

Análise de Tendência de Tráfego Origem Destino em Redes IP utilizando Estatística Multivariada

Arnoldo N. da Silva¹, Jorge L. C. Silva², J. E. Bessa Maia², Paulo R. F. Cunha¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7851 – 50732-970 – Recife – PE – Brazil

²Departamento de Estatística e Computação - Universidade Estadual do Ceará (UECE)
Av. Paranjana, 1700 60740-903 – Fortaleza, CE – Brazil
{ans2,prfc}@cin.ufpe.br, {jlcs,jmaia}@larces.uece.br

Abstract. *Anticipating the trend of the traffic behavior may help the network administration in many of the operational and planning tasks. Traffic through links and nodes is composed by superposition of origin destination (OD) flows, according to a routing matrix, which creates a multivariate structure and high dimension matrices. The analysis at one time of growth trend of OD flows with similar behaviors may help prevent future network congestion. This paper uses multivariate Statistics to analyze and forecast the traffic of OD flows. Previous papers have shown that OD flows can be modeled by using few components, called eigenflows in a principal components analysis. This work proposes a methodology to analyze the OD flows behavior and its growth trend in backbone networks. The experiments performed with actual data from the GEANT network show suitable results.*

Resumo. *O conhecimento antecipado da tendência de comportamento do tráfego auxilia o administrador de redes nas tarefas de operação e planejamento. O tráfego nos enlaces e nós da rede é formado pela superposição de fluxos origem-destino (OD), de acordo com uma matriz de roteamento, que gera uma estrutura multivariada, podendo atingir alta dimensionalidade. A análise simultânea das tendências de crescimento de grupos de fluxos OD com comportamentos semelhantes permite prever futuros congestionamentos na rede. Este trabalho aplica estatística multivariada na análise e previsão de fluxos de tráfego OD. Trabalhos anteriores mostraram que utilizando análise de componentes principais é possível modelar os fluxos usando um pequeno número de componentes, chamados de autofluxos. Este trabalho propõe uma metodologia para analisar o comportamento dos fluxos OD e sua tendência de crescimento em redes de backbone. Os experimentos realizados com dados reais da rede europeia GEANT mostram resultados satisfatórios.*

1. Introdução

Um fenômeno indesejável que prejudica o desempenho das redes e constitui uma classe de problema, particularmente importante para engenharia de tráfego, é o do congestionamento. Este problema tende a tornar-se crítico à medida que aumentam o

número de usuários e a quantidade e complexidade das aplicações. Uma das formas de controlar o congestionamento sem o desperdício de recursos é prever onde e quando ele ocorre, de modo que se possam tomar medidas pró-ativas.

O planejamento das necessidades futuras em termos de capacidade e dimensionamento em uma rede IP, para prevenir um possível colapso, é uma tarefa desafiadora. Uma prática bastante utilizada nesta área é baseada na experiência do administrador e no estado atual da rede. Um exemplo muito comum é observar o comportamento do tráfego em cada enlace a partir de gráficos gerados com a coleta de informações pelo SNMP. Esta prática motivou diversos trabalhos, tais como [Liu et al. 2006], [Papagiannaki 2005], [Silva 2004], que propuseram técnicas com o objetivo de automatizar esta tarefa. Este tipo de metodologia limita a análise da rede a enlaces isoladamente, pois, apesar de informar toda a carga de tráfego entre os dois pares de nós extremos do enlace, não informa a fonte geradora que está causando o aumento do tráfego. As soluções de *upgrades* de largura de banda em enlaces com tendência de sobrecarga, muitas vezes são executadas sem análises mais profundas. Decisões tomadas desta forma podem resultar em gastos desnecessários que não solucionam o problema.

Existe um conjunto de problemas importantes que requer uma modelagem e análise do tráfego na rede como um todo, ou seja, em todos os enlaces da rede simultaneamente. Nestes problemas estão incluídas a detecção de ataques e anomalias na rede, a estimação de matrizes de fluxos de tráfego OD e a previsão de tráfego para o planejamento de capacidade. Porém as dificuldades de análise crescem proporcionalmente ao tamanho da rede, exigindo maior custo de processamento. [Lakhina 2004] afirma que a análise do tráfego da rede como um todo é um objetivo difícil, que é ampliado ainda pelo fato de a modelagem de um único enlace, por si só, já se constitui uma tarefa complexa. Uma forma de abordar este tipo de análise é reconhecer que o tráfego observado em diferentes enlaces não é independente, e sim determinado por um conjunto de fluxos Origem-Destino (OD). Um fluxo OD é constituído por um conjunto de tráfego que é gerado por um ponto de entrada da rede e destinado a um ponto de saída comum.

De acordo com a abordagem referida acima, o planejamento de capacidade poderá apresentar resultados mais precisos através da análise de tendência do comportamento do tráfego em cada fluxo OD. Esta tarefa ainda se depara com problemas citados, pois o número de pares fluxos OD a serem analisados é formado pelo quadrado do número de nós da rede. Para redes WAN, como backbones de internet, analisar cada fluxo OD tornaria uma tarefa custosa em termos de tempo e processamento, pois além da alta dimensionalidade da amostra, fluxos que não apresentam comportamento crítico, seriam analisados desnecessariamente. A solução investigada neste trabalho busca aplicar a análise de tendência em um conjunto de fluxos OD que apresentam o mesmo padrão de comportamento. A estatística multivariada fornece meios que conduzem à solução proposta. Através da redução da dimensionalidade sem perda das propriedades dos elementos da amostra, são geradas estruturas que apresentam características comuns entre um conjunto de fluxos OD. A técnica de Análise dos Componentes Principais, PCA (do inglês “Principal Component Analysis”) [Jolliffe 1986] será usada neste trabalho para obtenção desta redução.

. Este trabalho propõe uma metodologia para analisar a tendência do comportamento dos fluxos de tráfego OD em redes de backbone. Com esta finalidade, a metodologia segue os seguintes passos: (i) Medição de Tráfego; (ii) Obtenção da Matriz de Tráfego; (iii) Pré-processamento da Matriz de Tráfego; (iv) Análise Estrutural; (v) Análise de Tendência; (vi) Identificação dos Fluxos OD. Os experimentos realizados com dados reais obtidos da rede GEANT mostram resultados satisfatórios.

Este artigo está organizado da seguinte forma. A seção 2 apresenta os trabalhos relacionados. Na seção 3 são descritas as etapas metodológicas para análise do comportamento do tráfego e das técnicas utilizadas. A seção 4 mostra o estudo de caso e a discussão dos resultados obtidos. As conclusões e a bibliografia estão respectivamente nas seções 5 e 6.

2. Trabalhos Relacionados

Existem diversas propostas no que se refere à previsão de tráfego na rede, apresentando distintas finalidades e diferentes técnicas estatísticas.

Pesquisas iniciais em previsão de longo prazo de tráfego de redes IP podem ser encontradas em [Groschwitz 1994], onde é usado um modelo de série temporal linear para o tráfego da rede National Science Foundation Network (NSFNET). Os autores mostram que a série obtida pode ser modelada com ARIMA, oferecendo projeção para acima de dois anos. Em [Sang 2000], os autores focalizaram a projeção do tráfego na Internet sobre pequenas escalas de tempo, como segundos ou minutos, que são relevantes para a alocação dinâmica de recursos. O uso de técnicas de inteligência artificial também foi explorado em [Liang 2002] onde foi utilizada a lógica fuzzy em previsão de tráfego para redes wireless. Enquanto que [Cortez 2006] [Ardhan et al 2007][Satsri et al 2007] abordam o uso de redes neurais.

Em [Silva et al. 2004] é proposto um novo método que prevê o volume de tráfego futuro da ordem de meses. O trabalho baseia-se no histórico do crescimento dos volumes máximo e mínimo do tráfego medidos no enlace de interesse. [Silva 2004] propõe um mecanismo chamado ProgCon para prever possíveis congestionamentos, através da identificação da componente de tendência em séries de tempo que representam o tráfego real de redes, utilizando a transformada de wavelet discreta. Em [Papagiannaki 2005], os autores introduzem uma metodologia para prever quando e onde se faz necessário a adição ou a atualização de um enlace em uma rede IP de *backbone*, projetando o volume de tráfego na rede meses no futuro, os dados de uma série são obtidos para um único enlace entre dois PoPs através do SNMP. [Liu et al. 2006] aplica um provisionamento de largura de banda em modelos de tráfego em larga escala de tempo, com comportamento periódico, forte correlação e distribuição não-gaussiana.

Todos os trabalhos citados acima apresentam soluções voltadas tanto para análise de enlaces isoladamente a partir de dados coletados via SNMP como também de fluxos IP. No entanto, não são considerados problemas que ocorrem devido à alta dimensionalidade da rede, que podem elevar o custo da gerência. Também não são levados em conta possíveis falhas na tomada de decisão do administrador devido ao uso apenas das informações fornecidas pelo SNMP. Nosso trabalho diferencia-se por propor uma solução para tratar tais problemas.

3. Metodologia

Este trabalho faz um estudo da análise de tendência a partir de fluxos OD. Modelos de previsões podem ser construídos a partir destas estruturas e prever o tráfego para todo um conjunto de fluxos OD. Até o presente trabalho, não foi encontrado na literatura nenhum estudo de previsão utilizando tal abordagem sugerida. A figura 1 mostra os passos da metodologia a ser seguida para a aplicação da análise de tendência.

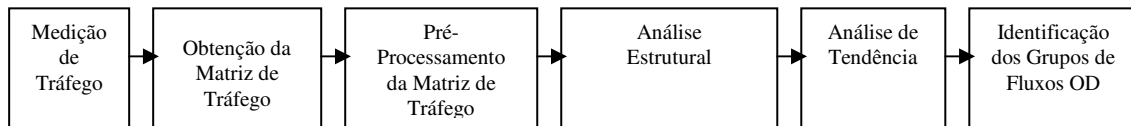


Figura 1: Metodologia para Análise de Tendência com Estatística Multivariada (PCA)

3.1. Medição de Tráfego

A etapa inicial da metodologia se dá através da obtenção de informações sobre os fluxos da rede. Este processo é feito por medição passiva, que consiste em um processo de monitorar o tráfego sem modificar ou injetar algum novo tráfego na rede. Um fluxo é definido como sendo uma agregação do tráfego que se enquadra em determinada regra, e elas devem descrever as características existentes naquele conjunto de pacotes que se deseja medir. Um fluxo pode ser, por exemplo, todo o tráfego que passe pelo coletor, cujo campo endereço IP de origem e de destino sejam semelhantes aos indicados na tabela ou também utilizando uma porta específica.

Alguns equipamentos já possuem ferramentas nativas de fábrica. Um exemplo são os equipamentos da Cisco com a ferramenta NetFlow [Cisco 1999], essa ferramenta possibilita a criação de regras para a identificação de fluxos no próprio roteador da rede, exercendo a função de medidor. Os dados coletados sobre estes fluxos, assim definidos devem ser encaminhados para um servidor, a fim de que possam ser tratados por um software específico de coleta.

3.2. Obtenção da Matriz de Tráfego (MT)

A obtenção da MT a partir de medições de fluxos é feita através do tratamento dos registros de fluxos nos traços contendo informações de coleta entre IPs origem e destino. A partir dos traços coletados, esta etapa consiste em selecionar as informações referentes à quantidade de bytes trafegados entre os pares de nós da rede. Estes dados serão dispostos em uma matriz bidimensional cruzando os nós de origem e destino. Para gerar a MT, serão extraídas as informações referentes aos IPs de origem, de destino e a quantidade de bytes registrada na coleta. Na MT resultante as colunas contém os dados dos pares de nós OD e cada linha, um instante de tempo para cada coleta. Nesta fase, o administrador deverá estimar o tempo de coleta dos traços e espaço de armazenamento necessário, para que políticas sejam estabelecidas em função dos recursos disponíveis. A fim de reduzir o custo das tarefas, o administrador pode decidir, por exemplo, retirar apenas algumas amostras de traços a cada hora de coleta disponível.

Considerando a gerência de um backbone internet, a MT será formada pelos pares de pontos de presença (PoPs) monitorados, onde cada PoP possui sua faixa de IPs. A montagem da MT PoP a PoP segue, assim, os seguintes passos: (i) Selecionar apenas

os IPs pertencentes aos PoPs monitorados; (ii) Agrupar dos pares idênticos de IPs origem-destino de cada instante de coleta; (iii) Contabilizar a quantidade de tráfego de cada par de PoP; (iv) armazenar a MT PoP a PoP.

3.3. Pré-processamento da Matriz de Tráfego

Esta etapa trata de um ajuste dos dados, de modo que os mesmos se tornem adequados para o tipo de análise a ser realizado. Uma das características da amostra a ser tratada é o seu tamanho, que pode acarretar em um maior custo de processamento. Para a redução de séries de longo período, os dados são agregados, onde o intervalo entre duas medidas deverá ser aumentado através da média dos dados dentro de cada um destes intervalos. Outro caso diz respeito à presença de dados atípicos ou *outliers*, que são caracterizadas por serem notavelmente diferentes dos demais.

Métodos estatísticos podem ser empregados, de modo a tornar o processo de detecção mais automatizado. Existem diversos métodos de detecção de outliers, entre eles podemos citar a verificação de valores que estiverem três vezes acima ou abaixo do desvio-padrão. Outra técnica comumente utilizada é através de percentis, verificando os pontos mais distantes do terceiro quartil. A Análise de Componentes Principais também pode ser utilizada para a identificação destes valores atípicos. Para isso, é feito o gráfico da dispersão dos pontos entre a primeira e a segunda componentes principais, os pontos que ficarem fora do intervalo onde há concentração de pontos serão candidatos a *outliers*.

Um processo de previsão trabalha com maior precisão quando sua série está suavizada, ou seja, livre destes tipos de anomalias. Portanto, para o caso de análise tendência, os *outliers* detectados deverão ser substituídos por valores que tornem a série adequada a um processo de previsão.

3.4. Análise Estrutural

A estrutura gerada pela MT pode apresentar altas dimensões, pois uma quantidade n de nós fornece n^2 pares de fluxos OD e a quantidade de instantes de tempo depende do período total de coleta e é inversamente proporcional ao intervalo entre duas coletas sucessivas. Quando surge a necessidade de analisar uma estrutura de alta dimensão, é preciso buscar uma alternativa de aproximação para uma estrutura com baixa dimensionalidade, desde que preserve suas propriedades mais importantes. A técnica de estatística multivariada chamada Análise dos Componentes Principais (PCA) tem por finalidade básica a redução da dimensionalidade a partir de combinações lineares das variáveis originais.

3.4.1 Análise de Componentes Principais

PCA é um método de transformação de coordenadas que mapeiam os dados medidos em novo sistema de eixos, chamados de componentes principais (CP). Cada CP tem a propriedade de apontar para a direção de variância máxima restante nos dados, considerando a energia já representada nos componentes anteriores. Os eixos principais são ordenados pela quantidade de energia nos dados por eles capturados.

Considerando uma matriz X , calcular os CPs é equivalente a resolver o problema de autovalores simétricos para a matriz $X^T X$. A matriz $X^T X$ é a medida de covariância

entre os fluxos. Cada componente principal v_i é o i -ésimo autovetor calculado a partir da decomposição espectral de $X^T X$.

$$X^T X v_i = \lambda_i v_i \quad i = 1, \dots, p \quad (1)$$

onde λ_i é o autovalor correspondente a v_i . Devido ao fato $X^T X$ ser definida positiva simétrica, seus autovetores são ortogonais e os autovalores correspondentes são reais não-negativos. Por convenção, os autovetores têm norma unitária e os autovalores são ordenados de forma decrescente, ou seja, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.

Uma vez que os dados foram mapeados em um espaço de componentes principais, os dados transformados em uma dimensão no tempo podem ser examinados. Considerando os dados mapeados em CPs, temos que a contribuição (equação 2) do eixo principal i em função do tempo é dada por $X v_i$.

$$u_i = X v_i \quad i = 1, \dots, p \quad (2)$$

Os vetores u_i têm tamanho t e são ortogonais por construção. A equação acima mostra que todos os fluxos OD, quando pesados por v_i , produzem uma dimensão de dados transformados. Deste modo u_i captura a variação temporal comum a todos os fluxos por todo o eixo principal i . Como os eixos principais estão em ordem de contribuição da energia total, u_1 captura a mais forte tendência comum temporal para todos os fluxos OD, u_2 captura o próximo mais forte, e assim por diante. Devido ao conjunto $\{u_i\}_{i=1}^p$ capturar a tendência comum de variação no tempo para os fluxos OD, este conjunto é referido como autofluxo de X . Assim a quantidade $r < p$ eixos principais que explicam a maior parte da variabilidade definirá o número de componentes.

O conjunto de CPs $\{v_i\}_{i=1}^p$ pode ser organizado na ordem de colunas de uma matriz principal V , cujo tamanho é $p \times p$. Da mesma forma, podemos formar uma matriz U de dimensões $t \times p$, onde a coluna i é u_i . Assim temos:

$$X_i = U (V^T)_i \quad i = 1, \dots, p \quad (3)$$

onde X_i é a série temporal do i -ésimo fluxo OD e $(V^T)_i$ é a i -ésima linha de V . A equação mostra que cada fluxo OD X_i está em torno de uma combinação linear de autofluxos, com pesos $(V^T)_i$ associados. Considerando as r 's primeiras componentes principais, podemos aproximar o X original como:

$$X \approx \sum_{i=1}^r u_i v_i^T \quad (4)$$

A aplicação de PCA em fluxos de tráfego desenvolvido por [Lakhina 2004] observou a classificação dos autofluxos em três categorias: *d-autofluxo*; *s-autofluxo* e *n-autofluxo*. O grupo de autofluxos que apresentam fortes tendências e periodicidade é classificado como *d-autofluxo*, ou seja, autofluxos determinísticos. A segunda categoria, *s-autofluxo*, diz respeito aos autofluxos que exibem altos valores de curto período, ou seja, contém picos (*spikes*) que são identificados por valores que estão, por exemplo, 4 ou 5 desvios-padrões acima da média do autofluxo. Finalmente, os autofluxos que se

apresentam aproximadamente estacionários e com distribuição Gaussiana são classificados como *n-autofluxo*, ou seja, representando ruídos (*noises*).

Cada um destes autofluxos carrega em sua estrutura características dos fluxos OD, como tendências de crescimento ou queda, períodos de estabilidade e comparações entre períodos de carga no tráfego.

3.4.2. Análise dos fluxos

Fluxos OD com um padrão de comportamento comum estão caracterizados nos autofluxos. Observando os pesos nas CPs referentes a cada fluxos, aqueles com valores acima de um limiar pré-definido são classificados com a presença da característica do autofluxo correspondente. Na figura 2, é mostrado um exemplo de um autofluxo u_i e seus correspondentes eixos principais v_i . Neste caso, a característica deste autofluxo é encontrada em vários fluxos OD, sendo mais fortemente presente nos fluxos OD 196, 201 e 245 por apresentarem maiores pesos. O valor do limiar é calculado pela a equação $1/\sqrt{p}$, definida em [Lakhina 2004], onde p é o número de variáveis aleatórias.

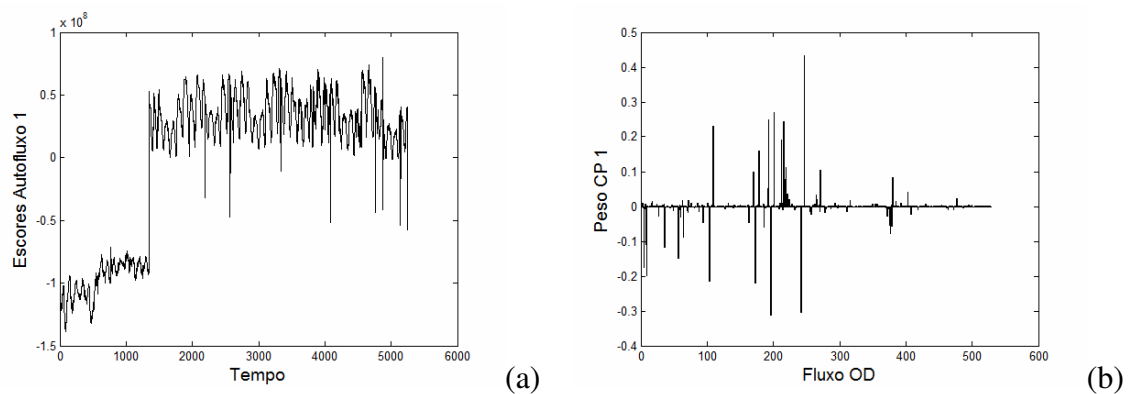


Figura 2: (a) Gráfico do Autofluxo e (b) Gráfico dos pesos para cada fluxo OD neste autofluxo

3.5. Análise de Tendência

A análise de tendência abordada neste trabalho utiliza regressão linear e tem por objetivo encontrar uma função que represente a série temporal. O método dos mínimos quadrados é utilizado para calcular os coeficientes de uma curva que melhor representa um conjunto de dados observados. A função é obtida a partir dos pontos que minimizam a soma dos quadrados dos erros em relação aos pontos originais.

O polinômio de grau k , que compõe a função, representa a possível tendência. A curva estimada é calculada para o autofluxo, pois sua tendência de comportamento deve refletir nos fluxos OD que contém as características deste autofluxo. Com base neste cálculo, é possível verificar a tendência para os próximos períodos do tráfego em vários fluxos simultaneamente.

3.6. Identificação dos Grupos de Fluxos OD

Nesta etapa, são identificados os fluxos que corresponderão à análise feita no autofluxo. A partir da identificação dos fluxos com tendências de crescimento, queda ou

estabilidade é feita a localização dos enlaces pelos quais passam os fluxos considerados críticos de acordo com o interesse do administrador.

4. Estudo de Caso

4.1. Descrição da Rede e Dados

A amostra de dados usada para a aplicação da metodologia proposta refere-se à rede GEANT [Uhlig et al. 2006], uma rede que conecta universidades e instituições de pesquisas pertencentes às Redes Nacionais de Ensino e Pesquisa da Europa (NRENs). Os dados fornecidos foram coletados a partir de uma topologia formada por 23 PoPs (nós) com 74 enlaces. Os dados foram coletados durante 4 meses em intervalos de 15 minutos através do Netflow [Cisco 1999], sendo que a amostra apresenta ausência de coleta durante um período de 6 dias. Na matriz de tráfego a ser analisada, as colunas contêm cada par de fluxos OD e cada linha equivale a um período de coleta. Assim, a MT gerada para esta rede possui 529 colunas por 10773 linhas. A série para cada fluxo OD está com o tempo indexado de 1 a 10773, equivalente aos meses de janeiro(1 a 2941), fevereiro(2942 a 5252), março(5253 a 8017) e abril(8018 a 10773). O software Matlab [Matlab 1991] é usado para implementação e aplicação das operações de PCA e a estimativa da curva média para a análise de tendência.

4.2. Resultados

As duas primeiras etapas consistem na coleta de dados via medição passiva seguido pela obtenção da MT. Os dados da rede GEANT foram fornecidos em arquivos XML, gerado através do *Totem toolbox* [Totem 2005] para o formato de MT. Os cruzamentos entre cada linha e coluna contêm a quantidade de tráfego do fluxo de um nó origem a um nó destino. A MT recebida foi adaptada para um formato trabalhado pelos nossos *scripts*, onde as linhas representam um instante de tempo de coleta e cada coluna contém os dados um par de fluxo OD. A terceira fase prossegue com o pré-processamento dos dados, sendo que os mesmos apresentam dados atípicos.

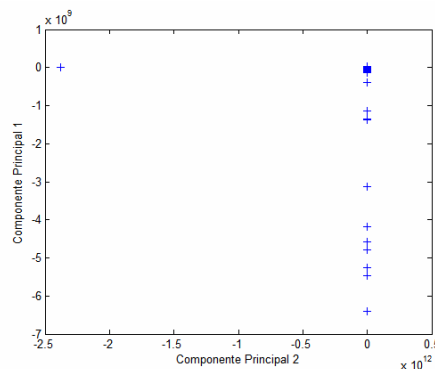


Figure 3: Gráfico CP 1 x CP 2

A detecção de outliers é feita com a aplicação do PCA sobre os dados originais. O gráfico da figura 3 mostra a dispersão dos pontos entre a primeira e a segunda CP. O procedimento matemático utilizado para definir os pontos distantes da área da concentração é auxiliado, neste estudo de caso, pelo cálculo dos percentis. Os dados atípicos encontrados são substituídos nos fluxos por seus valores vizinho anteriores.

Uma nova amostra é gerada para a aplicação da análise multivariada e posterior estudo de análise de tendência.

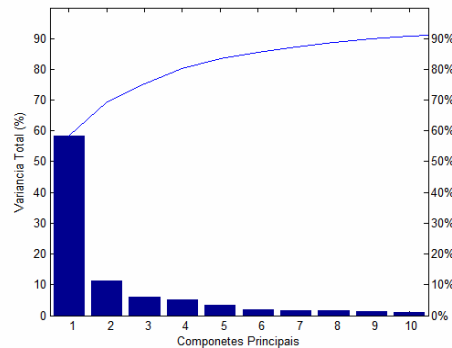
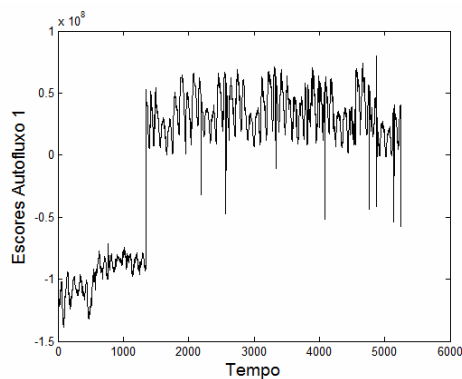


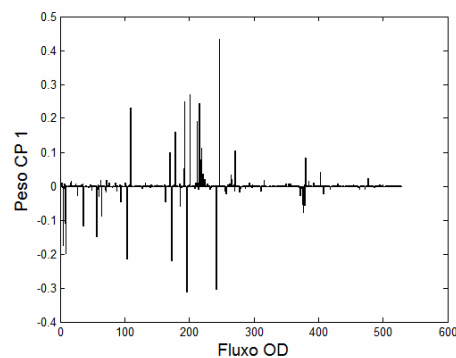
Figura 4: Variabilidade das Componentes Principais.

A quarta etapa da metodologia inicia o processo de redução da dimensionalidade da MT, com a aplicação do PCA. Com este procedimento, a amostra é reduzida para 10 componentes não correlacionados e ortogonais. O gráfico da figura 4 mostra o percentual de variabilidade, em ordem decrescente, explicada por estas 10 componentes. Os resultados apontam que as 10 primeiras CPs explicam 90% da variabilidade total, permitindo obter uma boa representatividade dos dados de tráfego. Com isso, podemos avaliar os autofluxos gerados e seus correspondentes vetores de CP com os pesos de cada fluxo OD que permitem identificar se estes contêm as propriedades do autofluxo.

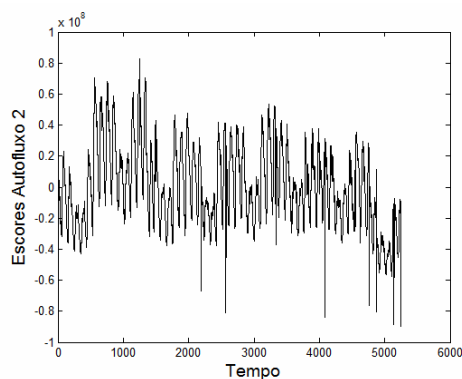
Na figura 5(a,b,c) são mostrados, respectivamente, os autofluxos 1,2 e 3 e seus correspondentes gráficos de pesos (d,e,f) para as três primeiras componentes principais.



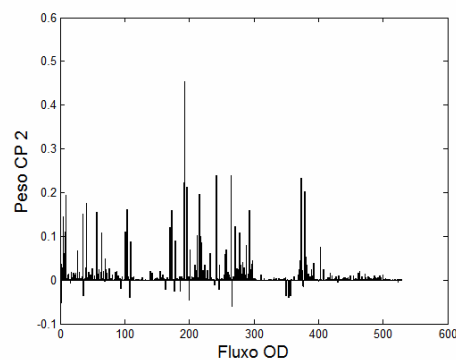
(a)



(d)



(b)



(e)

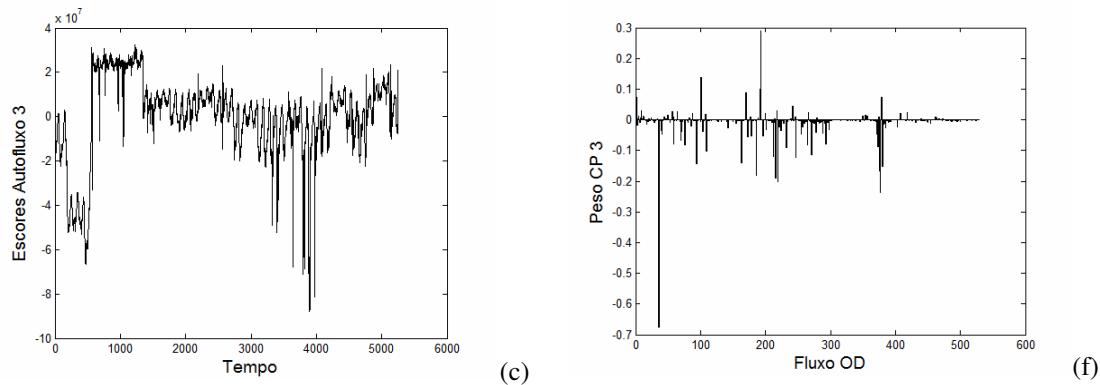
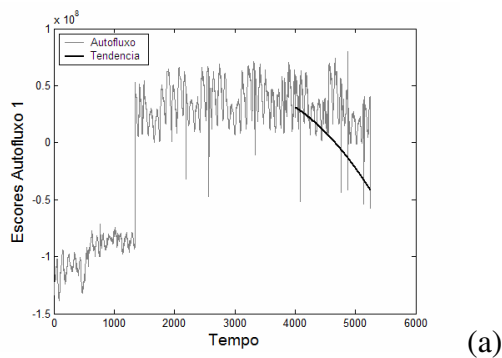


Figura 5: (a)(b)(c) Autofluxos 1,2 e 3; (d)(e)(f) Gráficos dos pesos para as CPs.

A quinta e sexta etapas dizem respeito à análise de tendência dos autofluxos e identificação dos fluxos simultaneamente avaliados. Devido à primeira CP apresentar o maior percentual de variabilidade, utilizamos neste estudo de caso o autofluxo 1, referente ao tráfego caracterizado pelo lado positivo no gráfico. Um polinômio de grau 2 é usado para estimar a curva média na projeção a partir do tempo indexado por 4000 até o final da série no ponto 5252, que corresponde a 13 dias de projeção. A figura 6(a) mostra que, no autofluxo 1, a projeção da curva de ajuste segue uma tendência de queda acompanhando o comportamento da série original no intervalo de tempo analisado.

A etapa de identificação dos fluxos OD, feita com base nos pesos referentes à primeira CP 1, permite encontrar os fluxos OD que ultrapassam em valor absoluto o limiar para os valores negativos e positivos do autofluxo 1.

As figuras 6(b,c,d,e) apresentam alguns dos gráficos dos fluxos OD e suas respectivas tendências no mesmo intervalo de tempo analisado no autofluxo. Verifica-se que a tendência de queda corresponde ao resultado encontrado na análise do autofluxo, o que permite que a partir de uma única estrutura seja possível analisar a tendência de comportamento simultaneamente em um grupo de fluxos OD. A identificação do(s) fluxo(s) OD responsáveis pela tendência de mudança no comportamento do tráfego permite melhor decisão sobre qual enlace sofrerá atualização. Da mesma forma que foi detectada queda no comportamento do tráfego, projeções de crescimento também poderão ser estimadas.



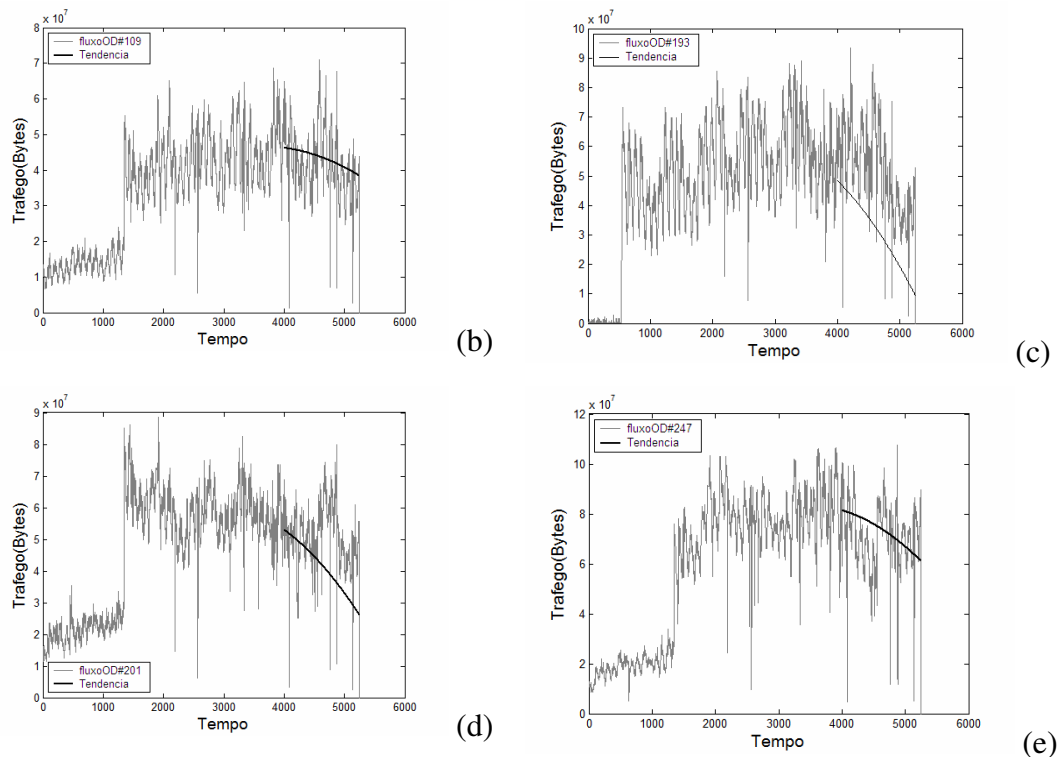


Figura 6: (a) Tendência no Autofluxo 1 e (b)(c)(d)(e) Tendência nos Fluxos OD.

5. Conclusão

Este trabalho mostrou a viabilidade de se realizar uma Análise da Tendência do tráfego em grandes redes IP através da redução da dimensionalidade das séries utilizando Análise de Componentes Principais e aplicá-la na antecipação dos efeitos sobre os enlaces e nós da rede. O processo consistiu em: estimar ou medir a série multivariada de fluxos de tráfego OD, reduzir a dimensionalidade via PCA, realizar a análise de tendência nas CPs significativas, identificar os fluxos OD com tendências de alteração para, finalmente, identificar os ajustes no roteamento ou, via matriz de roteamento, os enlaces candidatos a redimensionamento, seja ao aumento ou diminuição das suas taxas. Os resultados mostraram que estudos de projeção de tráfego nos autofluxos podem refletir nos vários fluxos OD que apresentam características comuns, auxiliando o administrador de grandes redes no planejamento da capacidade. É claro que esta é apenas uma das variáveis consideradas pelo administrador da rede na tomada desta decisão, mas, certamente, uma das mais importantes. Outras técnicas de estatística multivariada e de projeção serão experimentadas e avaliadas para investigar melhorias na precisão dos resultados.

6. Referências

Ardhan, S. et al (2007). Improved Model for traffic fluctuation prediction by Neural network. In International Conference on Control, Automation and Systems 2007. Outubro 2007, Seoul, Coreia.

- Bremner-Barr, A., Cohen, E., Kaplan, H., e Mansour, Y. (2003). Predicting and bypassing end-to-end Internet service degradations. *IEEE Journal on Selected Areas in Communications*, 961–978.
- Cisco NetFlow. <http://www.cisco.com/warp/public/732/Tech/nmp/netflow/>. 1999.
- Cortez, P. et al. Internet Traffic Forecasting using Neural Networks. In: International Joint Conference on Neural Networks. Vancouver, BC, Canada. July 16-21, 2006
- Daubechies, I. (1992). Ten Lectures on Wavelets. SIAM, Philadelphia.
- Groschwitz, N. K. e Polyzos, G. C. (1994). A time serie model of long-term NSFNET backbone traffic. Proc. IEEE ICC'94 pp. 1400-1404.
- Hair, J. et al. (1998) Multivariate Data Analysis. Editora Prentice Hall, New Jersey.
- Jolliffe, J. T. (1986). Principal component analysis". SpringerVerlag, New York.
- Lakhina, A., Papagiannaki, K., Crovella, M., Diot, C., Kolaczyk, E. D. and Taft, N. (2004). Structural Analysis of Network Traffic Flows. In: *Proceedings of ACM SIGMETRICS / Performance 2004*. pp. 61-72.
- Liu, Z. et al. (2006). Periodic Data Traffic Modeling and Prediction-Based Bandwidth Allocation. In Proceedings of the 4th Annual Communication Networks and Services Research Conference (CNSR'06), 131 - 138.
- Liang, Q. (2002). Ad hoc wireless network traffic-self-similarity and forecasting. In *Communications Letters, IEEE*, Volume 6, Issue 7, Page(s):297 - 299.
- MATLAB for Windows User's Guide, The Math Works Inc., 1991.
- Papagiannaki, K., Taft, N., Zhang, Z., e Diot, C. (2005). Long-term forecasting of Internet backbone traffic. In *IEEE Transactions on Neural Network*, Vol. 16, No. 5.
- Sang, A. e Li, S. 2000. A predictability analysis of network traffic. In *Proc. of the IEEE INFOCOM'2000*, Tel Aviv, Israel.
- Satsri, S. et al. (2007). ANN based NGN IP Traffic Prediction in Thailand. In International Conference on Control, Automation and Systems 2007. Outubro 2007, Seoul, Coreia.
- Silva, J. L. C. (2004), ProCon – Prognóstico de Congestionamento de Tráfego de Redes usando Wavelets. Tese de Doutorado, (UFPE).
- Silva, E. et al. (2004) Um método para projeção de tráfego usando wavelets e fecho convexo. 21º. Simpósio Brasileiro de Telecomunicações, Belém, PA.
- Totem, "A Toolbox for Traffic Engineering Methods," <http://totem.info.ucl.ac.be>, 2005.
- Uhlig, S., Quoitin, B., Balon, S. and Lepropre, J. (2006). Providing public intradomain traffic matrices to the research community. *ACM SIGCOMM Computer Communication Review*, 36(1).
- Zhao, G; Tang, H; Xu, W; Zhang, Y. (2004). Application of neural network for traffic forecasting in telecom networks. *Machine Learning and Cybernetics. Proceedings of 2004 International Conference on Volume 4*, 26-29.