Análise de Mecanismos de CAC para Redes ATM*

Michel Ferreira Abdalla
Department of Computer Science and Engineering
University of California, San Diego – EUA
9500 Gilman Dr. – La Jolla, CA 92093
E. mail: mabdalla@cs.ucsd.edu

Otto Carlos M. B. Duarte
Universidade Federal do Rio de Janeiro
COPPE/PEE – Brasil
C.P. 68504 – 21945-970 Rio de Janeiro, RJ
E. mail: otto@gta.ufrj.br

Resumo

Os recursos dos atuais sistemas de comunicação multimídias de alta velocidade são compartilhados nos seus nós de acesso e de intercomunicação. Para que este compartilhamento ocorra de forma ordenada, de modo a otimizar a utilização dos recursos e atender aos requisitos de qualidade de serviço, é necessário um controle de admissão eficaz sobre o tráfego submetido a este sistema. Este artigo faz um estudo sobre os mecanismos de controle de admissão de conexão em redes ATM. Uma classificação destes mecanismos baseada no princípio implícito neles utilizado é também proposta. Através de resultados obtidos analiticamente e por simulação, alguns destes mecanismos são analisados e comparados. Os resultados obtidos demonstram que, na maioria dos casos que se espera encontrar em redes ATM, estes mecanismos apresentam resultados bastante superestimados para a capacidade equivalente do agregado das conexões, subestimando o ganho potencial de multiplexação estatística destas redes.

1 Introdução

As Redes Digitais de Serviços Integrados de Faixa Larga (RDSI-FL) foram desenvolvidas para oferecer toda a infra-estrutura necessária para suportar uma grande variedade de classes de serviços compostos dos mais diversos tipos de tráfego com diferentes requisitos de qualidade de serviço [9]. Tais redes devem, portanto, atender de forma eficiente à qualquer tipo de aplicação, ou seja, as que exigem taxas baixas, altas, fixas, variáveis e de tempo real. Por satisfazer tais requisitos e permitir alcançar uma alta taxa de utilização em altas velocidades de transmissão, o Modo de Transferência Assíncrono (ATM - *Asynchronous Transfer Mode*) foi proposto como a tecnologia a ser utilizada nas RDSI-FL.

Uma das principais vantagens na utilização das redes ATM está na economia de recursos que pode ser obtida através da multiplexação estatística de suas chamadas, pois, com isto, uma banda passante menor que a taxa máxima de transmissão pode ser alocada para cada fonte e um número maior de conexões pode ser aceito. Além da rede ser eficiente, ela deve atender às expectativas do seu usuário. Como estas necessidades são diversas e dependem de cada aplicação em particular, denomina-se Qualidade de Serviço (QoS - *Quality of Service*) ao conjunto de parâmetros de desempenho que a rede deve garantir à aplicação usuária. Estes parâmetros são normalmente negociados durante o estabelecimento da conexão.

No entanto, o compartilhamento de recursos ou a multiplexação estatística, sem a devida monitoração e controle, podem causar situações de congestionamento, com perdas e atrasos de células, que degradam a QoS requerida [6]. Os mecanismos de controle de admissão de conexão (CAC - *Connection Admission Control*), juntamente com outros mecanismos de controle de tráfego e de congestionamento, reduzem a ocorrência destas situações abusivas, por parte do usuário, no compartilhamento de recursos da rede.

Este artigo faz um estudo sobre os mecanismos de CAC em redes ATM. Na seção 2, diversos mecanismos propostos na literatura são apresentados. Uma classificação destes mecanismos baseada no princípio implícito neles utilizado é proposta também nesta seção. A análise de alguns destes mecanismos é apresentada na seção 3. Por último, na seção 4, as principais conclusões são apontadas.

2 Mecanismos de CAC

Tipicamente, o problema de decisão sobre a aceitação ou não de uma nova conexão pode ser formulado como um problema de teoria de filas envolvendo um multiplexador ATM, representado por um conjunto de fontes ATM alimentando uma fila de capacidade finita que é servida por um servidor com um tempo de serviço constante [7, 8]. Neste problema, procura-se encontrar o número máximo de conexões que podem ser aceitas enquanto estando de acordo com os objetivos de QoS contratados. Entretanto, a solução deste problema de filas é muito árdua e difícil de ser realizada em tempo de execução. Em vista disto, diversos mecanismos de CAC tem sido propostos baseados em diferentes aproximações ou em esquemas diferentes que não necessitam da solução de tal problema de filas.

Para atender a diversidade de aplicações encontrada em redes ATM, várias classes de serviço foram definidas [2]: o serviço de taxa binária constante (CBR - *Constant Bit Rate*), o serviço de taxa binária variável (VBR - *Variable Bit Rate*), o serviço de taxa binária indefinida (UBR - *Undefined Bit Rate*) e o serviço de taxa binária disponível (ABR - *Available Bit Rate*). O serviço CBR foi definido para suportar aplicações com requisitos bem definidos de atraso e que geram tráfego a uma taxa constante. O serviço VBR também foi definido para suportar aplicações com requisi-

*Este trabalho foi realizado com recursos da UFRJ, FUJB, PROTEM-CC, CNPq, FAPERJ, CAPES e C OFECUB.

tos bem definidos de atraso, mas que geram tráfego a uma taxa variável. O serviço UBR foi definido para suportar aplicações sem requisitos predefinidos de serviço e que possam tolerar qualquer nível de QoS que a rede possar prover. O serviço ABR é adequado para aplicações com requisitos mínimos de vazão e atraso. Cada uma destas classes de serviço possui as suas próprias características e o seu próprio conjunto de parâmetros descritores de tráfego. Portanto, o procedimento de CAC deve ser específico para cada uma delas. Entretanto, uma combinação entre fontes destas classes pode ser feita tendo em vista uma melhora da eficiência da rede.

Os CACs para chamadas das classes CBR e UBR são bastante simples. Enquanto no primeiro caso, a alocação é feita pelo taxa de pico, nenhuma banda passante é alocada no segundo. Neste último caso, O CAC deve apenas verificar se esta facilidade é oferecida por todos os nós da rota escolhida.

No CAC para chamadas da classe ABR, vários parâmetros de tráfego são especificados e a taxa de transmissão a ser fornecida pela rede para a fonte pode ser variável, sendo apenas conhecidos a priori os seu valores mínimo e máximo. Um pedido de conexão será aceito caso haja banda passante disponível para prover a taxa binária mínima requerida, sendo este o valor previamente alocado para a fonte. No entanto, a taxa binária fornecida pela rede pode aumentar ou diminuir conforme o estado da rede.

O CAC para chamadas da classe VBR é bastante diferente dos anteriores, pois fontes com as mais diversas características são superpostas nesta classe. Em vista disto, vários mecanismos de CAC podem ser encontrados na literatura. Alguns destes mecanismos requerem um modelo específico de tráfego, enquanto outros apenas necessitam dos parâmetros de tráfego, tais como as taxas binárias média e de pico. Devido a esta diversidade, aqui é proposta uma classificação dos mecanismos desta classe para uma melhor compreensão dos mesmos. Esta classificação está baseada no princípio implícito que foi usado no desenvolvimento destes mecanismos, similar àquela feita em [7]. As classes de CAC estão representadas pelos seguintes grupos: método de alocação de capacidades, método de reservas rápidas, aproximação de tráfego intenso, aproximação não-paramétrica, método dinâmico e método híbrido.

O método de alocação de capacidades tem por objetivo determinar a capacidade efetiva de cada fonte de tráfego de acordo com a taxa de perda de células (CLR - *Cell Loss Ratio*) desejada. Um pedido de conexão só é aceito se a capacidade disponível do canal for maior do que a capacidade efetiva da fonte. Caso contrário, o pedido é rejeitado.

O método de reservas rápidas foi desenvolvido para fontes de tráfego em rajadas. Neste método, quando uma conexão é estabelecida, nenhum recurso de rede é previamente alocado, sendo apenas determinadas as rotas através da rede de cada conexão. No momento em que a fonte deseja transmitir, a rede tenta alocar os recursos necessários apenas para a duração da rajada.

A aproximação de tráfego intenso [10] está baseada no comportamento assintótico da cauda da distribuição do tamanho de fila quando a intensidade de tráfego é alta. Um pedido de conexão só é aceito caso a CLR, aproximada pela cauda da distribuição do tamanho da fila, seja menor que o valor desejado.

A aproximação não-paramétrica não assume nenhum modelo específico de tráfego e utiliza apenas os parâmetros de tráfego. A decisão de aceitação ou não de uma nova conexão é feita com base no limite superior que a CLR pode assumir.

O método dinâmico é baseado na distribuição do número de células que chegam durante um determinado período de estimação. Cada período de estimação é composto por vários períodos de medição e é chamado de período de renovação. Um pedido de conexão só é aceito se sua capacidade equivalente for menor do que a capacidade equivalente estimada disponível no canal.

O método híbrido lida com fontes de tráfego com objetivos de QoS diferentes. São providas classes de alta e de baixa qualidades e um mecanismo de prioridades no *buffer* de saída. A decisão de aceitação de um pedido de conexão é feita com base apenas na sua taxa de pico e na medição da taxa de chegadas de células do tráfego agregado. Para uma conexão de alta qualidade é feita a alocação pelo pico. A banda passante que não é usada pelas conexões de alta qualidade pode, no entanto, ser usada pelas conexões de baixa qualidade.

Nas subseções subseqüentes, os métodos de alocação de capacidades, reservas rápidas e aproximação não-paramétrica são descritos em maiores detalhes visando a análise destes na seção 3. Estes métodos foram escolhidos pelo fato dos mesmos apresentarem o mesmo subconjunto de parâmetros descritores de tráfego.

2.1 Alocação de Capacidades

Várias estratégias são adotadas no cálculo das capacidades efetivas de cada fonte. Enquanto algumas destas estratégias estão baseadas em métodos de simulação, outras se tratam apenas de métodos analíticos. O método de simulação usualmente adotado é o orientado a eventos discretos [1]. No caso do ATM, a simulação deve ser feita a nível de célula, de modo a representar o mais fielmente possível o comportamento do multiplexador, apesar de também ser possível a simulação a nível de rajada. Os eventos correspondem ou à chegada ou à partida de uma célula.

A solução por métodos analíticos está baseada na caracterização e modelagem das fontes de tráfego e na respectiva resolução dos modelos de filas que representam os multiplexadores. Dentre estes, destaca-se o método da capacidade equivalente, atribuído a Guérin et al. [5]. O objetivo deste método é obter uma expressão unificada para a representação da banda passante efetiva, ou capacidade equivalente, de uma conexão, assim como para o agregado das conexões.

Duas aproximações diferentes são utilizadas no cálculo da banda passante efetiva: uma baseada em modelo de fluxo contínuo e outra, em uma aproximação da distribuição estacionária da taxa binária. A primeira aproximação estima de um modo acurado a capacidade equivalente quando o impacto das características de uma fonte individual é crítica. A segunda aproximação é representativa nos casos em que a multiplexação estatística das conexões é significativa.

Na aproximação de fluxo contínuo, uma única fonte de tráfego é modelada por uma cadeia de Markov de 2 estados (fontes em rajadas), correspondentes aos períodos de atividade e silêncio. A caracterização de uma fonte é feita pela sua taxa binária de pico (R), pela utilização (ρ), e pelo tempo médio de duração da rajada (L). Note que a taxa de utilização é igual ao inverso do índice de rajada (b). A distribuição do tamanho da fila, quando tal fonte está alimentando uma fila com um servidor de taxa de serviço constante, pode ser derivada usando-se técnicas padrão [5]. Com esta distribuição, 'e possível determinar a capacidade equivalente necessária para alcançar a QoS desejada. Assumindo uma fila de tamanho finito k, a equação satisfeita pela capacidade equivalente c para uma probabilidade de transbordo ε é da forma:

$$\varepsilon = \beta \ \exp\left(\frac{c - R\rho}{L(1-\rho)(R-c)c}\ k\right) , \tag{1}$$

onde $\beta = \frac{(c-\rho R)+\varepsilon\rho(R-c)}{(1-\rho)R}$.

A capacidade equivalente para uma única fonte pode ser obtida resolvendo-se a equação 1 para c. Entretanto, em tal sistema, a capacidade equivalente só pode ser resolvida numericamente. Devido a isto, Guérin et al. propuseram uma aproximação fixando o valor de β em 1. Neste caso, a capacidade equivalente $\hat{c}$ de uma única fonte isolada é dada por:

$$\hat{c} = \frac{f_i - k + \sqrt{(f_i-k)^2 + 4k\rho_i f_i}}{2\alpha\, b_i(1-\rho_i)} , \tag{2}$$

onde $\alpha = \ln(1/\varepsilon)$ e $f_i = \alpha\, b_i(1-\rho_i)R_i$.

No caso de múltiplas fontes, esta aproximação equivale a não considerar o ganho de multiplexação estatística. Com isto, a capacidade equivalente do agregado de fontes, $\widehat{C_f}$, é dada pela soma das N capacidades equivalentes individuais das fontes.

Na aproximação estacionária, a distribuição das taxas de transmissão do tráfego agregado no estado estacionário é aproximado por uma curva gaussiana. Neste caso, a capacidade equivalente do agregado de fontes é dada por

$$C_g = m + \alpha\sigma , \tag{3}$$

onde

$$\begin{aligned} \alpha &= \sqrt{-2\log(\varepsilon) - 2\log(2\pi)} ; \\ m &= \textstyle\sum_{i=1}^{N} \rho_i R_i ; \\ \sigma &= \sqrt{\textstyle\sum_{i=1}^{N} \rho_i R_i (R_i - \rho_i R_i)} . \end{aligned}$$

Por último, o valor da capacidade equivalente para o agregado das conexões será o mínimo entre os dois valores calculados

2.2 Reservas Rápidas

O método de reservas rápidas utilizam basicamente duas diferentes técnicas de reserva: banda passante e espaço no *buffer*.

O método de reserva rápida de banda passante foi proposto por Tranchier et al. para fontes em rajadas [7]. A alocação da banda passante requerida pela fonte é feita através de pedidos de incremento ou decremento, podendo seu valor variar entre zero e a sua taxa de pico. Os pedidos de incremento são feitos através de uma célula especial de pedido de reserva. O aumento requerido só será aceito por um nó da rede caso a soma do tráfego total requerido não exceda a capacidade do canal, sendo portanto baseado na alocação pelo pico. Os pedidos de decremento são feitos através de células de gerenciamento e são sempre aceitos.

O método de reserva rápida de *buffer* é atribuído a Turner [11]. Neste método, o espaço em *buffer* é que é alocado ao invés da banda passante, como no método anterior. Com isto, fontes cujas taxas de pico são próximas à capacidade do canal podem ser aceitas. Cada fonte é modelada por uma cadeia de Markov de 2 estados (fontes em rajada), alternando entre os estados ativos e inativos. Quando a fonte está no seu estado ativo, um número predeterminado de compartimentos é alocado no *buffer* por toda a duração do período ativo, sendo liberados logo após o final deste período. As transições entre estes estados se dão através de células especiais que determinam o início e o fim de uma rajada. Todas as outras células são tratadas como células intermediárias.

2.3 Aproximação Não-paramétrica

O objetivo do Controle de Admissão de Conexão é o de alocar a banda passante necessária para uma conexão de modo a garantir a QoS desejada. E isto deve ser feito para qualquer conjunto de parâmetros de tráfego especificados pelo usuário e para qualquer processo de chegada de células que satisfaça os valores destes parâmetros. O método de avaliação de desempenho que não assume um modelo para o processo de chegada de células é chamado de aproximação não-paramétrica [9]. Sua idéia básica deste método está em calcular o valor máximo que a CLR pode assumir para um dado conjunto de parâmetros de tráfego e utilizá-lo como critério de aceitação para um pedido de uma nova conexão.

Sejam ANA e MNA respectivamente os números médio e máximo de células que chegam durante um intervalo de tempo fixo. O comprimento deste intervalo é escolhido como o tempo necessário para a transmissão de $K/2$ células pelo canal, onde K representa o tamanho do *buffer* em células.

Seja $\theta_i = \{\theta_i(j)\}$, descrita na equação 4, a distribuição do tipo Bernoulli obtida através dos parâmetros de tráfego (a_i, R_i) da conexão i, onde a_i é o ANA e R_i é o MNA.

$$\theta_i(j) = \begin{cases} a_i/R_i & , j = R_i ; \\ 1 - a_i/Ri & , j = 0 ; \\ 0 & , \text{caso contrário.} \end{cases} \tag{4}$$

Saito [9] mostra que esta distribuição representa o tráfego do pior caso para a conexão i e que, num canal de transmissão células de N conexões estão sendo multiplexadas, o limite superior da probabilidade de perdas de células (CLP - *Cell Loss Probability*) é dado por:

$$\begin{aligned} CLP &\leq B^*(\theta_1, \cdots, \theta_N; K/2) \\ &= \frac{\displaystyle\sum_{k=0}^{\infty} [k - K/2]^+ \theta_1^* \star \cdots \star \theta_N^*(k)}{\displaystyle\sum_{k=0}^{\infty} k \cdot \theta_1^* \star \cdots \star \theta_N^*(k)} \end{aligned} \tag{5}$$

onde $\star$ denota a convolução e $[x]^+ = \begin{cases} x & , x \geq 0 ; \\ 0 & , x < 0 . \end{cases}$

O controle de admissão que se baseia neste limite superior da CLP é bastante simples. Uma nova conexão $N+1$, que faz um pedido de admissão, será aceito se

$$B^*(\theta_1, \cdots, \theta_{N+1}; K/2) \leq \tilde{B} \quad (6)$$

e rejeitado em caso contrário, onde N é o número de conexões já conectadas e $\tilde{B}$ é a taxa máxima de perdas de células aceitável.

3 Análise de Desempenho

A figura 1 apresenta uma comparação entre os resultados de ambas as aproximações utilizadas no método da capacidade equivalente com os resultados obtidos com o modelo UAS [1] para uma CLR igual a 10^{-9}. Nela, o comprimento médio da rajada é variado de um valor bem pequeno até um valor igual ao tamanho do *buffer*. Os valores do tamanho do *buffer* e do número de fontes foram escolhidos de modo que a variação da aproximação de fluxo contínuo pudesse ser notória. A aproximação gaussiana, como já mencionado, não varia com este parâmetro. Pode-se observar nesta figura que, para comprimentos de rajadas pequenos, a aproximação de fluxo contínuo é muito boa. Isto se deve ao fato de que o impacto de uma única fonte é grande e que a capacidade equivalente de uma única fonte é próxima à sua taxa média. Conseqüentemente, quase não há ganho de multiplexação estatística. Entretanto, quando o mesmo aumenta, o impacto de uma única fonte deixa de ser significativo e o ganho de multiplexação estatística passa a ser maior. Com isto, a aproximação de fluxo contínuo tende a superestimar bastante o valor da capacidade equivalente. Nestes casos, a aproximação gaussiana passa a ser utilizada. Porém, como pode ser observado, ela também tende a superestimar o valor da capacidade necessária. E esta superestimação é tão maior quanto menor for a CLR desejada [1].

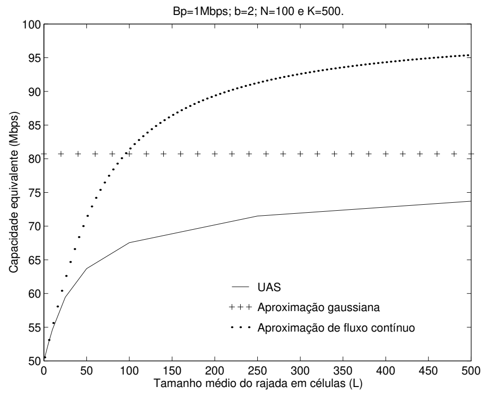

Figura 1: Comparação entre as aproximações de fluxo contínuo e gaussiana para um objetivo de CLR de 10^{-9}.

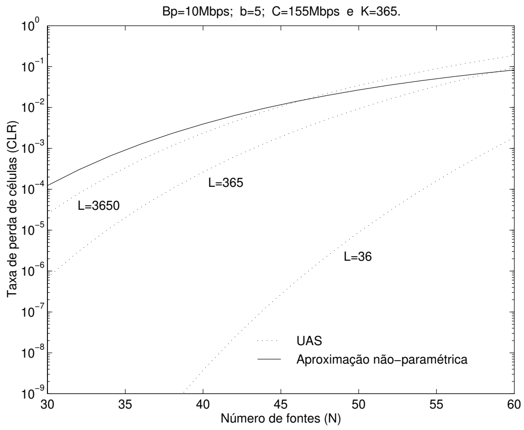

Figura 2: Comparação entre a aproximação não-paramétrica e o modelo UAS.

A figura 2 apresenta uma comparação entre os resultados do método de aproximação não-paramétrica com os resultados obtidos com o modelo UAS [1] para fontes com índice de rajada igual a 5, para diversos valores de comprimento médio de rajada (L). Esta comparação mostra que, para tráfegos com comprimentos médios de rajada menores, a aproximação superestima bastante a CLR. Isto se deve ao fato de que, para tais valores, a capacidade necessária para o tráfego agregado é bem inferior àquela estimada usando-se o tráfego de pior caso. Entretanto, para comprimentos maiores de rajada, esta aproximação passa a apresentar bons resultados.

Na análise do método de reserva de *buffer*, observou-se que seus resultados se tornam mais inacurados para tamanhos de *buffer* menores. Isto se deve ao fato de que, nestes casos, a quantidade de compartimentos que são reservados para cada conexão passa a ser muito pequena. Como a estimação da CLR é feita com base neste valor, conseqüentemente, a sua acurácia piora. Fato semelhante tam-bém foi observado na aproximação não-paramétrica. Além disto, como este método foi desenvolvido para fontes com períodos de rajada de longa duração, seus resultados são bastante superestimados quando fontes com períodos curtos de rajada são multiplexadas. Isto pode ser melhor observado na figura 3, onde os resultados deste método são comparados com os do modelo UAS para fontes com índice de rajada igual a 5, para diversos valores de comprimento médio de rajada (L). Resultados equivalentes podem ser obtidos para outros valores de índice de rajada.

4 Conclusões

Este artigo realizou um estudo e análise de diversos mecanismos de Controle de Admissão de Conexão (CAC) para redes ATM. Além disto, foi feita uma classificação destes mecanismos para que as suas diferenças pudessem ser melhor compreendidas.

O conceito apresentado nos mecanismos de alocação de capacidades é bastante útil para fins de engenharia, pois a decisão de aceitação ou não de uma nova conexão reside apenas em saber se a capacidade efetiva da nova conexão é maior ou não que a capacidade disponível no canal de transmissão. Esta simplicidade, contudo, como foi observado, pode reduzir sensivelmente o ganho de multiplexação estatística. Algumas extensões propostas para este método [3, 4] atentam para este detalhe e conseguem

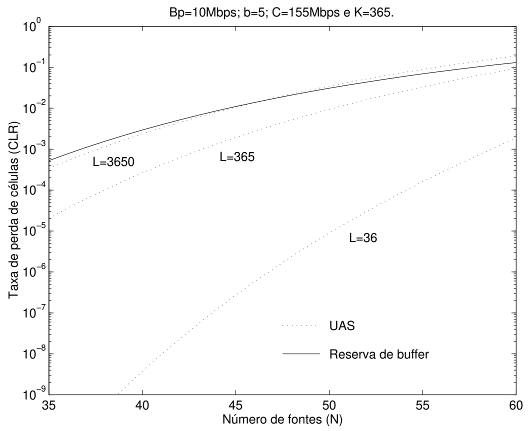

Figura 3: Comparação entre o método da reserva de *buffer* e o modelo UAS.

um melhor desempenho. Contudo, mais análises são necessárias para a validação destes modelos.

Os mecanismos de reservas rápidas apresentam um bom desempenho quando fontes com períodos de rajadas muito longos e pouco freqüentes são multiplexadas, pois, nestes casos, alocar recursos da rede por toda a duração da conexão pode ser muito dispendioso. Entretanto, como mostrou sua análise, seus resultados são bastante superestimados quando se multiplexa fontes com tamanhos de rajada menores. Em alguns casos, a CLR é superestimada em mais de 5 ordens de grandeza (figura 3). Em razão disto, a sua utilização em um ambiente heterogêneo tende a ser pouco eficiente.

A idéia apresentada na aproximação não-paramétrica é muito interessante, pois nenhum modelo de tráfego é assumido e a CLR é estimada com base apenas nos parâmetros de tráfego fornecidos pelo usuário. Isto vai de acordo com os objetivos das RDSI-FL em atender serviços com as mais diversas características. Contudo, assim como os mecanismos de reservas rápidas, seus resultados tendem a ser bastante superestimados, principalmente para tráfegos com períodos curtos de rajada. Em algumas vezes, este método chega a superestimar a CLR em várias ordens de grandeza (figura 2).

Finalmente, na análise de todos os mecanismos de CAC apresentados, constatou-se que a eficácia destes varia bastante de acordo com a faixa de parâmetros de tráfego utilizada e com os objetivos de QoS desejados. Enquanto alguns são mais adequados para os casos onde o ganho de multiplexação estatística é grande e/ou para CLR mais altas, outros o são para os casos onde o impacto de cada fonte isoladamente é mais significativo e/ou para CLR mais baixas. Portanto, a decisão sobre qual o mecanismo mais adequado depende bastante de qual o tráfego mais esperado, ou seja, com maior probabilidade de ocorrência. Entretanto, na maioria dos casos que se espera encontrar em redes ATM, estes mecanismos geralmente apresentam resultados bastante superestimados para a capacidade equivalente do agregado das conexões, subestimando, assim, o ganho potencial de multiplexação estatística destas redes. Conclui-se, portanto, que CAC continua a ser uma questão em aberto e que mecanismos mais eficazes ainda estão para ser encontrados. Além disto, a abrangência destes mecanismos deve ser a maior possível de modo a ser eficiente para uma maior faixa de valores possíveis.

Referências

[1] M. F. Abdalla. Análise de mecanismos de controle de admissão de conexão para redes ATM. Master's thesis, Universidade Federal do Rio de Janeiro, PEE/COPPE, 117 p., Rio de Janeiro, Sept. 1996.

[2] F. Banomi and K. W. Fendick. The rate-based flow control framework for the available bit rate ATM service. *IEEE Network Magazine*, 9(2):25–39, Mar. 1995.

[3] G. L. Choudhury, D. M. Lucantoni, and W. Whitt. Squeezing the most out of ATM. *IEEE Transactions on Communications*, 44(2):203–217, Feb. 1996.

[4] G. de Veciana, G. Kesidis, and J. Walrand. Resource management in wide-area ATM networks using effective bandwidths. *IEEE Journal on Selected Areas in Communications*, 13(6):1081–1090, Aug. 1995.

[5] R. Guérin, H. Ahmadi, and M. Naghshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE Journal on Selected Areas in Communications*, 9(7):968–981, Sept. 1991.

[6] D. Hong and T. Suda. Congestion control and prevention in ATM networks. *IEEE Network Magazine*, pages 10–15, July 1991.

[7] H. Perros. A literature review of call admission algorithms. Technical report, Dep. Elet. Comput. Eng. of North Carolina State Univ., Raleigh, Apr. 1994.

[8] H. Perros and K. Elsayed. Call admission control schemes: A review. *IEEE Network Magazine*, 34(11):82–91, Nov. 1996.

[9] H. Saito. *Teletraffic Technologies in ATM Networks*. Artech House, Boston, London, first edition, 1994.

[10] K. Sohraby. Flow and admission control of ON-OFF sources in high speed networks. In *IEEE GLOBECOM'94 - Tutorial 7*, San Francisco, Nov. 1994.

[11] J. S. Turner. Managing bandwidth in ATM networks with bursty traffic. *IEEE Network Magazine*, pages 50–58, 1992.