



## **Estimação Adaptativa de Parâmetros do Modelo ARMA Transformado Aplicado a Séries de Tráfego de Redes**

**Álison A. Cardoso**

Escola de Engenharia Elétrica, Mecânica e Computação (EMC)  
Universidade Federal de Goiás (UFG)  
alsnac@gmail.com

**Flávio H. T. Vieira**

Escola de Engenharia Elétrica, Mecânica e Computação (EMC)  
Universidade Federal de Goiás (UFG)  
flavio@emc.ufg.br

**Marcus V. Ferreira**

Instituto de Informática (INF)  
Universidade Federal de Goiás (UFG)  
marcusviniciusbr@gmail.com

### **RESUMO**

Neste trabalho, propomos um algoritmo para estimar adaptativamente os parâmetros de um modelo para geração de séries temporais sintéticas de modo a capturar a função de autocorrelação e a função de distribuição acumulada de séries de tráfego de redes reais. Comparações com os modelos autorregressivo tradicional, transformado e o proposto são apresentadas em termos de média, variância, momentos, autocorrelações e funções de distribuição de probabilidade. Simulações de um enlace de transmissão de um servidor também são realizadas para verificar a eficiência da modelagem proposta em gerar e descrever séries de tráfego de redes.

**PALAVRAS CHAVE.** Modelo Estatístico Adaptativo, Modelo Autorregressivo Transformado, Série Temporal Sintética.

**Modelos Probabilísticos, PO em Telecomunicações e Sistemas de Informações e Outras aplicações em PO**

### **ABSTRACT**

In this paper, we propose an adaptive model to generate synthetic time series based on the Transformed Autoregressive Moving Average (TARMA) model, with two independent process to capture the autocorrelation function and the cumulative density function of the desired time serie. Comparisons with the models traditional autoregressive model, transformed and the proposed are presented in terms of mean, variance, moments, autorrelations 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.

**KEYWORDS.** Adaptive Statistic Model, Transformed Autoregressive Model, Time Series.

**Probabilistic Models, OR in Telecommunications and Information Science and Other Application in OR.**



## 1. Introdução

O estudo de modelos estatísticos para geração de dados sintéticos é de fundamental importância para a modelagem e simulação de fenômenos naturais e de fenômenos realizados pelo ser humano. Alguns trabalhos destacam essa importância, como o estudo do tráfego de rede para o paradigma máquina-a-máquina (Laner et al. [2013]), verificação de anomalias no tráfego de rede baseado em predição para identificação de anomalias e ataques maliciosos (Wang [2015]), análise de desempenho de preditores de tráfego de rede na nuvem, o qual inclui o modelo ARMA (Autor-regressive Moving Average) como preditor (Dalmazo et al. [2017]).

O modelo autorregressivo de médias móveis transformado (*Transformed ARMA*) (Laner et al. [2014]) consiste de uma variação do modelo autorregressivo de médias móveis (ARMA), o qual acrescenta a capacidade de modelar a função de autocorrelação (ACF) e a função de distribuição acumulada (FDA) da série temporal desejada, através de um filtro linear e uma transformação polinomial.

Os autores em Laner et al. [2014] mostram que o modelo TARMA apresenta resultados em termos de Função de Distribuição de Probabilidade (FDP) e ACF próximos ao tráfego real. Entretanto, estes autores não apresentaram comparações com outros modelos da literatura, como por exemplo o modelo ARMA tradicional Montgomery et al. [2015]. No modelo TARMA são efetuadas uma regressão polinomial e uma estimação dos parâmetros do modelo ARMA contendo todos os dados 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, objetivamos neste trabalho propor um modelo adaptativo baseado no modelo TARMA.

Este trabalho está organizado da seguinte forma: Na seção 2, apresentamos o modelo TARMA. Na seção seguinte, apresentamos uma proposta de algoritmo para encontrar a função de distribuição acumulada aproximada. Na seção 5, apresentamos o algoritmo TARMA adaptativo proposto neste trabalho. Na sequência, apresentamos as simulações e os resultados obtidos. Por fim, concluímos este trabalho.

## 2. Modelo ARMA Transformado

O modelo TARMA (*Transformed ARMA*) é constituído por 3 processos independentes para modelar a série temporal. O primeiro é a geração de uma série temporal decorrelacionada no tempo, sendo uma variável aleatória normal com média 0 e variância 1. O segundo processo é um filtro linear invariante no tempo e o terceiro e último processo, 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.

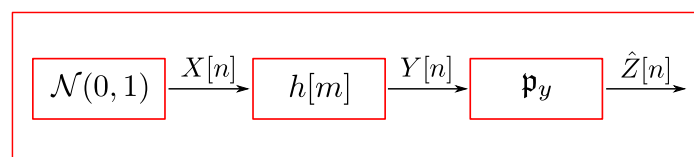


Figura 1: Blocos do Modelo TARMA adaptativo.

Na Figura (1), podemos visualizar o modelo TARMA, onde  $X[n]$  representa a série gerada pela variável aleatória normal  $\mathcal{N}$ ,  $Y[n]$  a saída do filtro linear  $h[m]$  e  $\hat{Z}[n]$  a saída da transformação polinomial, sendo uma estimativa do tráfego real  $Z[n]$ .

### 2.1. Transformação Polinomial

A transformação polinomial, última etapa realizada no algoritmo TARMA, realiza a conversão da Função de Distribuição Acumulada qualquer da variável  $Y[n]$  na FDA da série real a ser sintetizada. Assim, podemos escrever a transformação polinomial como sendo (Allen [2014]),

$$Z[n] = F_z^{-1}(F_y(Y[n])) = \mathbf{p}_y(Y[n]) \quad (1)$$



onde  $F_z^{-1}$  é função inversa da FDA da variável  $Z$  e  $F_y$  a FDA da variável  $Y$ . Escrevendo a Equação (1) na forma polinomial, temos,

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

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  $Y[n]$ , a qual representa um variável aleatória normal com média zero e variância 1 (Laner et al. [2014]). Assim, temos a seguinte expressão para  $F_y(Y[n])$  (Montgomery et al. [2015]):

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

### 3. Função da Distribuição Acumulada Inversa Discreta Adaptativa

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

Seja  $Z[n]$  a série temporal normalizada a ser sintetizada, contida no intervalo  $[0, 1]$ . Ao fracionarmos esse intervalo em  $N$  partes iguais, obtemos,

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

Com o intervalo dividido, podemos calcular a quantidade de ocorrências para o qual os valores da série temporal  $Z$  se encontram nos 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 dividido pela quantidade total de ocorrência do conjunto universo em análise, podemos obter através da divisão do vetor  $\mathcal{V}_Z$  pelo tamanho total da série temporal, uma aproximação da Função de Distribuição de Probabilidade, com  $N \rightarrow \infty$ .

Através de  $\mathcal{V}_Z$ , obtemos o histograma acumulado  $f_k$ , como sendo,

$$f_k = \sum_{i=j}^k v_i \quad (6)$$

Dividindo a Equação (6) pela quantidade total de amostras, obtemos a Função de Distribuição Acumulada, 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 realizar o histograma.

Uma vez obtida a Equação (7), podemos formalizar o **Algoritmo da FDA Inversa Adaptativa**, como se segue,

- **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:** Localize o intervalo para o qual a amostra  $Z[k]$  se encontra.

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



- **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:** Calcule uma estimativa, para a FDA inversa, através da 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 para o qual  $d$  está contido.

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

Com o algoritmo da FDA Inversa Adaptativa obtemos  $F_z^{-1}$  com poucos cálculos computacionais, sendo a complexidade computacional deste algoritmo da ordem de  $O(N)$ .

#### 4. Algoritmo RLS

Considere o problema de se encontrar um relação linear entre valores conhecidos (mensurados)  $\varphi = [\varphi_1, \dots, \varphi_n]$  e valores desconhecidos  $\theta = [\theta_1, \dots, \theta_n]$ , com objetivo de se encontrar uma resposta desejada  $y[t]$ . Na forma matemática, temos (Farhang-Boroujeny [2013]):

$$y[t] = \theta_1 \cdot \varphi_1[t] + \dots + \theta_n \cdot \varphi_n[t] = \varphi[t]^T \theta \quad (12)$$

onde o vetor  $\varphi[t] = [\varphi_1[t], \dots, \varphi_n[t]]$ , chamado de variáveis de regressão, são os valores para o qual se deseja obter e  $n$  a ordem do modelo de regressão.

Uma forma para obter o vetor de regressão  $\varphi$ , na Equação (12), é buscar a minimização da função do método dos mínimos quadrados (Farhang-Boroujeny [2013]), representado a seguir:

$$\xi = \frac{1}{2} \sum_{i=1}^t \lambda^{k-i} (y[i] - \varphi[i]^T \theta)^2 \quad (13)$$

onde  $\lambda$  é uma constante no intervalo  $(0, 1]$  utilizada como fator de esquecimento, ou seja, quanto mais próxima do valor 1, maior é o peso no mínimo quadrado das amostras mais recentes.

A seguir, apresentamos o **Algoritmo RLS** (*Recursive Least Squares*), utilizado neste trabalho:

- **Passo 1:** Inicialize o vetor de regressão como sendo zero,  $\varphi = 0$ .
- **Passo 2:** Inicialize o vetor  $P = \sigma I$ , onde  $\sigma$  é uma constante positiva e  $I$  a matriz identidade de tamanho  $n$ .
- **Passo 3:** Realize a seguinte sequência para atualizar o vetor de regressão.

$$\begin{aligned} \hat{\theta} &= \hat{\theta} + K(y(t) - \varphi^T \hat{\theta}) \\ K &= P\varphi(\lambda + \varphi^T P\varphi)^{-1} \\ P &= \lambda^{-1}(I - K\varphi^T)P \end{aligned}$$

- **Passo 4:** Retorne ao Passo 3 na próxima iteração, até o fim de todas as iterações.

O algoritmo descrito acima, apresenta uma complexidade computacional de ordem  $O(n^2)$ , onde  $n$  é a ordem do filtro adaptativo. A seguir, descrevemos o modelo adaptativo proposto.



## 5. Modelo TARMA Adaptativo

O modelo autorregressivo de médias móveis, leva este nome por considerar como saída a média das entradas anteriores (médias móveis), uma constante  $\mu$  e médias das saídas do modelo nos instantes anteriores (autorregressivo). Matematicamente, podemos representar o modelo ARMA pela seguinte equação:

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

ou na forma vetorial,

$$AY = \mu + BX \quad (15)$$

Nesta seção, propomos um algoritmo para estimação dos parâmetros do Modelo TARMA. Para tal, inicialmente apresentamos uma equação para a média de um modelo ARMA.

### 5.1. Média do Modelo ARMA

Aplicando o operador esperança matemática (média) na Equação (14), obtemos:

$$\begin{aligned} \mathbb{E}[y[t]] + a_1 \cdot \mathbb{E}[y[t - 1]] + \dots + a_p \cdot \mathbb{E}[y[t - n]] \\ = \mathbb{E}[\mu] + b_1 \cdot \mathbb{E}[x[t]] + \dots + b_m \cdot \mathbb{E}[x[t - m]] \end{aligned} \quad (16)$$

Considerando um filtro invariante no tempo temos  $\mathbb{E}[y[t]] = \mathbb{E}[y[t - 1]] = \dots$ . Sabendo que  $X[n]$  representa um variável aleatória normal com média zero, logo,  $\mathbb{E}[x[t]] = \mathbb{E}[x[t - 1]] = \dots = 0$ . Reescrevendo a Equação (17), obtemos a expressão da média do modelo ARMA,

$$\mathbb{E}[y[t]] = \frac{\mu}{(1 + a_1 + \dots + a_n)} \quad (17)$$

Neste trabalho, com o objetivo de se obter uma variável normal  $Y[n]$  com média igual a zero, fazemos  $\mu = 0$ .

### 5.2. Estimação Adaptativa dos Parâmetros ARMA

Os autores em Laner et al. [2014] apresentam algoritmos de regressão polinomial para estimação 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, empregamos o algoritmo RLS (*Recursive Least Squares*) (Farhang-Boroujeny [2013]).

Reescrevendo a Equação (14), temos:

$$y[t] = -a_1 \cdot y[t - 1] - \dots - a_p \cdot y[t - n] + b_1 \cdot x[t] + \dots + b_m \cdot x[t - m] \quad (18)$$

Nota-se, que a Equação (18) possui valores conhecidos ( $Y$  e  $X$ ) e valores que desejamos descobrir ( $A$  e  $B$ ) para obtermos a resposta desejada  $y[t]$ . Reescrevendo os parâmetros da Equação (18), temos:

$$\theta^T = [a_1, \dots, a_n, b_1, \dots, , a_m] \quad (19)$$

e

$$\varphi^T[t] = [-y[t - 1], \dots, -y[t - n], x[t], \dots, x[t - m]] \quad (20)$$

Assim, obtemos a equação para a saída do modelo, ou seja,

$$y[n] = \varphi^T[t]\theta \quad (21)$$

Obtidos  $\varphi^T$  e  $y[n]$ , pelas Equações (20) e (21), aplicamos o algoritmo RLS para obter os valores dos parâmetros do modelo ARMA através da Equação (19).



### 5.3. Obtenção dos Parâmetros da Transformação Polinomial

Podemos aplicar o algoritmo RLS para obter também os parâmetros da transformação polinomial. A seguir, descrevemos como isso pode ser feito.

Ao expandirmos a Equação (2), obtemos:

$$Z[n] = p_0(Y[n])^0 + p_1(Y[n])^1 + \dots + p_P(Y[n])^P \quad (22)$$

Nota-se, que assim como a Equação (12), possuímos valores conhecidos ( $Y$  e  $Z$ ) e valores que desejamos descobrir ( $p_i$  para  $i = 0, \dots, P$ ) para obtermos a resposta desejada  $Z[t]$ . Reescrevendo no mesmo formato do algoritmo RLS, obtemos os seguinte vetores.

$$\theta^T = [p_0, \dots, p_P] \quad (23)$$

Através do Algoritmo da FDA inversa Adaptativa e da Equação (3), chegamos na resposta desejada,

$$z[n] = F_z^{-1}(F_y(Y[n])) \quad (24)$$

Com as Equações (23) e (24), podemos aplicar o algoritmo RLS e a cada iteração obteremos a estimativa do polinômio.

Obtido o algoritmo para obtenção do polinômio através do algoritmo da FDA Inversa Adaptativa e da regressão polinomial pelo algoritmo RLS, podemos formalizar o **algoritmo TARMA Adaptativo**.

- **Passo 1:** Realize o algoritmo RLS para obtenção do filtro adaptativo ARMA, ordem  $O(n^2)$ .
- **Passo 2:** Realize o algoritmo da FDA Inversa Adaptativa para obter a resposta desejada, ordem  $O(N)$ .
- **Passo 3:** Realize o algoritmo RLS para obtenção do polinômio através da resposta desejada do Passo 2, ordem  $O(n^2)$ .
- **Passo 4:** Gere uma amostra da variável Normal  $\mathcal{N}(0, 1)$ , realize a filtragem e a transformação polinomial para gerar amostras do tráfego sintético, ordem  $O(1)$ .
- **Passo 5:** Retorne ao Passo 1, na próxima iteração, até o fim de todas iterações.

O algoritmo TARMA Adaptativo apresenta uma complexidade computacional de ordem  $O(2n^2 + N + 1) = O(n^2)$ , onde  $n$  é a ordem do filtro utilizada no algoritmo RLS.

Na sequência, apresentamos como foi realizada a simulação e o resultados obtidos.

## 6. Simulações e Resultados

As simulações foram realizadas levando em conta os modelos ARMA, TARMA e o modelo adaptativo TARMA proposto. A série de tráfego utilizada neste trabalho foi o tráfego de rede TCP/IP Waikato VIII 20110520-000000-0 (WITS [2014]) coletado na comunicação entre a Universidade de Waikato e o resto do mundo com escala de agregação de 10ms e 100ms. Neste trabalho, nomearemos esse tráfego de rede como Waikato.

Na comparação dos algoritmos, avaliamos os resultados 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 e a ACF das séries sintéticas em relação ao tráfego real, calculamos o erro quadrático médio (EQM).

$$EQM = \frac{1}{T} \sum_{i=1}^T (v_d(i) - \hat{v})^2 \quad (25)$$

onde  $v_d$  é o valor da propriedade (ACF ou FDP) real,  $\hat{v}$  o valor estimado (sintético) e  $T$  a quantidade total de amostras da propriedade estatística a ser comparada.



### 6.1. Simulação de Enlace de Transmissão com *Buffer*

Para verificar se o comportamento das séries sintéticas é semelhante ao da série de tráfego real, realizamos a simulação de um enlace de transmissão de um servidor simples, descrito na Figura (2), onde  $z_t$  é o valor observado (série temporal) de taxa de tráfego e  $\eta$  a taxa de transmissão fornecida no instante de tempo  $t$ .



Figura 2: *Buffer* no Enlace de Transmissão de um Servidor.

Para se efetuar uma análise de desempenho do modelo em gerar séries que alimentarão o servidor da Figura (2), as seguintes medidas de desempenho podem ser consideradas:

**Taxa de Utilização( $u$ ):** A utilização média mede a fração de banda usada para servir o fluxo de dados observados no período de tempo  $T$ , calculada por:

$$u = \frac{1}{T} \sum_{t=1}^T \min \left\{ \frac{z_t}{\eta}, 1 \right\} \quad (26)$$

**Taxa de Perda(TP):** A taxa de perda mede a quantidade de bytes perdidos devido à alocação de banda menor do que a necessária, dada por:

$$TP = \frac{1}{T} \sum_{t=1}^T \max \left\{ \frac{(z_t - \eta) - B}{z_t}, 0 \right\} \quad (27)$$

onde  $B$  é o tamanho máximo da fila no *buffer*.

Na sequência apresentamos os resultados obtidos com as simulações realizadas.

### 6.2. Série de Tráfego Agregada na Escala 10ms

Na Figura (3), visualizamos os resultados obtidos para a FDP dos tráfegos de rede agregados em 10ms. Os resultados indicam que todas as séries apresentaram maiores valores para a FDP no intervalo entre 0 e 0,4, com destaque para a FDP do modelo ARMA, o qual apresentou os valores mais dispersos em comparação à FDP do tráfego real. Destaca-se que as FDP do modelo TARMA e do modelo proposto mostram valores de probabilidade próximos ao desejado (do tráfego real).

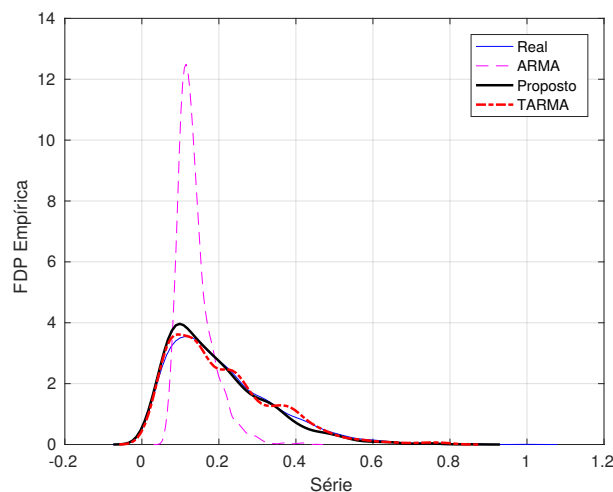


Figura 3: Função de Distribuição de Probabilidade para a Série Waikato, agregada na escala de 10ms.



Os resultados para a função de autocorrelação obtidos para o tráfego Waikato agregado na escala de  $10m.s$  podem ser visualizados na Figura (4). Nota-se que assim como os resultados da FDP, o modelo ARMA apresentou valores mais dispersos do desejado em comparação ao TARMA e o proposto. Destaca-se que para valores de atraso maiores que 20, o modelo proposto apresentou valores mais próximos da ACF da série real.

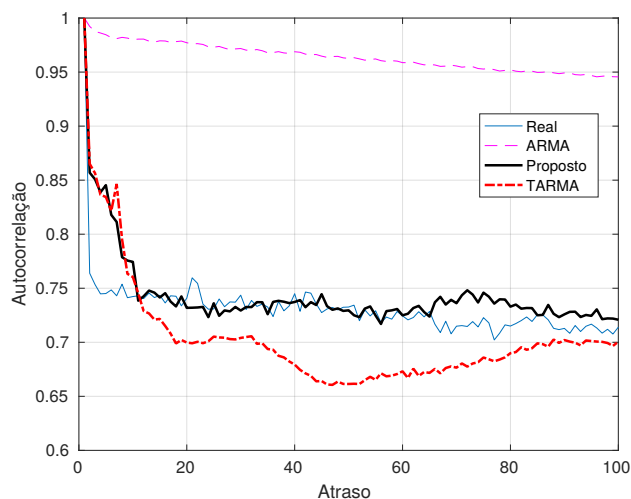


Figura 4: Autocorrelação com atraso até 100 amostras para a Série Waikato, agregada na escala de  $10m.s$ .

A Tabela (1) mostra os resultados estatísticos das séries sintéticas e real na escala de  $10m.s$ . Os valores mais próximos em relação ao tráfego real para as estatísticas de média, de variância e de momentos foram obtidos pelo modelo TARMA, seguidos pelo modelo Proposto e pelo modelo ARMA. Destaca-se, que apesar do algoritmo TARMA utilizar todos os dados do fluxo de rede real para estimar seus parâmetros, o mesmo apresentou valores estatísticos semelhantes aos do algoritmo TARMA adaptativo.

Tabela 1: Dados Estatísticos para a Série Waikato, agregada na escala de  $10m.s$ .

	Série	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 resultados do erro quadrático médio da FDP e da ACF na Tabela (1) indicam valores mais próximos do desejado pelo algoritmo TARMA proposto, seguidos do TARMA e do ARMA.

Os resultados obtidos para a relação Pico/Média na Tabela (1) indicam valores próximos ao do tráfego de rede real para o modelo TARMA adaptativo, seguidos dos modelos TARMA e do ARMA.



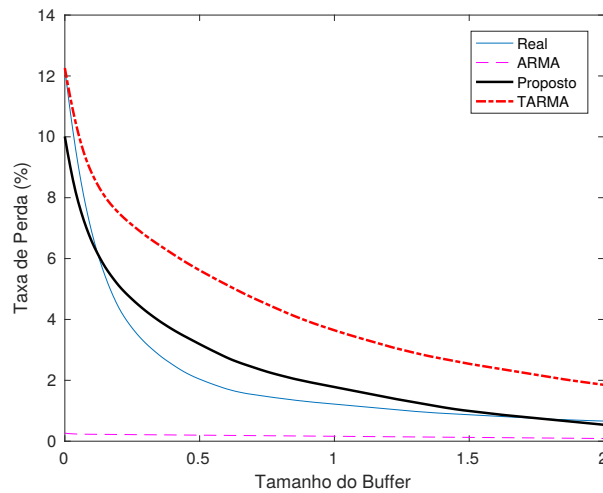


Figura 5: Taxa de Perda (%) no *buffer* em relação ao tamanho máximo da fila, com escala de agregação de 10ms.

A Figura (5) apresenta os resultados para a Taxa de Perda (%) em relação à variação do tamanho máximo do *buffer*  $B$ , onde quanto maior o valor de  $B$  menor será a Taxa de Perda, explicada pela maior capacidade de armazenamento de dados. Nota-se que os modelos TARMA e o modelo proposto apresentaram valores semelhantes aos do tráfego de rede real, enquanto os valores mais dispersos foi obtido pelo modelo ARMA. Observa-se que o modelo proposto produziu valores mais próximos do real.

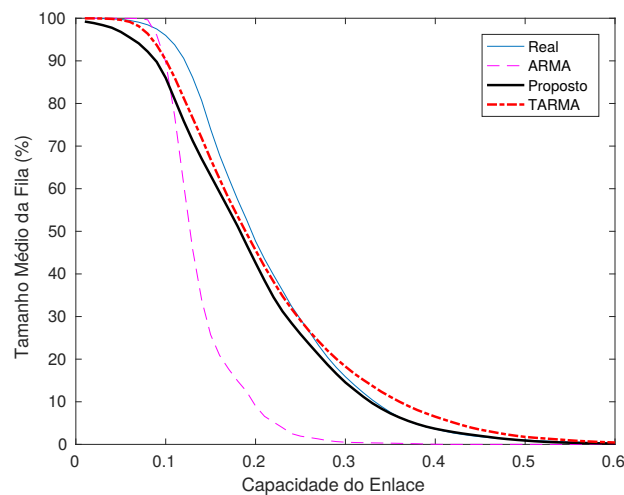


Figura 6: Tamanho Médio da fila no *buffer* em relação à capacidade de transmissão do enlace  $\eta$ , com escala de agregação de 10ms.

A Figura (6) apresenta os resultados para o Tamanho Médio da Fila (%) em relação à variação da capacidade do enlace de saída  $\eta$ . O comportamento apresentado pelo modelo proposto e pelo modelo TARMA foram os mais próximos da série real, enquanto o modelo ARMA apresentou valores inferiores para quase todos os valores de capacidade.

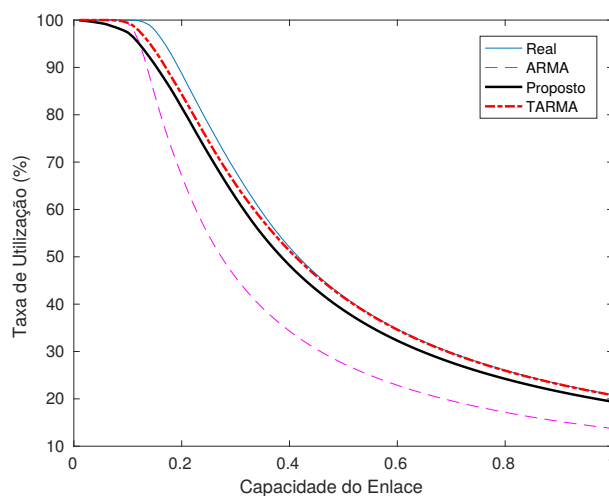


Figura 7: Taxa de Utilização no *buffer* em relação à capacidade de transmissão do enlace  $\eta$ , com escala de agregação de  $10ms$ .

Os resultados para a Taxa de Utilização no *buffer* em relação à capacidade de transmissão do enlace  $\eta$ , com escala de agregação de  $10ms$ , podem ser visualizados na Figura (7). Assim como os resultados para o Tamanho Médio da Fila, os modelos TARMA e proposto apresentam valores mais próximos em relação ao tráfego real, onde se repetiu o efeito de valores inferiores para o modelo ARMA.

Ao analisarmos os resultados para a simulação do enlace de transmissão, notamos que de forma geral, o modelo proposto apresenta comportamento semelhante ao fluxo de tráfego real, indicando que as amostras sintéticas podem ser utilizadas em simulações para representar a série desejada.

### 6.3. Série de Tráfego Agregada na Escala $100ms$

Na Figura (8), podemos visualizar os resultados obtidos para a FDP, na escala de agregação de  $100ms$  para a série Waikato. Nota-se valores concentrados de probabilidade próximos a 0,2, o que é comprovado na Tabela 2, onde média da série real é 0,2616. Os modelos TARMA e o modelo proposto apresentaram probabilidades mais próximas do desejado em comparação ao modelo ARMA.

Os resultados da ACF obtidos para o tráfego na escala de  $100ms$ , pode ser visualizado na Figura (8). Nota-se, valores de ACF do modelo ARMA mais distantes da autocorrelação da série real, principalmente para valores maiores de atraso.

A Tabela 2 apresenta os resultados estatísticos para a série Waikato em  $100ms$ . Destaca-se os valores mais próximos do desejado para o modelo proposto, com diferença mínima para o modelo TARMA. Nota-se os valores mais distantes do desejado para o modelo ARMA.

Os resultados do erro quadrático mínimo da FDP e da ACF, na Tabela 2, mostraram os menores valores para o modelo TARMA adaptativo, seguidos pelo modelo TARMA e do modelo ARMA. Para os resultados de Pico/Média, notamos os valores mais próximos da série real para o modelo TARMA.

As simulações realizadas para o tráfego de rede Waikato também foram realizadas para outros tráfegos de rede, onde destaca-se que os resultados obtidos foram semelhantes ao obtidos neste trabalho.

## 7. Conclusão

Neste trabalho, propomos um algoritmo para estimar adaptativamente os parâmetros do modelo TARMA para geração de séries temporais sintéticas de forma a capturar a autocorrelação e função de distribuição acumulada independentemente de séries de tráfego reais.

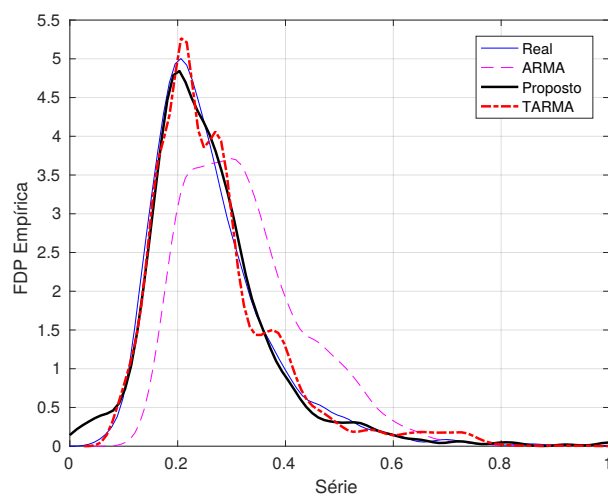


Figura 8: Função de Distribuição de Probabilidade para a Série Waikato, agregada na escala de 100ms.

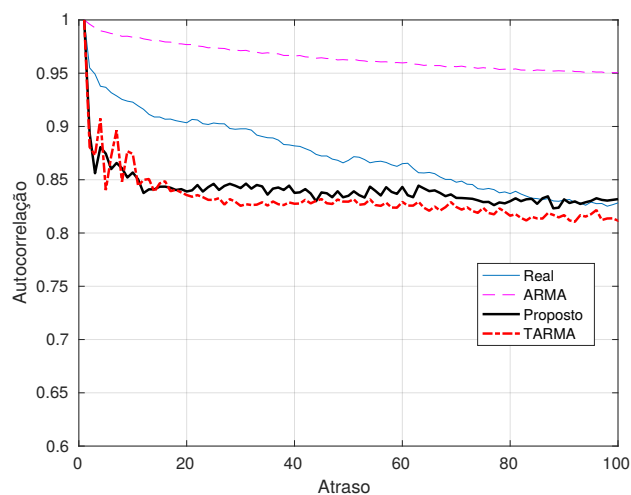


Figura 9: Autocorrelação com atraso de 100 amostras para a Série Waikato, agregada na escala de 100ms.

Tabela 2: Dados Estatísticos para a Série Waikato, agregada na escala de 100ms.

	Série	ARMA	TARMA	Proposto
Média	0,2616	0,3060	0,2683	0,2583
Variância	0,0126	0,0128	0,0142	0,0141
2º Momento	0,0810	0,1191	0,0808	0,0862
3º Momento	0,0303	0,0484	0,0337	0,0311
4º Momento	0,0137	0,0216	0,0160	0,0151
EQM FDP	0	0,7570	0,2515	0,2848
EQM ACF	0	$2,4 \cdot 10^6$	122,9	5.579,3
Pico/Média	3,8227	3,3037	3,8614	4,4148

Os resultados, de forma geral, mostraram que os modelos TARMA e o modelo TARMA adaptativo proposto apresentaram melhor desempenho para modelar a função de autocorrelação e



função de distribuição de probabilidade em comparação ao modelo ARMA.

Observa-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 mais próximos do desejado para o modelo proposto. Destaca-se que o modelo TARMA necessita de todas as amostras da série real, enquanto o modelo proposto é realizado adaptativamente.

Para trabalhos futuros, objetivamos avaliar o uso de outros algoritmos para obtenção do polinômio e do filtro linear, com menores complexidades computacionais.

### Referências

- Allen, A. O. (2014). *Probability, statistics, and queueing theory*. Academic Press.
- Dalmazo, B. L., Vilela, J. P., e Curado, M. (2017). Performance analysis of network traffic predictors in the cloud. *Journal of Network and Systems Management*, 25(2):290–320.
- Farhang-Boroujeny, B. (2013). *Adaptive filters: theory and applications*. John Wiley & Sons.
- Laner, M., Svoboda, P., Nikaein, N., e Rupp, M. (2013). Traffic models for machine type communications. In *ISWCS 2013; The Tenth International Symposium on Wireless Communication Systems*, p. 1–5.
- Laner, M., Svoboda, P., e Rupp, M. (2014). Parsimonious network traffic modeling by transformed arma models. *IEEE Access*, 2:40–55.
- Montgomery, D. C., Jennings, C. L., e Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
- Wang, H. (2015). Anomaly detection of network traffic based on prediction and self-adaptive threshold. *International Journal of Future Generation Communication and Networking*, 8(6):205–214.
- WITS (2014). Wits: Waikato internet traffic storage from university of waikato. <http://wand.net.nz/wits/waikato/8/>.

### Apêndice 1: Prova da Equação (11)

Seja o vetor  $[f_1, f_2, \dots, f_N]$ , o histograma acumulado e  $d$  o valor da FDA para o qual se deseja obter a inversa. Podemos estimar o valor de  $F_z^{-1}$ , através da média ponderada dos valores dos extremos no intervalo onde se encontra  $d$ , realizando a localização do intervalo de  $d$ , temos:

$$f_e = \sup_{d \leq f_i} \{f_1, f_2, \dots, f_N\} \quad (28)$$

e

$$e = \arg \left\{ \sup_{d \leq f_i} \{f_1, f_2, \dots, f_N\} \right\} \quad (29)$$

Sabendo que a FDA é aproximadamente o histograma acumulado dividido pela quantidade total de amostras  $L$  utilizadas da série. e através de proporcionalidade no intervalo  $[f_{e-1}, f_e]$ , onde  $d \in [f_{e-1}, f_e]$  temos:

$$\frac{(f_e/L - f_{e-1}/L)}{(d - f_{e-1}/L)} = \frac{(e - (e - 1))}{(F_z^{-1}(d) - (e - 1))} \quad (30)$$

Isolando  $F_z^{-1}(d)$ , obtemos a equação final para a inversa da FDA discreta,

$$F_z^{-1}(d) = (e - 1) + \frac{(L \cdot d - f_{e-1})}{(f_e - f_{e-1})} \quad (31)$$