria com os contadores de bytes e pacotes sendo acumulados ficando os mesmos acima do normal. Isso pode acontecer porque um fluxo pode representar a comunicação entre dois *hosts*, então para essa comunicação tem-se um único fluxo, porém seus contadores de bytes e pacotes sendo acumulados durante a troca de informação entre os *hosts*, justificando seu destaque nas Figuras 5.15 e 5.16 e não apesentando alterações na Figura 5.17 porque é apenas um único fluxo.

Em paralelo à pesquisa, procurou-se identificar por meio das informações disponíveis nos fluxos coletados, a origem, destino e motivo desses picos ou *outliers*. Após análise, foi encontrado um único *host* que estava executando um cliente Torrent em que solicitava conexões com endereços de destino variados e porta de destino 8080 via protocolo

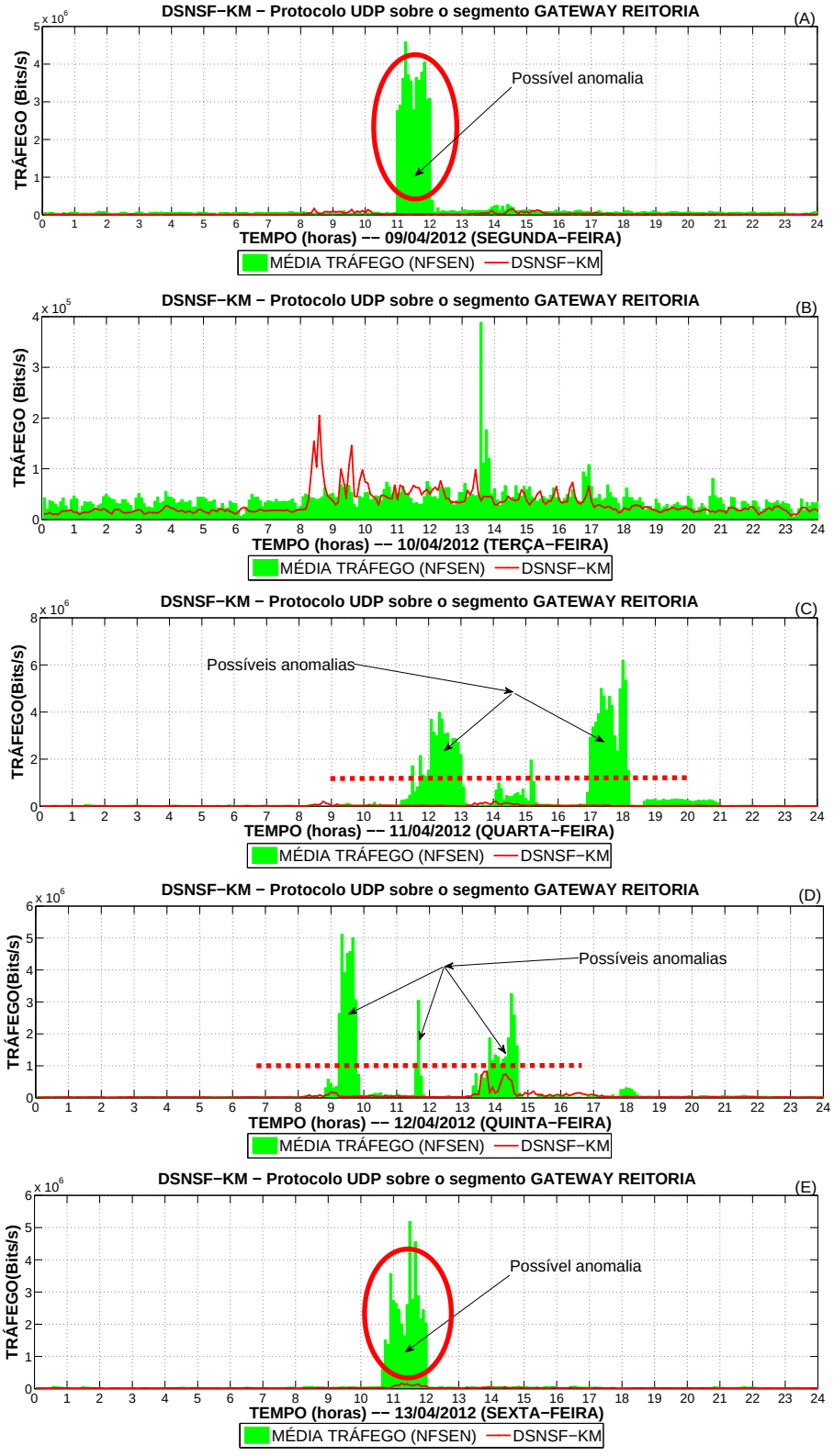

Figura 5.15: DSNSF-KM sobre o movimento registrado pelo NfSen para o protocolo UDP.

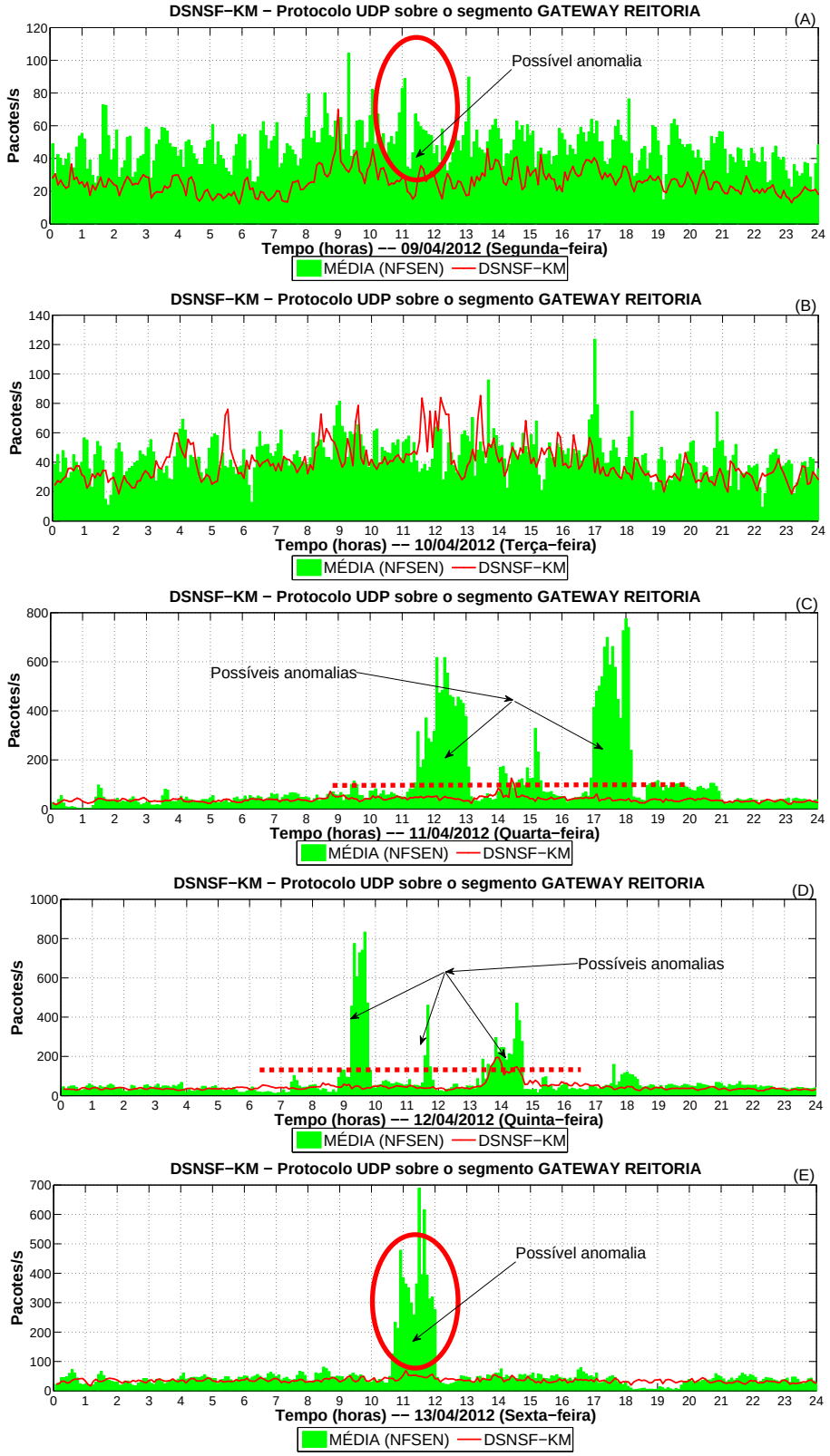

Figura 5.16: DSNSF-KM sobre os pacotes registrados pelo NfSen para o protocolo UDP.

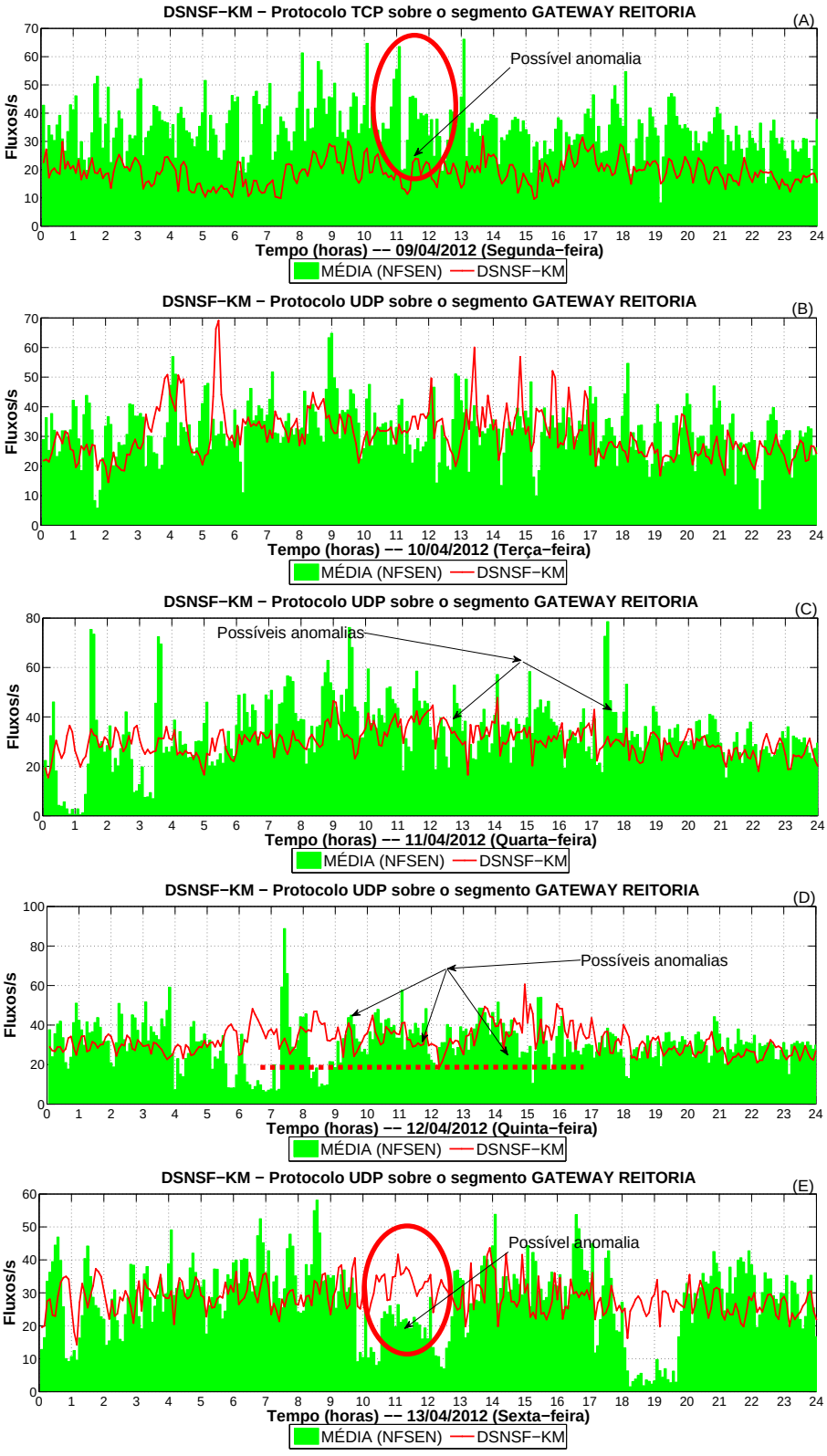

Figura 5.17: DSNSF-KM sobre os fluxos registrados pelo NfSen para o protocolo UDP.

Tabela 5.11: Correlação entre DSNSF-KM e o movimento do NfSen para os protocolos TCP e UDP.

protocolo TCP					
	Seg	Ter	Qua	Qui	Sex
Bits/s	0,8266	0,7203	0,8255	0,8047	0,6561
Pacotes/s	0,8039	0,7344	0,8154	0,8217	0,5931
Fluxos/s	0,6860	0,6402	0,7325	0,6227	0,7406
Protocol UDP					
	Seg	Ter	Qua	Qui	Sex
Bits/s	-0,0327	0,2562	0,0231	0,2706	0,7930
Pacotes/s	0,3497	0,2198	0,1623	0,3213	0,4820
Fluxos/s	0,1754	0,1078	0,3566	-0,1058	0,1848

UDP.

Para uma análise mais detalhada do modelo proposto, realizamos o cálculo da correlação para indicar como o modelo está relacionado com o movimento de cada dia tanto para o protocolo TCP quanto para UDP. Os resultados são apresentados na Tabela 5.11.

Na avaliação do DSNSF-KM, se tratando da correlação para o protocolo TCP os resultados foram bons tendo como menor valor 0,5931 para sexta-feira e o melhor 0,8266 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 apresentou como melhor índice 0,7930, no entanto para quinta-feira apresentou um valor negativo de -0,1058 indicando uma pequena descorrelação.

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 5.12. Para esse quesito, pode-se interpretar que quanto menor o valor obtido mais próximo o modelo está do movimento.

Como é observado na Tabela 5.12, o protocolo TCP apresentou para segunda-feira melhor resultado com 0,1964 e o resultado mais distante para quarta-feira de 1,6050. Para o protocolo UDP, a terça-feira foi melhor com 0,1862 e para quinta-feira o mais distante com o valor de 28,8021.

A fim de indicar se o DSNSF-KM está subestimando ou sobre-estimando o movimento para cada dia da semana para os protocolos TCP e UDP, foi utilizado o desvio fracional padrão (FSD). A Tabela 5.10 apresenta os resultados. Para o FSD tem-se como medida ideal o valor zero.

Para os resultados alcançados na Tabela 5.13, para o protocolo TCP a sexta-

Tabela 5.12: Erro quadrático médio normalizado entre DSNSF-KM e o movimento do NFSEN para os protocolos TCP e UDP.

Protocol TCP					
	Seg	Ter	Qua	Qui	Sex
Bits/s	0,8242	0,9580	1,6050	1,5396	0,7401
Pacotes/s	0,5244	0,3643	0,4955	0,2051	0,3100
Fluxos/s	0,1964	0,2643	0,3700	0,4062	14,5221
Protocolo UDP					
	Seg	Ter	Qua	Qui	Sex
Bits/s	0,6214	0,3668	2,1514	4,2354	0,6567
Pacotes/s	0,2190	0,1862	5,54	0,4230	1,4642
Fluxos/s	0,2095	0,2811	28,8021	1,1287	4,8101

Tabela 5.13: Desvio fracional padrão (FSD) entre DSNSF-KM e o movimento do NFSEN para os protocolos TCP e UDP.

Protocol TCP					
	Seg	Ter	Qua	Qui	Sex
Bits/s	-0,6265	-0,3746	-0,4206	-0,3122	-0,1103
Pacotes/s	-0,8127	-0,3432	-0,4955	-0,2743	-0,1921
Fluxos/s	-0,4618	-0,1528	-0,3500	-0,1678	-0,3831
Protocolo UDP					
	Seg	Ter	Qua	Qui	Sex
Bits/s	-1,8494	-0,1165	-1,8945	-1,4720	-1,8803
Pacotes/s	-0,5235	-0,0732	-1,7298	-1,2851	-1,6892
Fluxos/s	-0,6621	-0,1023	-0,8276	-0,5131	-0,8101

feira apresentou o melhor índice com -0,1103 estando o DSNSF-KM levemente subestimando em relação ao movimento, a segunda-feira apresentou o maior índice de -0,8127. Para o protocolo UDP o melhor índice foi -0,0732 também levemente subestimando em relação ao movimento, na quarta-feira obteve-se o maior índice com -1,8945.

A aplicação do DSNSF-KM para criação da assinatura neste cenário, apresentou bons resultados, pois possibilita que o comportamento do segmento analisado seja descrito de forma automática. Possibilita também a identificação de possíveis anomalias ou mesmo ataques DoS gerados contra o segmento. O DSNSF-KM permite 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, tornando-se uma ferramenta útil na detecção de possíveis anomalias.

Dentre os cenários analisados 1, 2, 3 e 4, verifica-se que para o tráfego total e para o protocolo TCP a assinatura possui um melhor ajuste quando comparado com o protocolo UDP, pois o UDP possui menos tráfego agregado, e é mais difícil de prever

seu comportamento. No entanto pode-se perceber mais facilmente qualquer alteração de comportamento do protocolo UDP, pois em geral para os cenários utilizados o valor médio para bits/s é em torno de 80Kb/s a 500Kb/s e qualquer aumento de carga aparece facilmente ao se usar a assinatura como limiar base.

Observa-se também uma certa relação entre os bits e pacotes já para os fluxos essa relação é menor. Essas informações podem ser usadas ao se desenvolver algum sistema de detecção de anomalias baseado nesses relacionamentos.

Os testes de correlação, NMSE e FSD também permitem identificar esses relacionamentos e podem ser usados como limiares na identificação de anomalias.

5.5 CENÁRIO 5: DSNSF-KM - Total de Bits, Pacotes e Fluxos - sFlow

Para esse cenário procurou-se avaliar o DSNSF-KM na construção das assinaturas perante outro segmento de rede, nesse caso analisou-se um segmento na Universidade Estadual de Londrina (UEL). Utilizou-se todo o tráfego sem a separação por protocolo como nos cenários anteriores, extraindo-se as seguintes informações: Bytes/s - convertido em bits/s -, pacotes/s e fluxos/s para geração das assinaturas. O período de coleta foi durante o mês de setembro e outubro de 2012. Utilizou-se fluxos no padrão sFlow para esse cenário com taxa de amostragem de 1:256, diferentemente dos outros cenários em que os fluxos eram no padrão NetFlow.

O DSNSF-KM analisou sete segundas, terças, quartas, quintas e sextas-feiras de fluxos, com objetivo de extrair uma assinatura para cada dia útil da semana. Com essas informações o DSNSF-KM faz a clusterização, aplica a regra de pontos mínimos pertencentes a cada centro e em seguida calcula a média ponderada para cada intervalo de cinco minutos, resultando no valor que será considerado como comportamento normal ou tendencioso para aquele intervalo no segmento.

As Figuras 5.18, 5.19 e 5.20, representam a aplicação do DSNSF-KM sobre o movimento gerado pelo NfSen para duas semanas subsequentes, correspondente aos dias 22 a 26 de outubro, organizados na primeira coluna, e 29 de outubro a 02 de novembro de 2012 na segunda coluna. As figuras possuem um gráfico para cada dia útil da semana sendo, cada um identificado no canto superior direito como (A, B, C, D ou E) e também por extenso na base dos gráficos correspondente a cada dia. Como pode ser observado nas figuras, as assinaturas geradas pelo DSNSF-KM permitem descrever o comportamento para esse segmento de rede.

Ainda para as Figuras 5.18, 5.19 e 5.20, faz-se a seguinte observação, conforme histórico utilizado para geração da assinatura, recebeu-se a informação de que *backups* eram realizados durante a madrugada, aproximadamente a partir das três horas. Independentemente o DSNSF-KM identificou esse comportamento correspondente aos *backups* durante geração da assinatura, porém para os dias subsequentes, em que a assinatura foi comparada ao movimento, o processo de *backup* deixou de ser realizado, assim causando uma diferença entre o movimento real e o esperado para aquele período. No entanto, o DSNSF-KM percebe que o ciclo de *backup* se encerrou. Isso pode ser constatado ao se comparar um dia semana com o da semana seguinte, por exemplo na Figura 5.18, para segunda-feira 22/10, as quatro horas tem-se o valor de 13 Mbps, já para segunda seguinte no dia 29/10, o valor cai para 11,6 Mbps. Dessa forma identifica-se que a assinatura está se adaptando ao novo comportamento do segmento.

Em relação a Figura 5.18 percebe-se que a assinatura permite descrever o comportamento do segmento até as três e trinta da manhã, a partir desse horário a assinatura prevê que o movimento suba em média para 10 Mbps, no entanto o movimento se mantém uniforme para a maioria dos dias. Entre sete e oito horas a assinatura volta a descrever o comportamento do segmento. O movimento do tráfego se estabiliza as oito horas e segue até as vinte uma com poucas oscilações.

Para esse segmento apresentado na Figura 5.18 a assinatura gerada pelo DSNSF-KM permite descrever o comportamento do mesmo. Têm-se pequenos picos que ultrapassam a assinatura, no entanto a maior parte do movimento encontra-se levemente abaixo da mesma. Observa-se que na sexta-feira dia 02 de novembro o movimento ficou bem abaixo do esperado pela assinatura gerada pelo DSNSF-KM.

Nesse período para terça-feira dia 30 de outubro alguma anomalia pode ter acontecido pois o movimento estava bem acima do esperado pela assinatura no intervalo de vinte e vinte duas horas.

A Figura 5.19 apresenta o movimento previsto para os pacotes sobre o movimento real registrado pelo NfSen. Como já citado nos outros cenários, observa-se que o eixo y possui uma variação menor quando comparado a bits/s. No entanto o comportamento é similar, apresentando um aumento no volume entre as sete e as oito horas e em seguida estabiliza até as vinte uma horas aproximadamente, caindo em seguida.

Ainda na Figura 5.19, quinta-feira dia 25 de novembro, apresenta um leve aumento no volume entre as doze e treze horas aproximadamente. Período esse em que a assinatura poderia indicar uma possível anomalia. Nesse cenário a assinatura gerada para os

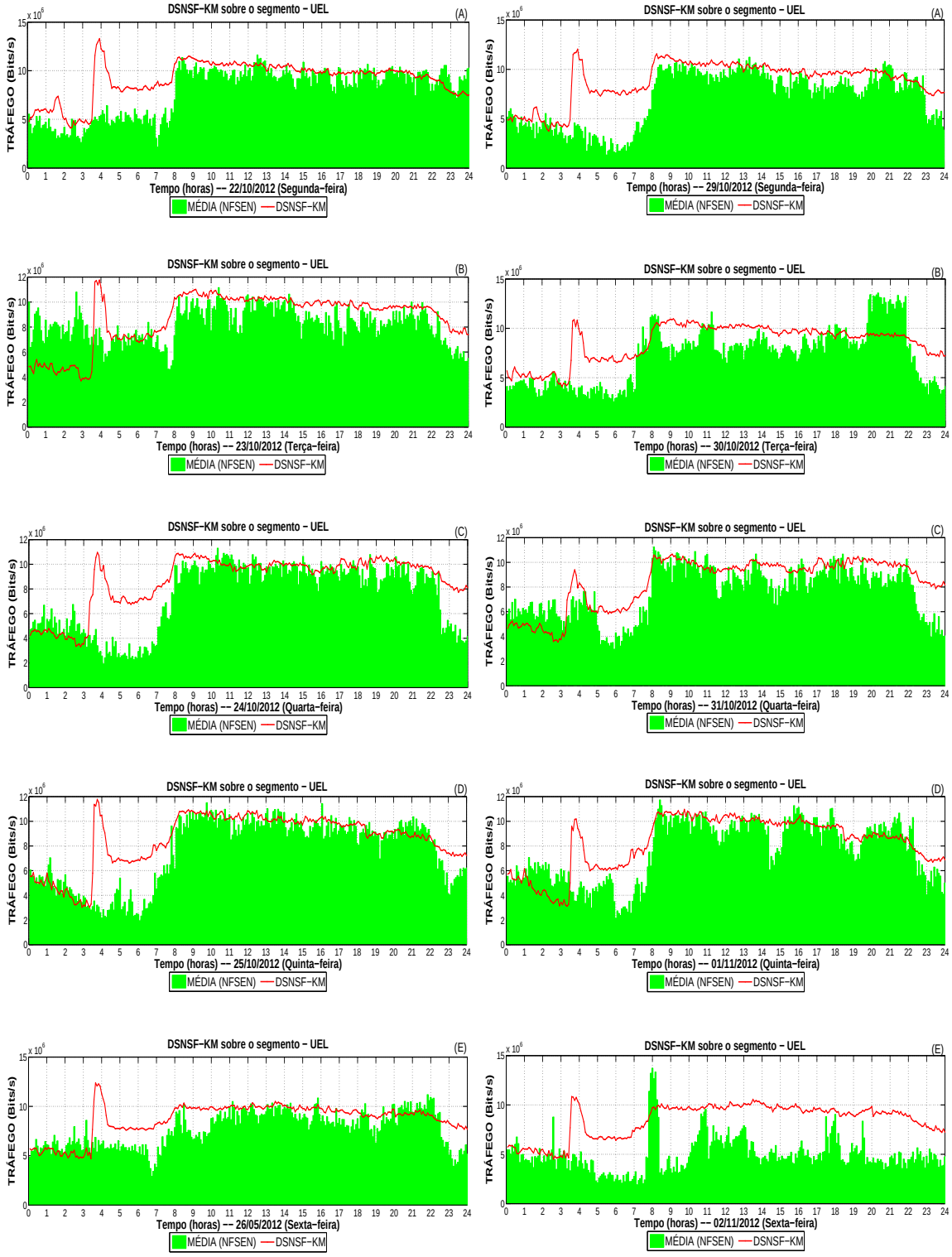

Figura 5.18: DSNSF-KM sobre o tráfego total em bits/s registrado pelo NfSen.

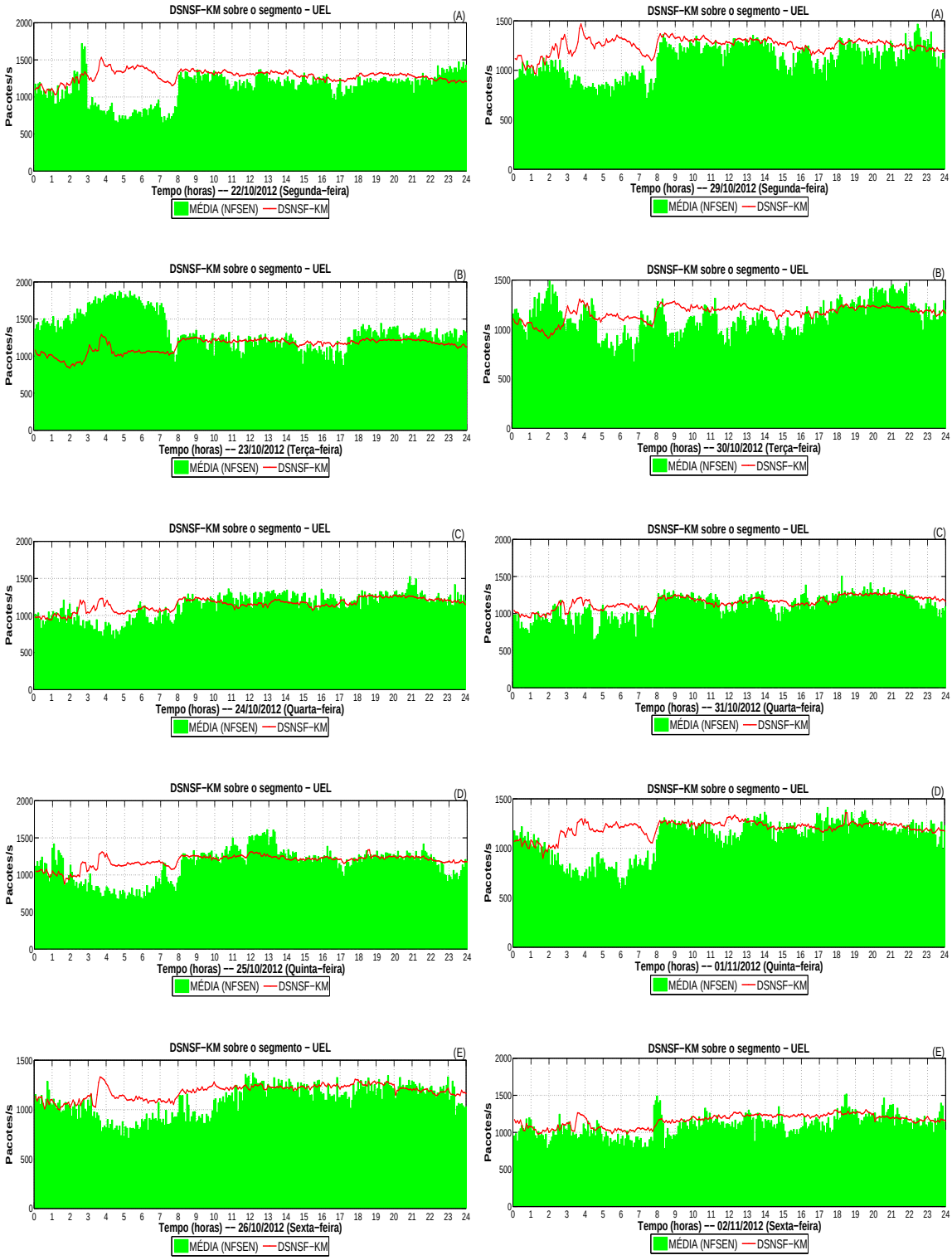

Figura 5.19: DSNSF-KM sobre o tráfego total de pacotes registrado pelo NfSen.

pacotes também permite descrever o comportamento do segmento de rede.

Os resultados do DSNSF-KM para criação da assinatura em se tratando do número total de fluxos são apresentados na Figura 5.20. Para esse segmento analisado, observa-se que os fluxos possuem uma boa relação com os pacotes e também com os bits apresentando um comportamento similar.

Tratando-se dos fluxos, a assinatura também possibilita prever o comportamento do movimento. Para esse período analisado, observa-se que na terça-feira dia 23 de outubro, entre zero e sete horas existe uma diferença maior entre a assinatura e o movimento que poderia indicar alguma anomalia no segmento. Também na quinta-feira dia 25 de outubro, entre doze e treze horas e trinta minutos. Na sexta-feira dia 02 de novembro há um pico em torno das oito horas que se destoa da assinatura, podendo também indicar um comportamento anômalo.

Como nos cenários anteriores, para uma análise mais detalhada das assinaturas construídas pelo DSNSF-KM perante o movimento gerado pelo NfSen, nesse caso coletando fluxos sFlow, realizou-se os testes de correlação, erro quadrático médio normalizado e o desvio fracional padrão conforme as Tabelas 5.14, 5.15 e 5.16. Sendo as mesmas organizadas como nos cenários anteriores.

Para correlação entre o DSNSF-KM e o movimento, a Tabela 5.14 indica como a assinatura está relacionada com o movimento para cada dia útil da semana para bits, pacotes e fluxos. Observa-se que quanto mais próximo de 1 melhor.

Tabela 5.14: Correlação entre DSNSF-KM e o movimento do NfSen.

	Seg	Ter	Qua	Qui	Sex
Bits/s	0,7144 : 0,7100	0,3970 : 0,6669	0,6929 : 0,7500	0,7366 : 0,7330	0,5869 : 0,2979
Pacotes/s	-0,0984 : 0,2815	0,4926 : 0,1810	0,6253 : 0,7351	0,5109 : 0,2651	0,9875 : 0,5661
Fluxos/s	-0,0984 : 0,2813	-0,4926 : 0,0122	0,6253 : 0,7351	0,5109 : 0,2651	0,5211 : 5660

Conforme a Tabela 5.14, temos para quinta-feira dia 25 de outubro o melhor índice chegando a 0,7500 para bits e o pior índice de -0,0984 para pacotes e fluxos na segunda-feira dia 22 de outubro. Pode-se concluir que tanto para bits, pacotes e fluxos os índices obtidos são considerados satisfatórios. Nota-se ainda que pacotes e fluxos obtiveram os mesmo valores nesse segmento para o período analisado.

Para definir a proximidade do modelo apresentado com o movimento, o erro quadrático médio normalizado foi utilizado, e os resultados são apresentados na Tabela 5.15. Para esse quesito, quanto mais próximo de zero, melhor.

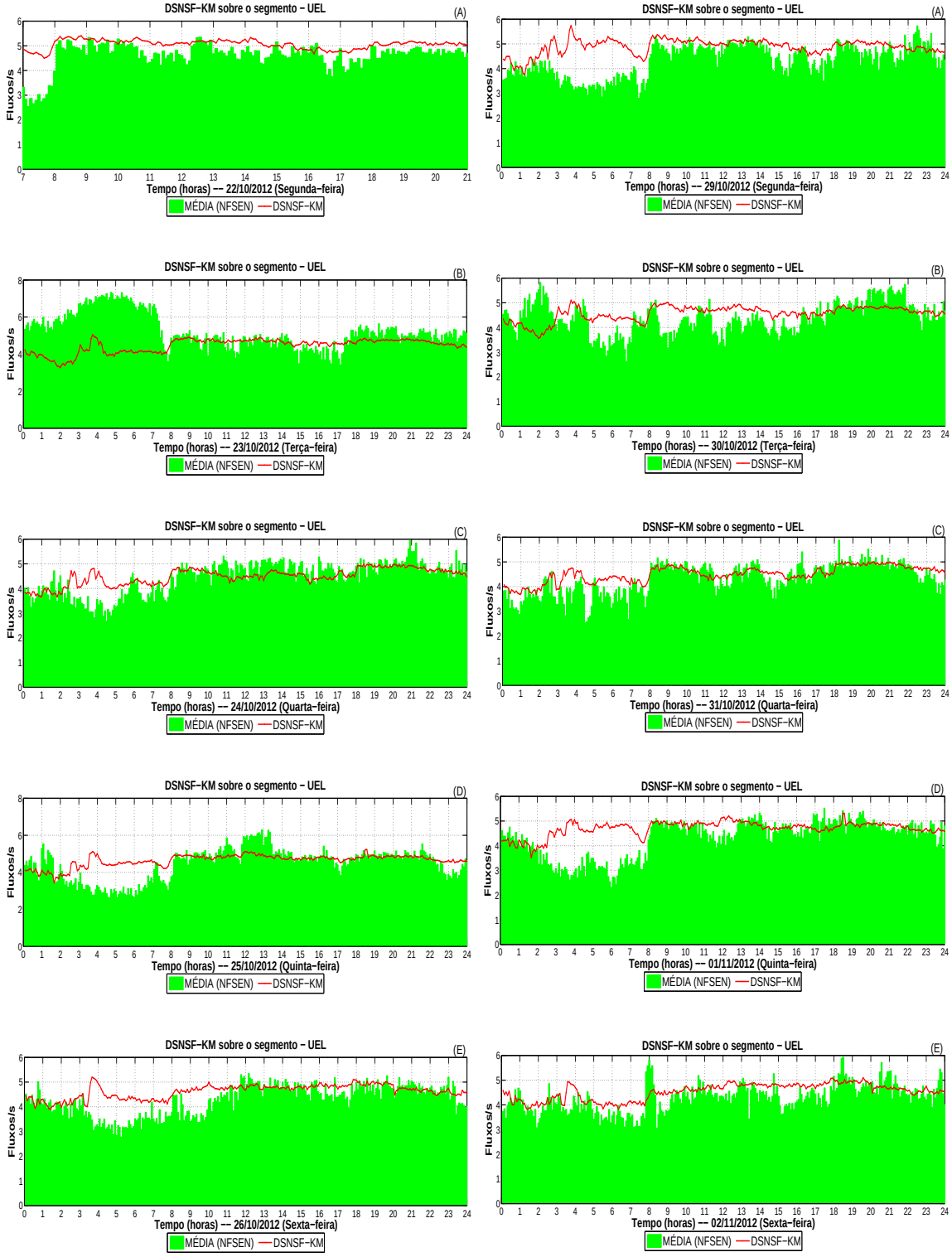

Figura 5.20: DSNSF-KM sobre o tráfego total de fluxos registrado pelo NfSen.

Tabela 5.15: Erro quadrático médio normalizado entre DSNSF-KM e o movimento do NFSEN.

	Seg	Ter	Qua	Qui	Sex
Bits/s	0,0775 : 0,0447	0,0316 : 0,0406	0,0169 : 0,0210	0,0121 : 0.0297	0,0119 : 0,8201
Pacotes/s	0,0238 : 0,0130	0,0083 : 0,0181	0,0058 : 0,0042	0,0066 : 0,0087	0,0087 : 0,0129
Fluxos/s	0,0238 : 0,0130	0,0083 : 0,0181	0,0043 : 0,0042	0,0066 : 0,0087	0,0125 : 0,0129

Tratando-se do NMSE na Tabela 5.15, quarta-feira dia 31 de outubro, obteve o melhor índice chegando a 0,0042 para pacotes e fluxos, e a segunda-feira dia 22 de outubro, o pior índice com 0,0775, concluindo-se que para esse quesito obteve-se bons resultados tanto para bits, pacotes e fluxos.

Para o desvio fracional padrão (FSD), a fim de indicar se o DSNSF-KM esta subestimando ou sobre-estimando o movimento para cada dia da semana. A Tabela 5.10 apresenta os resultados. Para o FSD tem-se como medida ideal o valor zero e variação de -2 a 2.

Tabela 5.16: Desvio Fracional Padrão entre DSNSF-KM e o movimento do NFSEN.

	Seg	Ter	Qua	Qui	Sex
Bits/s	-0,2783 : -0,3284	0,3985 : -0,4208	-0,2758 : -0,0294	-0,2558 : -0,1419	-0,0911 : -0,0781
Pacotes/s	0,8894 : -0,7139	-0,8380 : -0,7259	-0,6656 : -0,6267	-0,9291 : -0,8700	0,7482 : -0,5343
Fluxos/s	-0,8897 : -0,7141	-0,8381 : -0,6658	-0,0719 : -0,6269	-0,9293 : -0,8702	0,7483 : -0,5342

Conforme os resultados apresentados na Tabela 5.16, concluímos que na quarta-feira dia 31 de outubro obtivemos o melhor índice com 0,0294 e o pior de -0,8897 para segunfa-feira dia 22 de outubro. Para a maioria dos dias o DSNSF-KM subestimou o movimento. Os valores apresentados pelo FSD indicam que o DSNSF-KM permite ter uma boa previsão do volume para bits, pacotes e fluxos, tendo em vista que a maioria dos índices ficaram próximos de zero.

Para esse cenário, concluímos que as assinaturas geradas pelo DSNSF-KM obtiveram um bom resultado frente as duas semanas analisadas, pois permitiram descrever o comportamento do segmento analisado tanto para bits, pacotes e fluxos. Além de identificar que um processo deixou de ser executado, como no caso do Backup, em que se esperava um aumento no tráfego para aquele período. Também identificou o baixo movimento causado por conta do feriado na sexta-feira, dia dois de novembro. Nesse segmento de maneira geral obtiveram-se os melhores índices tanto para correlação, NMSE e FSD, um dos fatos que contribuiu para esse resultado, se deve pelo fato do sFlow utilizar amostragem, dessa forma acontece indiretamente uma suavização dos contadores. Conclui-se também que o DSNSF-KM permite gerar assinaturas utilizando o padrão de fluxos sFlow.

5.6 Resumo dos Resultados Obtidos

Nesta seção, faz-se um panorama geral de avaliação dos resultados obtidos utilizando-se os testes de correção, NMSE e FSD em relação a todos os cenários. A Tabela 5.17 apresenta a média para cada teste em relação a bits/s, pacotes/s e fluxos/s.

Tabela 5.17: Média dos resultados obtidos entre todos os cenários para Correlação, NMSE e FSD.

	Corr	NMSE	FSD
Bits/s	0,5614	2,4572	-0,3683
Pacotes/s	0,5672	1,4069	-0,4508
Fluxos/s	0,4329	2,1695	-0,4343

Conforme pode-se observar na Tabela 5.17, para correlação obteve-se o melhor índice para pacotes/s com o valor de 0,5672, para NMSE o melhor índice obtido foi para pacotes/s com 1,4069 e por fim para o FSD o melhor índice foi de -0,3683 para bits/s.

Como foram utilizados todos os dias de todos cenários para a avaliação geral, observa-se o seguinte:

- No cenário 4, realizou-se ataques DoS em três dos cinco dias utilizados, também nesse cenário identificou-se a presença de clientes torrent, gerando um alto tráfego para o protocolo UDP;
- Feriados ou recessos, como o de 01 de maio, causam um comportamento anormal para o movimento previsto;
- No cenário 5, conforme histórico utilizado para geração da assinatura, eram realizados *backups* durante a madrugada, dessa forma o DSNSF-KM identificou-se esse comportamento. Porém para os dias subsequentes esse processo de *backup* deixou de ser realizado, dessa forma causando uma diferença entre o movimento real e o esperado para aquele período.

A Tabela 5.18, indica os valores obtidos excluindo-se os períodos em que realizou-se os ataques, também os feriados e o período em que o backup deixou de ser executado. Dessa forma pode-se ter uma avaliação mais adequada das assinaturas.

Conforme as observações citadas para os períodos de feriado, backup e ataques esses períodos influenciaram de forma negativa nos resultados obtidos. Ao se verificar

Tabela 5.18: Média dos resultados obtidos entre todos os cenários para Correlação, NMSE e FSD excluindo-se feriados, ataques e o backup.

	Corr	NMSE	FSD
Bits/s	0,7008	0,5597	-0,2073
Pacotes/s	0,7262	0,2320	-0,3281
Fluxos/s	0,6430	0,2181	-0,0027

os resultados obtidos pela Tabela 5.18, verifica-se que os valores são muito melhores que os apresentados na Tabela 5.17. Por fim, realizou-se mais um cálculo para média geral entre bits, pacotes e fluxos, excluindo-se os períodos citados, alcançando os seguintes valores para correlação = 0,7 para NMSE = 0,3 e para FSD = -0,17. Fazendo uma análise constata-se que para correlação obtiveram-se bons resultados, da mesma forma para o NMSE. Para o FSD obteve-se um bom índice e conclui-se para bits, pacotes e fluxos que todas as assinaturas em média ficaram um pouco abaixo do movimento real, ou seja, subestimando o movimento.

6 CONCLUSÕES

Neste trabalho, foi abordado o problema para caracterização de tráfego em segmentos de rede com objetivo de auxiliar os administradores na detecção de anomalias. Foi proposto um modelo para construção de uma Assinatura Digital de Segmento de Rede utilizando Análise de Fluxos e clusterização K-means (ou *Digital Signature of Network Segment Using Flow Analysis and K-means Clustering* - (DSNSF-KM)), modelo esse utilizado para descrever o comportamento básico/normal do segmento de rede atráves de uma análise do histórico dos fluxos, utilizando abordagem de clusterização. Com isso, espera-se auxiliar os administradores na detecção de problemas que possam ser identificados como tráfegos anômalos em um segmento.

Alterações no comportamento do movimento real perante o modelo DSNSF-KM são motivo de estudo e servem como motivação para descobrir o que ocasiona essas mudanças, sendo algumas opções: aumento do número de usuários, atualizações de software, ataques, problemas nos ativos de rede que pode estar causando, por exemplo, retransmissões. Com um bom modelo, o administrador pode perceber rapidamente que sua rede mudou de comportamento e investigar as possíveis causas.

Para avaliar as assinaturas geradas pelo DSNSF-KM, foram utilizados 5 cenários com base em fluxos coletados de segmentos de redes reais, sendo o cenário 1, 2, 3 e 4 na Universidade Tecnológica Federal do Paraná - Câmpus Toledo, utilizando-se fluxos NetFlow versão 9. Para o cenário 5, na Universidade Estatual de Londrina, utilizando-se fluxos sFlow versão 5. Em todos os cenários analizaram-se os contadores dos campos Bytes/s, Pacotes/s e Fluxos/s extraídos dos fluxos. O período de aprendizagem utilizado pelo método proposto para adaptação às características das redes foi de sete semanas.

Em todos os cenários estudados foram encontrados bons resultados, tanto para os testes de correlação, NMSE quanto FSD, demonstrando o potencial que o modelo proposto apresenta para auxliar a gerência/administração de redes.

6.1 Contribuições

As assinaturas geradas por meio do DSNSF-KM apresentaram resultados que auxiliam os administradores de redes e possibilitaram que o comportamento dos segmentos analisados sejam caracterizados de forma automática. As assinaturas permitem que o administrador da rede, observando os parâmetros de correlação, NMSE e FSD, possa realizar cruzamentos entre os mesmos ou utilizá-los individualmente a fim de checar se o comportamento da rede está normal ou se alguma anomalia esteja acontecendo no segmento. Por exemplo, se o movimento está distante do DSNSF-KM, porém correlacionado, pode indicar ao administrador um aumento no número de usuários, serviços, alguma rotina específica sendo executada, no entanto, o comportamento do segmento continua o mesmo. Nesse caso, a correlação tem a função de indicar se o movimento deve diminuir ou aumentar para aquele instante, enquanto o NMSE indica em qual faixa de volume o tráfego deve estar.

O DSNSF-KM poderá ser utilizado na automatização do processo de detecção por meio de um sistema para geração de alarmes baseado no(s) limite(s) definido(s) pelas assinaturas para o segmento de rede, podendo-se chegar como resultado final, em que o administrador não precisará mais monitorar visualmente todos os seus segmentos, tarefa essa que se torna difícil em casos onde existem muitos segmentos para serem monitorados. Com o DSNSF-KM, a automatização de alarmes permitirá ao administrador focar nos segmentos que apresentam alguma anomalia, dispensando a necessidade de que um operador ou o próprio administrador acompanhe cada gráfico de cada segmento constantemente.

O DSNSF-KM permite a criação de assinaturas utilizando-se tanto fluxos NetFlow quanto fluxos sFlow, em que obteve um bom desempenho tanto no segmento da Universidade Tecnológica Federal do Paraná - Câmpus Toledo, quanto na Universidade Estatual de Londrina, conforme se observa os cenários utilizados no capítulo 5. Por meio dos cenários utilizados, especificamente nos cenários 1, 2 e 4, onde gerou-se assinaturas separadas para os protocolos TCP e UDP, pôde-se identificar o comportamento dos mesmos e sua variação dentro de cada segmento. Durante a pesquisa identificou-se tráfegos anômalos para o protocolo UDP, e constatou-se que os mesmos originaram-se de clientes torrent que causavam um aumento exagerado no tráfego para esse protocolo.

A gerência de redes por meio da análise de fluxos do tipo IPFIX, NetFlow e sFlow, abre caminho para uma série de personalizações que podem ser realizadas, pois possuem uma grande variedade de informações disponíveis. Dessa forma, permite que sejam geradas assinaturas personalizadas para um determinado endereço ou porta, por exemplo.

Outra vantagem de se utilizar fluxos reside no fato de que, ao se monitorar um roteador de um segmento, dispensa-se a instalação de agentes nos dispositivos desse segmento, ou seja, de um único ponto é possível coletar informações de todo o segmento.

Como trabalho futuro, pretende-se realizar a combinação e o relacionamento de mais elementos dos fluxos como portas de serviços, endereços IP de origem e destino, bem como identificar por meio das *flags* ataques que possuem assinaturas.

REFERÊNCIAS

1 PROENCA, M. L. J. et al. Baseline to help with network management. In: *e-Business and Telecommunication Networks*. [S.l.]: Springer Netherlands, 2006. p. 158–166.

2 FATEMIPOUR, F.; YAGHMAEE, M. Design and implementation of a monitoring system based on ipfix protocol. In: *Telecommunications, 2007. AICT 2007. The Third Advanced International Conference on*. [S.l.: s.n.], 2007. p. 22.

3 SHAWKY, A. et al. characterization and modeling of network traffic. In: *Computer Engineering Conference (ICENCO), 2010 International*. [S.l.: s.n.], 2010. p. 72 –76.

4 DENNING, D. An intrusion-detection model. *Software Engineering, IEEE Transactions on*, SE-13, n. 2, p. 222 – 232, feb. 1987. ISSN 0098-5589.

5 PATCHA, A.; PARK, J.-M. an overview of anomaly detection techniques: existing solutions and latest technological trends. *Computer Networks*, v. 51, n. 12, p. 3448 – 3470, 2007. ISSN 1389 - 1286. Disponível em: <http://www.sciencedirect.com/science/article/pii/S138912860700062X>.

6 CHANG, S. et al. A flow-based anomaly detection method using sketch and combinations of traffic features. In: *Network and Service Management (CNSM), 2010 International Conference on*. [S.l.: s.n.], 2010. p. 302 –305.

7 MURALEEDHARAN, N.; PARMAR, A.; KUMAR, M. A flow based anomaly detection system using chi-square technique. In: *Advance Computing Conference (IACC), 2010 IEEE 2nd International*. [S.l.: s.n.], 2010. p. 285 –289.

8 CELENK, M. et al. anomaly detection and visualization using fisher discriminant clustering of network entropy. In: *Digital Information Management, 2008. ICDIM 2008. Third International Conference on*. [S.l.: s.n.], 2008. p. 216 –220.

9 MOLNAR, S.; MOCZAR, Z. three-dimensional characterization of internet flows. In: *Communications (ICC), 2011 IEEE International Conference on*. [S.l.: s.n.], 2011. p. 1 –6. ISSN 1550-3607.

10 YINGQIU, L.; WEI, L.; YUNCHUN, L. network traffic classification using k-means clustering. In: *Computer and Computational Sciences, 2007. IMSCCS 2007. Second International Multi-Symposiums on*. [S.l.: s.n.], 2007. p. 360 –365.

11 SINGH, M.; SUBRAMANIAN, N.; RAJAMENAKSHI. Visualization of flow data based on clustering technique for identifying network anomalies. In: *Industrial Electronics Applications, 2009. ISIEA 2009. IEEE Symposium on*. [S.l.: s.n.], 2009. v. 2, p. 973 –978.

12 ROSSI, D.; VALENTI, S. fine-grained traffic classification with netflow data. In: *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*. New York, NY, USA: ACM , 2010. (IWCMC '10), p. 479–483. ISBN 978-1-4503-0062-9. Disponível em: <http://doi.acm.org/10.1145/1815396.1815507>.

13 CLAISE, B. *Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information*. IETF, jan. 2008. RFC 5101 (Proposed Standard). (Request for Comments, 5101). Disponível em: <http://www.ietf.org/rfc/rfc5101.txt>.

14 CLAISE, B. *Cisco Systems NetFlow Services Export Version 9*. IETF, out. 2004. RFC 3954 (Informational). (Request for Comments, 3954). Disponível em: <http://www.ietf.org/rfc/rfc3954.txt>.

15 PHAAL, P.; PANCHEN, S.; MCKEE, N. *InMon Corporation's sFlow a Method for Monitoring Traffic in Switched and Routed Networks*. IETF, set. 2001. RFC 3176 (Informational). (Request for Comments, 3176). Disponível em: <http://www.ietf.org/rfc/rfc3176.txt>.

16 MACQUEEN, J. B. Some methods for classification and analysis of multivariate observations. In: CAM, L. M. L.; NEYMAN, J. (Ed.). *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*. [S.l.]: University of California Press, 1967. v. 1, p. 281–297.

17 SON, H.; LEE, Y. Detecting anomaly traffic using flow data in the real voip network. In: *Applications and the Internet (SAINT), 2010 10th IEEE/IPSJ International Symposium on*. [S.l.: s.n.], 2010. p. 253 –256.

18 TIN, H.-W.; LEU, S.-W.; CHANG, S.-H. Measurement of flow burstiness by fractal technique. In: *Computer Symposium (ICS), 2010 International*. [S.l.: s.n.], 2010. p. 722 –727.

19 PESCAPE, A. et al. On the impact of sampling on traffic monitoring and analysis. In: *2010 22nd International Teletraffic Congress (ITC 22)*. IEEE, 2010. p. 1–8. ISBN 978-1-4244-8837-7. Disponível em: <http://dx.doi.org/10.1109/ITC.2010.5608718>.

20 TCHRAKIAN, T.; BASU, B.; O'MAHONY, M. Real-time traffic flow forecasting using spectral analysis. *Intelligent Transportation Systems, IEEE Transactions on*, v. 13, n. 2, p. 519 –526, june 2012. ISSN 1524-9050.

21 HSU, C.-M.; LIAN, F.-L. characterizing highway traffic dynamics using gmm and phase transition analysis. In: *System Science and Engineering (ICSSE), 2011 International Conference on*. [S.l.: s.n.], 2011. p. 530 –535.

22 MARQUES NT., H. T. et al. Characterizing broadband user behavior. In: *Proceedings of the 2004 ACM workshop on Next-generation residential broadband challenges*. New York, NY, USA: ACM, 2004. (NRBC '04), p. 11–18. ISBN 1-58113-935-7. Disponível em: <http://doi.acm.org/10.1145/1026763.1026767>.

23 MOGHADDAM, S. et al. data-driven co-clustering model of internet usage in large mobile societies. In: *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*. New York, NY, USA: ACM, 2010. (MSWIM '10), p. 248–256. ISBN 978-1-4503-0274-6. Disponível em: <http://doi.acm.org/10.1145/1868521.1868562>.

24 LAKHINA, A.; CROVELLA, M.; DIOT, C. Mining anomalies using traffic feature distributions. In: *Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*. New York , NY, USA : ACM , 2005. (SIGCOMM '05), p. 217–228. ISBN 1-59593-009-4. Disponível em: <http://doi.acm.org/10.1145/1080091.1080118>.

25 TRAMMELL, B.; BOSCHI, E. An introduction to ip flow information export (ipfix). *Communications Magazine, IEEE*, v. 49, n. 4, p. 89 –95, april 2011. ISSN 0163-6804.

26 QUITTEK, J. et al. *Requirements for IP Flow Information Export (IPFIX)*. IETF, out. 2004. RFC 3917 (Informational). (Request for Comments, 3917). Disponível em: <http://www.ietf.org/rfc/rfc3917.txt>.

27 DERI, L. *Ntop*. [Online; accessed 28-Jul-2011]. Disponível em: <http://www.ntop.org/solutions/flow-based-monitoring/>.

28 CISCO. *Cisco IOS XE Flexible NetFlow Overview*. jul. 2012. Disponível em: <http://www.cisco.com/en/US/docs/ios-xml/ios/fnetflow/configuration/xe-3s/fnetflow-overview-xe.html>.

29 CISCO. *NetFlow Services Solutions Guide v4*. jan. 2007. Disponível em: <http://www.cisco.com/en/US/docs/ios/solutions_docs/netflow/nfwhite.html>.

30 GERALD, C. *Wireshark - Network Protocol Analysis*. 1998. [Online; accessed 10-Jun-2011]. Disponível em: <http://www.wireshark.org/download.html>.

31 SFLOW.ORG. *sFlow Overview*. 2003. [Online; accessed 28-Sep-2011]. Disponível em: <http://www.sflow.org/sFlowOverview.pdf>.

32 PHAAL, P.; LAVINE, M. *sFlow version 5*. jul. 2004. [Online; accessed 02-Oct-2011]. Disponível em: <http://www.sflow.org/sflow_version_5.txt>.

33 LEINEN, S. *Evaluation of Candidate Protocols for IP Flow Information Export (IPFIX)*. IETF, out. 2004. RFC 3955 (Informational). (Request for Comments, 3955). Disponível em: <http://www.ietf.org/rfc/rfc3955.txt>.

34 QUITTEK, J. et al. *Information Model for IP Flow Information Export*. IETF, jan. 2008. RFC 5102 (Proposed Standard). (Request for Comments , 5102). Updated by RFC 6313. Disponível em: <http://www.ietf.org/rfc/rfc5102.txt>.

35 CASE, J. et al. *Simple Network Management Protocol (SNMP)*. IETF, maio 1990. RFC 1157 (Historic). (Request for Comments, 1157). Disponível em: <http://www.ietf.org/rfc/rfc1157.txt>.

36 HAAG, P. *NetFlow visualisation and investigation tool*. mar. 2005. Disponível em: <//nfsen.sourceforge.net>.

37 HAAG, P. *NFDUMP - NetFlow processing tools*. set. 2004. Disponível em: <//nfdump.sourceforge.net>.

38 FU, H. A novel clustering algorithm with ant colony optimization. In: *Computational Intelligence and Industrial Application, 2008. PACIIA '08. Pacific-Asia Workshop on*. [S.l.: s.n.], 2008. v. 2, p. 66 –69.

39 TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Cluster Analysis: Basic Concepts and Algorithms*. [S.l.]: Addison-Wesley, 2006. 769p p. ISBN 9780321321367.

40 ROUSSEEUW, P. J. silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* , v. 20, n. 0, p. 53 – 65, 1987. ISSN 0377-0427. Disponível em: <http://www.sciencedirect.com/science/article/pii/0377042787901257>.

41 MILLER, D. *Softflowd - Traffic flow monitoring*. 2010. Disponível em: <http://manpages.ubuntu.com/manpages/maverick/man8/softflowd.8.html>.

42 TECHNOLOGIES, P. *Low Orbit Ion Cannon*. 2006. [Online; accessed 20-Jun-2011]. Disponível em: <http://sourceforge.net/projects/loic/>.

Trabalhos publicados pelo autor

1. ZACARON, A. M. ; CARVALHO, L. F. ; ADANIYA, M. H. A. C. ; Taufik Abrão ; PROENCA JR., M. L. . Assinatura Digital de Segmento de Rede Utilizando Análise de Fluxos e Clusterização K-means. In: SEMISH - 39th Seminar on Hardware and Software, 2012, Curitiba. Anais do XXXII Congresso da Sociedade Brasileira de Computação (CSBC), 2012. (Qualis CC 2012, B4)

2. ZACARON, A. M. ; CARVALHO, L. F. ; ADANIYA, M. H. A. C. ; Taufik Abrão ; PROENCA JR., M. L. . Digital Signature of Network Segment using Flow Analysis. In: The International Joint Conference on e-Business and Telecommunications (ICETE 2012) / DCNET 2012, 2012, Roma. Proceedings of DCNET, ICE-B and OPTICS 2012 International Conference on Data Communication Networking, e-Business and Optical Communication Systems, 2012. p. 35-40. (Qualis CC 2012, B4)

3. FERNANDES JR, G.; ZACARON, A. M.; RODRIGUES, J. P. C., PROENÇA JR, Mario Lemes. Digital Signature to Help Network Management Using Principal Component Analysis and K-Means Clustering. Communications (ICC), 2013 IEEE International Conference, Budapest, Hungary. (Qualis CC 2012, A2).

Trabalhos Publicados com Outros Autores do Grupo de Pesquisa

4. CARVALHO, L. F. ; ZACARON, A. M. ; ADANIYA, M. H. A. C. ; PROENÇA JR, Mario Lemes . Ant Colony Optimization for Creating Digital Signature of Network Segments using Flow Analysis. In: XXXI International Conference of the Chilean Computer Science Society (SCCC'2012), 2012, Valparaíso. Proceedings of XXXI International Conference of the Chilean Computer Science Society (SCCC'2012), 2012. (Qualis CC 2012, B3)

5. ASSIS, Marcos V. O. de; ZACARON, A. M. ; PROENÇA JR, Mario Lemes . Time Series Forecasting Methods for Creating Digital Signature of Network Segments Using Flow Analysis.. In: XXXI International Conference of the Chilean Computer Science Society (SCCC'2012), 2012, Valparaíso. Proceedings of XXXI International Conference of the Chilean Computer Science Society (SCCC'2012), 2012. (Qualis CC 2012, B3)