

Generation of Synthetic Network Traffic Series Using a Transformed Autoregressive Model Based Adaptive Algorithm

Á. Cardoso, and F. Vieira

Abstract—Modeling, testing and simulation of communications systems require to generate synthetic traffic series that present characteristics and behaviors as similar as possible to those of real network traffic traces. In this paper, we propose an adaptive algorithm to estimate the parameters of the Transformed Autoregressive Moving Average (TARMA) model in order to capture the autocorrelation function and the cumulative density function of the desired network traffic trace. Different from other works involving TARMA, we consider the adaptive estimation of its parameters. We compare the performance of the proposed on-line modeling approach to those of the Autoregressive Moving Average Model (ARMA) and of the on-batch Transformed Model in terms of mean, variance, moments, autocorrelation and probability density function. A transmission link composed of a single server with buffer is also simulated, which proves the efficiency of the proposed model in describing real traffic traces. The simulations carried out in this work show that the adaptive TARMA model outperforms in general the other considered autoregressive models.

Index Terms—Adaptive Algorithms, Adaptive Estimation, Communication system traffic, Time Series Analysis.

I. INTRODUÇÃO

MODELOS estatísticos e geração de dados sintéticos são fundamentais para o estudo de fluxos de tráfego de redes [1]. Dentre os trabalhos relacionados à modelagem do tráfego de redes, pode-se destacar o estudo do tráfego de rede envolvendo o paradigma máquina-a-máquina [2], verificação de anomalias no tráfego de rede baseada em predição para identificação de anomalias e ataques maliciosos [3], análise de desempenho de preditores de tráfego de rede na nuvem, a qual inclui o modelo ARMA (Autoregressive Moving Average) como preditor [4], etc.

O Modelo Autorregressivo de Médias Móveis Transformado (TARMA-*Transformed Autoregressive Moving Average*) [5] consiste de uma variação do modelo autorregressivo de médias móveis (ARMA) que acrescenta a este último uma melhoria na capacidade de modelar a função de autocorrelação (ACF) e a função de distribuição acumulada (FDA) da série temporal desejadas através de um filtro linear e uma transformação polinomial [5].

Os autores em [5] mostram que o modelo TARMA apresenta resultados em termos de Função de Distribuição de Probabilidade (FDP) e ACF próximos aos do tráfego real. Entretanto, estes autores não apresentaram comparações com

outros modelos da literatura, como por exemplo o modelo ARMA tradicional [6]. No modelo TARMA são efetuadas uma regressão polinomial e estimação dos parâmetros do modelo ARMA utilizando-se todas as amostras da série temporal. A regressão polinomial e a estimação de parâmetros podem demandar um alto custo computacional e inviabilizar seu uso em aplicações de tempo real. Devido a tais características, objetiva-se neste trabalho propor um algoritmo adaptativo para modelagem e geração de séries de tráfego de redes baseado no modelo TARMA, que denominamos de TARMA adaptativo. De fato, um algoritmo adaptativo permite atualizar as informações do modelo à medida que dados de tráfego são disponibilizados na rede. Assim, obtém-se a cada instante pré-determinado, um modelo com parâmetros atualizados que pode ser utilizado para se realizar estimativas relativas ao tráfego e ao desempenho da rede.

Os trabalhos [7] e [8] publicados previamente em conferências apresentam a utilização do algoritmo RLS e do algoritmo de estimação adaptativa da função de distribuição inversa para atualização adaptativa dos parâmetros do modelo TARMA. Neste trabalho, diferentemente de [7] e [8], encontra-se os valores ótimos de duas variáveis via algoritmo de busca direta para a inicialização do algoritmo adaptativo TARMA proposto. A primeira variável é a quantidade N de intervalos utilizada no algoritmo de Estimação Adaptativa da FDA Inversa. A segunda variável otimizada é a ordem P do polinômio utilizado para a modelagem da função de distribuição da série de tráfego de rede a ser modelada. O processo de otimização é realizado variando N e P em valores discretos a fim de se obter o menor valor possível do erro de função de distribuição de probabilidade, ou seja, os melhores valores possíveis de N e P que minimizem o erro de FDA. No presente artigo, também são abordadas de forma mais aprofundada a Estimação Adaptativa da Função de Distribuição Acumulada Inversa e a aplicação do algoritmo RLS para obtenção dos Parâmetros da Transformação Polinomial. Além disso, são apresentados resultados do TARMA proposto para mais séries diferentes de tráfego.

Este trabalho está organizado da seguinte forma: Na seção II, são apresentados os trabalhos relacionados. Na seção III, o modelo TARMA é apresentado. Na seção seguinte, propõe-se um algoritmo para encontrar a função de distribuição acumulada aproximada. Na seção V, apresenta-se o algoritmo TARMA adaptativo proposto neste trabalho. Na sequência, as simulações e os resultados obtidos são apresentados. Por fim, concluímos este trabalho.

II. TRABALHOS RELACIONADOS

Um dos primeiros modelos a utilizar o método de transformação de variáveis aleatórias combinado com o ajuste de função de autocorrelação é o modelo ARTA (*AutoRegressive-To-Anything*) [9] [10]. No modelo ARTA, utiliza-se a função de autocorrelação da série temporal para obtenção dos parâmetros do modelo autorregressivo normalizado cujos valores seguem uma distribuição gaussiana. Após a modelagem da função de autocorrelação, é realizada a transformação da função de distribuição gaussiana para a distribuição da série temporal desejada. Embora o modelo ARTA apresente semelhanças com o TARMA, pode-se afirmar que a função de transformação não é obtida através de um polinômio, o que faz com que no modelo ARTA sejam utilizadas apenas funções de distribuições com fórmulas explícitas (por exemplo: Gaussiana, Exponencial, Lognormal e etc) [9].

Outros trabalhos como [11][12], realizam o processo de geração e modelagem de séries temporais com o uso de modelos transformados partindo da função de autocorrelação e função de distribuição obtidos de forma a priori, ou seja, tem-se o conhecimento dessas funções sem necessariamente a obtenção das amostras das séries temporais. Para tal, os autores em [11] analisam o uso de diferentes funções para realizar o processo de transformação, como por exemplo: fórmulas explícitas, expansões de séries, polinômios de Hermite e etc.

Buscando aperfeiçoar os modelos autorregressivos de médias móveis, alguns trabalhos foram propostos [13][14]. Em [13], é considerada uma extensão do modelo ARMA para se modelar séries temporais não estacionárias através do uso do método algébrico polinomial. O trabalho apresentado em [14], propõe uma melhoria para o modelo ARMA através de séries funcionais para identificação de séries temporais. Embora os resultados apresentados em [13][14] sejam melhores do que os do modelo ARMA, faz-se necessário uma regulagem precisa dos parâmetros do modelo para obter a convergência adequada. Outra desvantagem apresentada em [14] é a demanda de um alto custo computacional podendo inviabilizar seu uso em aplicações de tempo real.

Em relação à geração de amostras sintéticas de tráfego de rede utilizando o modelo autorregressivo de médias móveis, pode-se citar os seguintes trabalhos [15] [16], [17] e [18]. Outros trabalhos como [19][20], buscam modelar o tráfego de rede capturando características de dependência à longo prazo, autossimilaridade e rajadas em diferentes escalas através do uso de análise multifractal realizada no domínio *wavelets* através da transformada de Haar. Destaca-se o trabalho [6], que utiliza a análise multifractal para gerar amostras de tráfego de rede através da realização aproximada do Ruído Gaussiano Fracionário via transformada de Haar e do uso de um filtro linear de resposta ao impulso infinito para adicionar relações de dependência na função de autocorrelação.

III. MODELO ARMA TRANSFORMADO

O modelo TARMA (*Transformed ARMA*) é constituído por 3 etapas ou procedimentos independentes para modelar a série temporal [5]. A primeira etapa é a geração de uma série

temporal decorrelacionada no tempo, sendo uma variável aleatória normal com média 0 e variância 1. A segunda etapa é um filtro linear invariante no tempo e a terceira e última etapa é a transformação polinomial da saída do filtro linear. Ao final destas etapas espera-se gerar uma série sintética com as mesmas características da série real.

Seja $Z[n]$, o valor da série real de tráfego de rede a ser modelada no instante de tempo n . A Figura 1 apresenta os blocos correspondentes do modelo TARMA, onde $X[n]$ representa a série gerada pela variável aleatória normal \mathcal{N} , $Y[n]$ é a saída do filtro linear $h[m]$ e $\hat{Z}[n]$ é a saída da transformação polinomial p_y , sendo esta uma estimativa do tráfego real $Z[n]$.

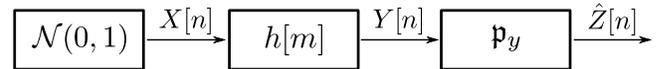


Fig. 1. Blocos do Modelo TARMA adaptativo - Adaptado de [5].

Nas próximas seções, serão abordados esses blocos do modelo TARMA antes de se apresentar a proposta de TARMA Adaptativo.

A. Transformação Polinomial

A transformação polinomial, última etapa realizada no algoritmo TARMA, tem a função de converter a Função de Distribuição Acumulada (FDA) qualquer da variável $Y[k]$ na FDA da série real a ser sintetizada. Assim, pode-se escrever a transformação polinomial como sendo [21],

$$Z[k] = F_z^{-1}(F_y(Y[k])) = p_y(Y[k]) \quad (1)$$

onde F_z^{-1} é função inversa da FDA da variável Z , F_y a FDA da variável Y e p_y o polinômio que representa o mapeamento da FDA de Y em Z . Escrevendo a Equação (1) na forma polinomial, tem-se:

$$Z[k] = \sum_{i=0}^P p_i \cdot (Y[k])^i \quad (2)$$

Ainda de acordo com a Equação (1), para se realizar a transformação de FDA é necessário a obtenção da Função de Distribuição Acumulada da variável $Y[k]$, sendo que esta é uma variável aleatória normal com média zero e variância 1 [5]. Assim, tem-se a seguinte expressão para $F_y(Y[k])$ [6]:

$$F_y(Y[k]) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{Y[k]}{\sqrt{2}} \right) \right] \quad (3)$$

IV. ESTIMAÇÃO ADAPTATIVA DA FUNÇÃO DE DISTRIBUIÇÃO ACUMULADA INVERSA

Nesta seção, propomos um algoritmo para estimar adaptativamente a Função de Distribuição Acumulada através do cálculo do histograma da série a ser sintetizada.

Seja $Z[k]$ a série temporal normalizada a ser sintetizada, cujos valores estão contidos no intervalo $[0, 1]$. Ao fracionarmos esse intervalo em N partes iguais, obtém-se:

$$[0, 1] = \left[0, \frac{1}{N} \right), \left[\frac{1}{N}, \frac{2}{N} \right), \dots, \left[\frac{N-1}{N}, \frac{N}{N} \right] \quad (4)$$

Com o intervalo dividido, pode-se calcular a quantidade de ocorrências de valores da série temporal Z nestes intervalos, ou seja:

$$\mathcal{V}_Z = [v_1, v_2, \dots, v_N] \quad (5)$$

onde v_i representa quantos valores da série que se encontram no intervalo $[(i-1)/N, i/N]$. Como a probabilidade de um evento ocorrer pode ser vista como a quantidade de ocorrência de um determinado evento dividida pela quantidade total de ocorrência do conjunto universo em análise, pode-se obter através da divisão do vetor \mathcal{V}_Z pelo tamanho total da série temporal, uma aproximação da Função de Distribuição de Probabilidade.

Através de \mathcal{V}_Z , obtém-se o histograma acumulado f_k como sendo:

$$f_k = \sum_{i=0}^k v_i, \quad k = 1, \dots, N \quad (6)$$

A Função de Distribuição Acumulada pode ser obtida dividindo a Equação (6) pela quantidade total de amostras, ou seja:

$$\hat{\mathcal{F}} = \left(\frac{1}{L}\right) \cdot [f_1, f_2, \dots, f_N] \quad (7)$$

onde L é a quantidade total de amostras da série temporal utilizadas para se obter o histograma.

Uma vez obtida a Equação (7), pode-se formalizar um algoritmo para se estimar adaptativamente a FDA Inversa, como se segue:

Algoritmo 1: Estimção Adaptativa da FDA Inversa

- **Passo 1:** Inicialize todos os valores do vetor \mathcal{V} iguais a zeros, $v_i = 0$ para $i = 1, \dots, N$.
- **Passo 2:** Inicialize o vetor $\hat{\mathcal{F}}$, de tamanho N , sendo igual a zero, $f_i = 0$ para $i = 1, \dots, N$.
- **Passo 3:** Inicialize a quantidade de amostras utilizadas L como sendo igual a zero, $L = 0$.
- **Passo 4:** Encontre o intervalo no qual a amostra $Z[k]$ se encontra.

$$r = \left\lfloor \frac{Z[k]}{1/N} \right\rfloor \quad (8)$$

- **Passo 5:** Atualize a quantidade de ocorrência do intervalo no histograma \mathcal{V}_Z :

$$v_r = v_r + 1 \quad (9)$$

- **Passo 6:** Atualize o histograma acumulado através da seguinte equação:

$$f_i = f_i + 1 \quad i = r, \dots, N \quad (10)$$

- **Passo 7:** Atualize a quantidade de amostra utilizadas $L = L + 1$.
- **Passo 8:** A FDA inversa pode ser estimada utilizando-se a seguinte equação (Ver Apêndice 1):

$$F_z^{-1}(d) = (e-1) + \frac{(L \cdot d - f_{e-1})}{(f_e - f_{e-1})} \quad (11)$$

onde d é o valor para o qual se deseja avaliar a função e e o valor inteiro da posição do intervalo no qual d está contido.

- **Passo 9:** Retorne ao Passo 4 na próxima iteração, até o fim de todas iterações.

O Algoritmo 1 permite que a FDA Inversa Adaptativa F_z^{-1} seja obtida com poucos cálculos computacionais, sendo a complexidade computacional deste algoritmo da ordem de $\mathcal{O}(N)$ para cada iteração.

V. MODELO TARMA ADAPTATIVO

O modelo autorregressivo de médias móveis recebe este nome por considerar como saída a soma dos valores das entradas anteriores (médias móveis), uma constante μ e das saídas do modelo nos instantes anteriores (autorregressivo). Matematicamente, pode-se representar o modelo ARMA pela seguinte equação [6]:

$$y[t] + a_1 \cdot y[t-1] + \dots + a_n \cdot y[t-n] = \mu + b_1 \cdot x[t] + \dots + b_m \cdot x[t-m] \quad (12)$$

ou na forma vetorial:

$$AY = \mu + BX \quad (13)$$

Neste trabalho, com o objetivo de se obter uma variável normal $Y[k]$ e de forma a manter a independência entre os processos do TARMA, faz-se a média igual a zero, ou seja $\mu = 0$ [5].

A. Estimção Adaptativa dos Parâmetros ARMA

Os autores em [5] apresentam algoritmos de regressão polinomial para estimção dos parâmetros do modelo ARMA utilizando a quantidade total de amostras do tráfego de rede real. Neste trabalho, com o objetivo de estimar adaptativamente os parâmetros do modelo ARMA, emprega-se o algoritmo RLS (*Recursive Least Squares*) [22].

Reescrevendo a Equação (12), tem-se:

$$y[t] = -a_1 \cdot y[t-1] - \dots - a_n \cdot y[t-n] + b_1 \cdot x[t] + \dots + b_m \cdot x[t-m] \quad (14)$$

Nota-se que a Equação (14) possui valores conhecidos (Y e X) e valores que se deseja descobrir (A e B) para obtermos a resposta desejada $y[t]$. Esta característica permite aplicar o algoritmo RLS para estimção dos parâmetros desconhecidos. Para tal, pode-se escrever a saída do modelo TARMA como:

$$y[k] = \varphi^T[t] \theta \quad (15)$$

onde

$$\theta^T = [a_1, \dots, a_n, b_1, \dots, b_m] \quad (16)$$

e

$$\varphi^T[t] = [-y[t-1], \dots, -y[t-n], x[t], \dots, x[t-m]] \quad (17)$$

Como os valores de φ^T e $y[k]$ são conhecidos, propõe-se aplicar o algoritmo RLS para obtenção dos valores dos parâmetros θ do modelo ARMA.

B. Obtenção dos Parâmetros da Transformação Polinomial

Pode-se aplicar o algoritmo RLS para se estimar também os valores dos parâmetros da transformação polinomial. A seguir, descreve-se como isso pode ser feito.

Ao expandirmos a Equação (2), tem-se:

$$Z[k] = p_0 \cdot (Y[k])^0 + p_1 \cdot (Y[k])^1 + \dots + p_P \cdot (Y[k])^P \quad (18)$$

A Equação (18) possui valores conhecidos (Y e Z) e valores que se deseja descobrir (p_i para $i = 0, \dots, P$) para obtermos a resposta desejada $Z[k]$. Os parâmetros da Equação (18) podem ser representados por um vetor θ_p^T da seguinte forma:

$$\theta_p^T = [p_0, \dots, p_P] \quad (19)$$

e

$$\varphi_p^T[t] = [(Y[k])^0, (Y[k])^1, \dots, (Y[k])^P] \quad (20)$$

Assim, uma estimativa de $Z[n]$ pode ser dada por:

$$\hat{Z}[n] = \varphi_p^T[t]\theta_p. \quad (21)$$

Através da Função de Distribuição Acumulada da variável $Y[k]$ dada pela Equação (3) e pelo Algoritmo 1 (Estimação Adaptativa da FDA Inversa), obtém-se a resposta desejada $Z[k]$ como:

$$Z[k] = F_z^{-1}(F_y(Y[k])) \quad (22)$$

Ao se unir a obtenção do polinômio através do algoritmo da FDA Inversa Adaptativa descrito na seção IV e da regressão polinomial pelo algoritmo RLS, pode-se formalizar o **algoritmo TARMA Adaptativo**.

Algoritmo 2: TARMA Adaptativo para Geração de Tráfego Sintético

- **Passo 1:** Aplicar o algoritmo RLS para obtenção dos parâmetros do filtro adaptativo ARMA cuja complexidade computacional tem ordem $\mathcal{O}(n^2)$ [22].
- **Passo 2:** Efetuar o algoritmo da FDA Inversa Adaptativa para obter a inversa da FDA do fluxo de tráfego de rede Z cuja complexidade computacional tem ordem $\mathcal{O}(N)$.
- **Passo 3:** Efetuar o algoritmo RLS para obtenção dos parâmetros do polinômio através da Equação (3) e da inversa obtida no Passo 2. A complexidade computacional do algoritmo RLS tem ordem $\mathcal{O}(n^2)$ [22].
- **Passo 4:** Gerar uma amostra da variável Normal $\mathcal{N}(0, 1)$, efetuar a filtragem e a transformação polinomial para gerar amostras do tráfego sintético cuja complexidade computacional tem ordem $\mathcal{O}(P)$.
- **Passo 5:** Retornar ao Passo 1, na próxima iteração, até o fim de todas iterações.

O algoritmo TARMA Adaptativo apresenta uma complexidade computacional de ordem $\mathcal{O}(2n^2 + N + P) = \mathcal{O}(n^2)$ para cada iteração, onde n é a ordem do filtro utilizado no algoritmo RLS, que corresponde ao número de saídas passadas do filtro.

Na próxima seção, são apresentados os resultados das simulações obtidos para os modelos considerados para geração de séries de tráfego sintética.

VI. SIMULAÇÕES E RESULTADOS

Simulações foram realizadas levando em conta os modelos ARMA, TARMA e o modelo adaptativo TARMA proposto. As séries de tráfego utilizadas neste trabalho são referentes ao tráfego de rede TCP/IP Waikato VIII 20110520-000000-0 [23] oriundo na comunicação entre a Universidade de Waikato e o resto do mundo, o tráfego de rede SNU20100315 [24] obtido na rede sem fio WiMAX da Universidade Nacional de Seoul e tráfego de rede MAWI-201804011400 [25] agregado na escala de $1ms$. As séries de tráfego MAWI representam dados diários de tráfegos de diferentes aplicações coletadas no *backbone* da internet do grupo de pesquisa *Measurement and Analysis on the WIDE Internet* (MAWI). Neste trabalho, essas séries de tráfegos de rede são nomeadas como Waikato, SNU e MAWI, respectivamente.

Os desempenhos de modelagem dos algoritmos foram comparados em termos de média, variâncias, momentos até a quarta ordem, função de distribuição de probabilidade, autocorrelação normalizada e relação pico/média. Para comparar a FDP (Função Densidade de Probabilidade) e a Função de Autocorrelação (ACF - *Autocorrelation Function*) das séries sintéticas em relação às do tráfego real, calculou-se o erro quadrático médio (EQM).

Buscando-se obter o valor ótimo para a variável N (quantidade de intervalos) do algoritmo da FDA Inversa Adaptativa pela análise do valor do EQM entre a FDA obtida e FDA real foram realizadas simulações com N variando de 2 até 10000, como pode ser verificado na Figura 2. Pode-se notar que os valores do EQM da estimação da FDA diminuem com o aumento da quantidade N de intervalos. Com base neste comportamento de decaimento de EQM a quantidade de intervalos considerada nas simulações foi de $N = 100$, uma vez que para um valor maior não houve diminuição significativa do EQM.

A. Série de Tráfego Agregada na Escala de 10ms.

Nesta seção, são apresentados os resultados de desempenho dos modelos autorregressivos para descrever a série Waikato agregada na escala de $10ms$.

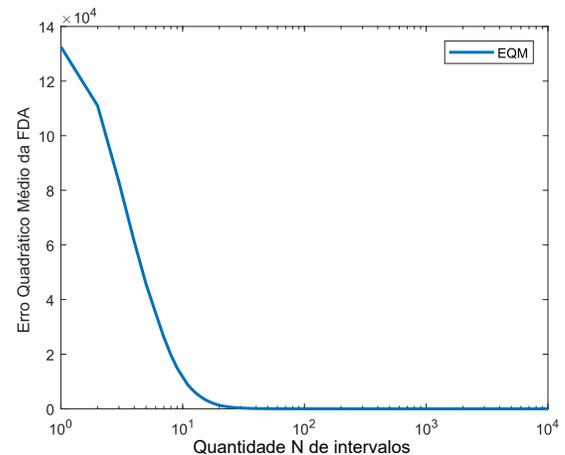


Fig. 2. Erro Quadrático Médio de estimação da FDA em relação à quantidade N de intervalos.

A ordem P do Polinômio influencia diretamente o EQM de estimação da FDA, podendo impactar o desempenho da geração de amostras sintéticas do tráfego de rede [5]. Um polinômio com poucos parâmetros pode não ser suficiente para representar a FDA do tráfego de rede. Enquanto para valores grandes de P o modelo também pode não representar adequadamente a FDA devido ao surgimento de oscilações [5]. Assim, se faz necessário obter o valor P para o qual o algoritmo apresenta o menor EQM entre a FDA estimada e a FDA original. Na Figura 3, pode-se verificar o EQM de estimação da FDA em relação à P . Nota-se que o menor EQM foi obtido para P igual a 10.

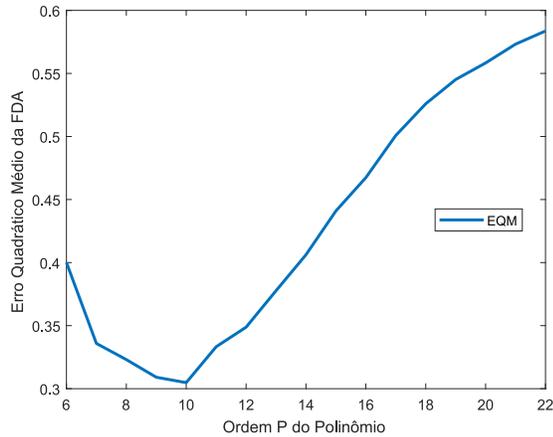


Fig. 3. Erro Quadrático Médio de estimação da FDA em relação à ordem P do Polinômio, com $N=100$.

A Figura 4 apresenta os valores da função densidade de probabilidade obtidos para as séries sintéticas geradas pelos modelos autorregressivos considerados com relação à série Waikato de tráfego de rede agregada em $10ms$. Os resultados indicam que todas as séries apresentaram maiores valores para a FDP no intervalo entre 0 e 0,4, sendo que a FDP do modelo ARMA apresentou os valores mais dispersos em comparação à FDP do tráfego real. Pode-se observar também que as FDPs do modelo TARMA e do modelo proposto apresentaram valores de probabilidade próximos ao desejado (do tráfego real).

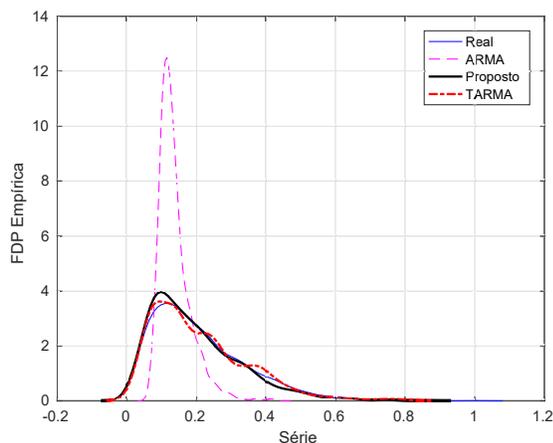


Fig. 4. Função de Distribuição de Probabilidade para a Série Waikato, agregada na escala de $10ms$.

Os valores para a função de autocorrelação obtidos para o tráfego Waikato agregado na escala de $10ms$ podem ser visualizados na Figura 5. Nota-se que assim como os resultados da FDP, o modelo ARMA apresentou valores mais discrepantes do desejado em comparação ao TARMA e ao algoritmo proposto. Por outro lado, para valores de atraso (amostras em diferentes instantes de tempo) maiores que 20, o modelo proposto apresentou valores mais próximos da ACF da série real.

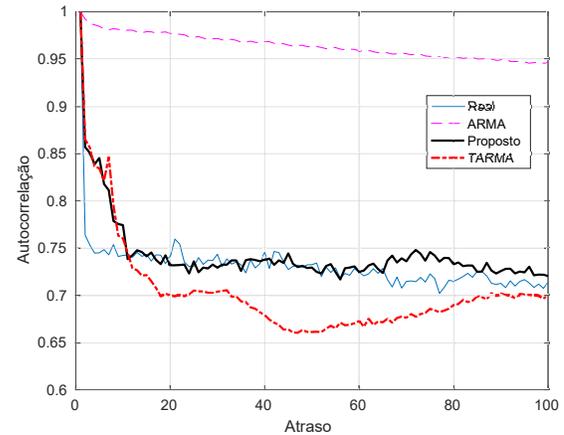


Fig. 5. Autocorrelação com atraso até 100 amostras para a Série Waikato, agregada na escala de $10ms$.

A Tabela I mostra os resultados estatísticos das séries sintéticas e real na escala de $10ms$. Os valores mais próximos em relação ao tráfego real para as estatísticas média, variância e momentos foram obtidos pelo modelo TARMA, seguidos pelo modelo Proposto e pelo modelo ARMA. Destaca-se que o algoritmo proposto TARMA adaptativo proporcionou estatísticas mais semelhantes aos do tráfego real do que o TARMA tradicional, mesmo neste último serem utilizados todos os dados do fluxo de rede real para se estimar seus parâmetros.

TABELA I
DADOS ESTATÍSTICOS PARA A SÉRIE WAIKATO, AGREGADA NA ESCALA DE $10ms$

	Real	ARMA	TARMA	Proposto
Média	0,2083	0,1572	0,2076	0,1938
Variância	0,0188	0,0051	0,0187	0,0162
2º Momento	0,0622	0,0409	0,0618	0,0537
3º Momento	0,0239	0,0076	0,0234	0,0191
4º Momento	0,0111	0,0024	0,0106	0,0081
EQM FDP	0	8,1634	0,0938	0,0646
EQM ACF	0	4,284,5	1,471,9	309,3
Pico/Média	4,7999	3,3037	3,8614	4,4148

Os valores do erro quadrático médio de estimação da FDP e da ACF na Tabela I indicam que o algoritmo TARMA proposto provê valores mais próximos dos desejados do que os algoritmos TARMA tradicional e ARMA.

Os resultados obtidos para a relação Pico/Média na Tabela I indicam valores próximos aos do tráfego de rede real para o modelo TARMA adaptativo, seguidos dos modelos TARMA e do ARMA.

B. Série de Tráfego SNU Agregada na Escala de 100ms.

Para se verificar o desempenho dos modelos considerados em descrever diferentes séries de tráfego de redes, apresenta-se nesta seção os resultados para a série SNU.

TABELA II
DADOS ESTATÍSTICOS PARA A SÉRIE SNU AGREGADA NA ESCALA DE 100ms

	Real	ARMA	TARMA	Proposto
Média	0,6352	0,6016	0,6398	0,6356
Variância	0,0231	0,0125	0,0217	0,0142
2º Momento	0,4266	0,3744	0,4310	0,4182
3º Momento	0,2949	0,2355	0,2818	0,2989
4º Momento	0,2088	0,1495	0,2122	0,1941
EQM FDP	0	3,2525	0,6019	0,1571
EQM ACF	0	1.507,2	1.502,2	1.333,2
Pico/Média	1.5742	1,4194	1,4194	1,4880

Os resultados estatísticos obtidos para o tráfego de rede SNU podem ser visualizados na Tabela II. Os valores mais próximos de média, 4º momento e Pico/Média em relação ao tráfego real foram obtidos pelo algoritmo proposto, seguidos pelo algoritmo TARMA. Os valores mais próximos dos desejados de variância, 2º e 3º momentos foram obtidos pelo modelo TARMA. Nota-se que os valores mais discrepantes foram obtidos pelo modelo ARMA. Destaca-se os menores erros quadráticos da FDP e ACF obtidos pelo modelo adaptativo.

C. Série de Tráfego MAWI Agregada na Escala de 1ms.

Na Figura 4, pode-se visualizar os resultados para a Função de Distribuição de Probabilidade para a Série MAWI, agregada na escala de 1ms. Nota-se os valores mais próximos de Função de Distribuição de Probabilidade para os modelos TARMAs em comparação aos da série de tráfego real.

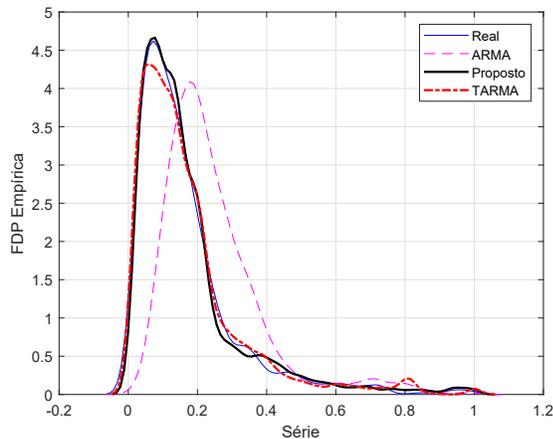


Fig. 6. Função de Distribuição de Probabilidade para a Série MAWI, agregada na escala de 1ms.

Os resultados para a função de autocorrelação com atraso até 100 amostras para a Série MAWI, agregada na escala de 1ms, podem ser visualizados na Figura 7. Nota-se o mesmo comportamento obtido nos resultados da função de autocorreção das outras séries de tráfegos, onde os valores dos modelos TARMAs se apresentaram mais próximos dos valores da função de autocorrelação da série de tráfego real.

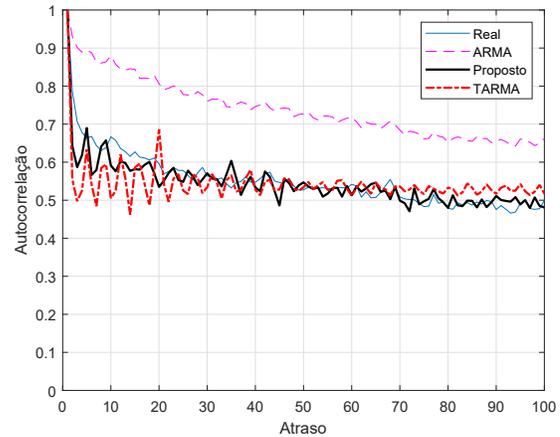


Fig. 7. Autocorrelação com atraso até 100 amostras para a Série MAWI, agregada na escala de 1ms.

Os dados estatísticos para a Série MAWI agregada na escala de 1ms podem ser visualizados na Tabela III. Os valores dos dados estáticos mais próximos em comparação os da série de tráfego real foram obtidos pelo modelo TARMA adaptativo, seguidos pelos modelos TARMA e ARMA, respectivamente. Em relação aos EQMs de FDP e ACF, destaca-se que os maiores valores foram obtidos pelo modelo ARMA quando comparados aos modelos TARMA e TARMA Adaptativo.

TABELA III
DADOS ESTATÍSTICOS PARA A SÉRIE MAWI AGREGADA NA ESCALA DE 1ms

	Real	ARMA	TARMA	Proposto
Média	0,1648	0,2455	0,1713	0,1700
Variância	0,0221	0,0212	0,0255	0,0248
2º Momento	0,0493	0,0814	0,0548	0,0537
3º Momento	0,0231	0,0362	0,0279	0,0265
4º Momento	0,0143	0,0205	0,0186	0,0170
EQM FDP	0	1,1262	0,0770	0,0796
EQM ACF	0	1100,1	841,02	799,07
Pico/Média	6,0680	4,0727	5,8368	5,8809

D. Simulação de Enlace de Transmissão com Buffer

Para verificar se o comportamento de fila das séries sintéticas é semelhante ao da série de tráfego real, realiza-se a simulação de um enlace de transmissão de um servidor simples onde o tráfego de rede é dado como entrada do servidor.

Os resultados para a Taxa de Utilização no *buffer* em relação à capacidade de transmissão do enlace η , com escala de agregação de 10ms, podem ser visualizados na Figura 8. Os modelos TARMA e proposto apresentam valores de taxa de utilização mais próximos em relação ao tráfego real, onde valores inferiores foram obtidos pelo modelo ARMA.

Os valores de Taxa de Perda para o tráfego SNU são apresentados na Figura 9. Observam-se comportamentos semelhantes obtidos pelos modelos TARMA e proposto, com valores de taxa de perda mais distantes do real obtidos pelo modelo ARMA.

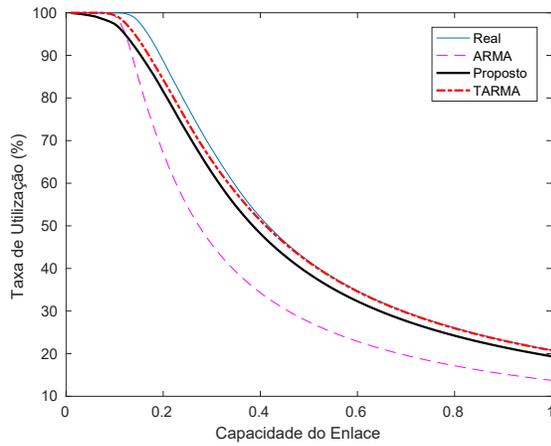


Fig. 8. Taxa de Utilização no *buffer* em relação à capacidade de transmissão do enlace η para o tráfego Waikato com escala de agregação de $10ms$ e $B = 0.6$.

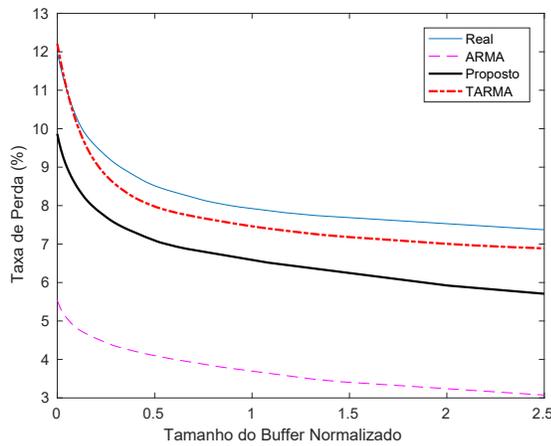


Fig. 9. Taxa de Perda no *buffer* em relação ao tamanho do *Buffer* para o tráfego SNU com escala de agregação de $10ms$ e $\eta = 0.6$.

Os resultados para a Taxa de Utilização e Taxa de perda na simulação do enlace de transmissão considerando o tráfego de rede MAWI agregado na escala de $1ms$ podem ser visualizados nas Figuras 10 e 11, respectivamente. Os resultados indicam que o modelo proposto TARMA adaptativo, assim como o modelo TARMA, apresentaram comportamento semelhantes ao do apresentado pelo tráfego rede real MAWI, destacado pelos valores semelhantes. Para o modelo ARMA, nota-se valores mais distantes dos valores da série de tráfego de rede real quando comparado aos modelos TARMAs.

Ao analisarmos os resultados para a simulação do enlace de transmissão, nota-se que de forma geral, o modelo proposto apresenta comportamento semelhante aos dos fluxos de tráfegos reais, indicando que as amostras sintéticas podem ser utilizadas em simulações para representar as séries desejadas.

Foram realizadas simulações com várias outras séries de tráfego de rede onde se constata desempenho semelhante dos modelos considerados neste trabalho para essas séries.

VII. CONCLUSÃO

Neste trabalho, propomos um algoritmo para estimar adaptativamente os parâmetros do modelo TARMA para geração de

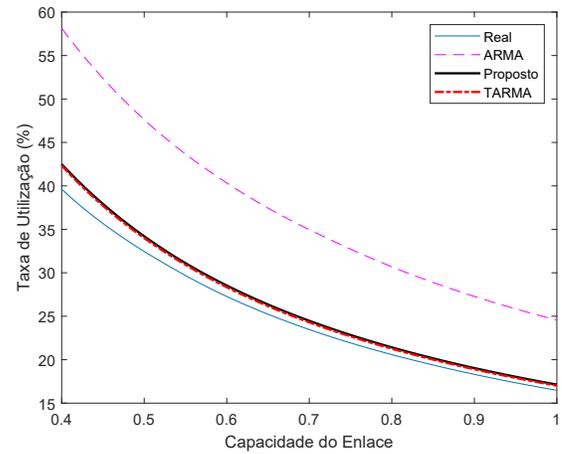


Fig. 10. Taxa de Utilização no *buffer* em relação à capacidade de transmissão do enlace η para o tráfego MAWI com escala de agregação de $1ms$ e $B = 0.6$.

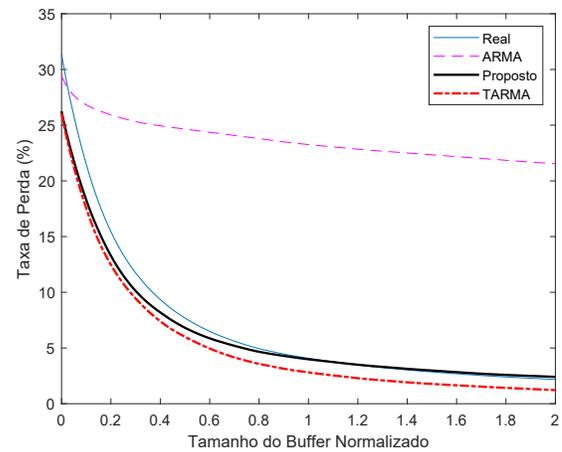


Fig. 11. Taxa de Perda no *buffer* em relação ao tamanho do *Buffer* para o tráfego MAWI com escala de agregação de $1ms$ e $\eta = 0.3$.

séries temporais sintéticas de forma a capturar a autocorrelação e função de distribuição acumulada independentemente de séries de tráfego reais.

Os resultados mostraram que o modelo TARMA adaptativo proposto apresentou um melhor desempenho para descrever as séries reais de tráfego de rede SNU e MAWI com relação aos parâmetros estatísticos: média, momentos de 2º a 4º ordem, EQM da ACF e relação Pico/Média. Para o tráfego de rede real WAIKATO, o algoritmo proposto apresentou melhor desempenho para os dados estatísticos de EQM de FDP, EQM de ACF e relação Pico/Média em comparação aos demais modelos autorregressivos comparados.

Observou-se na simulação do enlace de transmissão que as séries sintéticas do modelo TARMA e do modelo proposto apresentaram resultados semelhantes aos do tráfego de rede real, com alguns valores, tais como taxa de perda, mais próximos do desejado para o modelo proposto. Destaca-se que o modelo TARMA necessitou de todas as amostras da série real enquanto no algoritmo proposto a estimação dos parâmetros foi realizada adaptativamente.

Para trabalhos futuros, objetiva-se avaliar o uso de outros

algoritmos para obtenção do polinômio e do filtro linear, com menores complexidades computacionais.

APÊNDICE 1: PROVA DA EQUAÇÃO (11)

Seja \hat{F} o histograma acumulado (Equação (7)) e d o valor para o qual se deseja avaliar a função inversa $F_z^{-1}(d)$. Defina-se a variável e como sendo um número inteiro que represente o intervalo fracionado da Equação (4), na qual d está contido, ou seja:

$$e = \arg \left\{ \sup_{d \leq f_i} \{f_1, f_2, \dots, f_N\} \right\} \quad (23)$$

onde $e = 1, \dots, N$.

Define-se também f_e e f_{e-1} como sendo os extremos do intervalo fracionado da Equação (4) em que $d \in [f_{e-1}, f_e]$. Para o caso $e = 1$, tem-se que $f_{e-1} = 0$.

Neste trabalho, considera-se a função F_z^{-1} linear no intervalo $[f_{e-1}, f_e]$. Assim, através de proporcionalidade entre os intervalos $[f_{e-1}, f_e]$ e $[f_{e-1}, d]$, tem-se a seguinte expressão:

$$\frac{(f_e/L - f_{e-1}/L)}{(d - f_{e-1}/L)} = \frac{(e - (e - 1))}{(F_z^{-1}(d) - (e - 1))} \quad (24)$$

Isolando $F_z^{-1}(d)$, obtém-se a equação para a estimação da FDA inversa:

$$F_z^{-1}(d) = (e - 1) + \frac{(L \cdot d - f_{e-1})}{(f_e - f_{e-1})} \quad (25)$$

como se queria demonstrar.

REFERÊNCIAS

[1] F. L. de Mello, M. L. Augusto, A. B. de Lima, and J. R. de Almeida Amazonas, "Generation of gaussian self-similar series via wavelets for use in traffic simulations," *IEEE Latin America Transactions*, vol. 5, no. 1, pp. 9–20, 2007.

[2] M. Laner, P. Svoboda, N. Nikaein, and M. Rupp, "Traffic models for machine type communications," in *ISWCS 2013; The Tenth International Symposium on Wireless Communication Systems*, Aug 2013, pp. 1–5.

[3] H. Wang, "Anomaly detection of network traffic based on prediction and self-adaptive threshold," *International Journal of Future Generation Communication and Networking*, vol. 8, no. 6, pp. 205–214, 2015.

[4] B. L. Dalmazo, J. P. Vilela, and M. Curado, "Performance analysis of network traffic predictors in the cloud," *Journal of Network and Systems Management*, vol. 25, no. 2, pp. 290–320, 2017.

[5] M. Laner, P. Svoboda, and M. Rupp, "Parsimonious network traffic modeling by transformed arma models," *IEEE Access*, vol. 2, pp. 40–55, 2014.

[6] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

[7] A. A. Cardoso, F. H. T. Vieira, and M. V. Ferreira, "Estimação adaptativa de parâmetros do modelo arma transformado aplicado a séries de tráfego de redes," in *XLIX Simpósio Brasileiro de Pesquisa Operacional*, 2017.

[8] A. A. Cardoso, F. H. T. Vieira, R. A. P. Franco, and M. V. Ferreira, "Algoritmo de estimação adaptativa dos parâmetros da modelagem autorregressiva transformada aplicada a tráfego de rede," in *XIII Simpósio Brasileiro de Automação Inteligente*, 2017.

[9] M. C. Cario and B. L. Nelson, "Numerical methods for fitting and simulating autoregressive-to-anything processes," *INFORMS Journal on Computing*, vol. 10, no. 1, pp. 72–81, 1998.

[10] —, "Autoregressive to anything: Time-series input processes for simulation," *Operations Research Letters*, vol. 19, no. 2, pp. 51–58, 1996.

[11] H. Helgason, V. Pipiras, and P. Abry, "Synthesis of multivariate stationary series with prescribed marginal distributions and covariance using circulant matrix embedding," *Signal processing*, vol. 91, no. 8, pp. 1741–1758, 2011.

[12] M. Grigoriu, "Existence and construction of translation models for stationary non-gaussian processes," *Probabilistic Engineering Mechanics*, vol. 24, no. 4, pp. 545–551, 2009.

[13] R. B. Mrad, S. D. Fassois, and J. A. Levitt, "A polynomial-algebraic method for non-stationary tarma signal analysis—part i: The method," *Signal Processing*, vol. 65, no. 1, pp. 1–19, 1998.

[14] A. G. Poulimenos and S. D. Fassois, "Output-only stochastic identification of a time-varying structure via functional series tarma models," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1180–1204, 2009.

[15] M. F. Iqbal, M. Zahid, D. Habib, and L. K. John, "Efficient prediction of network traffic for real-time applications," *Journal of Computer Networks and Communications*, vol. 2019, 2019.

[16] M. Ring, D. Schlör, D. Landes, and A. Hotho, "Flow-based network traffic generation using generative adversarial networks," *Computers & Security*, vol. 82, pp. 156–172, 2019.

[17] A. Cao, Y. Qiao, K. Sun, H. Zhang, and J. Yang, "Network traffic analysis and prediction of hotspot in cellular network," in *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*. IEEE, 2018, pp. 452–456.

[18] M. JIANG, C.-m. WU, M. Zhang, and D.-m. HU, "Research on the comparison of time series models for network traffic prediction [j]," *Acta Electronica Sinica*, vol. 11, pp. 2353–2358, 2009.

[19] R. H. Riedi, M. Crouse, V. Ribeiro, and R. Baraniuk, "A multifractal wavelet model with application to network traffic," *Information Theory, IEEE Transactions on*, vol. 45, no. 3, pp. 992–1018, Apr 1999.

[20] S. Molnár, T. D. Dang, and I. Maricza, "On the queue tail asymptotics for general multifractal traffic," in *International Conference on Research in Networking*. Springer, 2002, pp. 105–116.

[21] A. O. Allen, *Probability, statistics, and queueing theory*. Academic Press, 2014.

[22] B. Farhang-Boroujeny, *Adaptive filters: theory and applications*. John Wiley & Sons, 2013.

[23] WITS, "Wits: Waikato internet traffic storage from university of waikato," <http://wand.net.nz/wits/waikato/8/>, April 2014.

[24] S. Kim, "Dataset of bittorrent traffic on korea telecom's mobile wimax network," <http://crawdad.cs.dartmouth.edu/snu/ bittorrent/>, April 2011.

[25] MAWI, "Mawi working group traffic archive," <http://mawi.wide.ad.jp/mawi/>, April 2018.



Alisson A. Cardoso Possui graduação em Engenharia de Computação pela Universidade Federal de Goiás (2012), mestrado em Engenharia Elétrica e de Computação pela Universidade Federal de Goiás (2014) e doutorado em Engenharia Elétrica e de Computação pela Universidade Federal de Goiás (2019). Possui experiência na área de Telecomunicações em análise e desempenho de redes móveis, modelagem de tráfego de redes e redes sem fio, sistemas inteligentes, reconhecimento de padrões e inteligência artificial.



Flávio H. T. Vieira possui graduação em Engenharia Elétrica pela Universidade Federal de Goiás (2000), mestrado em Engenharia Elétrica e de Computação (EEEC) pela Universidade Federal de Goiás (2002), doutorado em Engenharia Elétrica pela Universidade Estadual de Campinas (2006) e pós-doutorado em Engenharia Elétrica pela Universidade Estadual de Campinas (2008). Atualmente é professor Associado da Universidade Federal de Goiás (UFG), Coordenador de Pesquisa da Engenharia Elétrica e de Computação da EEEC-UFG. Tem experiência na área de Engenharia Elétrica e de Computação, com ênfase em Telecomunicações, atuando principalmente nos seguintes temas: modelagem e controle de tráfego, redes de computadores, redes sem fio e sistemas inteligentes.