

Modelagem de Tráfego VoIP Utilizando Misturas Gaussianas Ajustadas por Algoritmo de Maximização de Esperança e Comprimento de Descrição Mínimo

Cheila Mendes de Oliveira ¹
Rosana de Paula Lucas ¹
Flávio Henrique Teles Vieira ¹

Resumo: O aumento do número de aplicações que geram tráfego VoIP (Voice over IP) requer que modelos adequados para este tipo de tráfego sejam empregados para um melhor dimensionamento dos recursos da rede. Neste artigo, propomos a aplicação de Misturas Gaussianas ajustadas por um algoritmo EM (*Expectation – Maximization*) com seleção de número de classes utilizando o princípio do Comprimento de Descrição Mínimo (*MDL-Minimum Description Length*) para descrever o tráfego VoIP. Dados reais de VoIP foram coletados e séries sintéticas foram geradas de acordo com esses dados. O algoritmo proposto é comparado com o algoritmo EM e com algoritmo baseado em MDL em termos do desempenho em modelar as séries de tráfego VoIP consideradas.

Abstract: The increase in the number of applications that generate VoIP (Voice over IP) traffic requires that suitable models for this kind of traffic be used to better dimension and predict network resources. In this paper, we propose the application of Gaussian mixture adjusted by an EM algorithm (Expectation-Maximization) with selection of classes using the principle of Minimum Description Length (MDL-Minimum Description Length) in order to describe VoIP traffic. Real VoIP traces were collected and synthetic time series were generated according to these data. The proposed algorithm is compared to the EM algorithm and the MDL (Minimum Description Length) based algorithm in terms of performance in modeling the considered VoIP traffic traces.

1 Introdução

Entre as vantagens da transmissão de voz através de redes IP pode-se destacar a utilização de enlaces (cabos metálicos, fibras, etc.) e tecnologias já existentes e o fato de oferecer oportunidade de novos negócios ao setor de telefonia. Entretanto, o sistema VoIP, também possui obstáculos como a latência e a perda de pacotes que devem ser controlados para que os

¹Escola de Engenharia Elétrica, Mecânica e de Computação (EMC) da Universidade Federal de Goiás (UFG) Av. Universitária, n. 1488 - Quadra 86 - Bloco A - 3o piso 74605-010 - Setor Leste Universitário - Goiânia - GO {cheila@eee.ufg.br, rosana.rpl@hotmail.com, flavio@eeec.ufg.br}

serviços VoIP sejam adequados e de qualidade [3]. Para dimensionar corretamente as redes é interessante utilizar modelos que capturem as características dos fluxos de informação com um número razoável de parâmetros, sendo a estimativa dos mesmos o mais simples possível [2].

Um modelo de tráfego adequado é uma ferramenta valiosa para o planejamento da capacidade exata de uma rede [2]. O Modelo de Poisson, comumente utilizado em aplicações que se tem um grande número de fluxos de tráfego independente, o Modelo de Markov que apresenta a característica de que o estado futuro depende apenas do estado presente e o Modelo ON-OFF que é inspirado na natureza da voz que é composto por períodos de silêncio e som são exemplos de técnicas que modelam o tráfego VoIP.

A principal motivação para o estudo da aplicação de Misturas Gaussianas para modelar o tráfego VoIP foi a constatação de que suas densidades de probabilidade poderiam ser representadas pela soma de distribuições Gaussianas. As Misturas Gaussianas oferecem uma modelagem precisa para vários tipos de fenômenos [5]. Neste trabalho, propomos um algoritmo de ajuste de parâmetros para o modelo de Misturas Gaussianas baseado no algoritmo de estimação de maximização da esperança (EM) [5], e no conceito de Comprimento de Descrição Mínimo (MDL) [1], para representar eficientemente o tráfego VoIP.

Este artigo está dividido em quatro seções. A seção 2 introduz conceitos sobre Misturas Gaussianas. A seção 3 apresenta os algoritmos de ajuste de Misturas Gaussianas e os resultados das análises realizadas. Finalmente, a seção 4 expõe as conclusões obtidas neste trabalho.

2 Misturas Gaussianas

A Mistura Gaussiana é uma forma de modelagem de séries temporais que consiste no agrupamento de densidades de probabilidade gaussianas com características individuais. Este modelo armazena pequeno volume de informação e utiliza pouca carga computacional para efetuar as estimativas de densidade [8], [5].

Os parâmetros da mistura de distribuição Gaussiana são o número de distribuições que formam a mistura, as proporções que cada distribuição da mistura possuem, a média e a matriz de covariância de cada componente da mistura [8]. O objetivo é determinar o número de distribuições da mistura que melhor represente a densidade de probabilidade dos dados coletados, ou seja, automatizar o procedimento de escolha de um número inicial de Gaussianas para o algoritmo EM.

$$\hat{f}(x) = \sum_{i=1}^c \hat{\rho}_i \phi(x; \hat{\mu}_i, \hat{\Sigma}_i) \quad (1)$$

onde c é o número de Gaussianas da mistura, $\hat{\rho}_i$ é o coeficiente de ponderação da mistura, x é um vetor de n -dimensional, $\hat{\mu}_i$ é um vetor de médias de n -dimensional e $\hat{\Sigma}_i$ é uma matriz de covariância $n \times n$.

3 Algoritmos de Ajuste de Misturas Gaussianas

Para se gerar uma série sintética que apresente densidade de probabilidade semelhante a série coletada é necessário estimar os parâmetros da Equação (1). Série Sintética, neste trabalho, refere-se as séries geradas pelo software MatLab utilizando os parâmetros estimados pelos algoritmos EM e MDL em função das características observadas nas séries VoIP coletadas. Para isso, foram realizadas simulações computacionais utilizando algoritmos baseados nas seguintes abordagens:

3.1 Algoritmo EM (*Expectation – Maximization*)

A aplicação do algoritmo EM é largamente utilizado na solução de problemas em que os dados observados são vistos como dados incompletos e dificultam a estimativa de máxima verossimilhança [7]. Em cada iteração o algoritmo EM realiza dois procedimentos, no primeiro (passo E) o algoritmo estima os valores, mais prováveis, dos parâmetros da função de densidade de probabilidade dos dados observados e gera uma nova sequência de dados. No segundo (passo M) o algoritmo determina o log da função de verossimilhança da nova sequência de dados gerada no passo E [6]. Estes procedimentos são realizados até que os valores dos parâmetros estimados apresentem pequenas ou nenhuma alteração. O algoritmo EM possui duas limitações: a taxa de convergência pode ser lenta e é sensível à etapa de inicialização [5].

Os passos do algoritmo EM para estimar os parâmetros para uma mistura gaussiana com componentes normais multivariados são dados a seguir [5].

Algoritmo 01:

1. Determinar o número c de funções de densidade de probabilidade Gaussianas da mistura. Neste momento, é necessário a utilização de conhecimento prévio ou de algum outro método para estimar o número de termos.
2. Determinar uma estimativa inicial para os parâmetros componentes utilizando o algoritmo *Kmeans*. Estes parâmetros são as proporções da mistura, as médias e matrizes de covariância para cada distribuição normal.

3. para cada ponto de dados x_j , calcular a probabilidade a posteriori utilizando a Equação (2).

$$\hat{\tau}_{ij} = \left(\frac{\hat{\rho}_i \phi(x_j; \hat{\mu}_i, \hat{\Sigma}_i)}{\hat{f}(x_j)} \right); i = 1, \dots, c; j = 1, \dots, n. \quad (2)$$

onde $\hat{\tau}_{ij}$ representa a estimativa de probabilidade posterior do ponto x_j pertencer ao i -ésimo termo, $\phi(x_j; \hat{\mu}_i, \hat{\Sigma}_i)$ é a densidade normal multivariada para o i -ésimo termo avaliado em x_j e a estimativa da Mistura Gaussiana no ponto x_j é dada pela Equação (1)

A probabilidade a posteriori é a probabilidade de cada elemento da série observada pertencer a cada uma das densidades componente da mistura. Pode-se usar essa probabilidade para obter uma atualização ponderada dos parâmetros (proporções da mistura, média e matriz de covariância) de cada componente da mistura.

4. Atualizar os coeficientes da mistura, as médias e as matrizes de covariância para os componentes individuais através das Equações (3), (4) e (5).

$$\hat{\rho}_i = \frac{1}{n} \sum_{j=1}^n \hat{\tau}_{ij}, \quad (3)$$

$$\hat{\mu}_i = \frac{1}{n} \sum_{j=1}^n \frac{\hat{\tau}_{ij} x_j}{\hat{\rho}_i}, \quad (4)$$

$$\hat{\Sigma}_i = \frac{1}{n} \sum_{j=1}^n \frac{\hat{\tau}_{ij} (x_j - \hat{\mu}_i)(x_j - \hat{\mu}_i)^T}{\hat{\rho}_i} \quad (5)$$

5. Repetir os passos 3 a 4 até as estimativas dos coeficientes da mistura Gaussiana, das médias e das matrizes de covariância convergirem.

3.2 Algoritmo MDL (*Minimum Description Length*)

O MDL é uma abordagem de inferência indutiva que proporciona uma solução genérica para o problema da seleção de modelo e lidar com ausência de dados ou dados inconsistentes. O Princípio do MDL baseia-se na seguinte constatação: as possíveis regularidades encontradas nos dados podem ser usadas para comprimir os mesmos, assim cada regularidade pode ser vista como uma componente da mistura, que os dados observados apresenta, com os parâmetros proporção da mistura, média e matriz de covariância. Os dados compactados apresentam um número menor de funções de densidade de probabilidade e conseqüentemente menos parâmetros serão necessários para caracterizar esses dados [4].

As etapas relacionadas abaixo resumem a abordagem de utilização do critério MDL proposto em [1] para encontrar os parâmetros da Mistura Gaussiana.

Algoritmo 02:

1. Inicializar com um valor grande para número de funções de densidade de probabilidade da mistura, c .

2. Inicializar ϕ^1 (onde $\phi^1 = \rho_i^1, \mu_i^1, \hat{\Sigma}_i^1$) usando (6), (7) e (8):

$$\rho_i^1 = \frac{1}{c}, \tag{6}$$

$$\mu_i^1 = x_n, \tag{7}$$

$$\hat{\Sigma}_i^1 = \frac{1}{N} \sum_{n=1}^N x_n x_n^t \tag{8}$$

onde x_n é a amostra da série que se deseja modelar usando a Mistura Gaussiana; N é o número de elementos da série; $n = [(i - 1)(N - 1)/c - 1] + 1$

3. Aplicar a Equação (9) até a mudança de MDL($c; \phi$) dado pelas Equações (10) e (11), seja menor que ε dado pela Equação (12).

$$\begin{aligned} (\rho^{(i+1)}, \mu^{(i+1)}, \hat{\Sigma}^{(i+1)}) &= \arg_{(\rho, \mu, \hat{\Sigma}) \in \Omega(\kappa)} \max Q(\phi; \phi^{(i)}), \\ (\rho^{(i+1)}, \mu^{(i+1)}, \hat{\Sigma}^{(i+1)}) &= (\bar{\rho}, \bar{\mu}, \hat{\Sigma}) \end{aligned} \tag{9}$$

$$Q(\phi; \phi^{(i)}) = E[\log \rho_{y,x}(y, X|\phi) | Y = y, \phi^{(i)}] - \frac{1}{2} L \log(NM), \tag{10}$$

$$MDL(c, \phi) = -\log \rho_y(y|c, \phi) + \frac{1}{2} L \log(NM) \tag{11}$$

$$\varepsilon = \frac{1}{100} \left(1 + M + \frac{(M+1)M}{2} \right) \log(NM) \tag{12}$$

e L é um termo de penalidade e M é a dimensão da série temporal que se deseja modelar.

4. Armazenar o parâmetro $\phi^{(C, i_{final})}$ e o valor MDL($c, \phi^{(C, i_{final})}$).

5. Se o número de subclasses c é maior que 1, aplicar a Equação (13) para reduzir o número de subclasses, conjunto $c \leftarrow c - 1$, e voltar para o passo 3.

$$(l^*, m^*) = \operatorname{arg}_{(l,m)} \min d(l, m) \quad (13)$$

onde l e m são exemplos de subclasses.

6. Escolher o valor de c^* e dos parâmetros $\phi^{(c, i_{final})}$ que minimizam o valor do MDL. Uma vez obtidos c^* e $\phi^{(c, i_{final})}$ é possível utilizar as Equações (14) e (15) para calcular a densidade de probabilidade da Mistura e posteriormente gerar uma série sintética.

$$\rho_{y_n|x_n}(y_n|c, \phi) = \frac{1}{(2\pi)^{\frac{M}{2}}} |R_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_n - \mu_i)^t R_c^{-1}(y_n - \mu_i)\right\} \quad (14)$$

$$\rho_{y_n}(y_n|c, \phi) = \sum_{i=1}^c \rho_{y_n|x_n}(y_n|c, \phi)\pi_i \quad (15)$$

3.3 Algoritmo Proposto MDL-EM

Como mencionado na seção 3.1, o algoritmo EM é sensível a etapa de inicialização. O algoritmo MDL tenta obter o menor número de subclasses que melhor descreve os dados da amostra através da compressão dos dados em função das regularidades encontradas.

O algoritmo proposto de Maximização da Esperança com Seleção de Classes utilizando Comprimento de Descrição Mínimo (MDL-EM) determina o número de distribuições da mistura através do algoritmo MDL que será utilizado pelo algoritmo EM. A utilização do algoritmo MDL-EM automatiza a etapa que determina o número de distribuições da mistura no algoritmo EM. O algoritmo é apresentado abaixo:

Algoritmo Proposto:

Utilizar os passos 1, 2, 3, 4, 5 e 6 do **Algoritmo 02** para obter o número distribuições da mistura e com este número utilizar os passos 2, 3, 4 e 5 do **Algoritmo 01** para determinar os outros parâmetros necessários para obter a função de densidade de probabilidade e gerar as séries sintéticas.

4 Avaliação dos Algoritmos

4.1 Dados Coletados

Coletaram-se os dados de chamadas realizadas com um software (Skype) que utiliza a comunicação VoIP entre dois terminais fim-a-fim. Para mostrar os resultados de modelagem

neste artigo, utilizou-se duas séries de tráfego que representam a maioria dos comportamentos do tráfego VoIP das séries coletadas. A Tabela 1 mostra o comportamento estatístico dos dados coletados para as duas chamadas selecionadas.

Tabela 1. Resumo dos dados coletados nas chamadas VoIP.

Chamada	<i>N</i> ° de <i>Payloads</i> (Amostra)	Média	Variância
01	15.337	97,6224	313,5448
02	41.240	455,8242	338890

Para encontrar o número adequado de gaussianas para as Misturas Gaussianas e realizar as comparações entre as abordagens, geramos séries sintéticas, calculando para as mesmas, a média, a variância e o desvio-padrão.

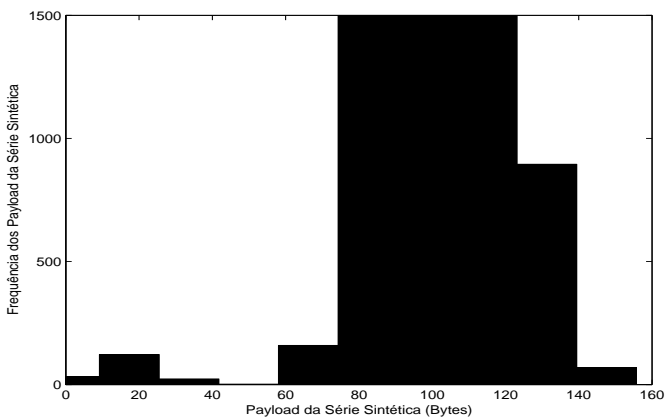
4.2 Algoritmo EM

Nesta seção, geramos séries com o número de gaussianas variando de 01 até 20 através do algoritmo mostrado na subseção 3.1. Fizemos 100 simulações para cada caso. Para cada série gerada, calculamos a média, a variância e o desvio padrão que armazenamos em um vetor, em seguida calculou-se as médias das mesmas. Calculamos também, o erro percentual entre as estatísticas para cada uma das chamadas. As Tabelas 2 e 3 mostram os erros percentuais calculados e as Figuras 1 e 2 mostram os gráficos obtidos a partir das séries sintéticas geradas pelo algoritmo EM. O número de gaussianas que faz com que a Mistura Gaussiana apresente o menor erro percentual para a média não é o mesmo que apresenta o menor erro percentual para variância, logo esta abordagem retorna mais de um valor para o número de gaussianas. A solução adotada para indicar o valor apropriado para o número de Gaussianas foi escolher o número de gaussianas que apresentou o menor valor de erro percentual para a média.

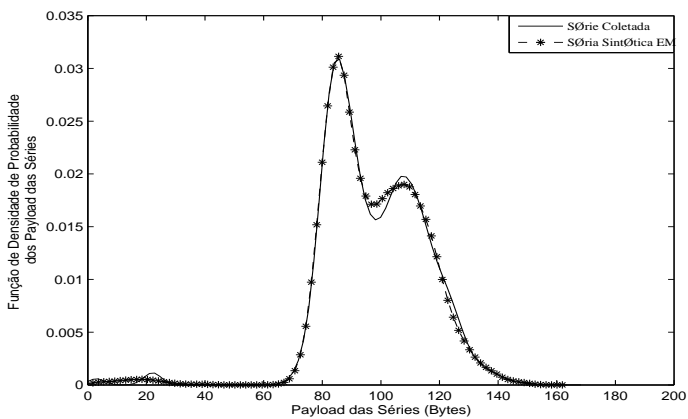
Tabela 2. Parâmetros estatísticos para chamada 01 usando EM.

	Série Coletada	Série Sintética 05 Gaussianas	Erro (%)
Média	97,6224	97,6436	0,0217
Variância	313,5448	313,8234	0,0888
Desvio Padrão	17,7072	17,7151	0,0446

A partir dos valores mostrados nas Tabelas 2 e 3, nota-se que os valores calculados para o número de gaussianas são próximos para todas as chamadas. O valor da média da série

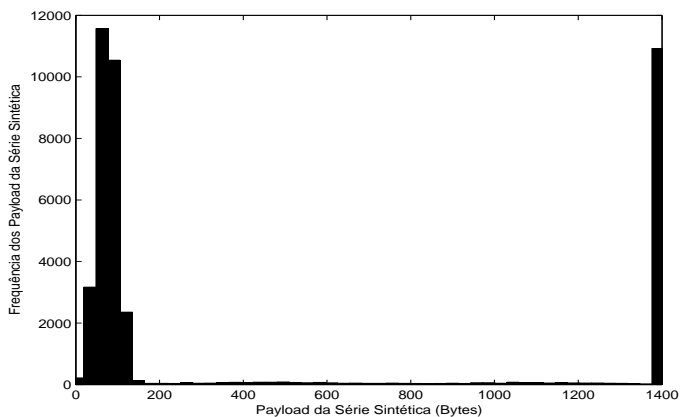


(a) Histograma série sintética

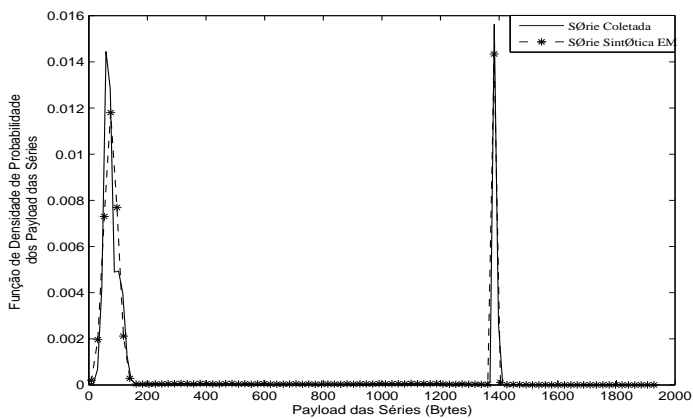


(b) pdf série coletada e sintética

Figura 1. Características da chamada VoIP 01 (abordagem EM)



(a) Histograma série sintética



(b) pdf série coletada e sintética

Figura 2. Características da chamada VoIP 02 (abordagem EM)

Tabela 3. Parâmetros estatísticos para chamada 02 usando EM.

	Série Coletada	Série Sintética 04 Gaussianas	Erro (%)
Média	455,8242	455,9194	0,0209
Variância	338890	338960	0,0206
Desvio Padrão	582,1422	582,2069	0,0111

coletada é bem próximo ao valor da média da série sintética apresentando assim pequenos erros percentuais. As figuras acima confirmam os resultados das tabelas, mostrando um histograma e um gráfico da função de densidade de probabilidade da série sintética semelhantes ao da série coletada.

4.3 Algoritmo MDL

Nesta seção, utilizamos o princípio do MDL para encontrar o número de gaussianas e os parâmetros necessários para gerar a série sintética. A partir da série gerada calculamos a média, a variância e o desvio padrão para comparar com os valores da série coletada. As Tabelas 4 e 5 mostram os valores estatísticos e as Figuras 3 e 4 mostram os gráficos construídos a partir das séries sintéticas geradas para as chamadas 01 e 02.

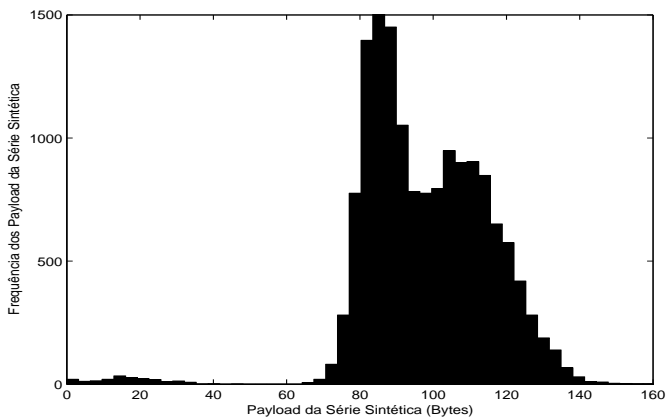
Tabela 4. Parâmetros estatísticos para chamada 01 usando MDL.

	Série Coletada	Série Sintética 03 Gaussianas	Erro (%)
Média	97,6224	97,6215	0,0009
Variância	313,5448	313,9180	0,1190
Desvio Padrão	17,7072	17,7177	0,0593

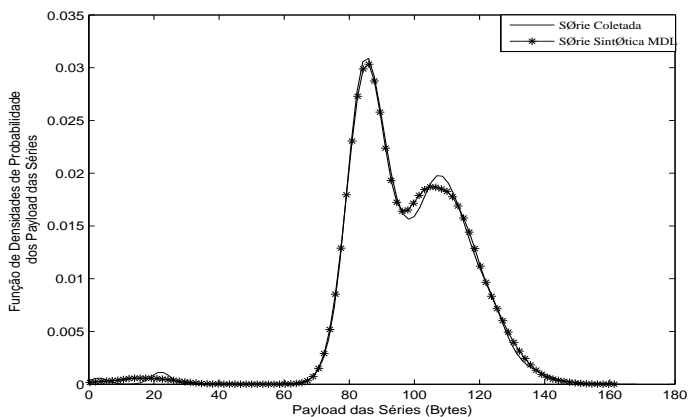
Tabela 5. Parâmetros estatísticos para chamada 02 usando MDL.

	Série Coletada	Série Sintética 09 Gaussianas	Erro (%)
Média	455,8242	456,4878	0,1456
Variância	338890	338610	0,0826
Desvio Padrão	582,1422	581,9055	0,0407

Nas Tabelas 4 e 5, observa-se que o valor da média da série coletada é bem próximo do valor da média da série sintética para a chamada 01, proporcionando um erro percentual



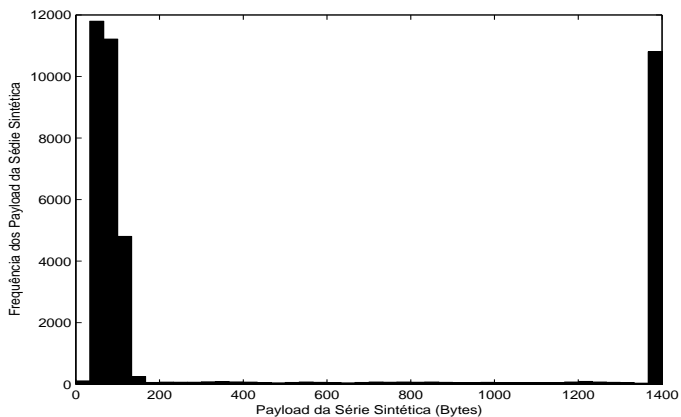
(a) Histograma série sintética



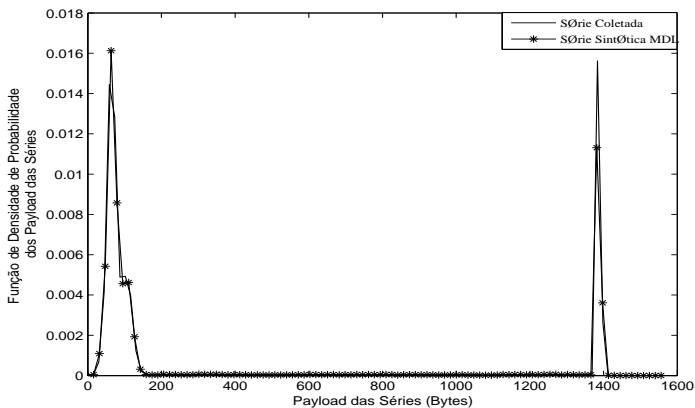
(b) pdf série coletada e sintética

Figura 3. Características da chamada VoIP 01 (abordagem MDL)

menor entre o valor da média, para a chamada 02 os valores da média da série coletada com a série sintética apresentam valores distantes e como conseqüência tem-se valores maiores para os erros percentuais calculados entre a média. As Figuras 3 e 4 confirmam os resultados das tabelas mostrando um histograma e um gráfico da função de densidade de probabilidade da série sintética, para a chamada 01, com uma maior semelhança a série coletada que o



(a) Histograma série sintética



(b) pdf série coletada e sintética

Figura 4. Características da chamada VoIP 02 (abordagem MDL)

histograma e o gráfico gerado com base na série sintética da chamada 02.

4.4 Algoritmo Proposto MDL-EM

Nesta seção, utilizamos o algoritmo MDL-EM para determinar a número de Gaussianas e os valores para os parâmetros necessários para a função de densidade de probabilidade e gerar a série sintética. Neste procedimento, utilizamos o algoritmo MDL para fornecer o número de grupos (subclasses) formados com os dados coletados, que corresponde ao número de gaussianas procurado e com este número de subclasse utilizou-se o algoritmo EM para obter os outros parâmetros da mistura para gerar a série sintética. As Tabelas 6 e 7 mostram os valores estatísticos e as Figuras 5 e 6 mostram os gráficos construídos a partir das séries sintéticas geradas para as chamadas 01 e 02 respectivamente.

Tabela 6. Parâmetros estatísticos para chamada 01 usando MDL-EM.

	Série Coletada	Série Sintética 03 Gaussianas	Erro (%)
Média	97,6224	97,6385	0,0166
Variância	313,5448	311,8561	0,5386
Desvio Padrão	17,7072	17,6594	0,2694

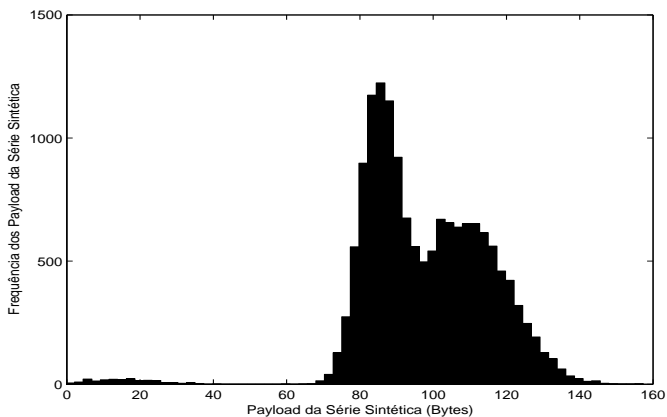
Tabela 7. Parâmetros estatísticos para chamada 02 usando MDL-EM.

	Série Coletada	Série Sintética 09 Gaussianas	Erro (%)
Média	455,8242	455,7431	0,0178
Variância	338890	338650	0,0700
Desvio Padrão	582,1422	581,9382	0,0350

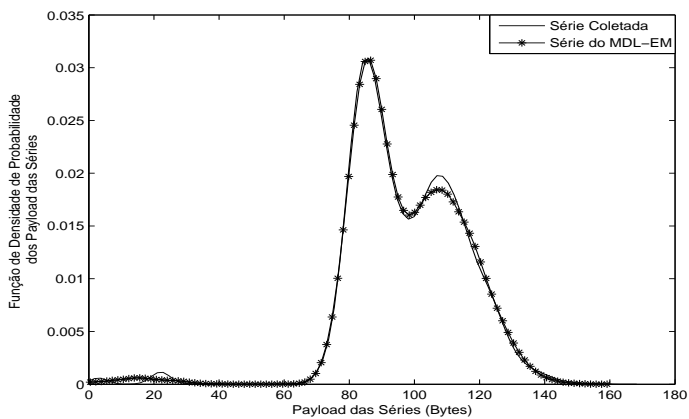
As Tabelas 6 e 7, apresentam valores próximos para o erro percentual entre a média da série sintética e a da série coleta para as chamadas 01 e 02 respectivamente.

Com a geração de séries sintéticas utilizando os valores calculados pelos algoritmos EM, MDL e MDL-EM para a Mistura Gaussiana, observa-se que o algoritmo MDL apresentou o menor valor para o erro percentual entre os valores da média da série coletada e da série sintética apenas para a chamada 01. Este resultado se deve as características da chamada 01, onde se obteve uma maior compressão dos dados observados do que para as outras séries com a utilização do algoritmo MDL. Para as outras chamadas, devido as suas características, o algoritmo MDL-EM apresentou os menores valores dos erros percentuais entre a média da série coletada e da série sintética.

Na avaliação do tempo de execução dos algoritmos na estimativa dos parâmetros do modelo de Misturas Gaussianas realizada, o algoritmo MDL-EM apresentou o menor valor



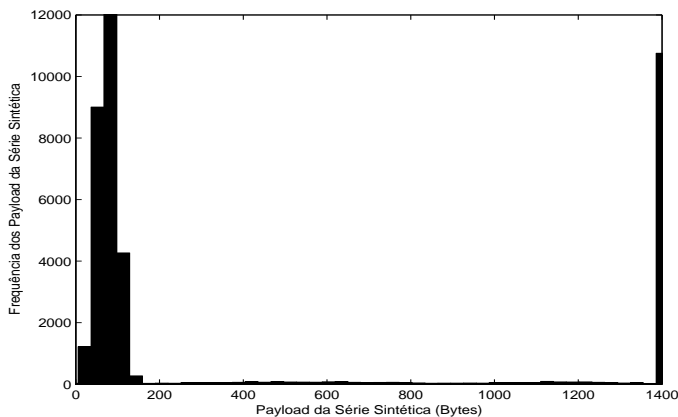
(a) Histograma série sintética



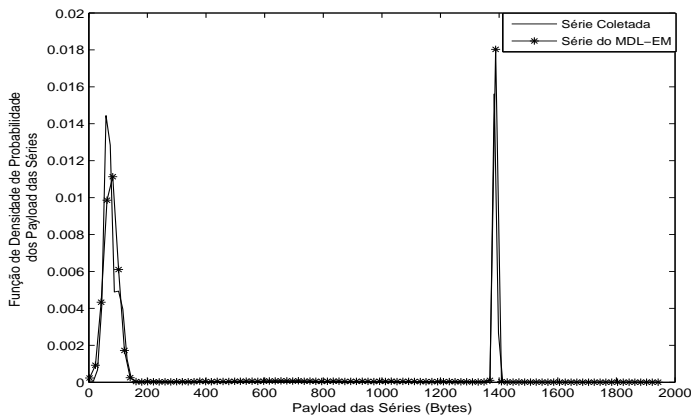
(b) pdf série coletada e sintética

Figura 5. Características da chamada VoIP 01 (abordagem MDL-EM)

para o tempo de execução do algoritmo. Verificou-se que a relação entre o tempo de execução do algoritmo e o tamanho da série (número de *payload*) é diretamente proporcional.



(a) Histograma série sintética



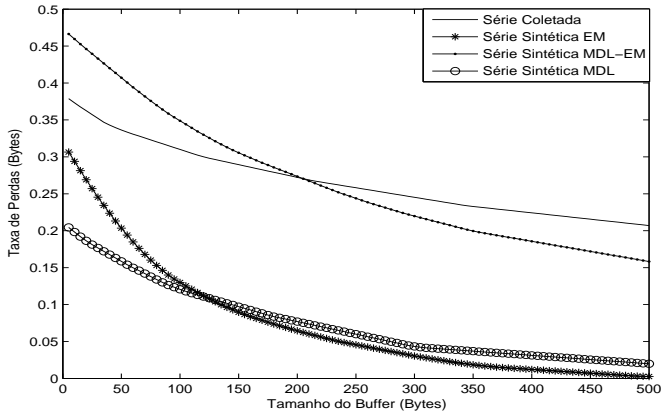
(b) pdf série coletada e sintética

Figura 6. Características da chamada VoIP 02 (abordagem MDL-EM)

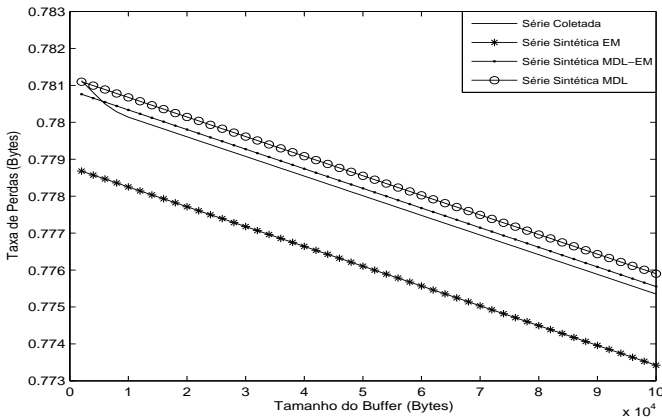
4.5 Taxa de Perda de Pacotes

Uma análise do comportamento de fila em um *buffer* também foi utilizada para comparar as séries sintéticas geradas com o da série VoIP original coletada. Para isso, considerou-se nas simulações, um enlace com um servidor e um *buffer* sendo alimentado pelas séries.

A relação entre a taxa de perda de pacotes e o tamanho do *buffer* para duas chamadas pode ser observada na Figura 7.



(a) Chamada 01



(b) Chamada 02

Figura 7. Relação entre taxa de perda de pacotes e o tamanho do *buffer*

Como mostra a Figura 7, as séries sintéticas geradas com os parâmetros calculados pelo algoritmo MDL-EM apresentaram o comportamento de fila bem próximo ao comportamento de fila das séries coletadas VoIP.

5 Conclusões

A abordagem MDL, para a chamada 01, apresentou o menor valor para o erro percentual da média e o menor número de gaussianas. Consequentemente, para a série chamada 01, o algoritmo MDL necessita de menor volume de informação a ser armazenado e de menor carga computacional para estimar os parâmetros da função de densidade. Para a chamada 02, obtivemos baixos valores para os erros percentuais com o método MDL, mas pode-se obter menores valores para os erros percentuais com outros métodos, como o EM por exemplo. O tempo de execução do algoritmo MDL é o maior se comparado com a execução do algoritmo EM depois de determinar o número de gaussianas ideal para gerar a série sintética. Com relação à taxa de perda de pacotes, o MDL apresenta valores aceitáveis em relação a série coletada VoIP.

A abordagem EM apresentou os menores valores para os erros percentuais da média para a chamada 02. O número de gaussianas, apresentado, também é menor que os apresentados pelo algoritmo MDL. Apesar dos valores apresentados, o método EM necessita de um processamento computacional maior para determinar o número correto de gaussianas que o método MDL. É necessário testar séries sintéticas geradas variando o número de gaussianas até encontrar a série sintética com menor erro percentual para média. O tempo de execução do algoritmo EM é o maior considerando o tempo para se determinar o número de gaussianas. Como o algoritmo MDL, o algoritmo EM, também apresentou valores para a taxa de perda de pacotes aceitáveis em relação à série coletada VoIP.

A abordagem MDL-EM apresentou melhores resultados do que a abordagem EM para a chamada 02. Para chamada 01, o mesmo ocorre, exceto que o algoritmo MDL consegue atingir um erro percentual inferior ao da abordagem MDL-EM. Avaliando os resultados obtidos, podemos dizer que as séries geradas utilizando os parâmetros calculados pelo algoritmo MDL-EM possui em geral características mais próximas das características das séries de VoIP coletadas. Esta semelhança com os dados reais também foi observada para a taxa de perda de pacotes.

Finalmente, conclui-se que é possível obter uma série sintética com características semelhantes das séries coletadas utilizando Misturas Gaussianas aplicando o algoritmo proposto. Como trabalho futuro, pode-se estender as Misturas Gaussianas de forma a atender outras características das séries de tráfego, não consideradas neste artigo, como por exemplo, seus momentos estatísticos.

Referências

- [1] C. A. Bouman, M. Shapiro, G. W. Cook, C. B. Atkins, H. Cheng, J.G. Dy, S. Borman, "CLUSTER: An unsupervised algorithm for modeling gaussian mixtures", School

of Electrical Engineering, Purdue University, West Lafayette, 2005.

- [2] B. Chandrasekaran, "Survey of network traffic models", Washington University in St. Louis CSE 567, 2009.
- [3] S. Colcher, A. T. A. Gomes, A. O. Silva, G. L. F. Souza e L. F. S. Soares, "VoIP - Voz Sobre IP", 4 ed., Editora Elsevier, Rio de Janeiro, 2005.
- [4] P. Grunwald, "A tutorial introduction to the minimum description length principle", Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands, 2004.
- [5] W. L. Martinez, A. R. Martinez, "Computational Statistics Handbook With MatLab", Chapman & Hall/CRC, Boca Raton, Florida, 2008.
- [6] G. J. McLachlan, T. Krishnan, "The EM algorithm and estensions", New York, John Wiley and Sons, 274p, 1997.
- [7] F. Picard, "An introduction to mixture models", Statistics for Systems Biology, Research Report No. 7, Paris, France, march 2007.
- [8] B. Resch, "Mixtures of gaussians", A Tutorial for the Course Computational Intelligence. Signal Processing and Speech Communication Laboratory. <http://www.igi.tugraz.at/lehre/CI/tutorials/MixtGaussian/MixtGaussian.html>, Infeldgasse 16c, last updated: 24. Jun 2010.
- [9] X., Zhuang; Y., Huang; K., Palaniappan; Y. Zhao, Gaussian mixture density modeling, decomposition, and applications, *Magazine IEEE TRANSACTIONS ON IMAGE PROCESSING*, Vol. 05, **No. 09**, 1996.