

Um método para projeção de tráfego usando wavelets e fecho convexo

Edmundo de Souza e Silva¹, Rosa M. M. Leão¹, Marcos B. Trindade²,
Antonio A. de A. Rocha¹, Bruno F. Ribeiro¹, Flávio P. Duarte¹ e Jorge A. Azevedo¹

Resumo—O conhecimento prévio do volume de tráfego de um enlace permite ao administradores da rede um melhor planejamento da expansão da capacidade da sua rede IP. Este trabalho propõe um novo método baseado em wavelets e fecho convexo para, a partir do histórico, prever a quantidade de tráfego em um enlace com meses ou anos de antecedência. O método proposto foi comparado com previsões feitas com métodos existentes de regressão linear e modelos de auto-regressão sobre dados reais da rede da RNP. Os resultados mostraram que o método proposto possui maior precisão e menor grau de complexidade na sua parametrização.

Abstract—The foreknowledge of a link's traffic volume allows network administrators to better plan IP network capacity growth. This work proposes a new methodology based on wavelets and convex hull in order to predict a link's amount of traffic with months or years ahead. The proposed method was evaluated against other existing forecast methods, such as linear regression and autoregressive models. Actual RNP network data was used in this evaluation. The results show that the proposed method allows greater precision and has much lower complexity in its parametrization.

I. INTRODUÇÃO

Prever o tráfego futuro para uma rede IP não é uma tarefa trivial. Porém, o conhecimento da projeção do volume de tráfego passante em um enlace permite que se possa planejar com antecedência a expansão ou redução da capacidade de um enlace da rede. Uma boa previsão do tráfego futuro do enlace necessita de um bom histórico do tráfego no mesmo. Por este motivo a qualidade das medidas obtidas é fundamental. Infelizmente, nem sempre ela é adequada, o que ocorre principalmente devido à falta de dados por problemas técnicos na sua coleta. Além da falta de dados, podemos ter eventos incomuns que afetem as medidas por um curto espaço de tempo mas que mudem sensivelmente o comportamento esperado dos dados. Outro problema que interfere diretamente nos resultados de previsão é a granularidade com que estas medidas são obtidas. Medições com granularidade relativamente fina (para o propósito de projeção), onde o intervalo entre duas amostras consecutivas é normalmente da ordem de minutos, possuem inconvenientes de armazenamento dos dados devido ao seu enorme volume e algumas vezes são complicadas de serem coletadas. Medições de granularidade grossa, onde o

intervalo é normalmente da ordem de horas ou mesmo dias, muitas vezes ocultam detalhes que podem ser importantes para a estimativa.

Para avaliar a precisão dos métodos foram usados dados fornecidos pela RNP (Rede Nacional de Pesquisa) de alguns enlaces de seu *backbone*. Os dados equivalem a coletas feitas da média diária do volume de tráfego passante em um enlace. Os dados cedidos pela RNP equivalem às medidas coletas nos dois sentidos para diversos enlaces. (RJ-AMPATH, RJ-DF, RJ-MG, RJ-SP e RJ-RS são alguns desses enlaces.) Todas as coletas foram divididas em duas partes: a primeira parte consiste nas amostras que foram dadas como entrada para os métodos de projeção. Esta parte selecionada para análise equivale a aproximadamente dez meses (300 dias) do total de 14 meses (415 dias) disponibilizados pela RNP; o restante das amostras foram utilizadas para validar a precisão da estimativa feita pelos algoritmos.

Neste trabalho foram avaliados alguns dos métodos de previsão existentes na literatura, mais notadamente os métodos de regressão linear e de auto-regressão. Os resultados obtidos destes métodos nem sempre se mostraram precisos quando aplicados ao conjunto de dados disponíveis. O método proposto neste trabalho foi avaliado sob o mesmo conjunto de dados e os suas previsões foram consistentemente precisas.

A organização deste trabalho é feita da seguinte forma. Na Seção II serão descritos alguns métodos de projeção de tráfego, com as vantagens e desvantagens particulares a cada um dos métodos avaliados. Na Seção III será apresentado um novo método para projeção de tráfego que utiliza filtros de wavelet e o fecho convexo. Diversos resultados obtidos com o novo método e comparações com os demais métodos avaliados neste trabalho serão apresentados na Seção IV. Por fim, na Seção V serão abordadas as conclusões deste trabalho.

II. ALGUNS MÉTODOS DE PROJEÇÃO DE TRÁFEGO

Um bom método de previsão precisa estar atento a todos os problemas inerentes às coletas que possam interferir em seus resultados. O problema de prever o tráfego é basicamente o de se estimar uma série temporal, provavelmente não estacionária, onde alguns pontos da série estão faltando. A seguir serão apresentados alguns métodos de filtragem da série temporal que auxiliam a sua previsão e depois alguns métodos de predição de séries temporais.

A. Filtragem dos dados.

Fatores como finais de semana, feriados, ou mesmo possíveis redirecionamentos de rotas, são responsáveis por

¹ Universidade Federal do Rio de Janeiro COPPE/Prog. de Eng. de Sistemas e Computação e Depto. de Ciência da Computação do IM. E-mails: {arocha, bruno, flaviop, allyson, rosam, edmundo}@land.ufrj.br.

² Fundação CPqD Centro de Pesquisa e Desenvolvimento em Telecomunicações. Rodovia Campinas - Mogi-Mirim, km 118,5 CEP 13086-902 - Campinas, SP - Brasil E-mail: trindade@cpqd.com.br.

Este trabalho foi parcialmente financiado pelo CNPq e CPqD.

“picos” (negativos ou positivos) na série temporal original que pode atrapalhar a estimativa da tendência de crescimento. Para reduzir as chances de que “picos” tenham influência significativa nas estimativas, a série temporal original pode passar por um processo de filtragem, onde estes “picos” são suavizados.

O processo utilizado é o mesmo usado em processamento de sinais. Desta forma podemos pensar na série temporal como um sinal. Um dos métodos mais eficientes para filtragem desses “picos” é a transformada de wavelet multi-resolução.

1) *Transformada de wavelets à-trous.*: A análise de wavelets multi-resolução descreve um processo de sintetização de um sinal discreto começando por um sinal de muito baixa resolução (na escala de tempo mais grosseira) e, sucessivamente, adicionando mais detalhes para criar resoluções mais altas do mesmo sinal. Uma destas wavelets de multi-resolução é a “à-trous wavelet transform”. Mais detalhes do método brevemente descrito abaixo podem ser encontrados em [3]. As propriedades de uma transformada de wavelet, mais especificamente da transformada “à-trous” (com buracos), podem ser encontradas em maiores detalhes em [5].

A j -ésima resolução da transformada é definida por

$$c_j(k) = \sum_l h(l)\phi(k - 2^{j-1}l),$$

onde $\phi(x)$ é uma função de escala e $h(l)$ é um filtro de passa baixa. A diferença de sinal w_j entre duas resoluções consecutivas é $w_j(k) = c_{j-1}(k) - c_j(k)$.

Para a função de escala $\phi(x)$, o filtro B-spline de grau 3 foi utilizado. A escolha deste filtro se deve pela necessidade de um filtro que não altere as principais características da série, como sua tendência de crescimento, e nem o formato da curva. Importante que o filtro não possua descontinuidades (como uma spline, por exemplo), não haja mudança de fase (linear), que não permita novos pontos de máximo ou mínimo, seja simétrico, possa ser usado como um banco de filtros, não amplifique o sinal (soma dos coeficientes seja 1) e tenha boa localização espacial (poucos coeficientes). Todas estas características são bem satisfeitas pelo filtro B-spline usado.

O sinal original pode ser reconstituído através da equação

$$c_0(k) = c_p(k) + \sum_{j=1}^p w_j(k).$$

Um cuidado especial é necessário com as “bordas” dos dados analisados. Quando os dados analisados possuem poucas amostras ou a wavelet possui uma resolução muito baixa (j muito alto), a quantidade de amostras da transformada de wavelet dos dados pode ser muito pequena. Isto porque o número de amostras eliminadas da j -ésima resolução da transformada de wavelet é $\sum_{i=1}^j 2^i$. Como exemplo tem-se a 6ª resolução de uma transformada à-trous wavelet, que já elimina 126 amostras dos dados iniciais. Quando há um número pequeno de amostras, não é possível usar resoluções muito altas. A Figura 1 ilustra o resultado da transformada de Wavelet sobre a média diária do tráfego no enlace RJ-MG da RNP (Rede Nacional de Pesquisa).

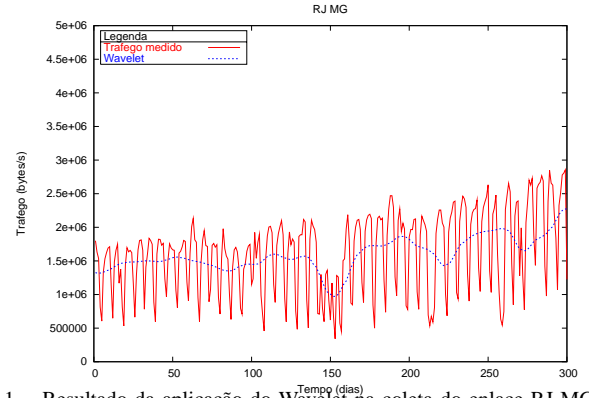


Fig. 1. Resultado da aplicação do Wavelet na coleta do enlace RJ-MG.

B. Previsão com Regressão linear.

O objetivo é tentar encontrar uma função que represente adequadamente a série temporal. Existem alguns métodos para realizar tal tarefa, entre eles um dos mais usados é a Regressão Linear, que utiliza o Método dos Mínimos Quadrados para calcular os coeficientes de uma curva que melhor representa um conjunto de dados observados.

A aplicação do Método dos Mínimos Quadrados tem como objetivo produzir uma função que se ajuste aos pontos da série. Essa função é obtida a partir dos pontos que minimizam a soma dos quadrados dos erros em relação aos pontos originais. Para melhorarmos a qualidade da previsão, precisamos também incluir elementos sazonais tais como os finais de semana.

Neste caso, a primeira parte da função $y(n)$ é composta por um polinômio, $f(n)$, de grau k , que representa a possível tendência da série (o que está sendo procurado), e a segunda composta por um somatório de cossenos e senos, representando a periodicidade ou sazonalidade da série (para diminuir a influência de efeitos sazonais na predição do tráfego).

O cálculo dos coeficientes da curva é feito através da resolução de um sistema linear de equações.

Vantagens do método: (a) Baixa complexidade de implementação e de manipulação da massa de dados. (b) Pode levar em conta fatores sazonais para melhorar a previsão futura.

Desvantagens do método: (a) “Buracos” nas coletas de dados podem prejudicar a avaliação da função por afetarem a periodicidade das medidas. (b) A determinação dos harmônicos dos efeitos sazonais mais importantes da função é por tentativa e erro. E uma escolha errada dos harmônicos reflete nos coeficientes da função $f(n)$.

A seguir revisitaremos um outro método de predição de séries temporais, de auto-regressão.

C. Previsão com modelos de auto-regressão

Modelos de predição de séries temporais que usam processos auto-regressivos e/ou com média móvel têm como objetivo reproduzir as características de auto-correlação dessas séries. A auto-correlação, de distância k , mede a influência média que a quantidade de tráfego medida na amostra n tem na quantidade de tráfego medida na amostra $n + k$. Definamos $\mathcal{X} : \{X_l : l = 0 \dots N\}$ como um processo estocástico, onde X_l é a variável aleatória que representa o tráfego do l -ésimo

instante de medição e N é o número total de medições. O nosso objetivo é estimar o processo \mathcal{X} através somente de um caminho amostral.

Caso possamos obter um modelo que contenha todas as características de auto-correlação do tráfego, poderemos a partir do conjunto de variáveis aleatórias de n_0 até a n_1 -ésima medição do processo estocástico \mathcal{X} , definir a $n_1 + 1$ ésima variável aleatória. Ou seja, poderemos definir X_{n_1+1} em função de X_{n_0}, \dots, X_{n_1} .

Entretanto, existem alguns requisitos necessários para a modelagem através da auto-correlação. O principal deles é a estacionaridade dos dados. Um processo estocástico é dito fracamente estacionário quando algumas características suas não se alteram no decorrer do tempo. Assim, para um processo fracamente estacionário as seguintes relações precisam ser satisfeitas para todo n : $E[X_n] = E[X_0]$, $E[X_n^2] = E[X_0^2]$ e $Cov(X_n, X_{n+k}) = Cov(X_0, X_k)$. Voltemos a exemplificar a não estacionaridade no caso do modelo de tráfego: todo tráfego em geral possui alguma tendência de crescimento, grandes diferenças de volume de tráfego entre os períodos noturno e diurno; o tráfego de uma rede acadêmica diminui no período de férias e o tráfego das rotas internacionais de uma rede de telefonia aumenta substancialmente no Natal. Exceto a tendência de crescimento do tráfego, todos os demais componentes citados acima são componentes sazonais, pois estão intimamente relacionados ao calendário.

Para modelar usando a auto-correlação, precisamos remover os componentes não estacionários. No caso específico dos modelos AR ou MA, é necessário remover estes componentes, juntamente com a tendência de crescimento do tráfego, antes de proceder à análise das auto-correlações. Uma infinidade de técnicas podem ser utilizadas para a remoção das sazonalidades e não existe um método eficiente para todas as situações. Na sessão de comparação dos resultados, estaremos analisando a metodologia utilizada por Papagiannaki et al. [3] para atingir a estacionaridade, juntamente com seus aspectos positivos e negativos.

Uma vez conseguida a estacionaridade dos dados, já podemos usar a auto-correlação para definir os parâmetros do modelo a ser utilizado. O componente auto-regressivo do modelo pode ser facilmente identificado através do gráfico da auto-correlação amostral. O leitor deve ter sempre em mente que a auto-correlação amostral não é equivalente à auto-correlação do processo estocástico, as duas somente são equivalentes se o número de amostras é infinito. Por ser a auto-correlação de um caminho amostral do processo estocástico \mathcal{X} , a auto-correlação amostral possui um intervalo de confiança $\pm 1.96/\sqrt{N}$ (que é o seu 95% percentil), onde N é o número de amostras. Desta forma, diz-se que a auto-correlação amostral é zero na distância k quando $-1.96/\sqrt{N} \leq Cor(X_0, X_k) \leq 1.96/\sqrt{N}$.

A autocorrelação é a principal medida para se estimar os parâmetros da auto-regressão. Um número alto de amostras para a autocorrelação é, portanto, fundamental para que possamos confiar no método. O parâmetro p da auto-regressão $AR(p)$ pode ser facilmente determinado pela auto-correlação amostral parcial dos dados. Já o parâmetro q de média móvel $MA(q)$ é obtido a partir da auto-correlação amostral. Os modelos $AR(p)$ e $MA(q)$ podem ser combinados para formar

o modelo $ARMA(p, q)$, que possui as duas propriedades, a auto-regressão e a média móvel.

1) *Processo ARIMA, SARIMA e outros*: Todos os demais processos baseados nos modelos AR e MA surgiram da necessidade da estacionaridade dos dados. O modelo mais popular é o $ARIMA(p, d, q)$, que é um modelo $ARMA(p, q)$ sobre as amostras X'_n , onde $X'_n = (1 - B)^d X_n$, onde B denota o operador de deslocamento para trás $BX_n = X_{n-1}$, $B^2 X_n = X_{n-2}$, etc. A diferenciação em muitos casos faz com que um modelo não estacionário se torne estacionário. Um conjunto de amostras cuja média cresce linearmente “pode” se tornar estacionário através da diferenciação com $d = 1$. Já o modelo $SARIMA(p, d, q)(P, D, Q)_s$ (Seasonal ARIMA) tenta tornar estacionária uma série temporal com componentes sazonais, onde a diferenciação $(1 - B^s)^D$ é incluída junto ou no lugar da diferenciação descrita acima. Para maiores detalhes sobre modelos ARMA, ARIMA e SARIMA, ver [1].

2) *Modelo proposto por Papagiannaki et al. [3] para amostras de tráfego em canais não congestionados*: A abordagem proposta por Papagiannaki et al. [3] se resume nos seguintes passos: (a) Obter a média de 1.5 hora do tráfego no canal. (b) A partir destes dados, obter o modelo ARIMA do resíduo $w(k)$ da transformada de wavelet de resolução de 12 horas. (c) Obter um modelo ARIMA para a média de tráfego semanal. (d) A previsão de tráfego será a a previsão do modelo ARIMA da média semanal \pm previsão do desvio padrão da semana.

Vantagens do método: (a) Bem parametrizado, pode conseguir boas estimativas do tráfego. (b) Por ser um modelo complexo, pode conseguir prever em detalhes o comportamento do tráfego por um curto espaço de tempo.

Desvantagens do método: (a) Alta complexidade de implementação e manipulação da massa de dados. (b) Dificuldade de fazer com que a série temporal fique fracamente estacionária. (c) Dificuldade de parametrizar corretamente o modelo. (d) A sazonalidade pode dificultar a predição do tráfego. (e) Não funciona bem com um pequena quantidade de dados. (f) Dificuldade de prever o comportamento do tráfego para longos períodos de tempo.

III. MÉTODO DE PROJEÇÃO COM FECHO CONVEXO

A alta complexidade de implementação e manipulação dos dados, necessidade de coletas com características apropriadas, tais como grande quantidade de amostras, curto período de tempo, série temporal fracamente estacionária e garantia da periodicidade, são alguns dos problemas inerentes aos métodos de projeção com regressão linear ou auto-regressão. Essas desvantagens, que já foram descritas na Seção II deste trabalho, motivam a definição de um novo método para previsão de tráfego que possua uma complexidade pequena e poucas restrições às coletas utilizadas como base para a projeção.

A nova proposta define um método simples e eficiente para prever o volume de tráfego de um enlace no futuro, com base no histórico do crescimento dos volumes máximos e mínimos do tráfego medido neste mesmo enlace. O objetivo é estimar uma curva que represente o crescimento médio entre os volumes máximos e mínimos identificados em uma coleta e utilizar esta mesma curva para estimar uma projeção do

tráfego em um período futuro. Experimentos mostrados na Seção IV mostram que bons resultados podem ser obtidos por este método.

O método proposto resume-se na execução de quatro processos: Filtragem da série através de transformada wavelet; identificação dos máximos e mínimos da coleta; estimativa da curva média dos máximos e mínimos; e projeção do tráfego. Cada um desses processos será detalhado nas próximas seções.

A. Filtragem da série

Para evitar que picos influenciem de forma significativa as estimativas, uma filtragem deve ser aplicada na coleta. Diversas espécies de filtros podem ser utilizados, no entanto, o filtro da transformada de wavelet foi utilizado por razões já descritas na Seção II-A. Devido ao número de amostras eliminadas pela transformada de Wavelet, o método proposto utiliza a terceira resolução do filtro. A Figura 1 mostrada na Seção II ilustra o resultado da transformada de Wavelet sobre uma coleta feita no enlace RJ-MG do *backbone* da RNP.

Com a filtragem dos dados, a identificação dos máximos e mínimos, assim como da curva média, será feita a partir dos dados filtrados.

B. Identificação dos máximos e mínimos

Após aplicada a transformada de Wavelet na coleta inicial, é necessário identificar curvas que representem o comportamento dos valores máximos e mínimos desta coleta. A identificação destes valores pode ser feita através da criação de um envelope que contorne todos os pontos da coleta. Considerando o envelope formado por duas curvas que têm início no menor valor da coleta na dimensão “x” e fim no maior valor da coleta desta mesma dimensão “x”, este envelope consiste no fecho convexo formado a partir das amostras da coleta.

O fecho convexo, FC, de um conjunto de pontos consiste no menor polígono convexo formado por um subconjunto desses pontos, onde todos os outros pontos deste conjunto se encontram na parte interior do polígono. Os pontos pertencentes ao subconjunto equivalem aos vértices do polígono definido. Esses vértices podem estar contidos em um dos dois limites do fecho convexo: superior ou inferior. O limite inferior de um fecho é formado pelos vértices inferiores do polígono entre o ponto de menor valor na dimensão “x” até o de maior valor na dimensão “x”. O limite superior de um fecho é formado pelos vértices superiores do polígono entre o ponto de menor valor na dimensão “x” até o de maior valor na dimensão “x”. (O fecho convexo tem sido usado por técnicas de medição ativa, para estimar o atraso de pacotes em um sentido [2], [4], [6].)

No método proposto o conjunto de pontos é representado pelas amostras da coleta após a aplicação do filtro, representada aqui por Ω . Seja então, $\Omega := [v_i = (i, m_i) | i = 1, \dots, N]$ o resultado da transformada, onde i é o dia da coleta, m_i o volume de tráfego médio referente ao dia i , e N o número total de dias de coleta. Dois algoritmos foram implementados para identificar os limites inferior e superior do fecho convexo de uma coleta, $FC(\Omega)$. Para identificar o limite inferior, o algoritmo faz uma análise de todos os pontos (v_i) da coleta

seguindo a ordem crescente, definida pelos dias das amostras. Para cada ponto da coleta o algoritmo determina se este v_i é um vértice inferior do polígono. O pseudo-código desse algoritmo está descrito em [2], [6]. De forma semelhante um algoritmo verifica os pontos (v_i) da coleta para identificar os vértices do limite superior do fecho convexo.

A Figura 2 ilustra as estimativas do fecho convexo para os pontos da curva resultante da transformada wavelet de resolução 3.

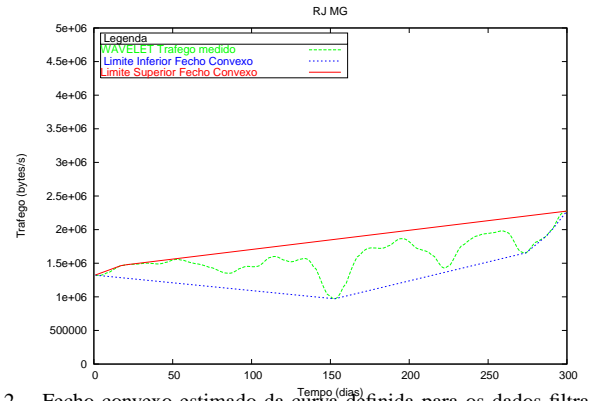


Fig. 2. Fecho convexo estimado da curva definida para os dados filtrados.

C. Curva média dos máximos e mínimos

Uma vez identificados os limites superiores e inferiores do fecho convexo, calculamos a “Curva Média do Fecho Convexo”. Essa curva representa a média entre as curvas formadas pelos vértices dos limites, inferior e superior, do fecho convexo. Para cada dia da coleta são estimados os volumes de tráfego definidos pelas curvas dos limites inferior e superior e calculado o valor de tráfego médio referente àquele dia. A curva média do fecho convexo é composta pelos vértices formados pelos dias da coleta e valores médios calculados para esses dias. O objetivo é utilizar esta curva para tentar estimar o crescimento do volume médio do tráfego medido.

A Figura 3 ilustra a curva média do fecho convexo estimativa a partir da diferença entre as tendências de crescimento dos limites inferiores e superiores do fecho convexo. Essa curva representa o crescimento médio do volume de tráfego no enlace RJ-MG identificado pelo método do fecho convexo.

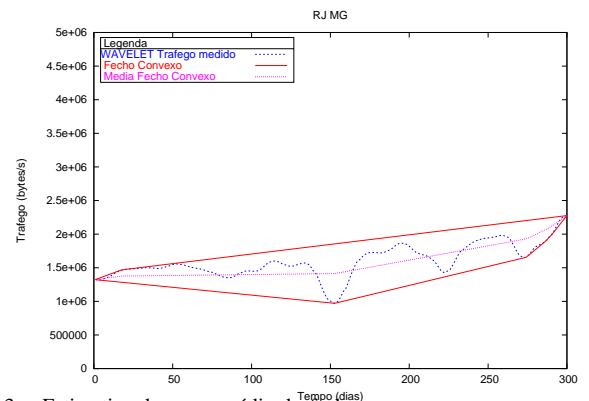


Fig. 3. Estimativa da curva média do fecho convexo.

D. Projeção de Tráfego

Para prever o volume de tráfego em um determinado enlace estima-se o polinômio que melhor represente a curva média

do fecho convexo. Utilizamos uma regressão linear de um polinômio de primeiro grau para estimar a reta que melhor representa a curva média do fecho convexo.

Polinômios de grau 1, 2 e 3 foram testados para fazer o *fitting* da curva média na projeção. Dentre os resultados obtidos, o polinômio de primeiro grau se mostrou o mais adequado para a projeção do tráfego. Os resultados estimados pelos polinômios de grau 1 se aproximaram mais dos valores reais do que os polinômios de grau 2 e 3. Isso se explica devido ao fato do tráfego medido ter apresentado um crescimento linear. Dessa forma, todas as projeções de tráfego apresentadas com este método consistem em estimativas com polinômios de primeiro grau.

Para determinar uma variação da projeção estimada, foram calculadas as variações máxima e mínima percentual do tráfego medido em relação à reta de projeção. Essas variações são identificadas nas figuras como “Percentil”. A variação máxima percentual é estimada por $S_{max} = \max\{|\frac{X_i - M_i}{X_i}|\}$, onde M_i é o ponto da reta de projeção estimada para o instante i , e X_i é o tráfego medido para o mesmo instante i . Analogamente, a variação mínima percentual pode ser descrita por $S_{min} = \min\{|\frac{X_i - M_i}{X_i}|\}$.

A projeção resultante obtida pelo método de wavelet com fecho convexo está ilustrada na Figura 4 onde observamos o tráfego RJ→MG. Nesta figura, além das curvas do tráfego medido e da wavelet, são apresentadas outras curvas: a reta de projeção, que equivale ao polinômio de primeiro grau estimado através da regressão linear; as retas percentil que equivalem às variações máximas e mínimas do volume para aquele ponto da reta de projeção; e curva utilizada para validação do método, chamada de tráfego validado, que equivale às amostras restantes do tráfego medido.

O resultado demonstra a precisão do método. A projeção para o volume médio de tráfego encontra-se bem próxima do volume médio real referente aos quatro meses subsequentes (tráfego validado). Assim como a média, a projeções para as variações máximas e mínimas do tráfego se mostraram, também, bastante precisas. Praticamente todos os pontos do “tráfego futuro” encontram-se entre as retas de variação do tráfego e estas estão bem ajustadas ao tráfego.

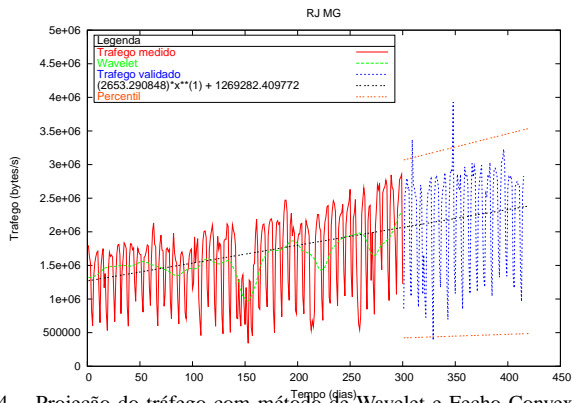


Fig. 4. Projeção do tráfego com método de Wavelet e Fecho Convexo.

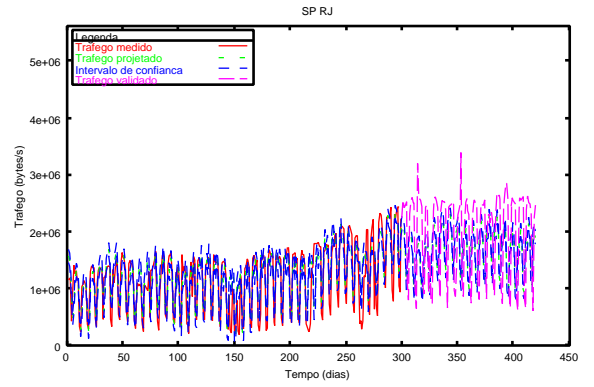
IV. RESULTADOS E COMPARAÇÃO DOS MÉTODOS

Nesta seção serão apresentados alguns resultados de testes realizados para projeção utilizando os métodos avaliados

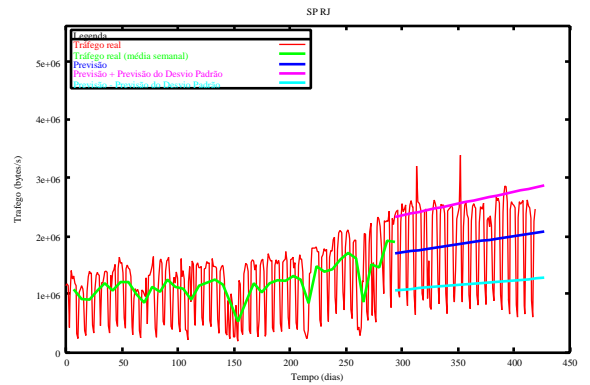
neste trabalho. O objetivo é possibilitar que seja feita uma comparação dos métodos estudados. A partir dos resultados mostrados é possível verificar as estimativas de previsão gerada pelos métodos avaliados para situações semelhantes. Cada uma das figuras apresenta resultados dos três métodos conforme a legenda abaixo:

- (a): projeção usando regressão linear
- (b): projeção usando auto-regressão
- (c): projeção usando wavelet com fecho convexo

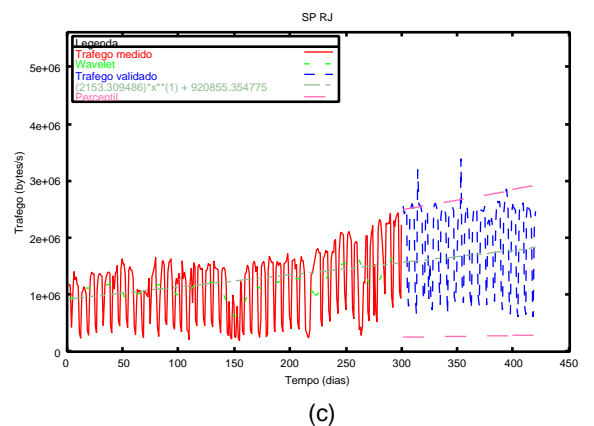
As figuras apresentam como valor máximo de escala do eixo-Y o valor real das capacidades do enlace de transmissão estudado.



(a)



(b)



(c)

Fig. 5. Resultados das projeção para o enlace SP-RJ

As estimativas foram feitas para todos as coletas fornecidas pela RNP. No entanto, por questão de espaço serão apresentados apenas alguns dos resultados. A Figura 5 ilustra os resultados obtidos pelos métodos aplicados às coletas do enlace SP-RJ. Pelo gráfico mostrado na figura, pode-se verificar que

os resultados obtidos para todos os métodos aplicados a este enlace resultaram em estimativas bastante razoáveis. Porém, esta precisão demonstrada em todos os métodos foi um caso único dos resultados obtidos neste trabalho.

Embora as projeções obtidas nos testes usando o método de auto-regressão tenham sido mais precisas do que utilizando o método de regressão linear, as previsões fornecidas pelo método de auto-regressão também não foram satisfatórias em diversas estimativas. Já as estimativas de projeção utilizando wavelets e fecho convexo se mostraram sempre mais precisas do que os resultados obtidos por ambos os métodos. A Figura 6 ilustra a previsão feita para o enlace RJ-RS, onde é possível verificar que apenas o método de projeção proposto neste trabalho fez a previsão corretamente.

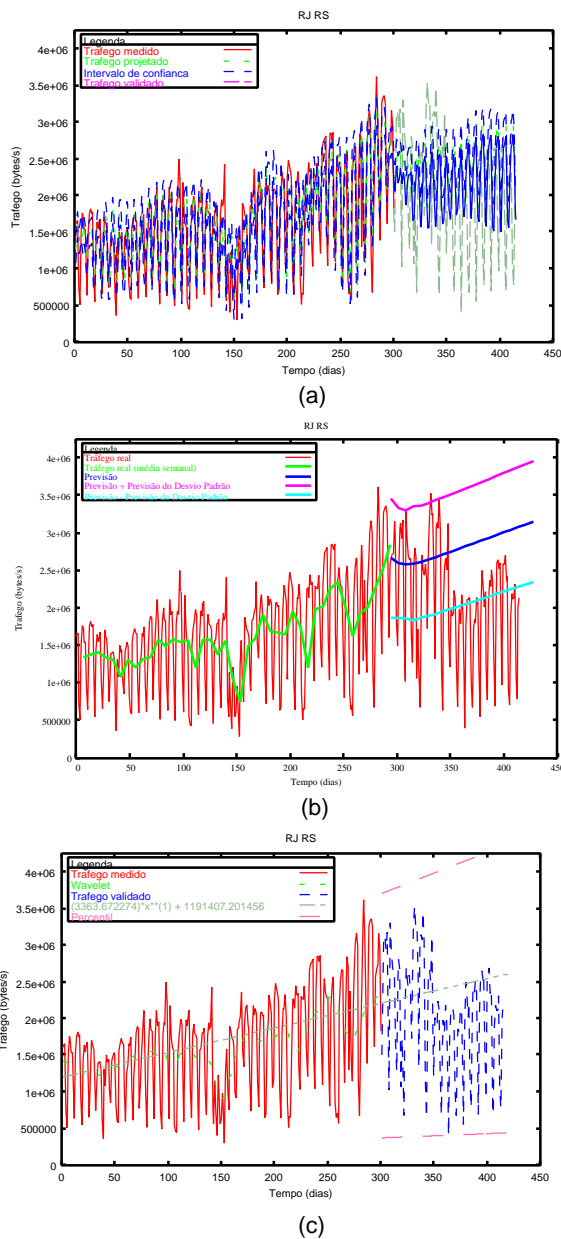


Fig. 6. Resultados das projeção para o enlace RJ-RS

V. CONCLUSÃO

Neste trabalho foi apresentada uma nova proposta para previsão de tráfego, baseada na identificação do fecho convexo

de medidas coletadas no enlace. Diferente das demais propostas existentes, o novo método é simples e de complexidade pequena o que facilita a implementação e automatização do método. Tanto as projeções usando regressão linear quanto aquelas usando auto-regressão são difíceis de automatizar, pois requerem uma análise cuidadosa para descobrir as possíveis periodicidades existentes na coleta tida como base da previsão.

A nova proposta apresentada não possui uma relação direta com a periodicidade das coletas. Ao contrário do que ocorre com os métodos de regressão linear e auto-regressão, os resultados obtidos pelo método de wavelet e fecho convexo não sofre influência pela falta de periodicidade que acontece com os dois métodos ou estacionaridade no caso da auto-regressão. A falta de dados proveniente de algum problema de medição não afeta o resultado estimado para projeção.

Os resultados das previsões feitas para enlaces da RNP com o método de projeção usando wavelets e fecho convexo foram melhores do que as previsões obtidas pelos métodos de regressão linear e auto-regressão. As validações através do tráfego dos quatro meses posteriores à coleta usada como base serviu para demonstrar a precisão obtida pelo método desenvolvido.

Agradecimentos

Agradecemos à Rede Nacional de Pesquisa (RNP) por disponibilizar os dados referentes ao volume de tráfego de entrada/saída de diversos enlaces de sua rede.

REFERÊNCIAS

- [1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis, Forecasting and Control*. Prentice Hall, 3rd edition, 1994.
- [2] Antonio Augusto de Aragão Rocha. *Medições ativas na Internet: Algoritmos Baseados em Retardo Fim-a-Fim e Experimentos*. Master's thesis, COPPE-UFRJ, 2003.
- [3] K. Papagiannaki, N. Taft, Zhang Zhi-Li, and C. Diot. Long-term forecasting of internet backbone traffic: observations and initial models. In *Proc. of INFOCOM 2003*, volume 2, pages 1178–1188, 2003.
- [4] A.A.A. Rocha, R.M.M. Leão, and E. de Souza e Silva. Metodologia para Estimar o Atraso em um Sentido e Experimentos na Internet. In *XXII SBRC*, Gramado, Brasil, Maio 2004.
- [5] J. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press (GB), 1998.
- [6] L. Zhang, Z. Liu, and C.H. Xia. Clock Synchronization Algorithms for Network Measurements. In *IEEE/Infocom*, pages 160–169, New York, USA, Junho 2002.