

Modelos analíticos em tempo discreto para a obtenção da probabilidade de perda de sistemas com diferenciação de serviço

João Carmo de Almeida Neto¹, Luís Felipe M. de Moraes¹

¹RAVEL - Laboratório de Redes de Alta Velocidade
Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro, Brazil

almeida@ravel.ufrj.br, Moraes@ravel.ufrj.br

Abstract. *This paper describes a discrete-time analytic model to obtain loss probability values in a diffserv mechanism called threshold dropping, taking into account a finite buffer size. The approach involves Markov chains to represent the system behaviour. The loss probability values are obtained through Bernoulli and MMBP (Markov Modulated Bernoulli Process) sources. The reliability of the model is tested, under different operation conditions by comparison with simulated results.*

Resumo. *Este artigo descreve um modelo analítico em tempo discreto para obter valores de probabilidade de perda de uma disciplina diffserv chamada threshold dropping, considerando uma fila finita. A abordagem envolve a utilização de Cadeias de Markov para representar o funcionamento do sistema. Os resultados de probabilidade de perda são obtidos considerando fontes de Bernoulli e MMBP (Markov Modulated Bernoulli Process). A confiabilidade do modelo é testada, sob diferentes condições de operação, por comparações com resultados simulados.*

1. Introdução

Com a motivação de fornecer diferentes qualidades de serviço aos tráfegos admitidos em uma rede de computadores, foram desenvolvidas diversas arquiteturas com o objetivo de suprir esta necessidade. Dentro deste contexto, surge o esquema chamado Diffserv [Blake 1998, Xiao 1999], que utiliza as disciplinas de filas dentro dos roteadores e comutadores para diferenciar o tratamento do tráfego de entrada. Neste caso, serão associadas prioridades diferentes aos pacotes que não possuem a mesma qualidade de serviço.

Da mesma forma que surgiram arquiteturas que provêem diferenciação de serviços, também aparece a necessidade de quantificar o desempenho destas arquiteturas. Esta tarefa pode ser realizada tanto por simulação numérica quanto por modelagem analítica. Ao estudar diversos artigos sobre Diffserv, como, por exemplo, [Aiello 2000, Alves 2000, Di 2000, Jia 2000], nota-se uma carência em termos de modelos analíticos capazes de fornecer valores de qualidade de serviço. Desta forma, o trabalho proposto está voltado a suprir parte desta deficiência. Em outras palavras, este

artigo realizará a modelagem analítica de uma arquitetura associada à diferenciação de serviços na Internet.

Em relação aos modelos propostos na literatura que avaliam as características de arquiteturas Diffserv, pode-se citar os trabalhos apresentados nas referências [May 1999, Nguyen 2000, Sahu 1999]. Em [May 1999, Sahu 1999], aborda-se os modelos de vários sistemas, considerando uma análise em tempo contínuo. Entre as disciplinas de fila encontradas nestas referências, estuda-se a arquitetura chamada *Threshold Dropping*, que é abordada neste artigo. Em [Nguyen 2000], analisa-se o caso em que mais de duas prioridades de descarte distintas são admitidas nesta disciplina de fila.

Devido às características da disciplina de fila abordada neste trabalho, é possível verificar que o método matemático utilizado pode ser direcionado para obter resultados de outras arquiteturas associadas à qualidade de serviço. Este fato ocorre, pois, os recursos da estrutura envolvida (tamanho do *buffer*, largura de banda, etc.) são divididos entre as prioridades de descarte existentes. Este funcionamento define uma das propostas encontradas em Diffserv: *Assured Forwarding* (AF). Desta forma, o trabalho apresentado representa um ponto de partida para diversos trabalhos futuros.

Em contraposição aos modelos geralmente encontrados na literatura, foi desenvolvida uma análise em tempo discreto. Em sistemas de transmissão contemporâneos com elevada taxa de transmissão, considera-se este tipo de modelagem analítica como um recurso útil e eficaz para a análise de desempenho. A partir da necessidade de novos métodos para caracterização de sistemas, procurou-se por uma abordagem original e com aplicações práticas.

2. Definições dos sistemas abordados:

Ao mesmo tempo que foram escolhidas variáveis aleatórias discretas, também determinou-se os tipos de fontes que representariam o tráfego de entrada. Desta forma, foram abordados dois tipos de chegada de unidades de informação: fontes que transmitem de acordo com uma variável aleatória de Bernoulli e fontes MMBP (“Markov Modulated Bernoulli Process” ou Processo de Bernoulli Modulado por Markov).

As fontes representadas por uma variável aleatória de Bernoulli geram, em um segmento de tempo, uma unidade de informação com probabilidade p ou não geram uma unidade de informação com probabilidade $1 - p$. No caso de fontes MMBP, tem-se um modelo muito similar ao da fonte “Liga-Desliga” ou “On-Off”. Neste caso, supondo que a fonte está no estado E_i , a chegada de uma unidade de informação em um segmento de tempo ocorre com probabilidade p_i . A representação deste modelo pode ser vista na figura 1. É importante mencionar que, enquanto as variáveis aleatórias de Bernoulli podem representar fontes de dados, as fontes MMBP estão associadas ao tráfego de vídeo.

A disciplina de fila considerada é a chamada *Threshold Dropping*. Neste sistema considera-se que, no máximo, B_H unidades de informação podem ser armazenadas e as unidades de informação de baixa prioridade podem ser aceitas juntamente com as de alta até um tamanho B_L , cujo valor é menor que B_H . Quando o número de unidades de informação armazenadas no sistema exceder B_L , somente o tráfego de alta prioridade é admitido enquanto houver espaço disponível. Após o armazenamento do tráfego admitido, as unidades de informação serão servidas de acordo com a ordem na fila.

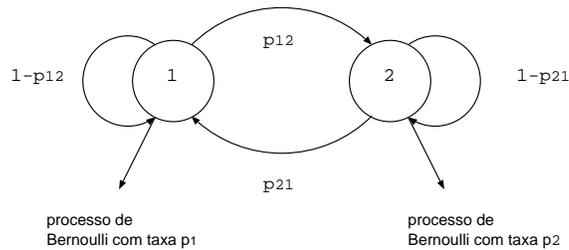


Figura 1: Representação do funcionamento de uma fonte pelo modelo MMBP

Através da Cadeia de Markov de cada modelo proposto, obtém-se a distribuição de probabilidade, em estado estacionário, do número de unidades de informação no sistema. A partir desta distribuição, diversas medidas de desempenho podem ser extraídas. Dentre as possibilidades, será enfatizado o valor conhecido como *probabilidade de perda* das unidades de informação, dado de extrema importância na área de qualidade de serviço.

3. Características dos Modelos Propostos

As unidades de informação armazenadas nos referidos sistemas serão, genericamente, denominadas usuários e pertencem à uma das prioridades de descarte existentes: alta e baixa. O tratamento de cada prioridade estará associado à disciplina de fila abordada.

De acordo com o que foi especificado no item 2., serão utilizadas fontes de Bernoulli e MMBP para representar o fluxo de entrada. No caso envolvendo fontes de Bernoulli, tem-se que os estados da Cadeia de Markov utilizada possuem somente uma variável de estado, ou seja, o número de usuários no sistema (número na fila + o que está sendo atendido). No caso de fontes MMBP considera-se, como variável de estado, não somente o número de usuários no sistema, mas também o número de fontes de alta ou baixa prioridade no estado E_1 (ver figura 1).

Dentro da disciplina de fila estudada, considera-se a chegada em grupo de usuários. Em outras palavras, duas ou mais unidades de informação, emitidas por fontes mutuamente independentes, podem chegar no sistema em um mesmo segmento de tempo. Desta forma, define-se que o sistema analisado armazena os usuários admitidos em segmentos de tempo distintos de acordo com a ordem de chegada e os que entram no mesmo segmento de tempo de forma aleatória.

Em relação à chegada e saída de usuários, considera-se que tanto a entrada quanto o término do serviço ocorrem no início do segmento de tempo, sendo que a saída de um usuário libera espaço para o armazenamento do tráfego de entrada. Este esquema é representado na figura 2. A partir desta figura, verifica-se as seguintes características:

- O instante em que o sistema é observado ocorre antes do momento em que uma unidade de informação está concluindo o seu atendimento;
- Caso o sistema possua i unidades de informação ($i \neq 0$) no momento em que é observado, o primeiro usuário (ou grupo de usuários) que entrar na fila, no próximo segmento de tempo, encontrará o sistema contendo $i - 1$ unidades de informação;
- Se o sistema estiver vazio quando for inicialmente observado e, em seguida, chegarem i usuários, serão encontradas i unidades de informação armazenadas quando o sistema for novamente observado no segmento de tempo consecutivo.

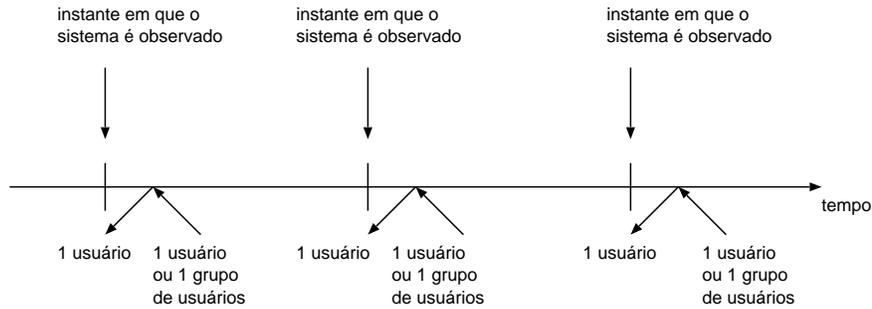


Figura 2: Esquema representando a entrada e saída de usuários

Através da equação $q = q\Pi$, onde q representa o estado do sistema e Π a matriz de transição de estados, obtém-se o vetor q em regime estacionário [Kleinrock 1975].

4. Sistema GEOM/D/1/ B_H com prioridades de descarte e chegadas em grupo

Nesta análise, a emissão de usuários ocorre a partir de fontes de Bernoulli. Desta forma, os valores p_a e p_b representam as probabilidades das fontes de alta e baixa prioridade transmitirem uma unidade de informação em um segmento de tempo. Dado que o tempo de serviço é determinístico, considera-se que o período de atendimento das unidades de informação é igual à um segmento de tempo.

4.1. Definições dos parâmetros utilizados:

Para obter os resultados desejados, foi preciso a utilização de vários parâmetros. Estes dados são necessários para determinar as características do sistema e facilitar a compreensão do modelo. A partir deste conceito, tem-se os seguintes valores:

n_a = número de fontes de alta prioridade;

n_b = número de fontes de baixa prioridade;

X = variável aleatória que representa o número de usuários no sistema em estado estacionário (incluindo o que está em serviço);

$q_i = P(X = i)$ = probabilidade de haver i usuários no sistema.

Através destes parâmetros podem ser definidos outros tipos de equações relacionadas ao funcionamento das fontes. Desta forma, dado que a_z^b e a_y^a representam as probabilidades de chegarem z e y usuários de baixa e alta prioridade no sistema, respectivamente, tem-se as seguintes distribuições Binomiais: $a_z^b = \binom{n_b}{z} p_b^z (1 - p_b)^{n_b - z}$ e $a_y^a = \binom{n_a}{y} p_a^y (1 - p_a)^{n_a - y}$.

A partir destas expressões, define-se que EAr^a e EAr^b representam os respectivos números médios de usuários de alta e baixa prioridade que chegam ao sistema.

4.2. Modelagem Analítica:

Nesta subseção, desenvolve-se a Cadeia de Markov que representa o funcionamento do sistema estudado. Uma vez esta estrutura esteja definida, é possível obter o valor procurado para probabilidade de perda.

4.2.1. Definição da Cadeia de Markov:

A fim de determinar a Cadeia de Markov, as expressões que caracterizam as transições de estado nesta estrutura devem ser especificadas. Assim como foi mencionado anteriormente, somente será utilizada uma variável de estado, representando o número de usuários armazenados no sistema. Desta forma, a matriz de transição de estados possui $(B_H + 1) \times (B_H + 1)$ elementos.

Considera-se π_{ij} como a probabilidade de haver a transição da variável de estado do valor i para j . A fim de definir as expressões utilizadas para obter π_{ij} , analisa-se quatro casos distintos:

- Os estados i e j são menores que B_L ;
- Os estados i e j são maiores que B_L e menores que B_H ;
- O estado i possui qualquer valor menor ou igual a B_H e j é igual a B_H ;
- O estado i é menor ou igual a B_L mas o estado j é maior ou igual a B_L e menor que B_H .

No primeiro caso, aplica-se diretamente as equações a_z^b e a_y^a para obter a probabilidade π_{ij} . Este fato ocorre devido à participação de ambas as prioridades de descarte na transição de estados. Desta forma, tem-se a seguinte expressão para i e j menores que B_L :

$$\pi_{ij} = \sum_{y=0}^r \binom{n_a}{y} p_a^y (1 - p_a)^{n_a - y} \binom{n_b}{r - y} p_b^{r - y} (1 - p_b)^{n_b - r + y} \quad (1)$$

onde,

$$\begin{aligned} \text{para } j \geq i \text{ e } i > 0 &\rightarrow r = j - i + 1; \\ \text{para } j \geq i \text{ e } i = 0 &\rightarrow r = j - i = j; \end{aligned}$$

No segundo tipo de expressão é encontrada uma similaridade à primeira. Neste caso, como somente os usuários de alta prioridade serão aceitos, considera-se que a mudança de estado ocorre somente devido à chegada deste tipo de tráfego. Portanto, a seguinte expressão é obtida:

$$\pi_{ij} = \binom{n_a}{y} p_a^y (1 - p_a)^{n_a - y} \quad (2)$$

onde $y = j - i + 1$.

A partir de 1 e 2 é possível verificar que a transição do estado i para $i - 1$ está sendo considerada. Isto pode ser observado se o valor de r for substituído por zero. Neste caso, as expressões de π_{ii-1} obtidas para $i \leq B_L$ e $i > B_L$ são, respectivamente, $(1 - p_a)^{n_a} (1 - p_b)^{n_b}$ e $(1 - p_a)^{n_a}$.

Para desenvolver o terceiro tipo de expressão utiliza-se uma abordagem mais simples, ou seja, baseada nas probabilidades de transição de um determinado estado i para todos os estados com menos de B_H usuários armazenados. Através deste método, encontra-se:

$$\pi_{iB_H} = 1 - \sum_{j=0}^{B_H - 1} \pi_{ij} \quad (3)$$

No quarto caso ocorre uma situação mais complexa. Em outras palavras, considera-se a influência da perda dos usuários de baixa prioridade. Este fato ocorre, pois, quando o tamanho do sistema ultrapassa B_L todo o tráfego de baixa prioridade gerado é descartado. Desta forma, encontra-se a seguinte expressão para $i > 0$:

$$\pi_{ij} = \sum_{r=j-i+1}^{n_b+j-i+1} \sum_{y=j-BL}^{j-i+1} \binom{n_a}{y} p_a^y (1-p_a)^{n_a-y} \binom{n_b}{r-y} p_b^{r-y} (1-p_b)^{n_b-r+y} c_1(r, y, j, i) \quad (4)$$

É importante observar que, nesta equação, as quantidades de usuários de alta e baixa prioridade emitidas pelas fontes são y e $r - y$, respectivamente.

A função $c_1(r, y, j, i)$ representa uma probabilidade diretamente ligada ao número de usuários admitidos no sistema e a distribuição deste número entre as prioridades existentes, ou seja, alta e baixa. Para cada caso envolvendo uma quantidade de unidades de informação aceitas e as respectivas parcelas de usuários de alta e baixa prioridade, existe um valor distinto para $c_1(r, y, j, i)$.

Como a transição de estado a partir de um sistema vazio não ocorre juntamente com o término do serviço de um usuário, as probabilidades π_{ij} devem ser recalculadas para este caso, ou seja, quando $i = 0$. Desta forma, o valor de $c_1(r, y, j, i)$ pode ser obtido de dois algoritmos: o primeiro considerando um sistema vazio antes da entrada de tráfego e o segundo quando o sistema está ocupado antes da chegada de usuários ($i > 0$). O algoritmo para obter o valor de $c_1(r, y, j, i)$, quando $i > 0$, pode ser visto abaixo.

- Caso $r = y = j - i + 1$

$$\implies c_1(r, y, j, i) = 1 \quad (5)$$

- Caso $r > j - i + 1$ e $j - i - y + 1 > 0$

$$\implies c_1(r, y, j, i) = \binom{r-y}{r-j+i-1} \prod_{t=0}^{r-j+i-2} \left(\frac{r-B_L+i-1-t}{r-t} \right) \prod_{t=0}^{j-i-y} \left(\frac{B_L-i+1-t}{j-i+1-t} \right) \quad (6)$$

- Caso $r > j - i + 1$ e $j - i - y + 1 = 0$

$$\implies c_1(r, y, j, i) = \prod_{t=0}^{r-j+i-2} \left(\frac{r-B_L+i-1-t}{r-t} \right) \quad (7)$$

- Caso $y \neq r = j - i + 1$

$$\implies c_1(r, y, j, i) = \prod_{t=0}^{j-i-y} \left(\frac{B_L-i+1-t}{j-i+1-t} \right) \quad (8)$$

Na expressão 5 ocorre somente a entrada de tráfego de alta prioridade. Portanto, como estas mensagens são aceitas no sistema, encontra-se $c_1(r, y, j, i) = 1$.

No caso da equação 6, vê-se que a mudança do número de usuários no sistema ($i - 1 \rightarrow j$) é menor que a quantidade de usuários produzidos pelas fontes (parâmetro r) e maior que o número de unidades de informação de alta prioridade aceitas. Desta

forma, uma parte do tráfego de baixa prioridade, representada pelo valor $r - j + i - 1$, é descartada e o restante é admitida. A combinação $\binom{r-y}{r-j+i-1}$ referencia o número de possíveis grupos de baixa prioridade distintos a serem perdidos. Em seguida, obtém-se as probabilidades dos usuários deste grupo serem perdidos e dos usuários restantes serem aceitos. Estes valores são representados pelos produtórios $\prod_{t=0}^{r-j+i-2} \left(\frac{r-B_L+i-1-t}{r-t} \right)$ e $\prod_{t=0}^{j-i-y} \left(\frac{B_L-i+1-t}{j-i+1-t} \right)$, respectivamente.

Na expressão 7, encontra-se uma situação onde todos os usuários de baixa prioridade são perdidos. Desta forma, a transição da variável de estado ocorre somente pela entrada do tráfego de alta prioridade. A probabilidade de ocorrer este caso está relacionada à probabilidade de perda de todos os usuários de baixa prioridade. Através desta conclusão, verifica-se que a expressão 7 é obtida a partir do componente, da expressão 6, relacionado com a perda do tráfego de baixa prioridade.

No caso da equação 8, analisa-se uma situação onde todos os tipos de usuários são aceitos. Portanto, o desenvolvimento desta expressão restringe-se à obtenção da probabilidade de todo o tráfego de baixa prioridade ser admitido. Como pode ser observado, a equação 8 é idêntica ao componente da expressão 6 referente à admissão dos usuários de baixa prioridade.

A versão da equação 4 para o caso em que $i = 0$ pode ser obtida se for desconsiderado o serviço de uma unidade de informação no início do segmento de tempo. Neste caso, o tráfego de entrada (parâmetro r) varia do valor j até $n_b + j$ e a quantidade de usuários de alta prioridade emitidos (parâmetro y) está entre os valores $j - B_L$ e j . Esta mesma característica também ocorre para a função $c_1(r, y, j, i)$.

5. Sistema MMBP/D/1/ B_H com prioridades de descarte e chegadas em grupo

Neste sistema, é analisado um caso análogo ao especificado na seção 4., mas aplicando fontes com características diferentes. Este fato ocorre, pois, é considerado o funcionamento das fontes MMBP (figura 1) e o seu impacto no sistema. Desta forma, define-se p_1^a e p_2^a como as probabilidades das fontes de alta prioridade produzirem usuários nos estados E_1 e E_2 , respectivamente. Analogamente, no caso das fontes de baixa prioridade, também utiliza-se p_1^b e p_2^b como probabilidades de produzir usuários nestes estados. Portanto, além dos parâmetros apresentados na subseção 4.1., considera-se os seguintes valores:

- n_a^1 = número de fontes de alta prioridade no estado E_1 ;
- n_b^1 = número de fontes de baixa prioridade no estado E_1 ;
- p_{12} = probabilidade da fonte MMBP passar do estado E_1 para o estado E_2 ;
- p_{21} = probabilidade da fonte MMBP passar do estado E_2 para o estado E_1 ;

A partir destes dados podem ser definidas as expressões para $a_z^{n_b^1}$ e $a_y^{n_a^1}$. Estes valores estão associados às probabilidades de, em um segmento de tempo, chegarem z e y usuários de baixa e alta prioridade dado que existem n_b^1 e n_a^1 fontes de baixa e alta prioridade no estado E_1 , respectivamente.

Dado que $F_1^a(t)$ e $F_1^b(t)$ representam as respectivas probabilidades de t fontes de alta e baixa prioridade estarem no estado E_1 , desenvolve-se, a partir de $a_z^{n_b^1}$ e $a_y^{n_a^1}$, as seguintes expressões para os valores a_z^b e a_y^a , apresentados na subseção 4.1.:

$$a_z^b = \sum_{t=0}^{n_b} \sum_{k=0}^z \binom{t}{k} (p_1^b)^k (1 - p_1^b)^{t-k} \binom{n_b - t}{z - k} (p_2^b)^{z-k} (1 - p_2^b)^{n_b - t - z + k} F_1^b(t) \quad (9)$$

$$a_y^a = \sum_{t=0}^{n_a} \sum_{k=0}^y \binom{t}{k} (p_1^a)^k (1 - p_1^a)^{t-k} \binom{n_a - t}{y - k} (p_2^a)^{y-k} (1 - p_2^a)^{n_a - t - y + k} F_1^a(t) \quad (10)$$

5.1. Modelagem Analítica:

Assim como foi mencionado na seção 3., a descrição do sistema deve considerar três variáveis de estado, ou seja, o número de unidades de informação no sistema e as quantidades de fontes de alta e baixa prioridade no estado E_1 . Desta forma, surge a necessidade de rever o esquema representando a entrada e saída de usuários, incluindo a transição de estado das fontes MMBP. A partir deste conceito, obtém-se a figura 3.

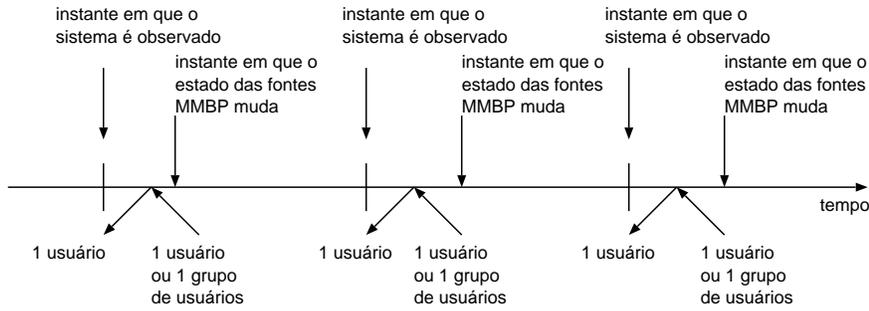


Figura 3: Esquema representando a entrada e saída de usuários ao utilizar uma fonte MMBP

O tráfego relativo a cada prioridade de descarte é obtido através de uma composição dos estados que as fontes podem assumir. Desta forma, define-se Q_{mn} como a probabilidade, genérica, de ocorrer uma transição do número de fontes no estado E_1 do valor m para n . Sendo f o número total de fontes envolvidas, a seguinte expressão é encontrada:

$$Q_{mn} = \begin{cases} \sum_{l=m-n}^{\min(m, f-n)} T_m(n - m + l) S_m(l) & \text{se } m \geq n \\ \sum_{k=n-m}^{\min(n, f-m)} T_m(k) S_m(m - n + k) & \text{se } m < n \end{cases} \quad (11)$$

Onde, $\min(x, y)$ representa o menor valor entre x e y . Na equação 11, $T_m(k)$ (ou $S_m(l)$) representa a probabilidade de k (ou l) fontes passarem do estado E_2 para o estado E_1 (ou vice versa), dado que m fontes estão no estado E_1 .

Portanto, encontra-se as equações: $T_m(k) = \binom{f-m}{k} T_{21}^k (1 - T_{21})^{f-m-k}$ e $S_m(l) = \binom{m}{l} T_{12}^l (1 - T_{12})^{m-l}$.

5.1.1. Definição da Cadeia de Markov:

Como são utilizadas três variáveis de estado, define-se q como o vetor linha que possui $(B_H + 1) \times (n_a + 1) \times (n_b + 1)$ componentes e representa o estado estacionário do

sistema. Em outras palavras, este sistema é constituído de $(B_H + 1)$ blocos, cada um com $(n_a + 1) \times (n_b + 1)$ elementos. A partir destas características, utiliza-se a relação $q_{ijk} = P(X = i, n_a^1 = j, n_b^1 = k)$.

Para obter os valores de todos os elementos da matriz de transição de estados, utiliza-se a seguinte expressão:

$$\Pi_{cd}^{ij} = \pi_{ij}^{hc} Q_{hk}^a Q_{cd}^b \quad (12)$$

Onde,

Π_{cd}^{ij} = probabilidade das variáveis de estado relacionadas ao número de unidades de informação armazenadas e às quantidades de fontes de alta e baixa prioridade no estado E_1 terem os seus valores alterados de i para j , de h para k e de c para d , respectivamente;

π_{ij}^{hc} = probabilidade de haver uma alteração de i para j na variável de estado referente ao número de usuários no sistema, dado que h e c fontes de alta e baixa prioridade estejam no período E_1 , respectivamente.

É possível observar que as probabilidades Q_{hk}^a e Q_{cd}^b são obtidas da expressão 11.

A fim de extrair as expressões que especificam π_{ij}^{hc} , aborda-se os mesmos casos encontrados na seção 4., mas considerando que o número de fontes de alta e baixa prioridade no estado E_1 é bem definido. Em outras palavras, são analisados os seguintes casos:

1. Os estados i e j são menores que B_L ;
2. Os estados i e j são maiores que B_L e menores que B_H ;
3. O estado i possui qualquer valor menor ou igual a B_H e j é igual a B_H ;
4. O estado i é menor ou igual a B_L mas o estado j é maior ou igual a B_L e menor que B_H .

Através de uma analogia com a seção 4., encontra-se expressões análogas às desenvolvidas para o sistema GEOM/D/1/ B_H .

6. Definição das expressões para probabilidade de perda dos usuários de alta e baixa prioridade:

Nesta seção, os valores referentes à probabilidade de perda dos tráfegos envolvidos serão caracterizados a partir das fontes e da Cadeia de Markov. Neste ponto, utiliza-se a idéia envolvendo a probabilidade de uma unidade de informação teste ser perdida ao tentar entrar no sistema [Murata 1990]. Inicialmente, será desenvolvida a expressão referente aos usuários de baixa prioridade e, em seguida, aos de alta prioridade.

Os fatores a serem considerados no caso dos usuários de baixa prioridade são: as quantidades de usuários de alta e baixa prioridade geradas e a probabilidade do usuário teste ser perdido dado que o tamanho inicial da fila (i) e a quantidade de tráfego de entrada são definidos. Enquanto no primeiro caso utiliza-se os componentes a_y^a e a_z^b , no segundo considera-se os termos $\frac{z+y-B_L}{z+y}$ e $\frac{i+z+y-B_L-1}{z+y}$, associados às situações em que o sistema está inicialmente vazio ($i = 0$) e ocupado ($i > 0$), respectivamente.

Através de uma análise, encontra-se o seguinte algoritmo para a probabilidade de perda dos usuários de baixa prioridade:

- Se $B_L - (n_a + n_b) \geq 0$:

$$P_{perda_{bp}} = \sum_{i=B_L-(n_a+n_b)+2}^{B_L} \sum_{y=\max(0, B_L-i-n_b+2)}^{n_a} \sum_{z=\max(1, B_L-i-y+2)}^{n_b} [X(i) a_y^a \frac{z a_z^b}{E A r^b} \left(\frac{i+z+y-B_L-1}{z+y} \right)] + \sum_{i=B_L+1}^{B_H} X(i) \quad (13)$$

- Se $B_L - (n_a + n_b) < 0$:

$$P_{perda_{bp}} = \sum_{i=1}^{B_L} \sum_{y=\max(0, B_L-i-n_b+2)}^{n_a} \sum_{z=\max(1, B_L-i-y+2)}^{n_b} [X(i) a_y^a \frac{z a_z^b}{E A r^b} \left(\frac{i+z+y-B_L-1}{z+y} \right)] + \sum_{i=B_L+1}^{B_H} X(i) + \sum_{y=\max(0, B_L-n_b+1)}^{n_a} \sum_{z=\max(1, B_L-y+1)}^{n_b} \left[X(0) \left(\frac{z+y-B_L}{z+y} \right) a_y^a \frac{z a_z^b}{E A r^b} \right] \quad (14)$$

Nestas equações, o termo $\frac{z a_z^b}{E A r^b}$ representa a probabilidade do usuário teste estar incluído em um grupo de tamanho z ao chegar [Murata 1990].

Para obter a perda dos usuários de alta prioridade, analisa-se dois casos distintos: o primeiro relacionado a um número de unidades de informação no sistema menor ou igual a B_L ($i \leq B_L$) e o segundo com este valor maior que B_L ($i > B_L$).

Ao considerar a variável i menor ou igual a B_L verifica-se a possibilidade dos tráfegos de alta e baixa prioridade aceitarem proporcionar a perda do usuário teste de alta prioridade. A partir deste conceito, as funções $c_2(k, z, y, i)$ e $c_3(k, z, y, i)$, definidas de forma análoga à $c_1(r, y, j, i)$ (subseção 4.2.1.), são utilizadas para obter a probabilidade de k unidades de informação de baixa prioridade serem admitidas a partir do tráfego produzido pelas fontes. Neste caso, enquanto $c_2(k, z, y, i)$ considera o término do atendimento de um usuário pelo servidor antes da chegada de tráfego, o termo $c_3(k, z, y, i)$ não permite que esta característica ocorra.

De forma análoga às equações 13 e 14, a condição $B_H - (n_a + n_b) < 0$ determina a possibilidade de um usuário de alta prioridade ser perdido quando o sistema está inicialmente vazio ($i = 0$). A partir destes dados, tem-se o seguinte algoritmo para $i \leq B_L$:

- Se $B_H - (n_a + n_b) \geq 0$:

$$P_{perda_{ap}(i \leq B_L)} = \sum_{i=B_H-(n_a+n_b)+2}^{B_L} \sum_{y=B_H-B_L+1}^{n_a} \sum_{k=\max(0, B_H-i-y+2)}^{\min(n_b, B_L-i+1)} \left[X(i) \frac{y a_y^a}{E A r^a} \left(\frac{i+k+y-B_H-1}{y} \right) \sum_{z=k}^{n_b} a_z^b c_2(k, z, y, i) \right] \quad (15)$$

- Se $B_H - (n_a + n_b) < 0$:

$$P_{perda_{ap}(i \leq B_L)} = \sum_{i=1}^{B_L} \sum_{y=B_H-B_L+1}^{n_a} \sum_{k=\max(0, B_H-i-y+2)}^{\min(n_b, B_L-i+1)} \left[X(i) \frac{y a_y^a}{E A r^a} \right]$$

$$\begin{aligned}
& \left(\frac{i+k+y-B_H-1}{y} \right) \sum_{z=k}^{n_b} a_z^b c_2(k, z, y, i) \Big] \\
& + \sum_{y=B_H-B_L+1}^{n_a} \sum_{k=\max(0, B_H-y+1)}^{\min(n_b, B_L)} \left[X(0) \frac{y a_y^a}{E A r^a} \right. \\
& \left. \left(\frac{k+y-B_H}{y} \right) \sum_{z=k}^{n_b} a_z^b c_3(k, z, y, 0) \right] \tag{16}
\end{aligned}$$

Ao considerar a variável i maior que B_L , observa-se uma situação mais simples, onde existe a admissão somente do tráfego de alta prioridade. Neste caso, obtém-se:

$$P_{perda_{ap}(i > B_L)} = \sum_{i=\max(B_L+1, B_H-n_a+2)}^{B_H} \sum_{y=B_H-i+2}^{n_a} \frac{y a_y^a}{E A r^a} X(i) \left(\frac{i+y-B_H-1}{y} \right) \tag{17}$$

Através das equações 15, 16 e 17, encontra-se o seguinte algoritmo:

- Caso $n_a < B_H - B_L + 1$:

$$P_{perda_{ap}} = \sum_{i=\max(B_L+1, B_H-n_a+2)}^{B_H} \sum_{y=(B_H-i+2)}^{n_a} X(i) \frac{y a_y^a}{E A r^a} \left(\frac{i+y-B_H-1}{y} \right) \tag{18}$$

- Caso $n_a \geq B_H - B_L + 1$:

Se $B_H - (n_a + n_b) \geq 0$:

$$\begin{aligned}
P_{perda_{ap}} = & \sum_{i=B_H-(n_a+n_b)+2}^{B_L} \sum_{y=B_H-B_L+1}^{n_a} \sum_{k=\max(0, B_H-i-y+2)}^{\min(n_b, B_L-i+1)} \left[X(i) \frac{y a_y^a}{E A r^a} \right. \\
& \left. \left(\frac{i+k+y-B_H-1}{y} \right) \sum_{z=k}^{n_b} a_z^b c_2(k, z, y, i) \right] + \\
& \sum_{i=B_L+1}^{B_H} \sum_{y=B_H-i+2}^{n_a} \left[X(i) \frac{y a_y^a}{E A r^a} \left(\frac{i+y-B_H-1}{y} \right) \right] \tag{19}
\end{aligned}$$

Se $B_H - (n_a + n_b) < 0$:

$$\begin{aligned}
P_{perda_{ap}} = & \sum_{i=1}^{B_L} \sum_{y=B_H-B_L+1}^{n_a} \sum_{k=\max(0, B_H-i-y+2)}^{\min(n_b, B_L-i+1)} \left[X(i) \frac{y a_y^a}{E A r^a} \right. \\
& \left. \left(\frac{i+k+y-B_H-1}{y} \right) \sum_{z=k}^{n_b} a_z^b c_2(k, z, y, i) \right] + \\
& \sum_{y=B_H-B_L+1}^{n_a} \sum_{k=\max(0, B_H-y+1)}^{\min(n_b, B_L)} \left[X(0) \frac{y a_y^a}{E A r^a} \left(\frac{k+y-B_H}{y} \right) \right. \\
& \left. \sum_{z=k}^{n_b} a_z^b c_3(k, z, y, 0) \right] + \sum_{i=B_L+1}^{B_H} \sum_{y=B_H-i+2}^{n_a} \left[X(i) \frac{y a_y^a}{E A r^a} \right. \\
& \left. \left(\frac{i+y-B_H-1}{y} \right) \right] \tag{20}
\end{aligned}$$

7. Resultados obtidos:

Neste ponto, os valores relacionados à probabilidade de perda ($P_{perda_{ap}}$ e $P_{perda_{bp}}$) são obtidos do modelo analítico e comparados com simulações. A fim de descrever o sistema, utiliza-se os seguintes parâmetros: taxa média de chegada dos usuários de alta e baixa prioridade (EAr^a e EAr^b), número de fontes de alta e baixa prioridade (n_a e n_b), B_H e B_L . No caso das fontes MMBP, serão incorporados os parâmetros p_{12} e p_{21} (ver figura 1).

A fim de realizar simulações, utilizou-se um programa desenvolvido em C++ que representa o funcionamento do sistema. Na obtenção dos valores simulados, considerou-se um intervalo de confiança de 95% e um erro percentual menor ou igual a 2.5%. Os resultados podem ser vistos nas figuras 4, 5, 6 e 7.

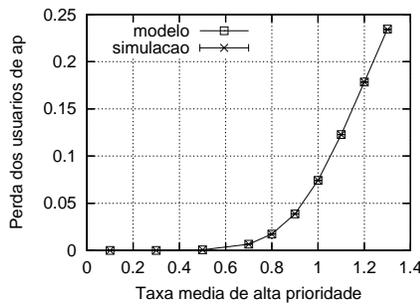


Figura 4: Simulação e modelo para $P_{perda_{ap}}$ em um sistema $GEOM/D/1/B_H$ com $EAr^b = 1.0$, $n_a = 2$, $n_b = 2$, $B_L = 5$ e $B_H = 8$

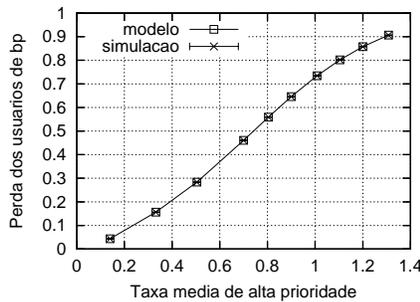


Figura 5: Simulação e modelo para $P_{perda_{bp}}$ em um sistema $MMBP/D/1/B_H$ com $EAr^b = 0.504$, $n_a = 2$, $n_b = 2$, $p_{12} = 0.6$, $p_{21} = 0.4$, $B_L = 2$ e $B_H = 3$

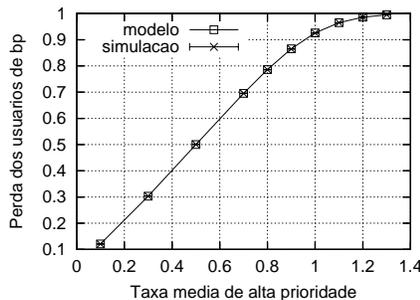


Figura 6: Simulação e modelo para $P_{perda_{bp}}$ em um sistema $GEOM/D/1/B_H$ com $EAr^b = 1.0$, $n_a = 2$, $n_b = 2$, $B_L = 5$ e $B_H = 8$

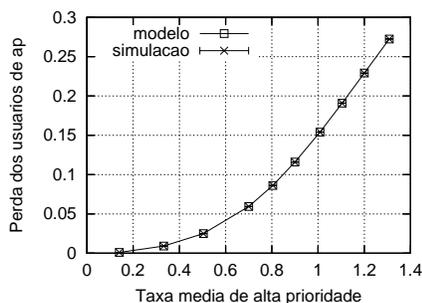


Figura 7: Simulação e modelo para $P_{perda_{ap}}$ em um sistema $MMBP/D/1/B_H$ com $EAr^b = 0.504$, $n_a = 2$, $n_b = 2$, $p_{12} = 0.6$, $p_{21} = 0.4$, $B_L = 2$ e $B_H = 3$

Como é possível observar, as figuras 4 e 6 apresentam características similares às apresentadas pelas figuras 5 e 7. Em outras palavras, os valores de probabilidade de perda para os usuários de alta prioridade apresentam-se muito inferiores aos valores obtidos para os usuários de baixa prioridade. Através de um estudo na disciplina de fila *Threshold Dropping*, verifica-se que esta característica é esperada. Portanto, este fato ajuda a validar o modelo desenvolvido.

8. Conclusões

Ao avaliar os gráficos apresentados, conclui-se que a validade do modelo desenvolvido foi confirmada devido à grande proximidade entre os valores teóricos e simulados. Desta forma, este artigo apresenta um novo recurso para estudar e projetar arquiteturas voltadas ao fornecimento de qualidade de serviço na Internet.

Apesar de ter sido abordado o caso envolvendo um valor determinístico para o tempo de serviço, a obtenção de resultados associados à uma distribuição geométrica é direta. Este fato pode ser comprovado se o modelo analítico for adaptado para considerar que o término do serviço de um usuário após um segmento de tempo ocorre com probabilidade β e não ocorre com probabilidade $1 - \beta$. Esta abordagem foi desenvolvida pelo autor deste artigo em [Almeida 2003].

Devido à utilização de um tempo determinístico para atender as unidades de informação, o modelo analítico apresentado não está diretamente associado a casos envolvendo uma arquitetura Diffserv, que não considera este tipo de tempo de serviço. Apesar desta característica, o trabalho desenvolvido pode ser utilizado em sistemas onde existe a transmissão de dados em alta velocidade e a utilização de qualidade de serviço. Neste caso, este tipo de modelagem proporciona uma boa aproximação da realidade. Um tipo de sistema com este tipo de funcionamento pode ser encontrado em [Callegati 2002].

É importante mencionar que, apesar deste artigo apresentar uma nova abordagem para o estudo de filas com diferenciação de serviço, os modelos apresentados possuem limitações. Durante a obtenção dos valores de probabilidade de perda, foi possível verificar que o método desenvolvido consome muitos recursos computacionais. Este fato foi comprovado quando, durante os testes com o modelo, houve um acréscimo na quantidade de fontes MMBP e um aumento no tamanho da fila (B_H). Desta forma, a obtenção de resultados analíticos foram restritos a um tamanho de oito pacotes para B_H e a uma quantidade de duas fontes MMBP.

Referências

- Aiello, W. A., Mansour, Y., Rajagopalan, S., et al. (2000), "Competitive Queue Policies for Differentiated Services" *Proceedings of the IEEE INFOCOM*, pp. 431-440, Tel-Aviv, Israel, March 2000
- Almeida Neto, J. C. (2003), *Modelos analíticos em tempo discreto para a análise de sistemas com diferenciação de serviço*. Tese de M.Sc., COPPE(UFRJ), Rio de Janeiro, RJ, Brasil.
- Alves, I. B. H. A., Rezende, J. F., Moraes, L. F. M., "Avaliando a Justiça na Marcação de Tráegos Agregados" (2000), *XVIII Simpósio Brasileiro de Redes de Computadores SBRC2000*, pp.119-134, Belo Horizonte, May 2000.
- Blake, S., Black, D., Carlson, M., et al., *An Architecture for Differentiated Services*, RFC - 2475, December 1998
- Borella, A., Chiaraluce, F., Meschini, F. (1996), "Statistical multiplexing of random processes in packet switching networks", *IEE Proc.- Commun.* Vol 143, No. 5, October 1996
- Callegati, F., Corazza, G., Raffaelli, C. (2002), "Exploitation of DWDM for Optical Packet Switching With Quality of Service Guarantees". *IEEE JSAC*, Vol.20, No.1, January 2002
- Di, Z., Mouftah, H. T. (2000), "Performance Evaluation of Per-Hop Forwarding Behaviours in the Diffserv Internet", *5th IEEE Symposium on Computers and Communications (ISCC 2000)*, pp. 334-339, France, July 2000.
- Jia, D., Dutkiewicz, E., Chicharo, J. F. (2000), "Performance Analysis of QoS Mechanisms in IP Networks", *5th IEEE Symposium on Computers and Communications (ISCC 2000)*, pp. 359-363, France, July 2000.
- Kleinrock, L. (1975) "Queueing Systems". Vol.1, 1 ed., New York, JohnWiley & Sons
- May, M., Bolot, J-C., Diot, C. e Jean-Marie, A. (1999), "Simple Performance Models of Differentiated Services Schemes for the Internet". *IEEE INFOCOM 1999*
- Murata, M., Oie, Y., Suda, T., Miyahara, H. (1990), "Analysis of a Discrete-Time Single-Server Queue with Bursty Inputs for Traffic Control in ATM Networks". *IEEE JSAC*, Vol.8, No.3, Abril 1990
- Nguyen, L. V., Eysers, T. e Chicharo, J. F. (2000), "Differentiated service performance analysis", *5th IEEE Symposium on Computers and Communications (ISCC 2000)*, pp. 328-333, France, July 2000.
- Ribeiro, M. R. N., O'Mahony, M. (2000), "On the Performance Benefits of WDM and Wavelength Conversion in Photonic Packet Switching Nodes". *Revista da Sociedade Brasileira de Telecomunicações*, Vol.15, No.1, Junho 2000
- Ross, S.M. (1972) "Introduction to Probability Models", Academic Press
- Sahu, S., Townsley, D., Kurose, J. (1999), "A Quantitative Study of Differentiated Services for the Internet", *Globecom 1999*.
- Xiao, X., Ni, L. M., "Internet QoS: A Big Picture", *IEEE Network*, Vol. 13, No 3, March/April 1999.