

Dimensionamento de Enlaces em Redes de Telecomunicações ATM

Tese apresentada à Faculdade de Engenharia Elétrica e de
Computação da Universidade Estadual de Campinas, como
parte dos requisitos exigidos para a obtenção do título de
Mestre em Engenharia Elétrica.

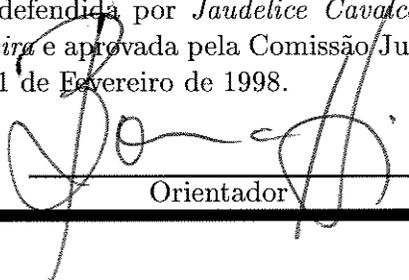
Jaudelice Cavalcante de Oliveira

Graduada em Engenharia Elétrica pela Universidade Federal do Ceará

Orientador:

Dr. Ivanil S. Bonatti

Este exemplar corresponde à redação final da
tese defendida por *Jaudelice Cavalcante de
Oliveira* e aprovada pela Comissão Julgadora
em 11 de Fevereiro de 1998.


Orientador

Fevereiro de 1998

UNIDADE	BC
N.º CHAMADA:	UNICAMP
	OL4d
V. Ex.	
T. MEIO BC/	33991
PROC.	395/98
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	28/05/98
N.º CPD	

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

OL4d

Oliveira, Jaudelice Cavalcante de
Dimensionamento de enlaces em redes de
telecomunicações ATM. / Jaudelice Cavalcante de
Oliveira.-Campinas,SP:[s.n.], 1998.

Orientador: Ivanil S. Bonatti
Dissertação (mestrado) - Universidade Estadual de
Campinas, Faculdade de Engenharia Elétrica e de
Computação.

1. Telecomunicações - Tráfego. 2. Rede digital de
serviços integrados. 3. Comunicações digitais. I. Bonatti,
Ivanil S. II. Universidade Estadual de Campinas. Faculdade
de Engenharia Elétrica e de Computação. III. Título

Banca Examinadora

- Prof. Dr. Ivanil Sebastião Bonatti (Presidente)
Doutor em Automática – 1981 – Toulouse - França
Local de Trabalho: DT-FEEC-UNICAMP

- Dr. Ing. Marcos Bafutto
Doutor em Engenharia Elétrica – 1995 – Stuttgart - Alemanha
Local de Trabalho: Agência Nacional de Telecomunicações - ANATEL

- Prof. Dr. Shusaburo Motoyama
Doutor em Engenharia Elétrica – 1982 – Tóquio - Japão
Local de Trabalho: DT-FEEC-UNICAMP

- Prof. Dr. Pedro Luis Dias Peres (Suplente)
Doutor em Automática – 1989 – Toulouse - França
Local de Trabalho: DT-FEEC-UNICAMP

Pedra Filosofal

António Gedeão

*Eles não sabem que o sonho
é uma constante da vida
tão concreta e definida
como outra coisa qualquer,
como esta pedra cinzenta
em que me sento e descanso,
como este ribeiro manso
em serenos sobressaltos,
como estes pinheiros altos
que em verde e oiro se agitam,
como estas aves que gritam
em bebedeiras de azul.*

*Eles não sabem que o sonho
é vinho, é espuma, é fermento,
bichinho álaçre e sedento,
de focinho pontiagudo,
que fuça através de tudo
em perpétuo movimento.*

*Eles não sabem que o sonho
é tela, é cor, é pincel,
base, fuste, capitel,
arco em ogiva, vitral,
pináculo de catedral,
contraponto, sinfonia,
máscara grega, magia,
que é retorta de alquimista,
mapa do mundo distante,
rosa-dos-ventos, Infante,
caravela quinhentista,
que é Cabo da Boa Esperança,
ouro, canela, marfim,
florete de espadachim,
bastidor, passo de dança,
Colombina e Arlequim,
passarola voadora,
pára-raios, locomotiva,
barco de proa festiva,
alto-forno, geradora,
cisão do átomo, radar,
ultra-som, televisão,
desembarque em foguetão
na superfície lunar.*

*Eles não sabem, nem sonham,
que o sonho comanda a vida,
e que sempre que um homem sonha
o mundo pula e avança
como bola colorida
entre as mãos de uma criança.*

Agradecimentos

Esta é, sem dúvida, uma das partes mais complicadas da tese. Estes quase dois anos de trabalho trazem influências de tantas pessoas que conheci aqui, durante o mestrado, e tantas outras que conheci antes. Há muito a agradecer.

Ao Prof. Ivanil Bonatti, muito mais que meu orientador, um amigo inestimável pronto para ajudar a qualquer hora, em qualquer assunto ou situação. Não foram raros os momentos em que vi alunos da graduação ou pós-graduação, que nunca sequer tinham feito alguma das disciplinas ministradas por ele, procurarem os seus conselhos. As defesas de tese de cuja banca ele participa sempre possuem alguns espectadores a mais, muitas vezes alheios ao assunto ou ao pós-graduando, mas ávidos pelo momento mais interessante da defesa: quando a palavra é passada ao Ivanil. Cada cafezinho gera um “novo paper”, cada manhã uma “frase do dia”. Por tudo o que aprendi, não só durante a composição da tese, mas em cada um destes momentos.

Ao Prof. Pedro Peres, outro exemplo de dedicação e competência, um dos primeiros a conquistar a minha amizade e admiração, logo que cheguei aqui (Hum... logo que o vi não imaginei que fosse o próprio professor, e sim um interessantíssimo orientado do Prof. Pedro Peres). Por todas as sugestões que deu a este trabalho, por tudo o que me ensinou, por todas as paixões que me despertou, principalmente a paixão pelo \LaTeX .

Aos Professores Marcos Bafutto e Shusaburo Motoyama, por terem aceitado o convite para participar da banca examinadora e por colaborarem para a melhoria deste trabalho.

A Amaury K. Budri, por todas as contribuições ao trabalho, pela parceria indispensável na produção dos artigos, pela paciência, e pelos momentos de pura descontração nas interessantíssimas reuniões com o Ivanil que pude presenciar.

A meu pai, pelo exemplo de constante batalha em busca de seus ideais e por ter me despertado o gosto pela música (minhas flautas foram grandes parceiras de trabalho). À minha mãe, por ter me ensinado a sorrir, haja o que houver, por ser uma das lutadoras mais fortes que já conheci, sem perder a ternura.

Ao meu irmão mais novo, Jauvane, e sua esposa, Carla, minhas “companhias virtuais” nos finais de semana solitários, nos longos feriados, Natal, Ano Novo, em quase todos os dias em que estive aqui na Unicamp. Mesmo sendo o irmão mais distante fisicamente, “virtualmente” foi o mais próximo. Por todo o apoio que me deu em todos os campos da minha vida: profissional, sentimental, etc., por todas as dicas e sugestões, que repassei a tantos outros colegas, na preparação dos papéis para realização do doutorado no exterior e por tudo mais que me ensinou.

Ao meu irmão mais velho, Jáuber, por ter iniciado o “desbravamento” deste mundo acadêmico que contagiou cada um dos nossos irmãos. Pelos longos bate-papos, embora escassos devido à correria do dia-a-dia de cada um.

À minha irmã Jaudênia, pela indispensável ajuda que me ofereceu desde a viagem pra vir morar em Campinas, a oportunidade de conhecer a Universidade de Ottawa e tudo mais que me proporcionou, desde o apoio financeiro ao emocional.

Ao Prof. Everardo, por ter ajudado a despertar em mim o gosto pela pesquisa, e a descobrir a paixão pela minha área de trabalho, e mais ainda, por ter me proporcionado a realização do sonho “impossível” que era vir estudar na Unicamp. Pela incrível amizade que temos, e também pelos deliciosos bate-papos e *e-mails*, acadêmicos ou não.

Aos Professores do Departamento de Engenharia Elétrica da Universidade Federal do

Ceará, pelo embasamento necessário ao bom desempenho no curso de mestrado. Ao Prof. Anthony Donegan por todo o apoio e incentivo à realização do meu curso de pós-graduação na Unicamp.

Aos meus padrinhos Vitoriano e Elídia, e à prima Eliete por todo o carinho e apoio em todas as fases da minha vida.

A Paulo Alexandre Aranda de Oliveira, um grande amigo que conheci em Jericoacoara. É também por causa dele que estou aqui me tornando mestre. Nunca esquecerei a forma como ele me ajudou, mesmo sem me conhecer. Espero que um dia, onde quer que esteja, ele possa ver este agradecimento. Pelo apoio que me foi dado na mesma época, agradeço à tia Miryam e à Socorro Braga.

Aos dois amigos que moraram comigo em Campinas: Matheus e Carlos Eduardo (Cacá), por todo o apoio e companhia, pelos momentos de descontração na bagunça que era o nosso apartamento, e nossos animados bate-papos que deixam tanta saudade. Fizeram muito mais por mim do que imaginam. Aos demais colegas de Fortaleza que vieram compartilhar a saudade e a realização de sonhos.

A Marcos Jesus, por toda a paciência, carinho e atenção quando eu o procurava para resolver “bugs” da rede, mandava uma mensagem pro pager bem na hora de um casório, etc. Agradeço também a deliciosa e inesquecível surpresa (um lindo bolo!) no meu aniversário.

À Denize Ricardi e a Ron Slotin pelo entusiasmo, motivação e admiração pelo meu trabalho, bem como pela revisão dos textos em inglês. Ainda à Denize, uma pessoa muito especial, por toda a energia e carinho que transmite a cada um que conhece, dando força para tornar possíveis as coisas que antes pareciam inatingíveis. Pelo quanto cresci desde que a conheci.

Aos colegas do DT, DCA, DECOM, DSIF, DENSIS, DMO e demais departamentos, além dos colegas da graduação que tive a oportunidade de conhecer e que fizeram parte desta família que formamos, aquecendo os corredores do prédio da elétrica, dando um brilho especial a isto tudo.

Aos demais professores com quem tive contato, mesmo que apenas em um cafezinho, por tudo o que me ensinaram.

Aos funcionários da Unicamp, especialmente aos funcionários da BAE, da Faculdade de Engenharia Elétrica e de Computação, e em particular, aos seguranças da portaria do prédio principal da Elétrica, pela simpatia e paciência de sempre terem ali a carteirinha daquela “menina misteriosa que dorme aqui no prédio”.

À Unicamp, à Faculdade de Engenharia Elétrica e de Computação da Unicamp e, especialmente, ao Departamento de Telemática, minha família neste período do mestrado. Aqui passei os melhores dias da minha vida.

À CAPES, pelo financiamento dos meus estudos durante o mestrado.

A você, leitor e a quem mais eu possa ter esquecido de mencionar.



Este trabalho é fruto de um grande sonho e foi feito com muito carinho.

A handwritten signature in black ink, consisting of a stylized, cursive script.

Resumo

Neste trabalho uma abordagem abrangente para a determinação da capacidade de enlaces ATM (*Asynchronous Transfer Mode*) é proposta, considerando o grau de serviço das chamadas e os parâmetros de qualidade de serviço das conexões. A partir dos parâmetros descritores de tráfego e de qualidade de serviço propostos pelos organismos de recomendação, ITU-T (*International Telecommunication Union - Telecommunication Standardization Sector*) e ATM Forum, uma banda efetiva é obtida para cada classe de serviço, utilizando modelos probabilísticos para a perda e modelos de pior caso para o atraso de células das conexões. Para serviço comutado, a probabilidade de aceitação de uma nova chamada no enlace é obtida através da banda efetiva e através de modelos para a demanda e tempo de retenção da conexão.

Abstract

In this work a broad approach regarding the determination of the ATM (*Asynchronous Transfer Mode*) link capacity is proposed, taking into account the grade of service of calls and connection quality of service parameters. An effective bandwidth is obtained for each of the class of services defined through the traffic descriptors and QoS parameters proposed by ITU-T (*International Telecommunication Union - Telecommunication Standardization Sector*) and ATM Forum. Such procedure uses probabilistic models for connection cell loss and worst-case models for connection cell delay. The call blocking probability for switched services in a link is obtained through the effective bandwidth and through models for the demand and holding time in the connection.

Conteúdo

1	Introdução Geral	1
2	Tecnologia ATM	5
2.1	Comutação Digital	5
2.1.1	Comutação de Circuitos	5
2.1.2	Comutação de Pacotes	6
2.2	Modo de Transferência Assíncrono ATM	6
2.3	Modelo de Referência para Protocolos	7
2.4	Camada Física	7
2.5	Camada ATM	7
2.5.1	Canal Virtual	8
2.5.2	Caminho Virtual	9
2.6	Camada de Adaptação	9
2.7	Controle de Tráfego	10
2.8	Categorias de Serviço - ATM Forum	11
2.8.1	Parâmetros de Tráfego	12
2.8.2	Parâmetros de Qualidade de Serviço	14
2.8.3	Definição das Categorias de Serviço	16
2.9	Capacidades de Transferência - ITU-T	17
2.10	Conclusão	19
3	Encaminhamento de Tráfego em Redes Públicas ATM	21
3.1	Introdução	21
3.2	Tecnologia ATM nas Redes Públicas	21
3.3	Encaminhamento do Tráfego	23
3.4	Formulação Matemática	24
3.5	Modelos para Critérios de Desempenho	26
3.6	Exemplo com Dois Enlaces	28
3.7	Estimativa do Ponto de Mínimo	30
3.8	Discussão de Algoritmos	31
3.9	Exemplo	31
3.10	Conclusão	33
4	Alocação de Banda em Enlaces ATM	35
4.1	Introdução	35
4.2	Perda de Células	37
4.3	Atraso de células	40

4.4	Adequação aos Órgãos de Recomendação	42
4.5	Exemplos	44
4.6	Atraso de Varredura	48
4.7	Conclusão	49
5	Probabilidade de Bloqueio de Chamadas em um Enlace	51
5.1	Introdução	51
5.2	Fórmula de Erlang	52
5.3	Bloqueio de Chamadas em Enlaces Multi-Serviço	53
5.4	Conclusão	60
6	Conclusão Geral	63
A	COST 242	65
B	Fila M/M/1/K e M/D/1/K	67
B.1	Fila M/M/1/K	67
B.1.1	Probabilidade de Perda de Células	68
B.1.2	Tamanho Médio da Fila	69
B.1.3	Tempo de Espera no <i>Buffer</i>	69
B.2	Fila M/D/1/K	69
B.2.1	Probabilidade de Perda de Células	73
B.2.2	Tamanho Médio da Fila	73
B.2.3	Tempo de Espera no <i>Buffer</i>	74
C	Fila MMPP/D/1/K	75

Capítulo 1

Introdução Geral

Por volta de 1988 teve início o desenvolvimento do conceito de Rede Digital de Serviços Integrados de Faixa Larga ou B-ISDN (*Broadband Integrated Services Digital Network*), sendo selecionada a tecnologia ATM para dar suporte aos serviços B-ISDN.

Em 1992 a indústria divulgou previsões de que o ATM atravessaria uma fase experimental no ano de 1993, iniciando a disponibilidade de produtos comerciais em 1997, e que o ano 2000 seria o ano da massificação da tecnologia. Em 1994 os primeiros produtos comerciais foram disponibilizados.

Nos últimos anos desenvolveu-se um consenso na indústria de telecomunicações de que o ATM será um padrão internacional.

O dimensionamento das redes ATM é uma questão ainda em aberto nos órgãos de recomendação de padrões para ATM, ITU-T (*International Telecommunication Union - Telecommunication Standardization Sector*) e ATM Forum, e tem se tornado um dos temas mais debatidos na literatura especializada.

Com o objetivo de obter o dimensionamento da rede, a determinação da taxa de transmissão que cada enlace deve possuir, para atender às requisições de qualidade de serviço (QoS) de cada um dos elementos das fontes de tráfego que o compartilham, é uma questão essencial.

O dimensionamento de uma rede ATM compreende diversas fases anteriores à determinação da capacidade dos enlaces que compõem a rede. A primeira fase consiste na definição da topologia da rede, na qual o projetista determina a localização dos nós de comutação, definindo a filiação e a existência ou não de entroncamentos entre os diversos nós.

Definida a topologia da rede, o próximo passo consiste na determinação do encaminhamento das demandas de tráfego entre origem e destino, decidindo pela abertura de rotas diretas ou rotas alternativas.

Segue-se a fase de dimensionamento dos enlaces da rede, definindo a capacidade que cada enlace deve possuir para atender à demanda de tráfego.

Uma vez em funcionamento, a rede ATM pode eventualmente necessitar de uma reconfiguração devido à instalação de novos equipamentos, alterações, etc.

O efeito da sazonalidade do tráfego sobre a rede em funcionamento requer uma estratégia de reencaminhamento da demanda, permitindo vantagens operacionais de adaptação da rede às variações do tráfego em relação ao previsto no planejamento. Uma contribuição deste trabalho é a formulação do problema de encaminhamento no contexto de implantação da tecnologia ATM nas redes públicas de telecomunicações, simplifican-

do ao máximo as funções de perda de células e de atraso, viabilizando a aplicação de algoritmos convencionais de otimização de fluxo em redes multiprodutos.

A Figura 1.1 ilustra as diversas fases do dimensionamento de uma rede ATM, e a Figura 1.2 mostra os detalhes do dimensionamento de cada enlace da rede.

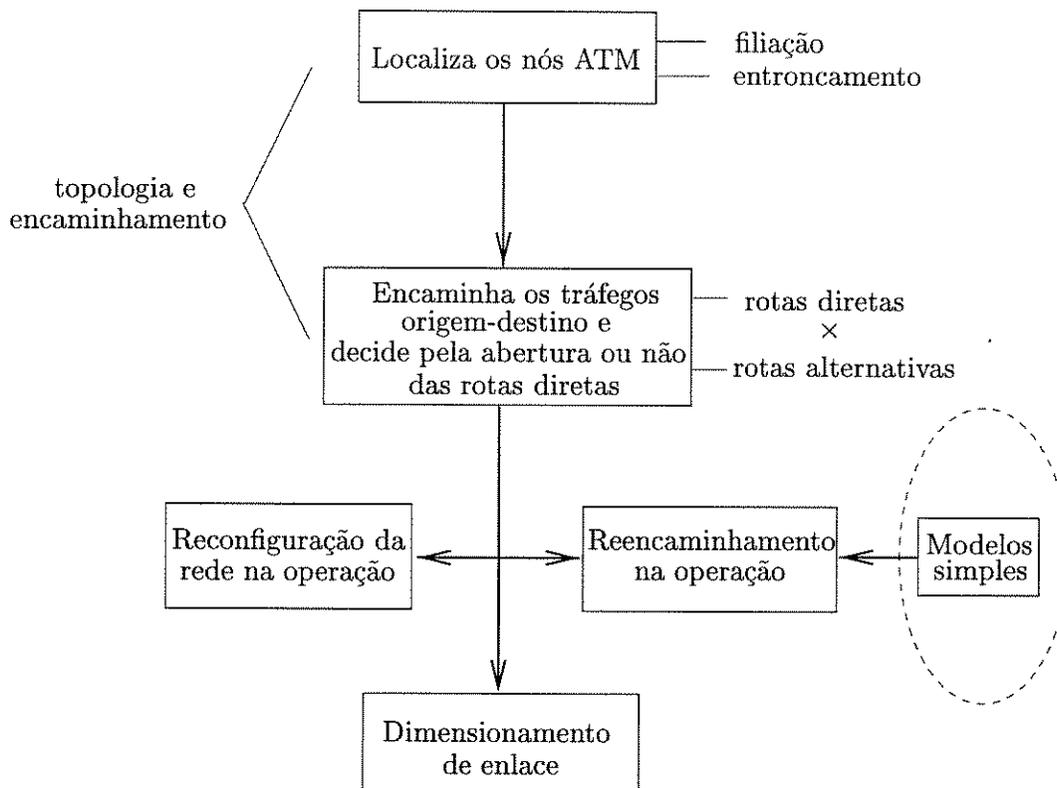


Figura 1.1: Diagrama do dimensionamento de uma rede ATM.

Dentre os vários modelos de cálculo da capacidade dos enlaces discutidos na literatura [7, 11, 15, 25, 31, 39] destaca-se o conceito de banda efetiva, que simplifica o dimensionamento, uma vez que a capacidade do enlace deve ser igual ou superior à soma das bandas efetivas dos tráfegos do enlace [11, 12, 13, 31].

Do ponto de vista do planejamento e gerência das redes ATM, os parâmetros de qualidade de serviço mais importantes são o atraso e a razão de perda de células. Assim, os serviços podem ser caracterizados, quanto aos parâmetros de QoS, em:

- sensíveis a atrasos e perda de informação (exemplo: vídeo interativo);
- sensíveis a atrasos, mas tolerantes a moderadas perdas (exemplo: voz);
- sensíveis a perdas, mas tolerantes a moderados atrasos (exemplo: dados interativos);
- relativamente tolerantes a atrasos e a uma limitada perda de informação (exemplo: transferência de arquivos).

O cálculo da banda efetiva depende das restrições a perda e atraso de células de cada serviço. Neste contexto, uma abordagem abrangente para a alocação de banda em enlaces ATM é proposta neste trabalho, onde cada um dos modelos propostos é adequado aos

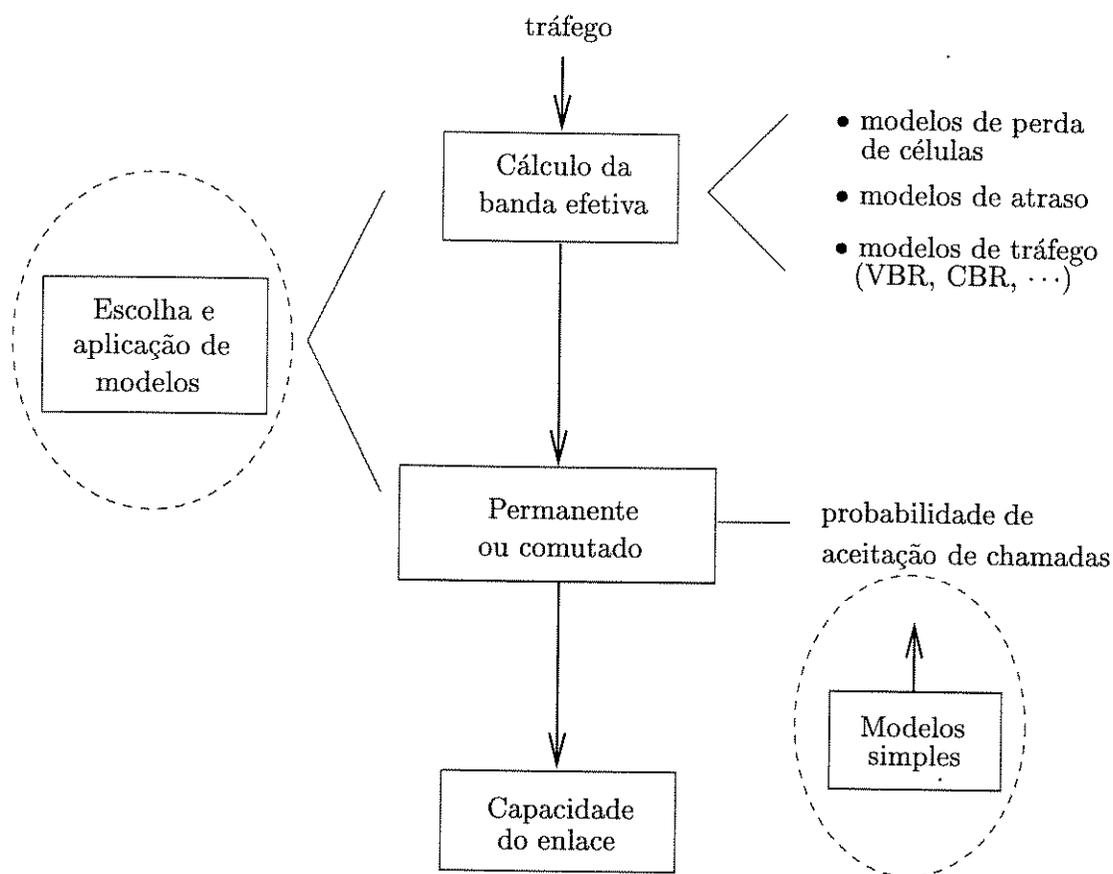


Figura 1.2: Diagrama do dimensionamento dos enlaces de uma rede ATM.

parâmetros de tráfego e de QoS do ITU-T e ATM Forum, sendo sugerida uma metodologia de cálculo da banda efetiva para cada uma das classes de serviço do ITU-T e ATM Forum.

Uma vez definida a banda efetiva do serviço, a capacidade do enlace pode ser encontrada. Para serviços com conexões permanentes, a capacidade do enlace será o limitante superior para a soma das bandas efetivas dos serviços do enlace. Para serviços comutados a alocação de uma banda efetiva às chamadas aceitas faz com que o sistema se comporte como um sistema comutado por circuito, e uma probabilidade de bloqueio de chamadas deve ser respeitada. Metodologias para o cálculo da probabilidade de bloqueio de chamadas em enlaces multi-serviço são discutidas neste trabalho.

A redação da tese está organizada da seguinte maneira: em primeiro lugar, conceitos básicos da tecnologia ATM são discutidos, seguidos pela definição dos parâmetros de tráfego e de qualidade de serviço das categorias de serviço do ATM Forum, e das capacidades de transferência do ITU-T.

No Capítulo 3 o problema do encaminhamento de tráfego em redes públicas ATM é formulado, simplificando ao máximo as funções de perda de células e de atraso. Algoritmos para solução do problema são discutidos e alguns exemplos ilustrativos são apresentados.

O problema da alocação de banda em enlaces ATM e o conceito de banda efetiva, calculada segundo modelos de perda e atraso de células, são apresentados no Capítulo 4. Modelos probabilísticos para perda de células e modelos de pior caso para o atraso são apresentados. Em seguida, propõe-se uma adequação dos modelos propostos aos parâmetros de tráfego e de QoS do ITU-T e ATM Forum, bem como uma metodologia para

o cálculo da banda efetiva para as categorias de serviço do ATM Forum e as capacidades de transferência do ITU-T. Uma seção de exemplos ilustra a obtenção da banda efetiva para serviços de taxa variável, evidenciando a sensibilidade dos serviços à perda e ao atraso de células.

A probabilidade de bloqueio de chamadas em enlaces multi-serviços é abordada no Capítulo 5. Após rememorar a fórmula de Erlang e o conceito de tráfego oferecido, o problema da determinação das probabilidades individuais de bloqueio dos diversos serviços é analisado.

A tese termina com as conclusões gerais do trabalho.

Capítulo 2

Tecnologia ATM

Neste capítulo apresenta-se a tecnologia do modo de transferência assíncrono ATM (*Asynchronous Transfer Mode*) que surgiu como técnica promissora para a construção de redes com integração de serviços e interconexão em alta velocidade.

O modo de transferência assíncrono é uma tecnologia de alto desempenho com comutação e multiplexação de células (pacotes de tamanho fixo) para transportar diferentes tipos de tráfego.

2.1 Comutação Digital

Comutação em uma rede de comunicação refere-se à alocação sob demanda dos recursos da rede para a transmissão de informação entre seus usuários. As principais formas de comutação são denominadas *comutação de circuitos* e *comutação de pacotes*.

2.1.1 Comutação de Circuitos

A comunicação através da comutação de circuitos se dá pelo estabelecimento de uma rota de comunicação dedicada entre a origem e o destino, e envolve três fases: o estabelecimento do circuito, a transmissão de informação e a desconexão.

A rota alocada durante o estabelecimento da conexão permanece dedicada de forma exclusiva aos terminais dos usuários, até que algum deles resolva desfazer a conexão. Para tráfegos não constantes a capacidade do meio físico é desperdiçada.

A rota pode ser formada por uma sucessão de enlaces físicos (comutação espacial), uma sucessão de canais de frequência (comutação de frequências), uma sucessão de canais de tempo (comutação do tempo) ou por qualquer combinação destes canais.

A comutação de circuitos rejeita novas chamadas em caso de congestionamento da rede, mas este congestionamento não degrada as conexões já estabelecidas. Os usuários têm disponível uma taxa de *bits* constante com um atraso fixo entre pontos terminais da conexão (*fim a fim* ou *ponto a ponto*).

Não há detecção de erros ou correção provida pela rede. Estas operações são de responsabilidade dos terminais. Portanto, não há *overhead* para correção de erros e a taxa de *bits* do circuito é totalmente disponível para a transmissão de informações do usuário.

2.1.2 Comutação de Pacotes

Na comutação de pacotes, as mensagens são segmentadas em unidades menores, denominadas pacotes, que podem ser de tamanho fixo ou variável. A principal vantagem dos pacotes de tamanho fixo é simplificar a comutação e diminuir o tamanho dos *buffers* nos nós de comutação. Entretanto, a comutação com pacotes de tamanho variável se adapta melhor à variada dimensão das mensagens.

A comutação de pacotes pode ser de dois tipos: com conexão, denominada de *circuito virtual*, e sem conexão, denominada de *datagrama*.

No modo sem conexão, os pacotes (datagramas) são tratados individualmente em cada nó de comutação, sendo entregues ao destino através do caminho mais adequado, determinado pelos algoritmos de roteamento. No caso de *encaminhamento dinâmico*, pacotes de uma mesma mensagem podem seguir caminhos distintos na rede, possibilitando a chegada ao destino numa seqüência diferente da transmitida.

Os pacotes caminham de nó em nó, sendo armazenados em *buffers* para serem retransmitidos (*store-and-forward*). Em cada nó, um pacote é recebido e o enlace de saída é escolhido em função do destino e condições de tráfego da rede. Pacotes recebidos com erro são retransmitidos. Um *overhead* considerável é necessário para realizar o controle de erro, reduzindo a taxa de *bit* efetiva disponível para a transmissão de informação.

Os mecanismos de espera nos *buffers* e de retransmissão introduzem um atraso variável de transmissão, sendo eventualmente necessária uma compensação desta variação estatística. Os atrasos de transferência na rede aumentam devido ao congestionamento e ao crescimento do tamanho das filas nos comutadores.

As redes de pacotes datagrama não podem garantir vazão constante e atraso máximo devido ao grande número de ações que devem ser desenvolvidas pelos nós de comutação. Cada nó recebe pacotes, armazena-os, processa algoritmos de detecção e controle de erros e de fluxo nos enlaces, toma decisões de roteamento, e espera que o enlace de saída esteja livre para encaminhar o pacote ao próximo nó.

Na comutação de pacotes orientada a conexão um mesmo caminho lógico (circuito virtual), definido durante a aceitação da chamada, é estabelecido para todas as células, enquanto durar a conexão. O circuito virtual contém mecanismos de controle de erro e de fluxo que garantem a seqüência de entrega dos pacotes ao destino, podendo ainda fazer uma pré-alocação de recursos nos nós, garantindo um melhor controle de congestionamento.

2.2 Modo de Transferência Assíncrono ATM

O modo de transferência assíncrono ATM é uma tecnologia de comutação de pacotes de tamanho fixo a altas taxas, padronizada pelo ITU-T (*International Telecommunication Union - Telecommunication Standardization Sector*) e pelo ATM Forum. O ATM provavelmente será o modo de comutação utilizado nas redes públicas e também em redes privadas [10, 33, 36, 37], pois suporta tráfegos heterogêneos com características diversas quanto à taxa de transmissão, tempo de transferência e outros parâmetros de qualidade de serviço.

Algumas características importantes da tecnologia ATM [10]:

- Opera em modo orientado a conexão estabelecendo uma conexão virtual entre a

fonte (assinante chamador) e o destino (assinante chamado).

- Os pacotes da rede ATM, chamados células, tem um comprimento fixo de 53 octetos. São 5 octetos de cabeçalho e mais 48 de informação. Este tamanho de célula, relativamente pequeno, reduz o tamanho dos *buffers* nos nós de comutação, bem como o tempo de transmissão de cada célula.
- O caminho virtual e a conexão virtual são identificados no cabeçalho de cada célula.
- Não há proteção contra erros no nível de enlace. Este controle deve ser feito, se necessário, entre a origem e o destino nas camadas superiores do protocolo.

2.3 Modelo de Referência para Protocolos

O ATM possui uma estrutura de protocolo em camadas. O modelo de referência para protocolos é composto de um *plano de usuário*, um *plano de controle* e um *plano de gerenciamento* [18].

O plano de usuário é responsável pelo transporte de informação, incluindo recuperação de erro e controle de fluxo. O plano de controle realiza a sinalização necessária à conexão e desconexão de chamadas. O plano de gerenciamento é subdividido em outros dois planos: o *plano de gerenciamento dos planos* e o *plano de gerenciamento das camadas*.

O plano de gerenciamento dos planos controla os elementos de rede dos planos de usuário, de controle e de gerenciamento, enquanto que o plano de gerenciamento das camadas controla o fluxo de informações de operação e manutenção relativo a cada camada.

Os planos de usuário e de controle têm em comum as duas camadas inferiores do protocolo de referência denominadas de *camada física* PHY (*Physical Layer*) e *camada ATM* (*ATM Layer*). As camadas física, ATM e a *camada de adaptação* AAL (*ATM Adaptation Layer*) correspondem aproximadamente às camadas física e de enlace do sistema OSI (*Open Systems Interconnection*).

2.4 Camada Física

A camada física do ATM é decomposta em duas subcamadas denominadas: *subcamada de meio físico* PM (*Physical Medium*), que trata da transmissão dos *bits* pelo meio físico incluindo a conversão eletro-óptica, e *subcamada de convergência de transmissão* TC (*Transmission Convergence*), que realiza o desacoplamento da taxa de transmissão em relação à taxa de geração de células, o controle de erros no cabeçalho e o delineamento de células.

2.5 Camada ATM

A camada ATM realiza a multiplexação e demultiplexação de células, a adição e remoção do cabeçalho das células, a comutação e o encaminhamento das células e o controle de fluxo nas interfaces usuário-rede UNI (*User-Network Interface*). As funções desta camada são especificadas pelas recomendações I.150, I.350 e I.356 [16, 19, 20].

A célula ATM contém 48 octetos para a informação (*payload*) e um cabeçalho de 5 octetos, sendo transmitidas através de conexões de circuitos virtuais com encaminhamento baseado em informações contidas no cabeçalho.

A composição dos *bits* do cabeçalho das células ATM, especificado na recomendação I.361 [21], difere na interface usuário-rede UNI e na interface NNI (*Network to Network Interface*). Os campos do cabeçalho são:

- GFC (*Generic Flow Control*), usado para o controle de fluxo nas interfaces usuário-rede.
- PT (*Payload Type*), que indica o tipo de informação contida na célula e se a célula sofreu atrasos devido à congestionamentos nos *buffers* dos nós da rede.
- CLP (*Cell Loss Priority*), que indica a prioridade de descarte das células por congestionamento. Um *buffer* considerado cheio para células com baixa prioridade pode ainda receber células de maior prioridade.
- HEC (*Header Error Check*), que é utilizado para a detecção de erros do cabeçalho, possíveis correções em erros de um único *bit* e delimitação de células.
- VPI (*Virtual Path Identifier*) e o VCI (*Virtual Channel Identifier*), que formam o rótulo no cabeçalho utilizado pelos comutadores para encaminhar as células ao enlace de destino.

2.5.1 Canal Virtual

Canal virtual é um termo genérico usado para descrever capacidade de comunicação unidirecional no transporte de células ATM [17]. O cabeçalho de cada célula possui um rótulo que a identifica explicitamente como pertencente a um determinado canal virtual VC (*Virtual Channel*). Este rótulo possui dois campos: o *identificador de canal virtual* VCI e o *identificador de caminho virtual* VPI, ambos relacionados ao enlace no qual a célula está sendo transmitida. Uma conexão fim a fim é traduzida, pela gerência da rede, em uma sucessão de enlaces em cascata entre o ponto origem e o destino da chamada. Em cada um destes enlaces, as células da mesma chamada recebem uma seqüência pré-determinada de identificadores VPI e VCI, que só têm validade durante a existência da chamada. Estes identificadores não têm significado fim a fim (usuário a usuário).

Na fase de comunicação (conversa, no caso telefônico) o VPI e o VCI, das células que chegam em um nó de comutação, são utilizados para localizar na tabela de encaminhamento, a porta de saída para as células e quais os novos valores de VPI e VCI que as células deverão ter. Os valores da tabela de encaminhamento são determinados no estabelecimento da conexão e permanecem constantes enquanto durar a conexão. As células seguem uma mesma rota dentro da rede e a integridade da seqüência de células é garantida.

Os comutadores que fazem a transposição dos valores de VCI, via tabela de encaminhamento, são denominados *comutadores de VC*. Um *enlace VC* é definido entre dois comutadores de VC consecutivos ou entre um terminal ATM e um comutador de VC. O enlace é originado pela designação de um valor de VCI, e terminado pela remoção do VCI.

A conexão entre dois pontos terminais é chamada *conexão de canal virtual* VCC (*Virtual Channel Connection*), e é formada por um conjunto de canais virtuais concatenados.

2.5.2 Caminho Virtual

Um VPI específico é associado à célula cada vez que o *caminho virtual* VP (*Virtual Path*) é comutado. Os nós de comutação que desconsideram o VCI e apenas operam sobre o valor do VPI são chamados *comutadores de VP*. Nem sempre que há uma comutação de VC o VPI muda, mas um novo VPI é sempre designado numa comutação de caminho virtual, mesmo que o valor do VCI não se altere. Os canais virtuais que compartilham o mesmo caminho virtual possuem o mesmo VPI.

Um *enlace VP* é definido entre dois comutadores de VP consecutivos, ou entre um terminal ATM e um comutador de VC, ou entre um comutador VC e outro comutador VP. O enlace é originado pela designação de um VPI por um terminal ATM, ou por um comutador de VC, ou por um comutador de VP, e é terminado pela remoção do valor de VPI.

Enlaces VP são concatenados para formar uma *conexão caminho virtual* VPC (*Virtual Path Connection*).

2.6 Camada de Adaptação

A camada de adaptação AAL (*ATM Adaptation Layer*) utiliza o transporte de células da camada ATM para oferecer serviços com requisitos específicos. Atualmente estão definidos cinco protocolos [23] para realizar a interface entre a camada ATM e as diferentes aplicações (serviços).

A AAL é dividida em duas subcamadas: a *subcamada de segmentação e remontagem* SAR (*Segmentation and Reassembly*) que particiona o fluxo de informação em fragmentos, para que caibam no campo de informação das células na transmissão, e remonta o fluxo de informação, a partir das células recebidas; e a *subcamada de convergência* CS (*Convergence Sublayer*) que detecta a perda de células e recupera a relação temporal da informação original no destino.

O ITU-T, na recomendação I.362 [22], dividiu as classes de serviço (tráfego) em relação a sua natureza VBR (*Variable Bit Rate*) ou CBR (*Constant Bit Rate*), considerando a necessidade ou não de manter a sincronização da informação entre a origem e o destino, e considerando ainda os serviços que são orientados ou não por conexão.

- **Classe A:** Emulação de circuitos com taxa constante, orientado a conexão. Exemplos: televisão e voz convencionais.
- **Classe B:** Orientado a conexão com taxa variável e necessitando sincronização entre a fonte e o destino, devendo compensar a variação estatística do atraso. Exemplo: televisão codificada.
- **Classe C:** Dados orientados a conexão. Exemplos: X.25, Frame Relay e DQDB (*Distributed Queue Dual Bus*).
- **Classe D:** Dados não orientados a conexão. Exemplos: TCP/IP (*Transmission Control Protocol/Internet Protocol*).

Na montagem de uma célula, nem todos os 48 octetos de carga útil são preenchidos com informações do usuário. Alguns octetos são utilizados pela AAL para enviar informações necessárias à remontagem da informação.

Os serviços da classe A são tratados pelo protocolo AAL1; a AAL2 trata dos serviços da classe B; os serviços das classes C e D são tratados pelo protocolo AAL3/4 (resultado da junção das especificações AAL3 e AAL4); os serviços da classe C podem também ser tratados pelo protocolo AAL5, que é uma versão mais simples (e portanto, com menos *overhead*) que a AAL3/4.

2.7 Controle de Tráfego

O ATM opera com multiplexação estatística, significando que o meio de transmissão é compartilhado por diferentes fontes de tráfego, não havendo garantia de que o meio escoar totalmente o tráfego oferecido. Pode haver congestionamento dos *buffers* da rede e, neste caso, as células esperam em filas até que possam ser transmitidas, podendo ainda a capacidade finita das filas ser excedida, ocorrendo perda de células.

Para as conexões estabelecidas, os parâmetros de grau de qualidade de serviço mais importantes, em relação ao planejamento e gerência de redes ATM são o *atraso* e a *probabilidade de perda* das células.

As fontes de tráfego com taxa fixa de transmissão estão perfeitamente caracterizadas pela própria taxa, enquanto que as fontes de tráfego com taxa variável exigem uma descrição mais detalhada. A recomendação I.350 [19] apresenta os parâmetros de qualidade de serviço (QoS), desempenho da rede e descritores de tráfego.

O ITU-T e o ATM Forum preconizam que o objetivo básico do gerenciamento de tráfego é evitar que a rede e os terminais sejam submetidos a congestionamentos, de forma a atingir o desempenho de qualidade de serviço contratado. Uma regra adicional é promover o uso eficiente dos recursos da rede [1].

As seguintes funções formam a base para o gerenciamento e controle de tráfego, podendo ser utilizadas em combinações apropriadas:

- *controle de admissão de conexões* CAC (*Connection Admission Control*), que é definido como o conjunto de ações tomadas pela rede durante o estabelecimento de uma conexão, de forma a determinar se a requisição de conexão pode ser aceita ou não;
- *controles baseados em realimentações*, definidos como o conjunto de ações tomadas pela rede e pelos terminais, para regular o tráfego submetido às conexões de acordo com o estado dos elementos da rede;
- *controle de parâmetro do usuário* UPC (*Usage Parameter Control*), definido como o conjunto de ações tomadas pela rede para monitorar e controlar o tráfego, em termos de tráfego oferecido e validação da conexão ATM. Seu principal objetivo é proteger os recursos da rede contra a má utilização, intencional ou não, por fontes de tráfego que excedam os limites especificados pelos descritores de tráfego, tomando atitudes apropriadas, como o descarte de células.
- *controle de prioridade de descarte de células* CLP (*Cell Loss Priority Control*). O usuário pode utilizar em cada célula um *bit* CLP para permitir que a rede faça um descarte seletivo de células, protegendo as células de maior prioridade.

O tratamento de uma chamada que utiliza a rede ATM pode ser descrito de maneira sucinta como: o usuário declara os valores de alguns parâmetros que descrevem o comportamento da conexão requerida, além do desempenho em relação a perdas e atrasos. A rede utiliza estes valores de tráfego e desempenho para tomar a decisão se aceita ou rejeita a conexão. Se aceitar, a rede policia a conexão para garantir que os valores de tráfego declarados não sejam ultrapassados. Este processo previne congestionamentos na rede e garante que o desempenho requerido para cada conexão seja atingido.

Os parâmetros de tráfego e desempenho acordados pelo usuário e pela rede formam um contrato de tráfego. A decisão CAC é tomada nos nós de comutação de uma rede ATM. O mecanismo que policia o usuário para que este não exceda os valores do contrato de tráfego é o UPC, que opera na entrada da rede. Para desenvolver algoritmos para os mecanismos CAC e UPC é necessário estudar as características das fontes de tráfego e sua multiplexação nos *buffers* da rede, em termos de atraso e perda de células.

A Tabela 2.1 ilustra os intervalos de ação das operações da rede ATM.

Operação	Controle de Tráfego e de Congestionamento	Ação
Tempo de inserção de Células $\approx 3 \mu s$	UPC, conformação do tráfego controle de prioridade descarte seletivo de células	Células
Tempo de propagação $\approx 10 ms$	Gerenciamento rápido de recursos	Controle de fluxo sobre uma célula na rede
Duração da conexão $\approx 180 s$	CAC	Aceitar ou não uma conexão
Médio prazo $\approx horas$	Gerenciamento de recursos da rede Reconfiguração do encaminhamento da rede	Qualquer conexão
Longo prazo $\approx meses$	Planejamento da rede • Dimensionamento do enlace	Rede • Banda efetiva

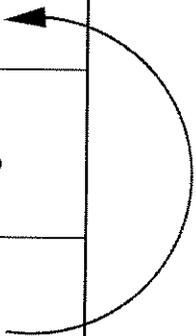


Tabela 2.1: Intervalos de ação na rede ATM.

2.8 Categorias de Serviço - ATM Forum

A tecnologia ATM suporta uma grande variedade de aplicações, provendo-lhes qualidades de serviço diferenciadas através do controle de tráfego. Os serviços são divididos em categorias para associar a cada uma delas os parâmetros de tráfego e de qualidade de serviço requeridos da rede.

2.8.1 Parâmetros de Tráfego

As características do tráfego devem ser descritas por um conjunto de parâmetros de tráfego padronizados, que possibilitem a operação das funções de controle de admissão de conexão CAC (*Connection Admission Control*) e do controle dos parâmetros do usuário UPC (*Usage Parameter Control*). Os procedimentos de admissão de conexão CAC utilizam os descritores de tráfego da conexão para alocar recursos da rede e para fornecer parâmetros para a operação do UPC.

Os parâmetros de tráfego quantitativos são: *taxa de pico de transmissão de células* PCR (*Peak Cell Rate*), *taxa média* SCR (*Sustainable Cell Rate*), *tamanho máximo do surto* MBS (*Maximum Burst Size*), e *taxa mínima de células* MCR (*Minimum Cell Rate*).

Um contrato de tráfego especifica as características negociadas de uma conexão e inclui o descritor de tráfego da fonte, a *tolerância à variação do atraso da célula* CDVT (*Cell Delay Variation Tolerance*) e a definição de conformidade das células.

A adequação do fluxo de células de uma conexão aos parâmetros SCR e PCR (e o CDVT associado), pode ser avaliada por um algoritmo denominado *algoritmo genérico de taxa de células* GCRA(T, Δ), onde T é o intervalo entre chegadas de células e Δ é a tolerância à variação deste intervalo.

Para a descrição do algoritmo GCRA (*Generic Cell Rate Algorithm*), considere a_k o tempo de chegada da k -ésima célula no ponto de medida, e r_k o tempo de chegada de referência. Os valores de r_k são atualizados conforme o seguinte procedimento:

Inicialização: $r_0 = a_0$; $r_1 = a_0 + T$; $k = 1$.

Operação:

$$r_{k+1} = \begin{cases} a_k + T & ; \quad a_k \geq r_k \\ r_k + T & ; \quad r_k - \Delta \leq a_k < r_k \\ r_k & ; \quad a_k < r_k - \Delta \rightarrow \text{célula não conforme} \end{cases}$$

$$k \leftarrow k + 1$$

O algoritmo é inicializado com a chegada da primeira célula. Supondo-se $a_0 = 0$, a próxima célula é esperada no instante de referência T . Se a célula chegar antes de $T - \Delta$, não estará em conformidade. Se chegar dentro do intervalo $T - \Delta$, ainda será considerada conforme. Neste caso, o algoritmo atualiza o instante referência, no qual a nova célula deverá chegar, adicionando T ao valor de referência. Se a célula chegar depois de T , estará em conformidade e o algoritmo atualiza o valor de referência como o instante de chegada da célula acrescido de T .

Seja $N(t)$ o número de células em conformidade com o algoritmo GCRA(T, Δ), no intervalo $[0, t]$, tem-se [1, 24]:

$$N(t) \leq \left\lfloor 1 + \frac{t + \Delta}{T} \right\rfloor \quad (2.1)$$

onde $\lfloor x \rfloor$ é o maior inteiro menor do que x .

Para t suficientemente grande, a taxa de células em conformidade é T^{-1} , não importando o valor da tolerância Δ .

As funções da camada ATM podem introduzir variação no atraso das células de uma conexão. As células de uma dada conexão podem ser atrasadas enquanto são inseridas células de uma outra conexão no multiplexador. Analogamente, algumas células podem ser atrasadas enquanto são inseridas células de controle da camada física ou células de manutenção OAM (*Operation and Maintenance*). Conseqüentemente, o tempo entre chegadas consecutivas de células pode diferir do intervalo mínimo entre chegadas de células PCR^{-1} , o mesmo ocorrendo para o tempo médio entre chegadas de células e o intervalo SCR^{-1} .

O teste de conformidade em relação à taxa de pico é obtido fazendo-se $T = PCR^{-1}$ e $\Delta = CDVT$, ou seja, para GCRA(PCR^{-1} , $CDVT$).

O teste de conformidade para a taxa média SCR é mais complexo do que para a taxa de pico, e depende dos parâmetros PCR, CDVT e do tamanho máximo do surto MBS. O cálculo do parâmetro Δ de tolerância à variação da chegada é dado por: $\Delta = BT + CDVT$, onde a *tolerância a surtos* BT (*Burst Tolerance*) é obtida através da expressão [1]:

$$BT = (MBS - 1) \left(\frac{1}{SCR} - \frac{1}{PCR} \right) \quad (2.2)$$

O número de células $N(t)$ que podem ser emitidas com intervalo entre chegadas de células maior que PCR^{-1} e ainda estar em conformidade com o GCRA (SCR^{-1} , $BT + CDVT$) é limitado por:

$$N(t) \leq \begin{cases} \lfloor 1 + SCR(t + BT) \rfloor & ; t \geq SCR^{-1} MBS \\ \lfloor 1 + t PCR \rfloor & ; \text{caso contrário} \end{cases} \quad (2.3)$$

O intervalo mínimo entre surtos sucessivos de tamanho MBS, t_{\min} , em conformidade com o GCRA(SCR^{-1} , $BT + CDVT$), é dado por [1]:

$$t_{\min} \geq \frac{1}{SCR} + BT \quad (2.4)$$

A Figura 2.1.a mostra o resultado da aplicação do algoritmo GCRA(3.0, 0) a um tráfego com taxa de pico de 1 célula por segundo. Nota-se que 2 em cada 3 células não são conformes. Para $t = 21$ o número de células em conformidade é $N(t) = 8$, obedecendo a equação (2.3).

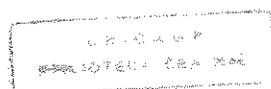
A Figura 2.1.b mostra a possibilidade de aceitação de surto considerando MBS = 3 e CDVT = 0, resultando em $BT = (3 - 1)(3 - 1) = 4$, conforme a equação (2.2). Nota-se que apenas um surto inicial foi aceito devido à entrada ser continuamente na taxa de pico. Observa-se que o número de células em conformidade $N(t) = 9$, com $t = 21$, obedece a equação (2.3).

A Figura 2.1.c mostra que surtos ocorrendo de 7 em 7 segundos ($t_{\min} = 7$) são aceitos, confirmando o resultado de 2.4.

A estratégia mais simples, utilizada pelo UPC para tratar as células que não satisfazem o contrato de tráfego, é descartá-las imediatamente.

A rede permite a demarcação de células para posterior descarte, em casos de congestionamento. Uma célula com o *bit* de prioridade de perda CLP = 0 (*Cell Loss Priority*) pode ser marcada com CLP = 1 e passada adiante.

O usuário pode negociar dois contratos de prioridade para a perda de células, conforme ilustrado na Figura 2.2, um para as células com CLP = 0 e outro válido para todas as células. A função UPC testa o fluxo com CLP = 0 e, em seguida, o fluxo combinado



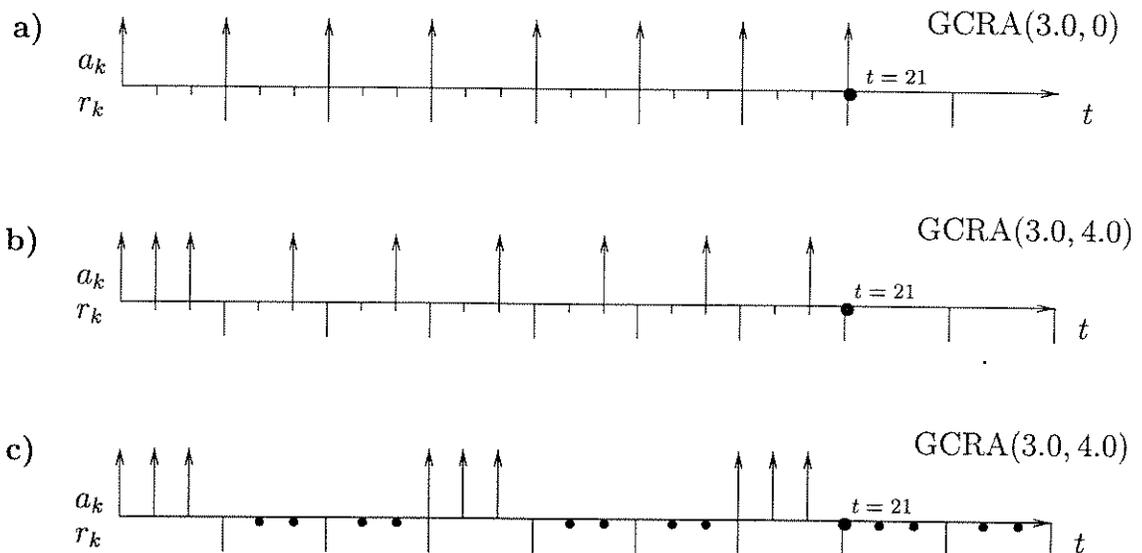


Figura 2.1: Interpretação do GCRA(T, Δ). As retas verticais curtas, abaixo da linha horizontal, indicam os instantes de chegadas de células, e as verticais longas os instantes de referência de chegada. As setas verticais indicam as células conformes. Os pontos pretos pequenos marcam o tempo em segundos e os pontos maiores marcam o tempo $t = 21$.

$CLP \in \{0, 1\}$. Se a opção de demarcação é utilizada, uma célula com $CLP = 0$ não concordante é marcada, mas ainda é considerada parte do fluxo $CLP \in \{0, 1\}$, sendo submetida ao segundo teste.

O objetivo é descartar células com menor prioridade, de forma a proteger as células de maior prioridade. A rede não diferencia as células marcadas pela fonte daquelas marcadas pela função UPC. A função UPC implementa as seguintes regras:

- Uma célula com $CLP = 0$ que satisfaz o contrato para $CLP = 0$ é passada adiante.
- Uma célula com $CLP = 0$ que não satisfaz o contrato para $CLP = 0$, mas satisfaz o contrato para $CLP \in \{0, 1\}$ é marcada com $CLP = 1$ e passada adiante.
- Uma célula com $CLP = 1$ que satisfaz o contrato para $CLP \in \{0, 1\}$ é passada adiante.
- Uma célula com $CLP = 1$ que não satisfaz o contrato para $CLP \in \{0, 1\}$ é descartada.
- Uma célula que não satisfaz ambos os contratos é descartada.

2.8.2 Parâmetros de Qualidade de Serviço

Os parâmetros de QoS (*Quality of Service*) negociados entre o terminal do usuário e a rede são [1]:

- *Variação pico a pico do atraso das células* - ppCDV (*Peak-to-Peak Cell Delay Variation*)
- *Atraso máximo de transferência de célula* - MaxCTD (*Maximum Cell Transfer Delay*)

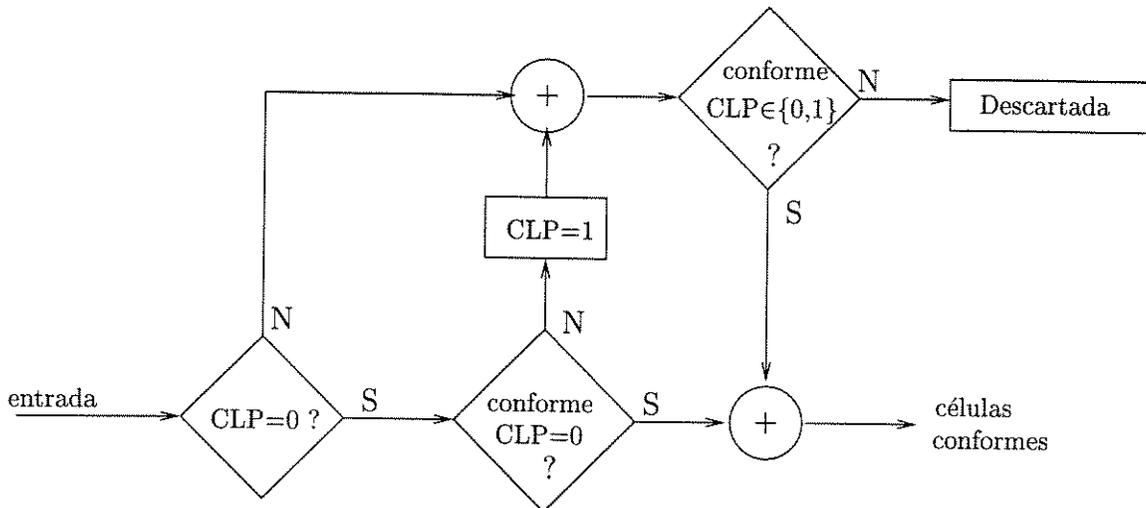


Figura 2.2: Contrato com dois níveis de prioridade

- Razão de perda de células - CLR (*Cell Loss Ratio*)

Os parâmetros de QoS são definidos através de um método de medida das características de transferência da rede observadas, em uma célula (ou numa seqüência de células), durante a existência da conexão.

O *atraso de transferência de célula* entre dois pontos da rede é definido como o tempo transcorrido entre o evento de saída de uma célula (fim da emissão do primeiro *bit*) no primeiro ponto, e o evento de entrada da célula (fim da recepção do último *bit*) no segundo ponto.

Este atraso de transferência é estatístico e possui um valor mínimo, composto pelo tempo de propagação no meio físico e pelas componentes fixas dos atrasos devido ao processamento nos comutadores.

O atraso máximo de transferência de célula MaxCTD, entre dois pontos de medida de uma conexão, é o valor de referência para o atraso que só pode ser excedido por uma proporção α de células da conexão. O parâmetro α deve ser sempre menor que a razão de perda de células CLR.

A variação do atraso das células pode se referir à observação em um único ponto ou entre dois pontos de gerenciamento. O desempenho negociado para a variação do atraso é expresso através do parâmetro ppCDV.

O parâmetro CLR é o valor de razão de perda que a rede concorda em oferecer durante a existência da conexão e é definido como a razão entre o número de células perdidas e o número total de células transmitidas.

Acúmulo de Parâmetros de QoS

O ATM Forum definiu um algoritmo de acúmulo, baseado em hipóteses de pior caso, para prover estimativas dos valores fim a fim dos parâmetros de atraso nos protocolos de sinalização com a finalidade de aceitação ou não de uma chamada.

Aplicando-se o algoritmo para determinação da estimativa da variação pico a pico do atraso ppCDV, cada comutador recebe o ppCDV acumulado e adiciona sua contribuição

de atraso. Da mesma forma, a estimativa do atraso máximo MaxCTD implica em cada comutador receber o MaxCTD acumulado e adicionar a sua contribuição de atraso.

Na sinalização, durante a fase de análise da chamada, o MaxCTD é acumulado somente na direção de ida (*forward*). Entretanto, o CDV é acumulado em ambas as direções (*forward and backward*). Devido ao fato de que o atraso fixo numa direção é igual ao da direção contrária, o MaxCTD para a direção de volta (*backward*) é calculado como:

$$\text{MaxCTD}_B = \text{CDV}_B + \text{MaxCTD}_F - \text{CDV}_F \quad (2.5)$$

O parâmetro CLR não é acumulado. Cada elemento de rede aceita ou rejeita a chamada baseado em uma comparação entre a razão de perda suportada pelo elemento de rede e o parâmetro CLR solicitado.

2.8.3 Definição das Categorias de Serviço

O ATM Forum definiu um conjunto de cinco categorias de serviço:

- CBR (*Constant Bit Rate*) taxa de bits constante
- rt-VBR (*Real-Time Variable Bit Rate*) taxa de bits variável em tempo real
- nrt-VBR (*Non-Real-Time Variable Bit Rate*) taxa de bits variável não em tempo real
- ABR (*Available Bit Rate*) taxa de bits disponível
- UBR (*Unspecified Bit Rate*) taxa de bits não especificada

Mecanismos de policiamento, encaminhamento e alocação de recursos, são definidos de maneira distinta para cada categoria. Uma vez estabelecida a conexão, a qualidade de serviço requerida é garantida a todas as células que estiverem em conformidade.

CBR - (taxa de *bits* constante)

O serviço CBR é utilizado por aplicações que podem requerer uma alocação de banda fixa continuamente disponível enquanto durar a conexão. A fonte CBR é caracterizada pela taxa de pico PCR e a tolerância à variação do atraso CDVT correspondente. Utilizando a capacidade CBR a fonte pode emitir células na taxa de pico negociada ou abaixo desta. Os parâmetros de qualidade de serviço negociados para esta categoria são: variação pico a pico do atraso de células ppCDV, atraso máximo de transferência de célula MaxCTD e a razão de perda de células CLR. Aplicações de áudio e vídeo comprimido em tempo real também podem utilizar CBR, mas esta categoria é ineficiente para tráfegos com taxa de *bit* variável.

rt-VBR - (taxa de *bits* variável em tempo real)

O rt-VBR é utilizado por aplicações que requerem fortes restrições ao atraso, sendo apropriado para voz e vídeo com taxa variável (videoconferência). As fontes transmitem a uma taxa que varia com o tempo, sendo caracterizadas através da taxa de pico PCR, da tolerância à variação do atraso CDVT, da taxa média SCR, e do tamanho máximo do surto MBS, além de negociarem os seguintes parâmetros de QoS: variação pico a pico do

atraso de células ppCDV, atraso máximo de transferência de célula MaxCTD e a razão de perda de células CLR.

nrt-VBR - (taxa de *bits* variável não em tempo real)

O serviço nrt-VBR é utilizado por aplicações com tráfego em surto e sem restrições ao atraso. O envio de um *email* multimídia é exemplo de serviço que requer esta categoria. O nrt-VBR é caracterizado pelos seguintes parâmetros de tráfego: PCR, CDVT correspondente, SCR e MBS. A categoria nrt-VBR negocia apenas a razão de perda de células CLR, uma vez que não há restrições de atraso.

ABR - (taxa de *bits* disponível)

Na categoria ABR a taxa de pico PCR (e CDVT correspondente) e a taxa mínima MCR são os únicos parâmetros negociados durante o estabelecimento da conexão. A rede informa as condições de tráfego à fonte, através de realimentação por um mecanismo de controle de fluxo, permitindo que a fonte se adapte, aumentando ou diminuindo a taxa de pico PCR, conforme haja ou não banda disponível, minimizando a perda de células. Nenhum parâmetro de QoS é negociado.

UBR - (taxa de *bits* não especificada)

O serviço UBR é utilizado por aplicações que não possuem restrições de atraso e que fazem o próprio controle de erro e de fluxo. Transferência de arquivos e *email* são exemplos destas aplicações.

As células UBR são transmitidas quando há banda disponível e descartadas em casos de congestionamento. Não há um mecanismo de realimentação das condições de tráfego para o usuário e nenhum parâmetro de QoS é negociado.

Os únicos parâmetros de tráfego especificados são o PCR e o CDVT correspondente. As funções de CAC e UPC podem ou não utilizar o valor especificado do PCR.

As Tabelas 2.2 e 2.3 apresentam algumas propriedades das categorias de serviço.

Características	CBR	rt-VBR	nrt-VBR	ABR	UBR
Garantia de banda disponível	sim	sim	sim	parcial	não
Adaptável para tempo real	sim	sim	não	não	não
Adaptável para tráfego em surto	não	não	sim	sim	sim
Controle de congest. por realimentação	não	não	não	sim	não

Tabela 2.2: Características das categorias de serviço do ATM Forum

2.9 Capacidades de Transferência - ITU-T

O ITU-T definiu quatro *capacidades de transferência* ATM [20, 24]:

	CBR	rt-VBR	nrt-VBR	ABR	UBR
Tráfego	PCR e CDVT				
		SCR e MBS	SCR e MBS		
				MCR	
QoS	ppCDV	ppCDV			
	MaxCTD	MaxCTD			
	CLR	CLR	CLR		

Tabela 2.3: Parâmetros das categorias de serviço do ATM Forum

- DBR (*Deterministic Bit Rate*) taxa de bits determinística
- SBR (*Statistical Bit Rate*) taxa de bits estatística
- ABT (*ATM Block Transfer*) transferência de bloco ATM
- ABR (*Available Bit Rate*) taxa de bits disponível

Algumas categorias de serviço do ATM Forum são equivalentes às capacidades de transferência do ITU-T, porém com nomes diferentes: CBR é chamada DBR e nrt-VBR é chamada SBR.

O ITU-T definiu os seguintes parâmetros de qualidade de serviço: *razão de perda de células* CLR, *atraso de transferência de células* CTD (equivalente ao MaxCTD do ATM Forum), e *variação do atraso das células* CDV (equivalente ao ppCDV do ATM Forum).

DBR - (taxa de bits determinística)

A capacidade de transferência com taxa de *bit* determinística DBR é utilizada por serviços que podem requerer uma quantidade estática de banda continuamente disponível, enquanto durar a conexão. É caracterizada pela taxa de pico de células PCR e o CDVT associado.

A capacidade DBR negocia os seguintes parâmetros de QoS: razão de perda de células CLR, atraso de transferência de células CTD, e variação do atraso das células CDV.

SBR - (taxa de bits estatística)

Na capacidade de transferência com taxa de *bits* estatística SBR, um conjunto maior de parâmetros é necessário para especificar o tráfego com taxa de *bits* variável. Além da taxa de pico PCR e CDVT associado, requer a especificação da taxa média SCR e CDVT associado, além da *tolerância intrínseca a surtos* IBT (*Intrinsec Burst Tolerance*).

Note que, diferentemente da categoria VBR do ATM Forum, a capacidade SBR requer a especificação da tolerância a surtos IBT, calculando o tamanho máximo do surto MBS, através de:

$$MBS = 1 + \lfloor IBT \left(\frac{1}{SCR} - \frac{1}{PCR} \right) \rfloor \quad (2.6)$$

O ITU-T não especificou o uso da SBR para aplicações em tempo real. A capacidade de transferência SBR não suporta restrições de atraso, tendo como único parâmetro de QoS negociado o CLR.

ABT - (transferência de bloco ATM)

A capacidade de transferência ABT provê serviços nos quais as características de transmissão são negociadas com base em blocos de células.

Um bloco é um grupo de células de uma conexão ATM, delimitadas por duas células de gerenciamento de recursos (*Resource Management Cells*), uma no início do bloco e outra no fim. A célula final de um bloco pode ser a célula inicial do bloco subsequente.

O ABT utiliza a taxa de pico de células PCR e o CDVT associado, a taxa média SCR com o CDVT associado e a tolerância a surtos IBT (que indiretamente define o tamanho máximo do surto MBS). Estes parâmetros não podem ser renegociados durante a transmissão de um bloco.

Durante a fase de estabelecimento da chamada, os parâmetros de qualidade de serviço negociados são: CDV e CLR. Durante a existência da conexão, cada novo bloco pode ter os parâmetros de QoS redefinidos.

ABR - (taxa de *bits* disponível)

A categoria ABR foi definida para suportar aplicações que possuem a habilidade de alterar a taxa de transferência em função do estado da rede.

Espera-se que um usuário que adapte o tráfego atinja uma razão de perda de células CLR aceitável. Os atrasos e suas variações não são controlados.

No estabelecimento de uma conexão ABR o usuário deve informar à rede a máxima banda requerida (PCR e o CDVT associado) e a mínima banda utilizável (MCR). O MCR também pode ser especificado com valor zero. A banda disponível pela rede pode variar, mas não deve ser menor que MCR.

	DBR	ABT	SBR	ABR
Tráfego	PCR e CDVT	PCR e CDVT	PCR e CDVT	PCR e CDVT
		SCR, IBT e CDVT	SCR, IBT e CDVT	
				MCR
QoS	CDV	CDV		
	CTD			
	CLR	CLR	CLR	

Tabela 2.4: Parâmetros das capacidades de transferência ATM do ITU-T

A Tabela 2.4 mostra os parâmetros de tráfego e de qualidade de serviço de cada capacidade de transferência ATM.

2.10 Conclusão

Os aspectos básicos da tecnologia ATM foram apresentados neste capítulo. O modo de transferência assíncrono é uma tecnologia de alto desempenho com comutação e multiplexação de células, utilizando pacotes de tamanho fixo para transportar diferentes tipos de tráfego.

O tráfego B-ISDN é estatisticamente heterogêneo (voz, vídeo, dados, sinalização da rede) solicitando uma variedade de qualidades de serviço, variando entre limites determinísticos e estatísticos para a probabilidade de perda de células e/ou atrasos. Do ponto de vista de uma rede ATM, cada conexão consiste de uma fila de pacotes de 53 *bytes* chamados células; células associadas a uma dada conexão seguem a mesma rota, chamada circuito virtual, e chegam ordenadas ao destino.

O ATM propiciará o surgimento de novas oportunidades de receita para seus provedores, através das múltiplas classes de serviço, interconexão de redes locais em alta velocidade, voz, vídeo e futuras aplicações multimídia no mercado comercial, a curto prazo, e a longo prazo no mercado residencial.

Capítulo 3

Encaminhamento de Tráfego em Redes Públicas ATM

Modelos para perda e atraso das células, em uma rede ATM, são discutidos neste capítulo visando a determinação de planos de encaminhamento para as conexões. A estratégia de encaminhamento é resultado da otimização de um critério de desempenho que considera concomitantemente a perda de células e o atraso de transferência ponto a ponto.

3.1 Introdução

A questão do encaminhamento de tráfego nas redes ATM (*Asynchronous Transfer Mode*) tem sido abordada sob vários aspectos [8, 26, 27, 40].

Dois pontos distinguem esta proposta das demais da literatura: a inserção do problema de encaminhamento no contexto de implantação da tecnologia ATM nas redes públicas de telecomunicações; e o objetivo de simplificar ao máximo as funções de interesse (por exemplo, perda de células e atraso nos *buffers*), com o intuito de viabilizar a solução do problema de otimização associado para redes de porte significativo.

O problema de encaminhamento semi-permanente nas redes ATM e as funções de perda de células e atraso nos *buffers* são discutidos, resultando na formulação de um problema de otimização para o qual a maior parte dos algoritmos de fluxo ótimo de multiprodutos pode ser aplicada [3].

A organização do capítulo é como segue: preliminarmente, discute-se a introdução da tecnologia ATM nas redes públicas de telecomunicações. Após a formulação matemática do problema de encaminhamento, discute-se o ponto essencial que são as funções de interesse e a estimativa do ponto de mínimo, terminando com uma análise do algoritmo de solução e um exemplo.

3.2 Tecnologia ATM nas Redes Públicas

A tecnologia ATM surgiu como resposta à necessidade de integração de serviços (telefonia, dados, etc.) das operadoras de telecomunicações que, em vários países, tais como o Brasil, se organizam na forma de empresas públicas.

As empresas operadoras são tradicionais fornecedores de dois tipos principais (em termos monetários e em volume de tráfego) de serviços de telecomunicações: telefonia e linhas privadas. É nesse cenário que a tecnologia ATM tem o seu maior trunfo.

A tecnologia ATM é uma rede de pacotes com características bem particulares. É orientada a conexão, podendo atender seus usuários de maneira individualizada, alocando recursos na medida do solicitado. Pode realizar conexões, ditas semi-permanentes, que emulam as linhas privadas, ou ainda operar por conexão sob demanda, emulando os serviços telefônicos comutados.

Os pacotes na rede ATM (denominados células) são de tamanho fixo de 53 *bytes*, sendo 5 de cabeçalho e 48 de carga útil. Este pequeno *overhead* só é possível devido à rede ser orientada a conexão e ser projetada para operar sobre meios de transmissão, tais como fibras ópticas, com probabilidades de erro de *bit* muito pequenas.

Por ser uma rede de pacotes operando em multiplex assíncrono, a rede ATM transporta apropriadamente diferentes serviços com taxas variadas. Sobre a mesma rede, através das centrais de comutação e suas rotas (especificadas pelo ITU-T (*Standardization Sector of Telecommunication of the International Telecommunication Union*) em múltiplos de 155.52 *Mbps*), trafegando sobre sistemas de transmissão SDH (*Synchronous Digital Hierarchy*) ou PDH (*Plesiochronous Digital Hierarchy*), viajam trens de *bits* da mais variada gama de serviços. A tecnologia ATM foi projetada para ser invariante com o meio de transmissão.

Este capítulo analisa a rede ATM sob a perspectiva de sua introdução nas redes das operadoras de telecomunicações. Devido à flexibilidade inerente às redes ATM e à dificuldade e demora dos órgãos normativos para especificar padrões de interconexão das redes ATM com a rede telefônica convencional, houve um atraso considerável na introdução da tecnologia ATM nas empresas operadoras. Assim, os grandes fabricantes de equipamentos voltaram-se para o segmento das redes locais de computadores, no qual está havendo uma expansão considerável de redes ATM.

A rede ATM é analisada como uma rede de transporte para grandes usuários, tipicamente a interligação entre centrais telefônicas, ou a interligação entre redes locais de computadores. A topologia considerada é mostrada na Figura 3.1.

Note que as centrais telefônicas e as redes locais não ATM necessitam de interfaces para a conversão dos protocolos, o que tem limitado a expansão do uso das redes ATM nas empresas públicas de telecomunicações.

Em um primeiro momento da introdução da tecnologia ATM nas operadoras, estas podem prescindir de centrais com capacidade de conexão sob demanda, apenas com conexões semi-permanentes. Isto implica no uso de centrais que apenas comutam VP (*Virtual Path*), ao invés de centrais que comutam VP-VC (*Virtual Path — Virtual Channel*), mais complexas e caras que as primeiras.

Logo que as operadoras amortizarem as centrais VP, concomitantemente com a disseminação de equipamentos terminais ATM, as centrais originais podem naturalmente expandir-se para centrais com comutação VP-VC.

Este capítulo aborda uma das principais questões das redes ATM, que é o encaminhamento das células com o objetivo de majorar o fluxo de informação circulando pela rede, atendendo os parâmetros de qualidade de serviço. A determinação do encaminhamento é formulada como um problema de programação matemática e os modelos de perda de células e de atraso apropriados são analisados.

Um problema associado e que de fato antecede o de encaminhamento é o de dimen-

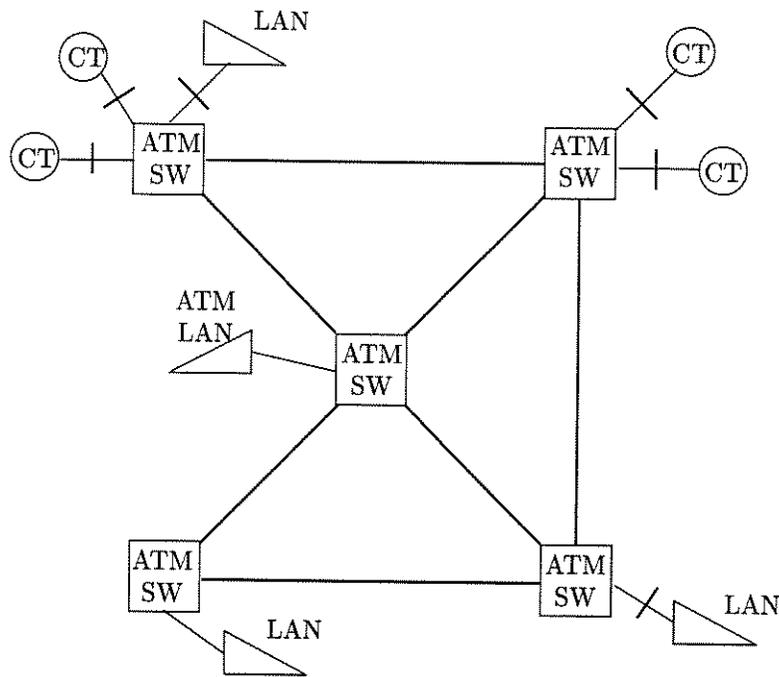


Figura 3.1: Configuração típica da introdução da tecnologia ATM nas redes das empresas operadoras de telecomunicações.

sionamento da rede ATM. Aqui, pressupõe-se que a rede tenha sido previamente dimensionada em função de uma demanda prevista para o ano horizonte. Durante a operação da rede, a demanda de tráfego é re-estimada com maior precisão e em intervalos menores de tempo. O re-encaminhamento é uma das técnicas para adaptar a rede a novas demandas sem exigir investimentos suplementares, aproveitando-se da flexibilidade das centrais ATM.

3.3 Encaminhamento do Tráfego

Considere uma rede descrita pelo grafo conexo $(\mathcal{N}, \mathcal{A})$, onde $\mathcal{N} = \{1, 2, \dots, N\}$ é o conjunto de índices dos nós e $\mathcal{A} = \{1, 2, \dots, M\}$ é o conjunto de índices dos arcos orientados. Associadas a cada arco, existem triplas que designam o nó origem, o nó destino e a capacidade em *Mbps* do enlace, denotadas $(o, d, C)_a$, $a \in \mathcal{A}$, $o, d \in \mathcal{N}$.

Em redes públicas ATM, os enlaces são modulares, com capacidades múltiplas de 155.52 *Mbps*. A determinação da topologia da rede, incluindo a localização e o porte das centrais de comutação, e as capacidades dos enlaces é considerada conhecida. Em geral, a estrutura da rede é resultado de uma análise técnico-econômica que dimensiona a rede para atender uma previsão de demanda. Uma vez implantada a rede, devido à sazonalidade do tráfego, a previsão de curto prazo para a demanda (por exemplo, hora de maior movimento) pode ser bastante distinta daquela utilizada durante a fase de planejamento.

Duas abordagens não excludentes são possíveis: reconfiguração da rede (por exemplo, usando a flexibilidade dos equipamentos *cross-connect*), e utilização de encaminhamentos alternativos àqueles inicialmente previstos.

Este capítulo considera apenas a segunda estratégia, obtendo novas propostas de enca-

minhamento através da resolução de problemas de programação matemática. A principal vantagem dessa abordagem é não implicar em custos adicionais em equipamentos (como seria o caso da utilização de *cross-connect*), mas apenas utilizar de maneira eficaz a flexibilidade do *software* de encaminhamento das centrais.

Seja $\mathcal{K} = \{1, 2, \dots, T\}$ o conjunto de índices das diferentes demandas (também denominadas produtos), cujos atributos são triplas $(o, d, \lambda)_k$, $k \in \mathcal{K}$, designando o nó origem, o nó destino e a taxa de informação (suposta constante e conhecida durante o período de encaminhamento), expressa em células por segundo, do produto k .

3.4 Formulação Matemática

O problema de programação matemática é formulado com o objetivo de repartir as demandas entre os possíveis caminhos que ligam cada origem ao respectivo destino, de maneira a otimizar uma função de interesse.

As funções consideradas são: perda total de células na rede, atraso médio de transporte, e suas combinações lineares. Estas duas funções estão entre os principais parâmetros de qualidade de serviço propostos pelo ITU-T e ATM Forum. Além disso, conforme será visto na seqüência, essas funções são convexas nas variáveis de decisão, e aditivamente separáveis por arco da rede. A variação do atraso de transporte não é tratada.

Para a formulação matemática do problema de otimização, algumas definições são necessárias.

Seja p_r um caminho factível para a demanda k , $k \in \mathcal{K}$. Um caminho é uma seqüência de arcos $[(i_o, i_1), (i_1, i_2), \dots, (i_{d-1}, i_d)]$ na qual o nó inicial de cada arco coincide com o nó final do arco precedente, com os nós i_o, \dots, i_d distintos entre si, em \mathcal{N} . Como por hipótese a rede é conexa, sempre existe pelo menos um caminho para cada um dos produtos. Os algoritmos de solução nas formulações arco-caminho possuem sempre uma fase inicial, que consiste na enumeração dos caminhos a serem considerados (para cada produto).

O conjunto de índices dos caminhos possíveis para cada produto k é denominado $\mathcal{P}_k = \{1, 2, \dots, \ell_k\}$, cada qual designando um caminho p_r , $r \in \mathcal{P}_k$. Denotando $x_{p_r}^k$ a taxa do produto k enviada pelo caminho p_r , $r \in \mathcal{P}_k$, tem-se:

$$x_a \triangleq \sum_{k \in \mathcal{K}} \sum_{\substack{r \in \mathcal{P}_k \\ a \in p_r}} x_{p_r}^k \quad (3.1)$$

que é a taxa de tráfego oferecido ao enlace a .

De forma geral, as perdas e atrasos nas redes dependem de todos os tráfegos de cada enlace da rede. Entretanto, em situações de carga leve (rede não muito congestionada), é razoável supor que o tráfego de cada enlace é uma variável estatisticamente independente dos demais tráfegos. Assim, a variável x_a e a capacidade C_a definem de maneira única os valores da perda de células e do atraso de transferência do enlace a , denotados pelas funções de interesse $f_a(x)$ (perda) e $g_a(x)$ (atraso).

A parte fundamental deste capítulo consiste na discussão e propostas de formulação das funções de interesse. Funções excessivamente simples podem não mais representar adequadamente o problema físico; por outro lado, funções muito complexas podem inviabilizar a resolução do problema de otimização associado. Nesse sentido, busca-se um compromisso entre esses dois extremos.

As variáveis de decisão $x_{p_r}^k$ representam a partição de carga do tráfego nos caminhos semi-permanentes. Nas redes ATM, cada enlace é caracterizado pelos parâmetros VP (*Virtual Path*) e VC (*Virtual Channel*). Cada caminho semi-permanente é constituído por uma seqüência de VP's (que são mantidos fixos dentro do período válido para um determinado encaminhamento). Para cada chamada, transação, etc., de um certo produto (demanda entre origem-destino) é feita uma conexão automática associando a cada enlace do caminho escolhido os VC's correspondentes. A implementação da política de repartição de carga, obtida como solução do problema de otimização, se faz através desta escolha dos VC's associados aos VP's segundo os percentuais resultantes da otimização. De fato, como o número médio de células geradas por uma chamada depende do seu tipo (telefonia, dados, etc.), essa repartição é aproximada.

É importante observar que a implementação desta estratégia necessita apenas de comutadores VP, menos complexos e mais baratos que os comutadores de VP-VC.

A Figura 3.2 ilustra, para uma rede de 4 nós e 4 enlaces, os dois caminhos associados a um único produto (com taxa λ) circulando na rede (entrando pelo nó 1, com destino ao nó 4).

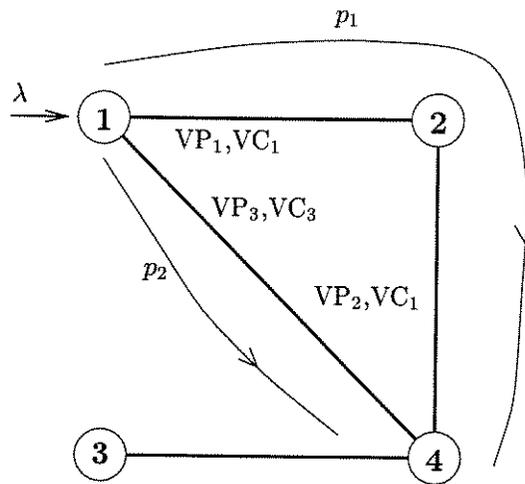


Figura 3.2: Exemplo de caminhos semi-permanentes na rede ATM.

A perda normalizada de células na rede é dada por:

$$F = \frac{1}{\gamma} \sum_{k \in \mathcal{K}} \sum_{r=1}^{\ell_k} x_{p_r}^k \sum_{a \in p_r \in \mathcal{P}_k} f_a(x_a) \quad ; \quad \gamma \triangleq \sum_{k \in \mathcal{K}} \lambda_k \quad (3.2)$$

O valor dado por (3.2) super-estima a perda, pois o tráfego oferecido em um enlace a qualquer de um caminho é um pouco menor do que x_a , devido a perdas nos enlaces anteriores do mesmo caminho. Como entretanto as probabilidades de perda de células são muito baixas na rede ATM, essa aproximação é bastante apropriada.

A expressão para a espera média na rede é dada por:

$$G = \frac{1}{\gamma} \sum_{k \in \mathcal{K}} \sum_{r=1}^{\ell_k} x_{p_r}^k \sum_{a \in p_r \in \mathcal{P}_k} g_a(x_a) \quad (3.3)$$

Observe que as funções $f(\cdot)$ e $g(\cdot)$ são sempre positivas, sendo que, quando $x \rightarrow 0$, $f(\cdot)$ necessariamente tende a zero e $g(\cdot)$ tende para o valor do atraso de propagação do enlace.

As restrições operacionais do problema podem ser traduzidas matematicamente por:

$$\sum_{r=1}^{\ell_k} x_{p_r}^k = \lambda_k ; x_{p_r}^k \geq 0 ; \forall k \in \mathcal{K} \quad (3.4)$$

Nota-se que as tradicionais restrições de capacidade (do tipo $x_a \leq C_a$)¹ não aparecem explicitamente na expressão (3.4). Como a rede ATM opera com perdas, $x_a > C_a$ é ainda uma solução factível, correspondendo entretanto a perdas e atrasos altíssimos, que são penalizados nas funções $f(\cdot)$ e $g(\cdot)$. Essa propriedade facilita a escolha dos algoritmos de solução.

Durante a operação da rede, coloca-se o seguinte problema de encaminhamento:

$$\min_{x_{p_r}^k \in \mathcal{X}} \beta F + (1 - \beta)G \quad (3.5)$$

onde \mathcal{X} representa o conjunto dos pontos factíveis (que satisfazem as restrições (3.4)) e β é um escalar entre 0 e 1, que pondera a relação de interesse entre perda e atraso.

3.5 Modelos para Critérios de Desempenho

As redes faixa larga, baseadas na transmissão de células ATM, suportam uma variedade de tipos de tráfego, diferindo consideravelmente do modelo tradicional de rede telefônica comutada por circuitos, projetada inicialmente para transportar voz. As redes ATM diferem também das redes comutadas por pacotes, que trafegam quase que exclusivamente dados.

Os modelos de tráfego possuem um papel importante no projeto destes tipos de rede. Em particular, chegadas poissonianas e tempo de serviço exponencial têm servido como excelentes modelos para tráfego de voz comutado por circuito. Chegadas poissonianas e modelos exponenciais de tamanho de pacote têm sido usados extensivamente no estudo do desempenho de redes comutadas por pacotes. Entretanto, estes modelos bem estabelecidos não são suficientes para a análise e síntese das redes de faixa larga do futuro. A integração de voz, vídeo, imagens, tráfego gerado por computadores (com tráfegos em rajadas e longas transferências de arquivos), cada qual com sua própria requisição de qualidade de serviço, solicita o desenvolvimento de novos modelos de tráfego.

No processo de síntese da rede, vários modelos complementares são necessários. No projeto utilizam-se modelos simples, que permitem a obtenção de uma versão preliminar da rede. Em seguida, durante a fase de análise, modelos mais elaborados possibilitam um ajuste fino. Numa fase final de validação, modelos minuciosos, em geral a eventos discretos, são usados para homologar o projeto [2, 29].

Neste capítulo são propostos modelos simples para as perdas e atrasos das células na rede, mas que permitem a determinação de estratégias de encaminhamento apropriadas para as redes ATM.

Para a obtenção do modelo apropriado para o problema de encaminhamento, considera-se a Figura 3.3 que mostra um comutador com *buffers* nas saídas.

Analisando um dos enlaces de saída com capacidade C , seja λ a taxa de células a ser transmitida por unidade de tempo e $1/\mu$ o tempo para transmitir uma célula. Como cada

¹Capacidade C_a em células por segundo.

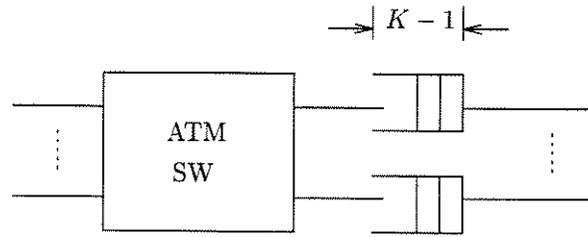


Figura 3.3: Comutador ATM com *buffers* nas saídas.

célula tem 53 *bytes* de comprimento, tem-se:

$$\mu = \frac{C}{53 \times 8} \quad (3.6)$$

Cada um dos enlaces de comunicação conectado nas saídas do comutador é modelado por uma fila M/M/1/K, isto é, chegadas poissonianas, duração exponencial negativa, um servidor e *buffer* de tamanho $(K - 1)$ [14]. Definindo o tráfego oferecido como:

$$\rho \triangleq \frac{\lambda}{\mu} \quad (3.7)$$

e denotando por P a probabilidade de perda de célula, tem-se:

$$P = \frac{1}{1 - \rho^{K+1}} (1 - \rho) \rho^K \quad (3.8)$$

e o tempo médio de espera no *buffer* é dado por:

$$E\{W\} = \frac{1}{\mu} \frac{\rho}{1 - \rho} \frac{[1 - K\rho^{K-1} + (K - 1)\rho^K]}{1 - \rho^K} \quad (3.9)$$

A fila M/D/1/K descreve as redes ATM (cujas células são de tamanho fixo) de maneira mais apropriada que a M/M/1/K, mas não dispõe de solução analítica fechada e deve ser tratada numericamente [2]. O Apêndice B apresenta de forma detalhada as filas M/M/1/K e M/D/1/K. O Apêndice C apresenta as filas MMPP/D/1/K que possuem uma modelagem ainda mais complexa que a das filas M/D/1/K.

As Figuras 3.4 e 3.5 mostram a comparação entre as filas M/D/1/K e M/M/1/K, para $\mu = 1$, considerando respectivamente o tempo médio de espera e a probabilidade de perda de células nos *buffers*. Nota-se (Figura 3.4) que o tempo de espera (para $\rho < 0.7$) nas filas M/M/1/K é praticamente o dobro do tempo nas filas M/D/1/K; na verdade, este resultado apenas confirma um comportamento largamente conhecido para filas com *buffers* de tamanho infinito [14].

A probabilidade de perdas no caso M/M/1/K é muito maior do que no caso M/D/1/K, como pode ser visto, para $\rho < 0.7$, na Figura 3.5. Assim, a hipótese de duração exponencial para o tempo de serviço majora significativamente a probabilidade de perda de células.

No contexto da determinação das estratégias de encaminhamento, os valores absolutos das probabilidades de perda e de espera não são relevantes em si, importando mais o fato de serem semelhantes as formas de dependência dessas quantidades em relação ao tráfego oferecido, para os dois modelos de duração do tempo de serviço aqui considerados.

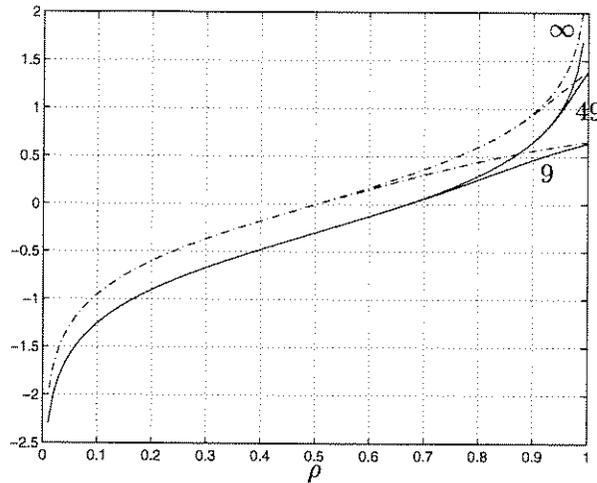


Figura 3.4: Logaritmo do tempo médio de espera nos *buffers* de tamanho 9, 49 e ∞ . Curvas cheias referentes à fila $M/D/1/K$, e as tracejadas referentes à fila $M/M/1/K$, versus o tráfego oferecido ρ .

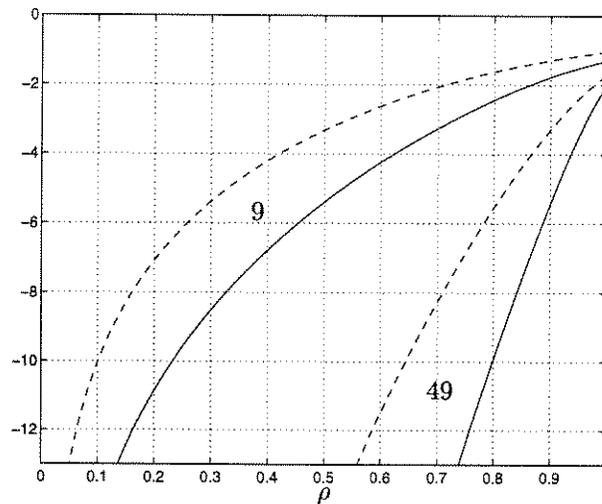


Figura 3.5: Logaritmo da probabilidade de perda de células com *buffer* de tamanho 9 e 49, curvas cheias referentes à fila $M/D/1/K$, e as tracejadas referentes à fila $M/M/1/K$, versus o tráfego oferecido ρ .

3.6 Exemplo com Dois Enlaces

Considerando-se uma rede com apenas dois enlaces em paralelo, um com capacidade $C_1 = 155.52 \text{ Mbps}$, e o outro com capacidade $C_2 = 2C_1$ e $\lambda = \mu_1$, mostrada na Figura 3.6, pode-se determinar o valor do tempo de espera e a perda de células em função da parcela α de tráfego enviada ao enlace 1.

A Figura 3.7 mostra o tempo médio de espera das células nos *buffers* em função da porcentagem α de envio das células pelo enlace C_1 , e $(1 - \alpha)$ por C_2 . Observa-se que os pontos de mínimo da função espera são praticamente os mesmos para os dois modelos considerados ($M/M/1/K$ e $M/D/1/K$).

Comportamento similar ocorre na Figura 3.8, que mostra a probabilidade de perda de células em função de α , para os dois modelos considerados.

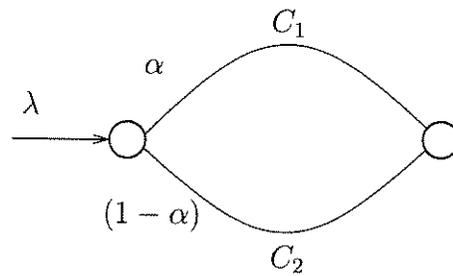


Figura 3.6: Exemplo de rede com dois enlaces com capacidades $C_1 = 155.52 \text{ Mbps}$ e $C_2 = 2C_1$, $\lambda = \mu_1$.

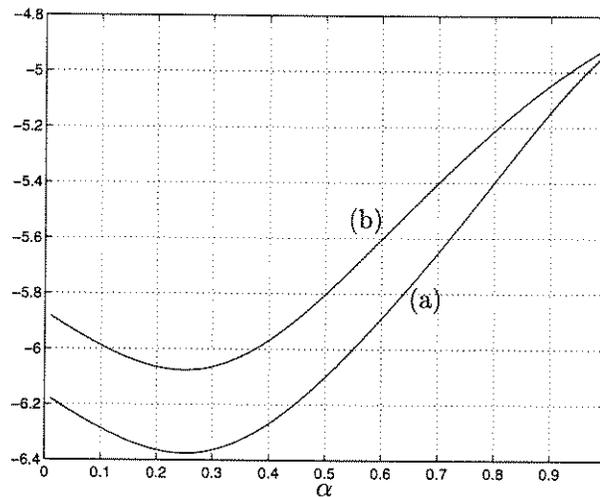


Figura 3.7: Logaritmo do tempo médio de espera referente às filas M/D/1/K (a) e M/M/1/K (b), versus α .

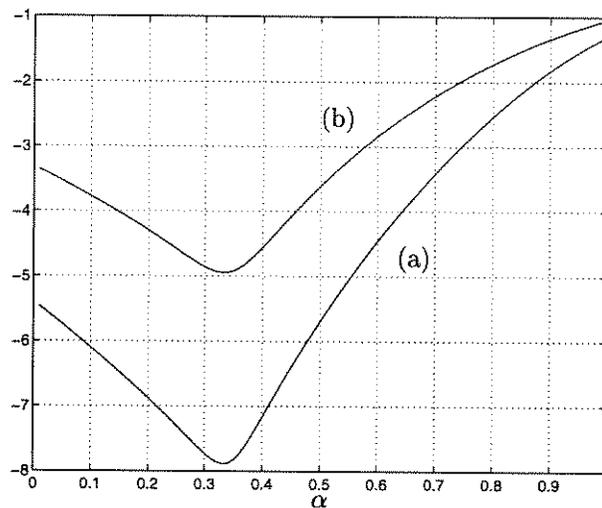


Figura 3.8: Logaritmo da probabilidade da perda de células referente às filas M/D/1/K (a) e M/M/1/K (b), versus α .

Pode-se concluir que o modelo M/M/1/K é apropriado para a determinação da estratégia de encaminhamento. Exclusivamente para o cálculo do tempo de espera, é possível simplificar ainda mais, utilizando-se um modelo M/M/1.

3.7 Estimativa do Ponto de Mínimo

Um cálculo aproximado para o ponto de mínima perda ou de mínimo atraso pode ser obtido analiticamente para uma rede composta de enlaces paralelos.

Mínima Perda

Sejam ρ_1 e ρ_2 os tráfegos oferecidos aos enlaces 1 e 2, respectivamente:

$$\rho_1 = \frac{\alpha\lambda}{\mu_1}; \quad \rho_2 = \frac{(1-\alpha)\lambda}{\mu_2} \quad (3.10)$$

A perda P de células é:

$$P = \alpha \frac{1}{1 - \rho_1^{k+1}} (1 - \rho_1) \rho_1^k + (1 - \alpha) \frac{1}{1 - \rho_2^{k+1}} (1 - \rho_2) \rho_2^k \quad (3.11)$$

$$P \cong \alpha^{k+1} \left(\frac{\lambda}{\mu_1} \right)^k + (1 - \alpha)^{k+1} \left(\frac{\lambda}{\mu_2} \right)^k \quad (3.12)$$

Derivando-se a equação (3.12) em relação a α , e considerando-se que $\sqrt[k]{x} = 1$, quando $x \rightarrow +\infty$, o mínimo de P ocorre para $\alpha \cong \frac{\mu_1}{\mu_1 + \mu_2}$.

A repartição do tráfego proporcional às capacidades dos enlaces garante o mínimo da perda de células. Para $\mu_2 = 2\mu_1$, tem-se α igual a $1/3$, conforme pode ser verificado na Figura 3.8.

Mínimo Atraso

O atraso médio $E\{W\}$ das células na rede da Figura 3.6 é dado por:

$$E\{W\} = \alpha \frac{1}{\mu_1} \frac{\rho_1}{1 - \rho_1} + (1 - \alpha) \frac{1}{\mu_2} \frac{\rho_2}{1 - \rho_2} \quad (3.13)$$

Para determinação do valor aproximado da parcela α que minimiza o atraso, a expressão (3.13) pode ser aproximada por:

$$\lambda E\{W\} \cong \alpha^2 \frac{\lambda^2}{\mu_1^2} + (1 - \alpha)^2 \frac{\lambda^2}{\mu_2^2} \quad (3.14)$$

O valor de α que minimiza a expressão (3.14) é dado por:

$$\alpha \cong \frac{\mu_1^2}{\mu_1^2 + \mu_2^2} \quad (3.15)$$

A repartição do tráfego proporcional ao quadrado das capacidades minimiza o atraso na rede de enlaces paralelos. Para $\mu_2 = 2\mu_1$ tem-se $\alpha \cong 0.2$, conforme pode ser verificado na Figura 3.7.

3.8 Discussão de Algoritmos

O problema de programação matemática (equação (3.5)) é um problema convexo com restrições lineares, e uma série de algoritmos da literatura podem ser aplicados para sua resolução [3]. Este capítulo não trata da solução exata da equação (3.5), mas pode-se afirmar que, para problemas de encaminhamento em telecomunicações, heurísticas simples apresentam resultados bastante satisfatórios, permitindo ainda incluir restrições operacionais que não são explicitadas na formulação matemática do problema.

Algumas indicações para a implementação dessas heurísticas são discutidas sucintamente a seguir.

Uma característica topológica do problema é a distância física entre os nós de comutação. Assim, o tempo de propagação nos enlaces ligando esses nós é derivado diretamente da localização dos nós e do conhecimento dos meios de transmissão. Desta forma, para cada produto, é possível calcular o caminho mínimo de propagação ligando seus nós origem-destino. Esse é o primeiro passo, que não depende do volume da demanda.

Portanto, para cada produto, determina-se o caminho de menor profundidade (profundidade de um caminho é o seu número de enlaces).

Os custos dos meios de transmissão são significativamente menores do que os dos meios de comutação, e essa é uma tendência histórica das telecomunicações. Assim, é possível afirmar que a rede é bastante malhada, significando que uma boa parte dos produtos possui caminhos de profundidade unitária. Além disso, a escolha de um caminho com $n + 1$ enlaces, em detrimento de um com n (quando ambos existirem), implica no uso de mais recursos da rede envolvidos na comunicação. Desta forma, o algoritmo deve analisar apenas os caminhos de profundidade um nível acima da profundidade mínima do produto k . Essa estratégia vai na direção de limitar severamente o porte do problema de programação matemática.

Definindo-se o parâmetro capacidade de caminho como sendo a menor das capacidades dos enlaces envolvidos no caminho, obtém-se uma solução inicial repartindo-se o produto proporcionalmente a essas capacidades. Nota-se que essa solução é sempre factível, pois no problema não há limitante superior para o tráfego oferecido ao enlace. A partir dessa solução inicial, um método clássico do tipo gradiente pode ser empregado.

3.9 Exemplo

Para ilustrar numericamente o problema de encaminhamento, considera-se a rede mostrada na Figura 3.9 composta por 4 nós e 4 enlaces. A demanda é composta por tráfegos bidirecionais poissonianos com taxas de tráfego de 1 *kcps* (*cps* - células por segundo) entre os produtos $A \leftrightarrow C$, $B \leftrightarrow D$, $E \leftrightarrow G$ e $F \leftrightarrow H$. O encaminhamento é definido da seguinte forma: tráfego $A \leftrightarrow C$ pelo nó 2; $B \leftrightarrow D$ pelo nó 3; $E \leftrightarrow G$ pelo nó 4; $F \leftrightarrow H$ pelo nó 1. Os enlaces 1 e 2 têm capacidade de transmissão bidirecional de 6 *kcps* e os enlaces 3 e 4 têm capacidade de 12 *kcps*. Os *buffers* possuem tamanho $b = 9$.

Cada produto segue um único caminho até seu destino. A taxa de tráfego oferecida a cada enlace, é igual a 2 *kcps*. A taxa de tráfego total oferecida à rede igual a 8 *kcps*.

As probabilidades de perda e atraso podem ser calculadas considerando-se os modelos $M/M/1/K$ para a perda e $M/M/1$ para o atraso, $K = b + 1$.

Para cada enlace, o tráfego oferecido é $\rho = \lambda/\mu$, onde o tempo médio de serviço $1/\mu$

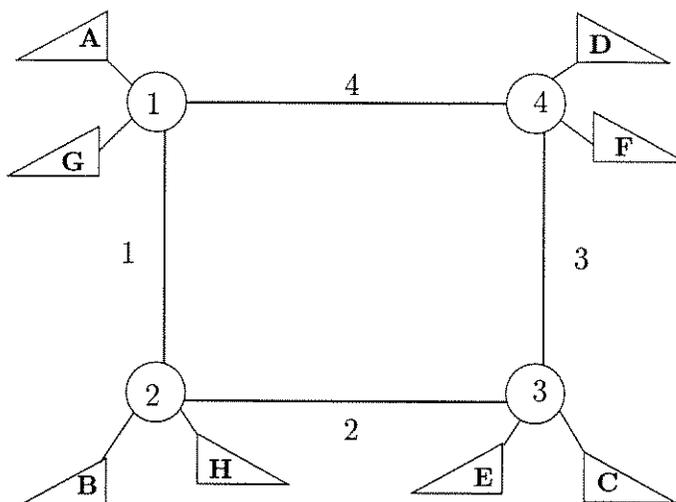


Figura 3.9: Exemplo de rede com 4 nós e 4 enlaces.

é dado pelo inverso da capacidade de cada enlace, em células por segundo. Denotando-se por P a probabilidade de perda de células tem-se:

$$P = \frac{1}{1 - \rho^{K+1}} (1 - \rho) \rho^K \quad (3.16)$$

e o tempo médio de espera no *buffer* é dado por:

$$E\{W\} = \frac{1}{\mu} \frac{\rho}{1 - \rho} \quad (3.17)$$

A Tabela 3.1 mostra as características dos enlaces da rede exemplo. As Tabelas 3.2, 3.3 e 3.4 mostram os valores de P e $E\{W\}$ para cada enlace, produto e para a rede, respectivamente, considerando o encaminhamento inicial, proporcional e proporcional ao quadrado das *capacidades dos caminhos* (menor capacidade de enlace no caminho)

	Enlaces	
	1 e 2	3 e 4
tráfego em <i>kcps</i>	1	1
capacidade em <i>kcps</i>	6	12
atraso em μs	83.3	16.7
perda ($\times 10^6$)	11.3	0.014

Tabela 3.1: Parâmetros de desempenho de cada enlace da rede mostrada na Figura 3.9.

A Tabela 3.3 mostra que para os produtos $A \leftrightarrow C$, $B \leftrightarrow D$ e $F \leftrightarrow H$, o encaminhamento proporcional à capacidade de cada caminho produziu uma redução na perda e o encaminhamento proporcional ao quadrado, uma redução do atraso. A discrepância no caso do produto $E \leftrightarrow G$ decorre deste inicialmente usar o caminho composto pelos enlaces 3 e 4, que são os maiores da rede exemplo.

	Enlaces			
	1 e 2		3 e 4	
	perda ($\times 10^6$)	atraso (μs)	perda ($\times 10^6$)	atraso (μs)
inicial	11.3	83.3	0.014	16.7
proporcional	1.97	64.1	0.062	20.1
proporcional ao quadrado	0.37	50.72	0.18	23.05

Tabela 3.2: Probabilidade de perda e atraso para cada enlace da rede nos vários encaminhamentos aplicados.

	Produtos							
	A \leftrightarrow C		B \leftrightarrow D		E \leftrightarrow G		F \leftrightarrow H	
	perda	atraso	perda	atraso	perda	atraso	perda	atraso
inicial	22.6	166.7	11.3	100	0.027	33.3	11.3	100
proporcional	1.4	69.6	2.04	84.22	1.4	69.6	2.04	84.22
proporcional ao quadrado	0.43	57.17	54.53	73.77	0.43	57.17	54.53	73.77

Tabela 3.3: Probabilidade de perda ($\times 10^6$) e atraso (μs) para cada produto nos vários encaminhamentos aplicados.

	Rede	
	perda ($\times 10^6$)	atraso (μs)
inicial	11.3	100
proporcional	1.72	76.9
proporcional ao quadrado	0.49	65.5

Tabela 3.4: Probabilidade de perda e atraso total na rede, nos vários encaminhamentos aplicados.

3.10 Conclusão

O problema de encaminhamento de tráfego nas redes públicas com tecnologia ATM, permitindo vantagens operacionais de adaptação da rede às variações do tráfego em relação ao previsto no planejamento, foi abordado neste capítulo. O desempenho efetivo, isto é, a capacidade da política de encaminhamento fazer o tráfego fluir com baixas perdas e atrasos pequenos, pode ou não estar em níveis satisfatórios (atendendo ou não as recomendações de qualidade de serviço do ITU-T). O mau desempenho, na verdade, é um excelente diagnóstico de um dimensionamento inapropriado da rede, indicando que esta deve ser replanejada ou reconfigurada.

A principal contribuição foi formular o problema de encaminhamento no contexto de implantação da tecnologia ATM nas redes públicas de telecomunicações, simplificando ao máximo as funções de perda de células e de atraso, viabilizando a aplicação de algoritmos convencionais de otimização de fluxo em redes multiprodutos.

Capítulo 4

Alocação de Banda em Enlaces ATM

O dimensionamento das redes ATM (*Asynchronous Transfer Mode*) compreende a determinação da capacidade dos enlaces para atender a demanda de tráfego com um certo grau de serviço (probabilidade de perda de células, tempo de transferência das células e máxima variação do atraso). Para calcular esta capacidade, surgiu na literatura especializada uma série de métodos baseados no conceito de banda efetiva, que simplifica o dimensionamento, uma vez que a capacidade do enlace deve ser igual ou superior à soma das bandas efetivas dos tráfegos do enlace. Uma abordagem abrangente da determinação da capacidade dos enlaces ATM é proposta neste capítulo. A partir dos parâmetros descritores de tráfego e de qualidade de serviço propostos pelos organismos de recomendação, ITU-T (*International Telecommunication Union - Telecommunication Standardization Sector*) e ATM Forum, uma banda efetiva é obtida para cada classe de serviço, utilizando modelos probabilísticos para a perda e modelos de pior caso para o atraso de células das conexões.

4.1 Introdução

As redes tradicionais de telecomunicações transportam serviços homogêneos quanto aos requisitos de qualidade de serviço. As redes telefônicas foram otimizadas para as características da voz humana, enquanto que as redes de pacotes foram projetadas para a transferência de arquivos.

A demanda de tráfego atual vem solicitando cada vez mais a implantação de uma rede que seja capaz de integrar os diferentes serviços. Estas redes devem tratar as requisições de qualidade de serviço durante a fase de aceitação das chamadas.

Para se definir os parâmetros de comunicação (taxa de transmissão, probabilidade de perda de *bits*, mecanismo de recuperação de erro, etc.) perante a rede, os serviços podem ser caracterizados, quanto aos parâmetros de qualidade de serviço, em:

- sensíveis a atrasos e perda de informação (por exemplo, vídeo interativo);
- sensíveis a atrasos, mas tolerantes a moderadas perdas (por exemplo, voz);
- sensíveis a perdas, mas tolerantes a atrasos moderados (por exemplo, dados interativos);
- relativamente tolerantes a atrasos e a uma limitada perda de informação (por exemplo, transferência de arquivos).

As redes de telecomunicações com tecnologia ATM destacam-se pela capacidade faixa larga de transportar serviços com diferentes requisições de qualidade de serviço. O modo de transferência assíncrono ATM comuta pacotes de 53 *bytes*, denominados células, que são transportados através de caminhos fixos, com multiplexação e compartilhamento de enlaces e nós ao longo da rede.

A determinação da taxa de transmissão que o enlace deve possuir, para atender às requisições de qualidade de serviço de cada um dos elementos das fontes de tráfego que o compartilham, é uma questão essencial no dimensionamento da rede ATM.

Diversas propostas para este dimensionamento têm sido abordadas na literatura especializada [4, 15, 30, 31], destacando-se a idéia de banda efetiva nas suas mais variadas formulações.

Banda efetiva é a capacidade que deve ser alocada em um enlace ATM, para um determinado tráfego, de forma que o grau de serviço seja satisfeito, independentemente dos demais tráfegos que compartilham o enlace [9].

Este capítulo trata da obtenção de um modelo de banda efetiva que atenda os requisitos de tráfego preconizados pelo ITU-T e ATM Forum e seja calculado de maneira numericamente simples. A banda efetiva deve ser suficientemente grande para acomodar as flutuações estatísticas do tráfego, e pequena o suficiente para assegurar uma ocupação satisfatória do meio de comunicação, garantindo retorno financeiro ao operador da rede.

O cálculo da banda efetiva depende de hipóteses assumidas em relação à classe de serviço: se possui ou não um tráfego de pico limitado ou uma duração máxima do surto limitada, etc. As diferentes propostas de banda efetiva encontradas na literatura não são estritamente comparáveis entre si, pois algumas consideram somente o tempo de atraso como parâmetro de qualidade de serviço e outras, apenas a perda de células. Entretanto, todas as definições alternativas devem satisfazer a condição de aditividade linear, quaisquer que sejam as características estatísticas do serviço, ou seja [9]:

$$c\left(\sum_{i=1}^k s_i\right) = \sum_{i=1}^k c(s_i) \quad (4.1)$$

onde c é a banda efetiva e s_i representa os parâmetros descritores de tráfego e de qualidade de serviço de cada componente.

Cada uma das fontes de tráfego, compartilhando um enlace de transmissão, pode ser modelada como sendo servida pela respectiva banda efetiva, sem comprometer o tráfego das demais fontes que compartilham o *buffer*. A banda efetiva é calculada a partir de parâmetros de tráfego e pode ser utilizada em um processo de aceitação de chamadas, semelhante a um circuito comutado, na rede ATM.

A redação subsequente do capítulo está organizada como se segue: primeiramente são apresentados os modelos de banda efetiva, determinados a partir da perda e do atraso das células; posteriormente uma adequação destes modelos aos parâmetros descritores de tráfego e de qualidade de serviço recomendados pelo ITU-T e ATM Forum é apresentada, além de uma metodologia para o cálculo da banda efetiva para cada uma das classes de serviço do ITU-T e ATM Forum; e o capítulo é finalizado com a apresentação de alguns exemplos ilustrativos.

4.2 Perda de Células

O cálculo da banda efetiva através do momento logarítmico, proposto por Kesidis [31], pressupõe a especificação da probabilidade de perda de células.

A banda efetiva calculada através do método do momento logarítmico é intrínseca à fonte de tráfego e à probabilidade de perda definida, não dependendo do número ou do tipo de fontes compartilhando o *buffer*. A apresentação que se segue é baseada na referência [31].

Seja uma fila com capacidade infinita e uma taxa de serviço de C cps (células por segundo) e X o número de células na fila no instante de chegada de uma nova célula. Deseja-se que a probabilidade de se encontrar mais de b células na fila seja inferior a um certo valor fornecido:

$$\Pr\{X > b\} \leq \exp(-b\delta) \quad (4.2)$$

Em situações nas quais a perda de células é pequena, como é o caso das redes ATM, a probabilidade de perda de células em um *buffer* de tamanho b pode ser aproximada pela probabilidade do número de células em um *buffer* de tamanho infinito ser maior que b .

Supondo-se que existem k fontes de tráfego multiplexadas no *buffer*, admite-se que existe uma função c_i , dependente exclusivamente da fonte i e de δ , tal que o grau de serviço solicitado em (4.2) é atingido se, e somente se:

$$\sum_{i=1}^k c_i \leq C \quad (4.3)$$

onde c_i é a banda efetiva calculada para a fonte i ; $i = 1, \dots, k$.

Para todo δ real, a função $h(\delta)$ geradora de momentos logaritmos assintóticos, para uma fonte estacionária e ergódica, é dada por:

$$h(\delta) \triangleq \lim_{t \rightarrow +\infty} \frac{\ln E\{\exp[N(t)\delta]\}}{t} \quad (4.4)$$

onde $N(t)$ é o número de células geradas no intervalo de tempo $(0, t)$.

Se $h(\delta)$ é finita e diferenciável para todo número real δ , então é convexa, positiva e crescente para $\delta > 0$ e a banda efetiva pode ser calculada através de:

$$c_i(\delta) = \frac{h_i(\delta)}{\delta} \quad (4.5)$$

Resumindo,

$$c = \frac{1}{\delta} \lim_{t \rightarrow +\infty} \frac{\ln E\{\exp[N(t)\delta]\}}{t} \iff c \leq C \quad (4.6)$$

A banda efetiva da superposição de fontes independentes é a soma das bandas efetivas individuais, pois para N_1 e N_2 variáveis aleatórias independentes, tem-se:

$$E\{z^{(N_1+N_2)}\} = E\{z^{N_1}\}E\{z^{N_2}\} \quad (4.7)$$

Assim,

$$\ln E\{z^{(N_1+N_2)}\} = \ln E\{z^{N_1}\} + \ln E\{z^{N_2}\} \quad (4.8)$$

resultando em $c = c_1 + c_2$.

Modelos de Fontes

Diversos tipos de fontes podem compor o tráfego submetido a um *buffer* em uma rede ATM. Para alguns tipos de fontes é possível encontrar expressões analíticas para o cálculo da banda efetiva.

Fonte de Taxa Constante

O caso mais simples de cálculo da banda efetiva é o de uma fonte com taxa de geração de células constante λ . Neste caso $N(t) = \lambda t$ é determinístico, portanto:

$$c = \frac{h(\delta)}{\delta} = \frac{1}{\delta} \lim_{t \rightarrow +\infty} \frac{\ln E\{\exp[\lambda t \delta]\}}{t} = \lambda \quad (4.9)$$

Fonte Poissoniana

Para fontes poissonianas, o número de células geradas possui a distribuição de probabilidade:

$$\Pr\{N(t) = n\} = \frac{(\lambda t)^n}{n!} \exp(-\lambda t) \quad (4.10)$$

Portanto,

$$E\{\exp[N(t)\delta]\} = \sum_{n=0}^{+\infty} e^{n\delta} \frac{(\lambda t)^n}{n!} \exp(-\lambda t) = \exp[\lambda t(e^\delta - 1)] \quad (4.11)$$

$$c = \lambda \frac{(e^\delta - 1)}{\delta} \quad (4.12)$$

Observa-se que para δ pequeno, isto é, alta probabilidade de perda de células, $c \cong \lambda$.

Fontes MMPP

Uma fonte é MMPP (*Markov-Modulated Poisson Process*) se a geração de células é poissoniana com taxa λ função de uma cadeia de Markov contínua no tempo.

O modelo MMPP consiste de uma fonte de tráfego com um certo número de fases. A fonte muda de fase conforme uma cadeia de Markov contínua no tempo. Quando a fonte se encontra em uma fase i , gera células aleatoriamente com taxa média λ_i , segundo um processo de Poisson.

A fonte MMPP é caracterizada por duas matrizes: Λ e R . A matriz Λ é uma matriz diagonal com as taxas médias de geração de células em cada fase. A matriz R apresenta as taxas de transição entre as fases ¹.

O vetor linha q de probabilidades estacionárias de fase da fonte é dado por:

¹

$$r_{ij} = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr\{J(t + \Delta t) = j | J(t) = i\}}{\Delta t}; \quad i \neq j$$

$$r_{ij} = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr\{J(t + \Delta t) = i | J(t) = i\} - 1}{\Delta t}; \quad i, j \in \{1, \dots, m\}$$

$$qR = 0; \quad qu = 1 \quad (4.13)$$

onde u é um vetor coluna unitário.

A matriz de probabilidades $P(n, t)$ da fonte gerar n células no intervalo de tempo t , dado que no início do intervalo a fonte estava na fase i , e no final está na fase j , é dada por:

$$P(n, t) = \frac{e^{(R-\Lambda)t} (\Lambda t)^n}{n!} \quad (4.14)$$

Portanto,

$$E\{\exp[N(t)\delta]\} = q \sum_{n=0}^{+\infty} e^{n\delta} \frac{e^{(R-\Lambda)t} (\Lambda t)^n}{n!} u = q e^{[R+(e^\delta-1)\Lambda]t} u \quad (4.15)$$

Assim, a banda efetiva para fontes MMPP é dada por:

$$c = \frac{1}{\delta} \text{eigmax}[R + (e^\delta - 1)\Lambda] \quad (4.16)$$

onde $\text{eigmax}[A]$ é o maior autovalor da matriz A .

Para uma fonte MMPP de duas fases, tem-se:

$$c = \alpha + \sqrt{\alpha^2 + \beta} \quad (4.17)$$

com

$$\alpha = \frac{1}{2\delta} \left((e^\delta - 1)(\lambda_1 + \lambda_2) - \frac{1}{T_1} - \frac{1}{T_2} \right) \quad (4.18)$$

$$\beta = \left(-\lambda_1 \lambda_2 (e^\delta - 1) + \frac{\lambda_1}{T_2} + \frac{\lambda_2}{T_1} \right) \frac{(e^\delta - 1)}{\delta^2} \quad (4.19)$$

onde T_1, T_2 são os tempos médios e λ_1, λ_2 são as taxas médias de cada fase, respectivamente.

Fontes On-Off

Uma das representações mais simples para fontes de taxa variável é o modelo on-off, no qual a fonte possui duas fases: uma em que transmite à taxa λ e outra em que fica em silêncio. A fonte permanece em cada fase durante um tempo aleatório exponencialmente distribuído. Assim:

$$\Lambda = \begin{bmatrix} \lambda & 0 \\ 0 & 0 \end{bmatrix}; \quad R = \begin{bmatrix} -1/T_{\text{on}} & 1/T_{\text{on}} \\ 1/T_{\text{off}} & -1/T_{\text{off}} \end{bmatrix} \quad (4.20)$$

onde T_{on} e T_{off} são os tempos médios de permanência em cada fase.

A banda efetiva para fontes MMPP on-off é determinada pela equação (4.17), com:

$$\alpha = \frac{1}{2\delta} \left((e^\delta - 1)\lambda - \frac{1}{T_{\text{on}}} - \frac{1}{T_{\text{off}}} \right) \quad (4.21)$$

$$\beta = \frac{\lambda}{T_{\text{off}}} \frac{(e^\delta - 1)}{\delta^2} \quad (4.22)$$

Fonte Fluido Markoviano

Uma fonte fluido markoviano é semelhante a uma fonte MMPP, diferindo apenas na taxa de geração de células que, neste caso, é constante em cada fase.

A banda efetiva para fontes do tipo fluido de Markov é dada por:

$$c = \frac{1}{\delta} \text{eigmax}[R + \delta\Lambda] \quad (4.23)$$

Fontes On-Off

Para fontes fluido markoviano on-off a banda efetiva é dada pela equação (4.17), com:

$$\alpha = \frac{1}{2\delta} \left(\lambda\delta - \frac{1}{T_{\text{on}}} - \frac{1}{T_{\text{off}}} \right); \quad \beta = \frac{\lambda}{\delta T_{\text{off}}} \quad (4.24)$$

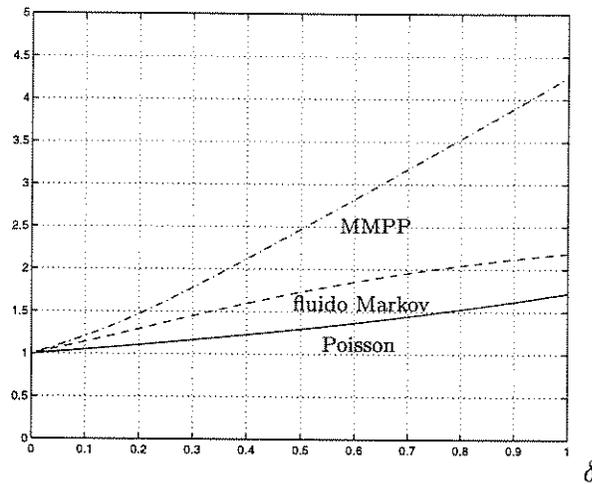


Figura 4.1: Banda efetiva para fonte poissoniana ($\lambda = 1$), fonte MMPP on-off e fonte fluido markoviano on-off ($\lambda = 3$, $T_{\text{on}} = 1$ e $T_{\text{off}} = 2$).

A banda efetiva é sempre menor que a taxa de pico e maior que a taxa média, conforme pode ser verificado em [31].

A Figura 4.1 mostra as bandas efetivas calculadas para as fontes poissoniana ($\lambda = 1$), MMPP on-off e fluido markoviano on-off para $\lambda = 3$, $T_{\text{on}} = 1$ e $T_{\text{off}} = 2$. Observa-se que, para valores grandes de δ as fontes on-off têm bandas efetivas significativamente maiores que a fonte poissoniana. Considerando-se δ igual a 1 e b variando de 1 a 10, a probabilidade de perda varia de 3.7×10^{-1} a 4.5×10^{-5} .

4.3 Atraso de células

Nesta seção a banda efetiva é calculada através de um modelo de atraso determinístico de pior caso, proposto por Boudec [5, 6]. O atraso máximo requerido é especificado.

Uma fila com *buffer* infinito servida pela taxa c garante um limite Δ para o atraso de células, se:

$$c = \max_t \frac{N(t)}{t + \Delta}; \quad t > 0 \quad (4.25)$$

onde $N(t)$ é o número de células geradas pela fonte no intervalo $(0, t)$.

O modelo de banda efetiva determinística não considera a perda de células, assim o *buffer* deve possuir um tamanho tal que:

$$b = \sum_i c_i \Delta_i \quad (4.26)$$

onde c_i e Δ_i são a banda efetiva e o atraso máximo suportado para cada uma das conexões do enlace, respectivamente.

A banda efetiva para fontes de taxa constante é a taxa de transmissão, pois:

$$\max_t \frac{\lambda t}{t + \Delta} \rightarrow \lambda \quad (4.27)$$

Uma fonte variável definida pelos parâmetros: taxa média $\bar{\lambda}$, taxa de pico λ e comprimento máximo do surto L , tem seu atraso de pior caso quando é modelada por uma fonte on-off com surto de tamanho máximo ocorrendo ininterruptamente, conforme mostrado na Figura 4.2.

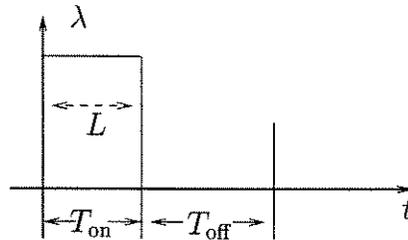


Figura 4.2: Ocorrência de pior caso para uma fonte on-off.

O tempo de duração do surto pode ser determinado pela relação:

$$T_{\text{on}} = \frac{L}{\lambda} \quad (4.28)$$

O intervalo T_{off} entre ocorrência de surtos é determinado por:

$$\bar{\lambda} = \frac{T_{\text{on}}}{T_{\text{on}} + T_{\text{off}}} \lambda \quad (4.29)$$

A Figura 4.3 mostra o preenchimento e o esvaziamento do *buffer* a uma taxa c , quando da ocorrência de um surto.

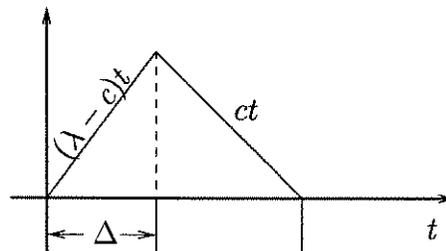


Figura 4.3: Número de células no *buffer* na ocorrência de um surto.

A banda efetiva c é sempre maior que a taxa média $\bar{\lambda}$, garantindo assim o esvaziamento do *buffer* após cada surto.

Nota-se que o *buffer* é preenchido segundo a função $(\lambda - c)t$ e esvaziado com a taxa c . O número de células no *buffer* no final do surto é $(\lambda - c)T_{\text{on}}$ e deverá ser esvaziado num tempo máximo Δ , a uma taxa c . Assim, a banda efetiva é dada por:

$$c = \frac{\lambda T_{\text{on}}}{T_{\text{on}} + \Delta} \quad (4.30)$$

O parâmetro Δ pode representar o máximo atraso permitido no *buffer*, ou a máxima variação permitida para este atraso.

A situação, aparentemente relaxada, com $\Delta > T_{\text{off}}$, na qual o atraso máximo se sobreporia a um novo surto, é evitada com a restrição $c > \bar{\lambda}$, pois:

$$(\lambda - c)T_{\text{on}} < (\lambda - \bar{\lambda})T_{\text{on}} = \frac{\lambda T_{\text{on}} T_{\text{off}}}{T_{\text{on}} + T_{\text{off}}} < c T_{\text{off}} \quad (4.31)$$

O modelo de banda determinística é conservador, feito a partir de uma análise de pior caso que prevê a colisão dos momentos de transmissão na taxa de pico das fontes multiplexadas.

4.4 Adequação aos Órgãos de Recomendação

A adequação dos modelos de cálculo da banda efetiva aos parâmetros de tráfego e de QoS definidos para as capacidades de transferência ATM do ITU-T ou categorias de serviço do ATM Forum, depende da sensibilidade individual à perda ou ao atraso de células dos serviços.

ITU-T

Os serviços são divididos pelo ITU-T [24] em capacidades de transferência ATM segundo parâmetros de tráfego e de qualidade de serviço característicos de cada serviço. Assim, algumas capacidades são sensíveis a atrasos e perdas e outras somente a atrasos ou somente a perdas.

- DBR

A taxa de *bits* determinística é utilizada por serviços que podem requerer uma quantidade de banda continuamente disponível, enquanto durar a conexão. É modelada como uma fonte de taxa constante, com banda efetiva dada pela taxa de pico de transmissão de células PCR.

- SBR

A capacidade de transferência SBR com taxa de *bits* estatística tem como único parâmetro de QoS negociado a CLR. A banda efetiva é dada por:

$$c = \alpha + \sqrt{\alpha^2 + \beta} \quad (4.32)$$

$$\begin{aligned}
\alpha &= \frac{1}{2\delta} \left(\lambda\delta - \frac{1}{T_{\text{on}}} - \frac{1}{T_{\text{off}}} \right) & \beta &= \frac{\lambda}{\delta T_{\text{off}}} \\
\lambda &= \text{PCR} & \exp(-b\delta) &= \text{CLR} \\
\text{MBS} &= 1 + \lfloor \text{IBT} \left(\frac{1}{\text{SCR}} - \frac{1}{\text{PCR}} \right) \rfloor \\
T_{\text{on}} &= \frac{\text{MBS}}{\text{PCR}} & T_{\text{off}} &= \text{MBS} \left(\frac{1}{\text{SCR}} - \frac{1}{\text{PCR}} \right)
\end{aligned} \tag{4.33}$$

- ABT

A capacidade ABT (*ATM Block Transfer*), transferência de bloco ATM, provê serviços nos quais as características de transmissão são negociadas com base em blocos de células. A taxa de células do bloco é constante. Os parâmetros de tráfego e de QoS são relativos a cada bloco. A recomendação I.371 [24] do ITU-T propõe uma banda efetiva (*committed bandwidth*) igual a taxa média de transmissão de células SCR.

- ABR

A capacidade de transferência com taxa de *bits* disponível suporta aplicações que possuem a habilidade de alterar a taxa de transferência em função do estado da rede. A banda efetiva é dada pela taxa mínima de células MCR.

ATM Forum

O ATM Forum definiu cinco categorias de serviço [1]:

- CBR

A categoria com taxa de *bits* constante é idêntica à DBR do ITU-T, portanto, a banda efetiva é dada pela taxa de pico de transmissão de células PCR.

- rt-VBR

A categoria rt-VBR com taxa de *bits* variável em tempo real, é utilizada por aplicações que requerem fortes restrições ao atraso e perdas, sendo apropriada para voz e vídeo com taxa variável. A banda efetiva é o maior dentre os valores encontrados utilizando-se os modelos de perdas e o de atrasos:

- ♣ Perdas

Cálculo da banda efetiva idêntico ao da SBR do ITU-T;

- ♣ Atrasos

$$c = \max\left\{ \text{SCR}, \frac{\text{MBS}}{\text{PCR}^{-1}\text{MBS} + \Delta} \right\}; \tag{4.34}$$

$$\Delta = \min\left\{ \text{MaxCTD} - t_{\text{prop}} - \frac{C}{53 \times 8}; \text{ppCDV} \right\}$$

onde t_{prop} é o tempo de propagação do enlace.

- nrt-VBR

A categoria com taxa de *bits* variável não em tempo real é utilizada por serviços com tráfego em surto, sem restrições ao atraso. O cálculo da banda efetiva é idêntico ao da SBR do ITU-T.

- ABR

Idêntica ao ABR do ITU-T.

- UBR

A categoria com taxa de *bits* não especificada, é utilizado por aplicações que não possuem restrições de atraso e que fazem o próprio controle de erro e de fluxo. As células UBR são transmitidas quando há banda disponível e descartadas em casos de congestionamento. Não há um mecanismo de realimentação das condições de tráfego para o usuário e nenhum parâmetro de QoS é negociado. No dimensionamento do enlace ATM o planejador pode ou não considerar a existência de tráfego UBR. Quando considerada, a banda efetiva para a categoria UBR é especificada pelo projetista.

4.5 Exemplos

Alguns serviços de taxa variável, sensíveis à perda e a atraso de células, são analisados nesta seção. Inicialmente considera-se apenas a perda de células para explicitar como a banda efetiva depende dos parâmetros de qualidade de serviço.

A taxa média e a taxa de pico em *Mbps* e o logaritmo na base 10 da razão de perda de células de alguns exemplos de serviços de taxa variável são mostrados na Tabela 4.1.

Serviço	Pico	Média	Perda
Videofone	2	0.4	-7
Videoconferência	10	2	-7
Imagem Médica	34	1.7	-7
CAD/CAM	34	1.7	-6
Dados	34	6.8	-6

Tabela 4.1: Taxa de pico e taxa média em *Mbps* e logaritmo da razão de perda de células de alguns exemplos de serviço.

b : MBS	1	10	100	1000
10	2461	4429		
100	1048	2461	4429	
1000	953	1048	2461	4429

Tabela 4.2: Banda efetiva em *cps*, segundo o modelo de perda de células, para o serviço videofone com taxa de pico de 4717 *cps*, taxa média de 943 *cps* e razão de perda de 10^{-7} .

Considerando o parâmetro CLR (razão de perda de célula) e utilizando o modelo da seção SBR (equações 4.32 e 4.33) determina-se a banda efetiva (vide Tabela 4.2) para o

serviço videofone. Observa-se que a banda efetiva não depende explicitamente do tamanho do *buffer* b ou do tamanho máximo do surto MBS, mas sim da razão entre eles. Este comportamento é justificado pelas equações (4.32) e (4.33). A grosso modo, pode-se concluir que a banda efetiva é igual à taxa de pico quando a relação b/MBS é da ordem de 1, e igual à taxa média quando b/MBS é maior do que 1000.

As Tabelas 4.3, 4.4, 4.5, 4.6, mostram a banda efetiva calculada em função da razão de perda de células para os demais serviços da Tabela 4.1. As considerações feitas para o serviço de videofone são igualmente válidas para os demais serviços.

$b : \text{MBS}$	1	10	100	1000
10	12305	22145		
100	5243	12305	22145	
1000	4766	5243	12305	22145

Tabela 4.3: Banda efetiva em *cps*, segundo o modelo de perda de células, para o serviço videoconferência com taxa de pico de 23585 *cps*, taxa média de 4717 *cps* e razão de perda de 10^{-7} .

$b : \text{MBS}$	1	10	100	1000
10	33996	75231		
100	4685	33996	75231	
1000	4069	4685	33996	75231

Tabela 4.4: Banda efetiva em *cps*, segundo o modelo de perda de células, para o serviço imagem médica com taxa de pico de 80189 *cps*, taxa média de 4009 *cps* e razão de perda de 10^{-7} .

$b : \text{MBS}$	1	10	100	1000
10	27878	74408		
100	4576	27878	74408	
1000	4060	4576	27878	74408

Tabela 4.5: Banda efetiva em *cps*, segundo o modelo de perda de células, para o serviço CAD/CAM com taxa de pico de 80189 *cps*, taxa média de 4009 *cps* e razão de perda de 10^{-6} .

$b : \text{MBS}$	1	10	100	1000
10	38142	74495		
100	17553	38142	74495	
1000	16180	17553	38142	74495

Tabela 4.6: Banda efetiva em *cps*, segundo o modelo de perda de células, para o serviço dados com taxa de pico de 80189 *cps*, taxa média de 16038 *cps* e razão de perda de 10^{-6} .

O parâmetro de atraso máximo MaxCTD é importante na definição de banda efetiva para conexões de muito longa distância, ou que passem por mais de uma dezena de nós

de comutação, ou ainda naquelas em cujo o tempo de empacotamento é significativo. Já o parâmetro $ppCDV$, que especifica a máxima variação do atraso aceitável, atua na determinação da banda efetiva qualquer que seja a conexão ATM.

A Tabela 4.7 mostra a banda efetiva para o serviço videofone considerando a variação do atraso Δ . Observa-se que a banda efetiva, para uma dada taxa de pico, depende da relação MBS/Δ e não de cada um dos parâmetros isoladamente, conforme pode ser evidenciado pela equação (4.34).

Δ : MBS	1	10	100	1000
0.1	3205	4504	4695	4715
1	943	3205	4504	4695
10	943	943	3205	4504
100	943	943	943	3205

Tabela 4.7: Banda efetiva em *cps*, segundo o modelo de atraso de células, para o serviço videofone com taxa de pico de 4717 *cps*, taxa média de 943 *cps* e atraso Δ em *ms*.

As Tabelas 4.8, 4.9, 4.10 e 4.11 mostram a banda efetiva, considerando o atraso, para os demais serviços da Tabela 4.1.

Δ : MBS	1	10	100	1000
0.1	7022	19084	23042	23530
1	4717	7022	19084	23042
10	4717	4717	7022	19084
100	4717	4717	4717	7022

Tabela 4.8: Banda efetiva em *cps*, segundo o modelo de atraso de células, para o serviço videoconferência com taxa de pico de 23585 *cps*, taxa média de 4717 *cps* e atraso Δ em *ms*.

Δ : MBS	1	10	100	1000
0.1	8891	44503	74236	79551
1	4009	8891	44503	74236
10	4009	4009	8891	44503
100	4009	4009	4009	8891

Tabela 4.9: Banda efetiva em *cps*, segundo o modelo de atraso de células, para o serviço imagem médica com taxa de pico de 80189 *cps*, taxa média de 4009 *cps* e atraso Δ em *ms*.

É apropriado supor que os comutadores ATM utilizarão grande capacidade de memória. A Tabela 4.12 apresenta a banda efetiva para o serviço videofone considerando simultaneamente a perda e o atraso de células para um *buffer* de tamanho $b = 1000$ células.

Notam-se, na Tabela 4.12, 2 regiões com comportamentos distintos: uma na qual a banda efetiva é determinada pela restrição de atraso e outra na qual a perda de células é o parâmetro mais significativo.

O Apêndice A apresenta a banda efetiva calculada pelo modelo COST 242 para efeito de comparação.

Δ : MBS	1	10	100	1000
0.1	8891	44503	74236	79551
1	4009	8891	44503	74236
10	4009	4009	8891	44503
100	4009	4009	4009	8891

Tabela 4.10: Banda efetiva em *cps*, segundo o modelo de atraso de células, para o serviço CAD/CAM com taxa de pico de 80189 *cps*, taxa média de 4009 *cps* e atraso Δ em *ms*.

Δ : MBS	1	10	100	1000
0.1	16038	44503	74236	79551
1	16038	16038	44503	74236
10	16038	16038	16038	44503
100	16038	16038	16038	16038

Tabela 4.11: Banda efetiva em *cps*, segundo o modelo de atraso de células, para o serviço dados com taxa de pico de 80189 *cps*, taxa média de 16038 *cps* e atraso Δ em *ms*.

Δ : MBS	1	10	100	1000
0.1	3205	4504	4695	4715
1	953	3205	4504	4695
10	953	1048	3205	4504
100	953	1048	2461	4429

Tabela 4.12: Banda efetiva em *cps*, segundo o modelo de perdas e de atraso de células, para o serviço videofone com taxa de pico de 4717 *cps*, taxa média de 943 *cps*, *buffer* de tamanho $b = 1000$ células e atraso Δ em *ms*.

As Tabelas 4.13, 4.14, 4.15 e 4.16 mostram comportamentos similares para os demais serviços da Tabela 4.1.

Δ : MBS	1	10	100	1000
0.1	7022	19084	23042	23530
1	4766	7022	19084	23042
10	4766	5243	12305	22145
100	4766	5243	12305	22145

Tabela 4.13: Banda efetiva em *cps*, segundo o modelo de perdas e de atraso de células, para o serviço videoconferência com taxa de pico de 23585 *cps*, taxa média de 4717 *cps*, *buffer* de tamanho $b = 1000$ células e atraso Δ em *ms*.

Δ : MBS	1	10	100	1000
0.1	8891	44503	74236	79551
1	4069	8891	44503	75231
10	4069	4685	33996	75231
100	4069	4685	33996	75231

Tabela 4.14: Banda efetiva em *cps*, segundo o modelo de perdas e de atraso de células, para o serviço imagem médica com taxa de pico de 80189 *cps*, taxa média de 4009 *cps*, *buffer* de tamanho $b = 1000$ células e atraso Δ em *ms*.

Δ : MBS	1	10	100	1000
0.1	8891	44503	74236	79551
1	4060	8891	44503	74408
10	4060	4576	27878	74408
100	4060	4576	27878	74408

Tabela 4.15: Banda efetiva em *cps*, segundo o modelo de perdas e de atraso de células, para o serviço CAD/CAM com taxa de pico de 80189 *cps*, taxa média de 4009 *cps*, *buffer* de tamanho $b = 1000$ células e atraso Δ em *ms*.

Δ : MBS	1	10	100	1000
0.1	16180	44503	74236	79551
1	16180	17553	44503	74495
10	16180	17553	38142	74495
100	16180	17553	38142	74495

Tabela 4.16: Banda efetiva em *cps*, segundo o modelo de perdas e de atraso de células, para o serviço dados com taxa de pico de 80189 *cps*, taxa média de 16038 *cps*, *buffer* de tamanho $b = 1000$ células e atraso Δ em *ms*.

4.6 Atraso de Varredura

A banda efetiva calculada neste capítulo não considera um componente de atraso que é fundamental em serviços de faixa estreita e irrelevante em serviços de faixa larga. Trata-se do atraso de varredura que é fortemente dependente da disciplina interna dos comutadores, tais como, FIFO (*First In First Out*), FIFO com prioridade, WRR (*Weighted Round Robin*), etc.

O atraso de varredura de pior caso é igual ao inverso da banda efetiva expressa em células por segundo.

Para serviços CBR de $N \times 64$ *kbps*, o atraso de varredura τ é igual a:

$$\tau \cong \frac{6}{N} \text{ ms} \quad (4.35)$$

O serviço CBR de $N \times 64$ *kbps* com banda efetiva igual a taxa de pico, atenderá à restrição de atraso de 1 *ms* se ao menos N for maior ou igual a 6.

Serviços de faixa estreita, como por exemplo telefonia, utilizam a tecnologia ATM eficientemente se os circuitos virtuais agregarem ao menos 6 canais telefônicos de 64 *kbps*.

4.7 Conclusão

A determinação da capacidade dos enlaces de uma rede ATM, através do cálculo da banda efetiva dos tráfegos do enlace, foi abordada neste capítulo.

Os modelos disponíveis na literatura especializada usualmente consideram somente a perda de células, ou o atraso, ou a variação do atraso de células. Alguns poucos trabalhos que consideram os três parâmetros propõem a aplicação de um único modelo probabilístico para o cálculo de tais parâmetros. Neste trabalho, uma estratégia abrangente para o cálculo da banda efetiva foi proposta, considerando modelos probabilísticos para a perda de células e modelos de pior caso para o atraso de transferência de células.

Ambos os modelos discutidos foram adequados aos parâmetros de tráfego e de qualidade de serviço do ITU-T e ATM Forum. Além disto, discutiu-se uma metodologia para o cálculo da banda efetiva de cada uma das capacidades de transferência ATM do ITU-T e para cada uma das categorias de serviço do ATM Forum. Os exemplos apresentados ao final do capítulo ilustraram a sensibilidade dos serviços à perda e ao atraso de células.

A utilização da ferramenta computacional que implementa a metodologia de cálculo apresentada neste capítulo, possibilita que o projetista de redes sentir-se tão confortável, no dimensionamento de uma rede ATM, quanto se estivesse de posse de uma tabela de Erlang no caso de um entroncamento telefônico.

Capítulo 5

Probabilidade de Bloqueio de Chamadas em um Enlace

A questão do dimensionamento de um enlace sob o ponto de vista do grau de serviço, isto é, da probabilidade de aceitação de novas chamadas, é abordada neste capítulo. Após rememorar a fórmula de Erlang e o associado conceito de tráfego oferecido, o problema de determinação das probabilidades individuais de bloqueio dos diversos serviços, atendidos por um enlace, é analisado e alguns exemplos ilustrativos são apresentados.

5.1 Introdução

O modo de transferência assíncrono ATM é uma tecnologia de comutação de pacotes de tamanho fixo (denominados células) orientada a conexão. Um mesmo caminho lógico (circuito virtual) é definido durante a aceitação da chamada, sendo estabelecido para todas as células, enquanto durar a conexão. O circuito virtual contém mecanismos de controle de erro e de fluxo que garantem a seqüência de entrega das células ao destino, podendo ainda fazer uma pré-alocação de recursos em cada nó, garantindo um melhor controle do congestionamento.

A comunicação através da comutação de circuitos virtuais se dá pelo estabelecimento de uma rota de comunicação entre a origem e o destino, e envolve três fases: o estabelecimento do circuito virtual, a transmissão de informação e a desconexão.

A comutação de circuitos virtuais rejeita novas chamadas em caso de congestionamento da rede através do mecanismo de controle de admissão de conexões CAC, uma vez que as conexões já estabelecidas não podem ser degradadas.

Os usuários devem ter disponível uma taxa de bits (na forma de uma banda efetiva) que garanta os parâmetros de qualidade de serviço negociados no contrato de tráfego. O conceito de banda efetiva permite que o problema de dimensionamento dos enlaces ATM possa ser formulado como um problema de bloqueio multitaxas (*Multi-Rate Loss Models*), tratando simultaneamente aplicações com diferentes parâmetros descritores de tráfego e de qualidade de serviço.

No caso particular de serviços com bandas efetivas idênticas, supondo-se ainda que o processo de demanda é poissoniano e que o tempo de retenção da conexão é exponencial negativo com taxas conhecidas, a determinação da probabilidade de bloqueio de novas chamadas é dada pela *fórmula de Erlang*.

5.2 Fórmula de Erlang

Com o intuito de introduzir o conceito de *tráfego oferecido*, considera-se um enlace, entre dois nós de comutação, com um número infinito de circuitos virtuais, cada um com a mesma capacidade C em *bps*, mostrado na Figura 5.1. A demanda é suposta poissoniana e a duração das chamadas é exponencial negativa, com parâmetros λ e μ , respectivamente.



Figura 5.1: Enlace interligando dois nós de comutação.

A probabilidade de nenhuma chamada terminar em um intervalo de duração t , dado que $X = m$ estão ativas, é dada por:

$$\Pr\{\tau > t\} = \exp(-m\mu t) \quad (5.1)$$

Nas condições estacionárias, definindo-se p_m como a probabilidade de m chamadas estarem ativas, tem-se as seguintes equações de equilíbrio de estado:

$$\begin{cases} \lambda p_0 & = \mu p_1 \\ \lambda p_1 & = 2\mu p_2 \\ & \vdots \\ \lambda p_{m-1} & = m\mu p_m \\ & \vdots \end{cases} \quad (5.2)$$

Resolvendo-se para p_m , tem-se:

$$p_m = \frac{\rho^m}{m!} p_0 \quad ; \quad \rho \triangleq \frac{\lambda}{\mu} \quad (5.3)$$

$$\sum_{m=0}^{+\infty} p_m = 1 \quad \implies \quad p_0 \sum_{m=0}^{+\infty} \frac{\rho^m}{m!} = 1 \quad \implies \quad p_0 \exp(\rho) = 1$$

Portanto,

$$p_m = \frac{\rho^m}{m!} \exp(-\rho) \quad (5.4)$$

As probabilidades de estado p_m têm distribuição poissoniana e $E\{X\} = \rho$ e $\sigma_X^2 = \rho$, isto é, a média é igual à variância.

O número médio de circuitos virtuais ocupados é denominado de *tráfego escoado*. No caso de um número infinito de circuitos, a probabilidade de bloqueio é nula e o tráfego escoado é igual ao tráfego oferecido, que é expresso em *Erlang*.

Definindo-se $A = XC$ (tráfego em *bps*), tem-se:

$$E\{A\} \triangleq a = \rho C; \quad \sigma_A^2 = \rho C^2 = aC \quad (5.5)$$

Para o caso de um enlace com M circuitos virtuais, tem-se:

$$p_m = \frac{\rho^m}{m!} p_0 \implies p_0 \sum_{m=0}^M \frac{\rho^m}{m!} = 1 \implies p_m = \frac{\rho^m / m!}{\sum_{i=0}^M \rho^i / i!}$$

A probabilidade B de bloqueio é dada por:

$$B = \Pr\{X = M / \text{cham.}\} = \frac{\Pr\{\text{cham.} / X = M\} \Pr\{X = M\}}{\Pr\{\text{cham.}\}} = p_M \quad (5.6)$$

A probabilidade de bloqueio é igual à probabilidade de todos os M circuitos estarem ocupados e é denominada fórmula de Erlang.

$$B = Er(\rho, M) = \frac{\rho^M / M!}{\sum_{i=0}^M \rho^i / i!} \quad (5.7)$$

A expressão (5.7) não é conveniente para valores grandes de M , e pode ser substituída pela seguinte recorrência:

$$Er(\rho, m) = \frac{\rho Er(\rho, m-1)}{m + \rho Er(\rho, m-1)} \quad ; \quad Er(\rho, 0) = 1 \quad (5.8)$$

A fila é sem espera, e portanto o tempo T de permanência na fila é igual ao tempo médio de serviço, resultando em:

$$E\{T\} = \frac{1}{\mu} \quad (5.9)$$

Como a taxa média de ingresso é igual à taxa de chamadas originadas menos a taxa de chamadas bloqueadas, tem-se:

$$E\{\lambda_m\} = \lambda(1 - B) \quad (5.10)$$

Assim, pelo teorema de Little [14], tem-se:

$$E\{X\} = E\{\lambda_m\} E\{T\} = \lambda(1 - B) \frac{1}{\mu} \quad (5.11)$$

Resultando, para o tráfego escoado:

$$E\{X\} = \rho(1 - B) = \rho[1 - Er(\rho, M)] \quad (5.12)$$

Se M tende para infinito, então $E\{X\}$ tende para ρ .

5.3 Bloqueio de Chamadas em Enlaces Multi-Serviço

Modelos multi-serviços são utilizados na análise de mecanismos de controle de admissão de conexão CAC e no dimensionamento das redes ATM. O grau de serviço é dado pelas probabilidades de bloqueio de chamadas das diferentes fontes de tráfego.

Um sistema multi-serviço comporta-se como um sistema comutado por circuito, se uma banda efetiva é alocada às chamadas aceitas.

Um número N de diferentes classes de tráfego compartilham um mesmo enlace de transmissão com capacidade fixa de C bps, não havendo portanto reserva de banda para qualquer classe. O processo de chegada da classe de serviço i é poissoniano com taxa λ_i . Para uma chamada da classe i , o tempo de retenção segue uma função de distribuição exponencial com média $1/\mu_i$. Para cada chamada aceita da classe i , uma taxa de bits constante c_i é reservada, e é liberada imediatamente após o término desta chamada.

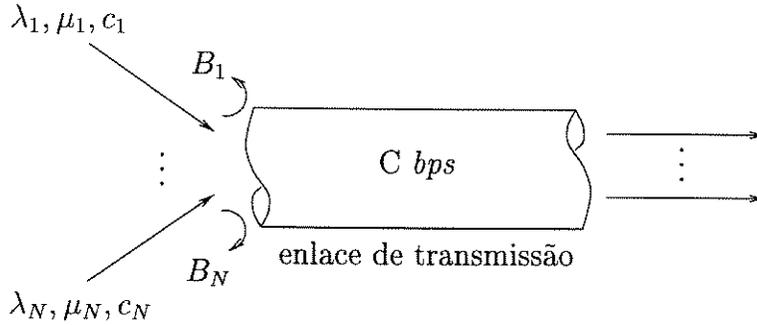


Figura 5.2: Modelo do enlace com compartilhamento de banda por N classes de serviço.

Os estados $m = [m_1 \ m_2 \ \dots \ m_N]'$, do enlace multi-serviço mostrado na Figura 5.2, podem ser modelados por uma cadeia markoviana, cujo espaço S dos estados factíveis é dado por:

$$S = \{m; m'c \leq C\} \quad (5.13)$$

onde $m'c = \sum_{i=1}^N m_i c_i$.

Os estados S_i de bloqueio da i -ésima classe de serviço são dados por:

$$S_i = \{m \in S; (m + 1_i)'c > C\} \quad (5.14)$$

onde $1_i' = [0 \ 0 \ \dots \ 1 \ \dots \ 0]'$ (vetor de zeros, com 1 na i -ésima posição).

As equações de equilíbrio dos estados podem ser escritas numa forma matricial para uma dada ordenação dos estados:

$$Q\beta = 0 \quad (5.15)$$

onde Q é a matriz de transição de estado e o vetor β representa as probabilidades não normalizadas dos estados, isto é:

$$p = \beta \left(\sum_{i=1}^M \beta_i \right)^{-1} \quad (5.16)$$

e M representa o número de estados factíveis. A solução da equação (5.15) é obtida pela determinação do espaço nulo da matriz Q .

A Figura 5.3 mostra a cadeia de Markov para $C = 7$ bps, $c_1 = 2$ bps e $c_2 = 3$ bps.

Considerando-se a seguinte ordenação dos estados factíveis:

$$S = \{00; 10; 01; 20; 11; 02; 30; 21\}$$

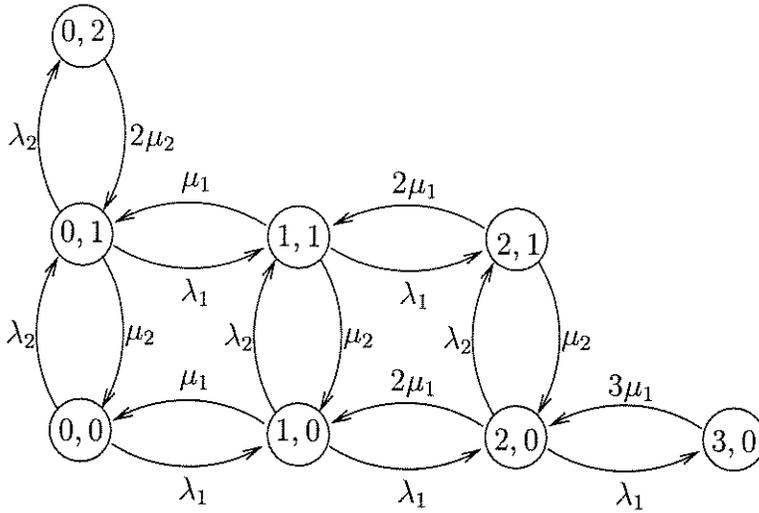


Figura 5.3: Cadeia de Markov para um enlace com $c_1 = 2$ bps, $c_2 = 3$ bps e $C = 7$ bps.

a correspondente matriz Q é dada por:

$$\begin{bmatrix}
 -\lambda_{12} & \mu_1 & \mu_2 & 0 & 0 & 0 & 0 & 0 \\
 \lambda_1 & -(\lambda_{12} + \mu_1) & 0 & 2\mu_1 & 0 & \mu_2 & 0 & 0 \\
 \lambda_2 & 0 & -(\lambda_{12} + \mu_2) & 0 & 2\mu_2 & \mu_1 & 0 & 0 \\
 0 & \lambda_1 & 0 & -(\lambda_{12} + 2\mu_1) & 0 & 0 & 3\mu_1 & \mu_2 \\
 0 & 0 & \lambda_2 & 0 & -2\mu_2 & 0 & 0 & 0 \\
 0 & \lambda_2 & \lambda_1 & 0 & 0 & -(\lambda_1 + \mu_{12}) & 0 & 2\mu_1 \\
 0 & 0 & 0 & \lambda_1 & 0 & 0 & -3\mu_1 & 0 \\
 0 & 0 & 0 & \lambda_2 & 0 & \lambda_1 & 0 & -(\mu_2 + 2\mu_1)
 \end{bmatrix}$$

onde $\lambda_{12} = \lambda_1 + \lambda_2$, $\mu_{12} = \mu_1 + \mu_2$.

Para $\lambda_1 = 0.3$, $\lambda_2 = 0.2$ e $\mu_1 = \mu_2 = 1$, tem-se:

$$p' = [610.3 \ 183.1 \ 122.1 \ 27.5 \ 12.2 \ 36.6 \ 2.7 \ 5.5] \times 10^{-3}$$

$S_1 = \{30; 21; 02\}$ e $S_2 = \{30; 21; 02; 11\}$ são os estados de bloqueio das classes 1 e 2 respectivamente, cujas probabilidades de bloqueio são, respectivamente $B_1 = 2.04\%$ e $B_2 = 5.71\%$.

Uma alternativa ao uso da equação (5.15) para o cálculo do vetor β é derivada da propriedade fundamental das cadeias de Markov de que cada elemento do vetor β pode ser escrito em função de produtos fundamentais [14], isto é:

$$\beta_{m_1, m_2, \dots, m_N} = \prod_{i=1}^N \frac{\rho_i^{m_i}}{m_i!} \tag{5.17}$$

Aplicando-se a equação (5.17) ao exemplo, tem-se:

$$\beta = [1 \ \rho_1 \ \rho_2 \ \frac{\rho_1^2}{2} \ \frac{\rho_2^2}{2} \ \rho_1 \rho_2 \ \frac{\rho_1^3}{6} \ \frac{\rho_1^2}{2} \rho_2]'$$

Definindo-se o índice $\Delta(S) = \sum_{m \in S} \beta_m$, as probabilidades de bloqueio das classes de serviço são dadas por:

$$B_i = \Delta(S_i) \Delta^{-1}(S); \quad i = 1, \dots, N \quad (5.18)$$

As equações (5.15) e (5.17) fazem uso explícito da determinação dos estados factíveis S da cadeia de Markov, e a obtenção das probabilidades de bloqueio B_i ; $i = 1, \dots, N$, necessita do conhecimento dos correspondentes estados S_i . A descrição da cadeia de Markov torna-se impraticável, quando o número N de serviços e a capacidade C do enlace crescem significativamente. A enumeração unidimensional, proposta por Kaufman em [28], alivia consideravelmente os cálculos envolvidos na determinação das probabilidades de bloqueio das classes de serviço.

Modelo Iterativo

A solução iterativa para o modelo de probabilidade é baseada no mapeamento do espaço de estados multidimensional em um espaço de estados unidimensional de acordo com uma discretização apropriada da largura de banda [28]. Sem grande perda de generalidade, pode-se considerar que as bandas efetivas, c_i ; $i = 1, \dots, N$, são dadas por números inteiros de *bps*. Assim, uma unidade de largura de banda δ pode ser definida como:

$$\delta = \text{mdc}\{c_1, \dots, c_N\} \quad (5.19)$$

onde $\text{mdc}\{\}$ é o máximo divisor comum.

O número máximo de unidades de banda disponíveis é dado por $M = \lfloor C/\delta \rfloor$ e o número de unidades de banda requerido pela classe i é expresso por $n_i = c_i/\delta$, onde $\lfloor x \rfloor$ representa o maior inteiro menor que x .

A probabilidade não normalizada de estados α é calculada através do seguinte algoritmo iterativo [28, 34, 35]:

$$\alpha_0 = 1; \quad m\alpha_m = \sum_{i=1}^N \tilde{a}_i \alpha_{m-n_i}; \quad m = 1, \dots, M \quad (5.20)$$

onde $\tilde{a}_i = a_i/\delta$, com $a_i = \rho_i c_i$, e $\alpha_{m < 0} = 0$.

As probabilidades de bloqueio B_i , $i = 1, \dots, N$, são dadas por:

$$B_i = \sum_{m=M-n_i+1}^M q_m, \quad i = 1, \dots, N \quad (5.21)$$

onde $q_m = \alpha_m \left(\sum_{m=0}^M \alpha_m \right)^{-1}$.

A equação (5.21) explicita o fato de que as classes têm probabilidades de bloqueio tanto maiores quanto maiores forem as suas bandas efetivas.

Para o caso particular de $c_i = \delta$, $i = 1, \dots, N$, α é dado por:

$$\alpha_0 = 1; \quad \alpha_m = \frac{\rho}{m} \alpha_{m-1} = \frac{\rho^m}{m!}; \quad m = 1, \dots, M \quad (5.22)$$

onde $\rho = \sum_{i=1}^N \rho_i$. Assim, a fórmula de Erlang se aplica também para serviços com tempo de retenção distintos, mas com mesma banda efetiva.

A aplicação do método iterativo (equação 5.20) ao exemplo resulta em $\delta = 1$, $M = 7$, $n_1 = 2$, $n_2 = 3$ e:

$$\alpha = [1 \quad 0 \quad \rho_1 \quad \rho_2 \quad \frac{\rho_1^2}{2} \quad \rho_1\rho_2 \quad \frac{\rho_1^3}{6} + \frac{\rho_2^2}{2} \quad \frac{\rho_1^2}{2}\rho_2]'$$

Observa-se que o mapeamento entre a cadeia de Markov original (bidimensional) e a unidimensional se faz agregando-se os estados que possuem o mesmo número de unidades de banda δ :

$$\alpha_0 = \beta_{00}; \quad \alpha_2 = \beta_{10}; \quad \alpha_3 = \beta_{01}; \quad \alpha_4 = \beta_{20}; \quad \alpha_5 = \beta_{11}; \quad \alpha_6 = \beta_{30} + \beta_{02}; \quad \alpha_7 = \beta_{21}$$

As probabilidades de bloqueio, para o exemplo, são dadas por:

$$B_1 = q_6 + q_7 = 2.04\% \quad B_2 = q_5 + q_6 + q_7 = 5.71\%$$

Outro Exemplo

A Figura 5.4 mostra um enlace com capacidade $C = 7$ bps e três classes serviços distintas (voz, dados e vídeo) com bandas efetivas $c_1 = 2$ bps, $c_2 = 3$ bps e $c_3 = 4$ bps. A taxa média de cada uma das classes é, respectivamente, $\lambda_1 = 0.3$, $\lambda_2 = 0.2$ e $\lambda_3 = 0.1$. O tempo de retenção é suposto ser igual para cada um dos serviços: $\mu_1 = \mu_2 = \mu_3 = 1$.

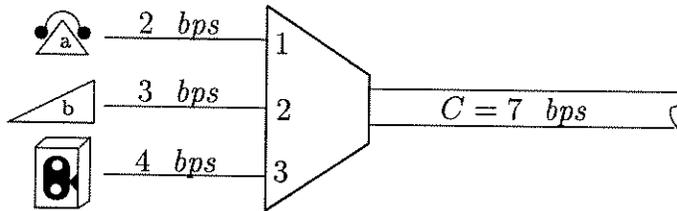


Figura 5.4: Exemplo de enlace com três classes de serviço distintas: voz, dados e vídeo.

A Figura 5.5 mostra os estados factíveis da cadeia de Markov deste exemplo. Considerando-se a seguinte ordenação dos estados factíveis:

$$S = \{000; 100; 010; 200; 001; 110; 300; 020; 101; 210; 011\}$$

pode-se escrever a matriz Q , tal que:

$$Q\beta = 0$$

Os estados factíveis de bloqueio das classes de serviço são:

$$S_1 = \{300; 020; 101; 210; 011\}$$

$$S_2 = \{110; 300; 020; 101; 210; 011\}$$

$$S_3 = \{200; 001; 110; 300; 020; 101; 210; 011\}$$

e as probabilidades de bloqueio individuais das classes de serviço são:

$$B_1 \cong 5\% \quad B_2 \cong 8\% \quad B_3 \cong 16\%$$

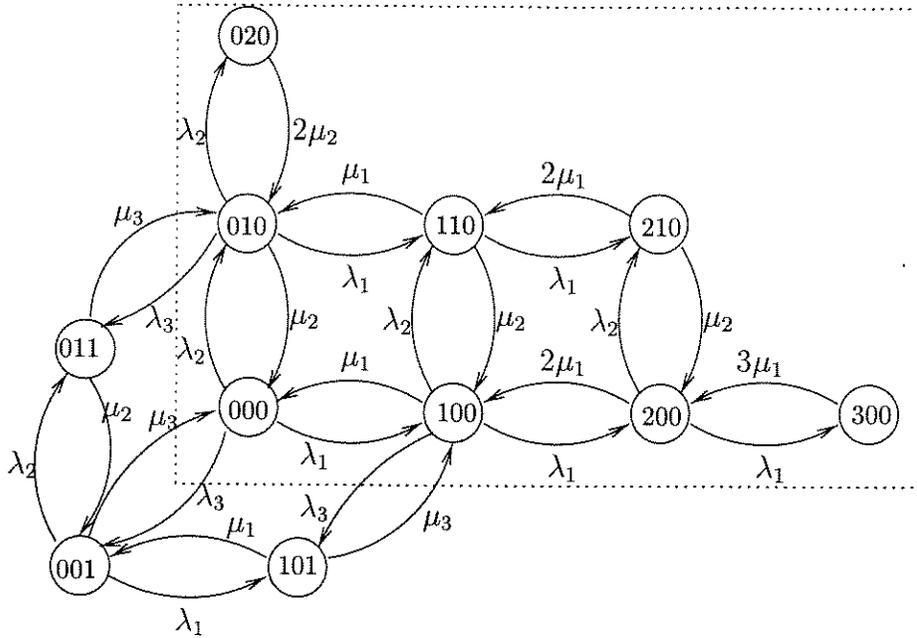


Figura 5.5: Cadeia de Markov para um enlace com $c_1 = 2 \text{ bps}$, $c_2 = 3 \text{ bps}$, $c_3 = 4 \text{ bps}$ e $C = 7 \text{ bps}$.

Utilizando-se o modelo iterativo de Kaufman:

$$\delta = \text{mdc}\{2, 3, 4\} = 1$$

$$M = C/\delta = 7; \quad n_1 = 2; \quad n_2 = 3; \quad n_3 = 4$$

A Figura 5.6 mostra os estados factíveis encontrados após a utilização do modelo iterativo de Kaufman. Observa-se uma redução no número total de estados factíveis em relação aos estados factíveis da cadeia de Markov para o exemplo, o que confirma a estratégia do método iterativo de Kaufman, agrupando os estados de bloqueio, conforme ilustrado na Figura 5.6.

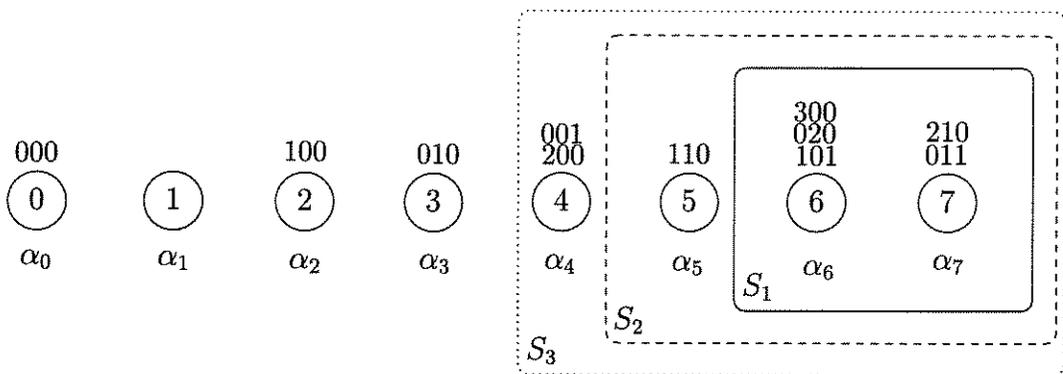


Figura 5.6: Estados factíveis utilizando o método iterativo de Kaufman para um enlace com capacidade $C = 7 \text{ bps}$ e três classes de serviços com bandas efetivas $c_1 = 2 \text{ bps}$, $c_2 = 3 \text{ bps}$ e $c_3 = 4 \text{ bps}$.

Os estados factíveis de bloqueio das classes de serviço são:

$$S_1 = \{\alpha_6, \alpha_7\}$$

$$S_2 = \{\alpha_5, \alpha_6, \alpha_7\}$$

$$S_3 = \{\alpha_4, \alpha_5, \alpha_6, \alpha_7\}$$

e as probabilidades de bloqueio individuais das classes de serviço são:

$$B_1 \cong 5\% \quad B_2 \cong 8\% \quad B_3 \cong 16\%$$

Os valores encontrados para a probabilidade de bloqueio individual das classes de serviço corresponde exatamente ao valor encontrado através da cadeia de Markov.

Solução Aproximada

O dimensionamento da capacidade C de um enlace multi-serviço requer que o cálculo das probabilidades de bloqueio seja feito repetidas vezes, em um processo de aproximação sucessiva da capacidade final (por exemplo, pelo método da bipartição). As abordagens previamente apresentadas são computacionalmente complexas para serem utilizadas em problemas de porte considerável. Esta seção apresenta um método, proposto no projeto COST 242 [34, 35], para o cálculo aproximado da probabilidade de bloqueio, baseado na média e variância do tráfego oferecido.

A superposição independente de N classes de tráfego oferecidos a um enlace tem média e variância dadas por:

$$a = \sum_{i=1}^N a_i; \quad v = \sum_{i=1}^N a_i c_i \quad (5.23)$$

A aplicação direta da fórmula de Erlang permite uma primeira estimativa das probabilidades de bloqueio das classes de serviço.

$$\bar{B} = E(\bar{\rho}, C/\bar{c}) \quad (5.24)$$

onde \bar{c} a banda efetiva média e $\bar{\rho}$ o tráfego oferecido médio, calculados por:

$$\bar{c} = v/a; \quad \bar{\rho} = a/\bar{c} = \sum_{i=1}^N \frac{\rho_i c_i}{\bar{c}} \quad (5.25)$$

A fórmula de Erlang pode ser estendida para um número não inteiro de circuitos através da interpolação dada por:

$$Er(\rho, x) = \frac{\rho Er(\rho, m)}{x + \rho Er(\rho, m)} \quad ; \quad m = [x] \quad (5.26)$$

A Figura 5.7 mostra as probabilidades de bloqueio B_1 e B_2 e a aproximação \bar{B} para o exemplo previamente considerado. Observa-se que \bar{B} subestima consideravelmente as probabilidades B_1 e B_2 .

No projeto COST 242 uma heurística (validada por simulação) é proposta para determinação aproximada das probabilidades de bloqueio individuais, dadas por:

$$\tilde{B}_i = \frac{a c_i (a + \gamma c_i)}{a^2 \bar{c} + \gamma \sum_{i=1}^N a_i c_i^2} \tilde{B} \quad (5.27)$$

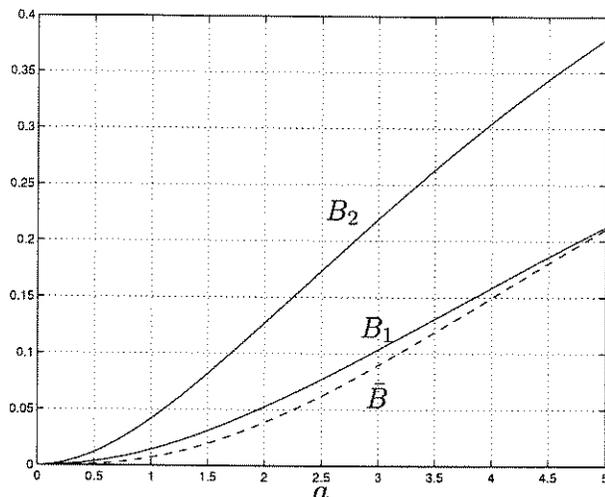


Figura 5.7: Probabilidades de bloqueio B_1 e B_2 e a aproximação \tilde{B} (equação 5.24) para um enlace com $c_1 = 2$ bps, $c_2 = 3$ bps e $C = 7$ bps.

onde $\gamma = \frac{C - a}{2\bar{c}}$ e $\tilde{B} = E\left(\bar{\rho}, \frac{C - \Delta}{\bar{c}}\right)$, com $\Delta = \bar{c} - \min_i c_i$.

A Figura 5.8 mostra as probabilidades de bloqueio B_1 e B_2 e a aproximação do projeto COST 242 para o exemplo considerado.

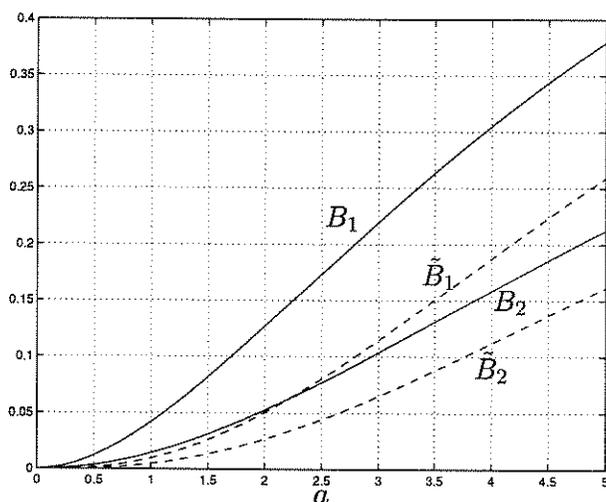


Figura 5.8: Probabilidades de bloqueio B_1 e B_2 e a aproximação do projeto COST 242 (equação 5.27) para um enlace com $c_1 = 2$ bps, $c_2 = 3$ bps e $C = 7$ bps.

5.4 Conclusão

A determinação da probabilidade individual de bloqueio para serviços comutados compartilhando um mesmo enlace foi discutida neste capítulo.

Alocando-se uma banda efetiva às chamadas aceitas, o sistema multi-serviço comporta-se como um circuito comutado. A probabilidade de bloqueio individual calculada através

das cadeias de Markov mostra-se impraticável quando o número de serviços ou a capacidade do enlace cresce significativamente. O modelo de Kaufman foi apresentado como solução simplificadora para o cálculo das probabilidades de bloqueio, uma vez que este modelo propõe um mapeamento unidimensional do espaço de estados, de acordo com uma discretização da largura de banda dos serviços, disponibilizando um método iterativo para realização do cálculo. Os exemplos apresentados ao longo do capítulo validaram e ilustraram a estratégia do modelo.

Uma solução aproximada, proposta no projeto COST 242 [34, 35], baseada na média e variância do tráfego oferecido é apresentada.

Conclusão Geral

O dimensionamento de enlaces em redes de telecomunicações ATM, atendendo um certo grau de serviço solicitado, foi analisado neste trabalho.

A determinação da capacidade de enlaces de uma rede ATM foi abordada, à luz do conceito de banda efetiva, utilizando modelos probabilísticos para a perda de células e modelos de pior caso para o atraso de transferência de células. Os modelos discutidos foram adequados aos parâmetros de tráfego e de qualidade de serviço do ITU-T e do ATM Forum. Discutiu-se uma metodologia para o cálculo da banda efetiva para cada uma das capacidades de transferência ATM do ITU-T e para cada uma das categorias de serviço do ATM Forum. Em particular, para serviços de taxa variável em tempo real, uma estratégia abrangente para o cálculo da banda efetiva considerando a perda e o atraso de células foi proposta, e uma série de exemplos foi analisada.

Concomitante à determinação da capacidade dos enlaces está a determinação da probabilidade individual de bloqueio para serviços comutados, compartilhando um mesmo enlace. Neste tópico, a contribuição deste trabalho consistiu na análise e escolha de modelos simplificadores que forneceram algoritmos para a solução do problema.

Outro problema abordado neste trabalho foi o encaminhamento de tráfego em redes públicas com tecnologia ATM. Neste contexto, a principal contribuição foi a formulação matemática do problema, simplificando ao máximo as funções de perda de células e de atraso, viabilizando a aplicação de algoritmos convencionais de otimização de fluxo em redes multiprodutos para a solução do problema.

Apêndice A

COST 242

O modelo COST 242 é um método empírico para o cálculo da banda efetiva desenvolvido pela *European Cooperation in The Field of Scientific and Technical Research* [34, 35]. O modelo tem sido adotado por diversas instituições, sendo recomendado inclusive pela TELEBRÁS [38].

O modelo mostra uma abordagem para o cálculo da banda efetiva essencialmente dependente de dois parâmetros de tráfego da fonte: taxa média e taxa de pico; além da taxa de serviço do enlace e a razão de perda de células desejada. Do ponto de vista do dimensionamento, possui a inconveniência de que a capacidade do enlace é uma das variáveis do modelo. Conseqüentemente, não é possível obter uma equação fechada para o cálculo da capacidade do enlace, sendo necessária a utilização de métodos recursivos.

O modelo COST 242 é um modelo de perda de células, não sendo considerados atrasos no *buffer*.

Seja $\bar{\lambda}$ a taxa média, λ a taxa de pico, C a taxa de transmissão e $P = 10^{-l}$ a razão de perda de células desejada, calcula-se:

$$v = \lambda/\bar{\lambda} \quad ; \quad a = C/\lambda \quad ; \quad d = 2l/a \quad ; \quad f = 1 + 2l/100 \quad (A.1)$$

A banda efetiva c é dada por:

$$c = \begin{cases} f\lambda & ; \quad v < 3d \\ f\bar{\lambda} \left[1 + 3d \left(\frac{v-1}{v} \right) \right] & ; \quad v \geq 3d \end{cases} \quad \text{para } d \leq 1$$
$$\begin{cases} f\lambda & ; \quad v < 3d^2 \\ f\bar{\lambda} \left[1 + 3d^2 \left(\frac{v-1}{v} \right) \right] & ; \quad v \geq 3d^2 \end{cases} \quad \text{para } d > 1 \quad (A.2)$$

O modelo COST 242 é fundamentado em aproximações, sendo válido para *buffer* de grande capacidade, taxa de pico muito menor que a capacidade do enlace ($15 < C/\lambda < 1000$) e em uma determinada faixa de razão entre a taxa de pico e a taxa média ($\lambda/c > 2$ e $\lambda/\bar{\lambda} < 20$).

A Tabela A.1 mostra os valores encontrados para a banda efetiva de diversos serviços de taxa variável utilizando o modelo COST 242.

Serviço	Pico	Média	Perda	Banda
Videofone	4717	943	-7	1542
Videoconferência	23585	4717	-7	17034
Imagem Médica	80189	4009	-7	91415
CAD/CAM	80189	4009	-6	89811
Dados	80189	16038	-6	89811

Tabela A.1: Taxa de pico e taxa média em *cps*, logaritmo da razão de perda de células e banda efetiva para os exemplos do Capítulo 4 para uma capacidade de enlace de 155.52 *Mbps*.

Apêndice B

Fila M/M/1/K e M/D/1/K

B.1 Fila M/M/1/K

Nesta seção são apresentados os resultados da literatura [14] para filas com disciplina de acesso ao serviço tipo FIFO (*first in, first out*), markovianas (intervalo entre chegadas poissoniano, duração do serviço com distribuição exponencial negativa)¹, um servidor e capacidade da fila (incluindo a célula que está sendo atendida) K , ou seja, filas M/M/1/K.

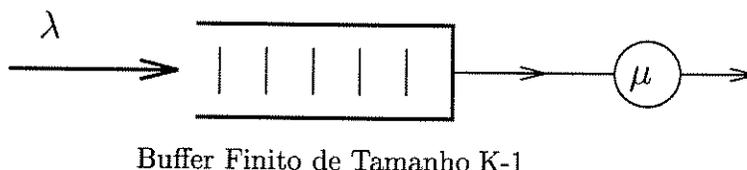


Figura B.1: Representação da Fila M/M/1/K.

A taxa média de chegadas, em células por unidade de tempo, é denotada pelo símbolo λ , e o número médio de células transmitidas por unidade de tempo é representado por μ . Assim, as células chegam espaçadas aleatoriamente no tempo, em média a cada $1/\lambda$ unidades de tempo e levam em média $1/\mu$ unidades de tempo para serem transmitidas. A Figura B.1 ilustra a representação desta fila. A Figura B.2 mostra o diagrama de estados para esta fila.

Enquanto uma célula está sendo transmitida, as que chegam são armazenadas no *buffer* e quando este atinge o seu limite ($K - 1$ células), as células seguintes são descartadas, enquanto o *buffer* permanecer cheio.

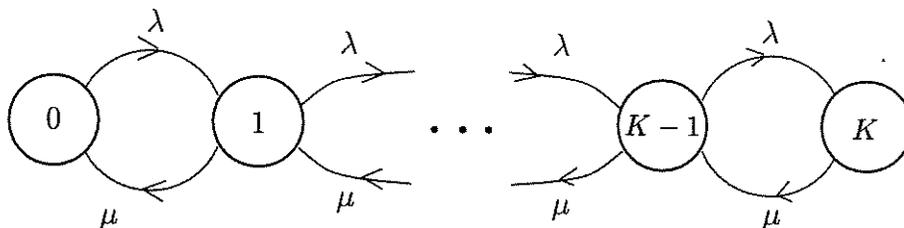


Figura B.2: Diagrama de Transição de Estados para a Fila M/M/1/K.

Denotando por $X(t)$ o número de células no sistema no instante de observação t , tem-se:

¹nesta seção as células são consideradas de tamanho contínuo com distribuição exponencial negativa.

$$p_n(t) \triangleq \Pr\{X(t) = n\} \quad (\text{B.1})$$

Supondo que a fila atinge seu estado estacionário, isto é, de equilíbrio estatístico,

$$\lim_{t \rightarrow +\infty} p_n(t) = p_n \quad (\text{B.2})$$

As equações de equilíbrio de fluxo são dadas por:

$$\lambda p_n = \mu p_{n+1} \quad ; \quad n = 0, \dots, K-1 \quad (\text{B.3})$$

Definindo o *tráfego oferecido* ρ :

$$\rho \triangleq \frac{\lambda}{\mu} \quad (\text{B.4})$$

resulta

$$p_n = \rho^n p_0 \quad ; \quad n = 0, \dots, K \quad (\text{B.5})$$

Trata-se de uma distribuição geométrica [14].

Como

$$\sum_{n=0}^K p_n = 1 \quad (\text{B.6})$$

e,

$$\sum_{n=0}^K \rho^n = \frac{1 - \rho^{K+1}}{1 - \rho} \quad ; \quad \rho \neq 1 \quad (\text{B.7})$$

então²

$$p_0 = \frac{1}{1 - \rho^{K+1}} (1 - \rho) \quad (\text{B.8})$$

Nota-se que para $\rho < 1$ e $K \rightarrow +\infty \implies p_0 = 1 - \rho$, confirmando o resultado para a fila M/M/1, conforme [14].

B.1.1 Probabilidade de Perda de Células

Definindo P como a *probabilidade de perda de células*, tem-se:

$$P = \frac{\Pr\{\text{buffer cheio}\} \Pr\{\text{chegar uma célula num intervalo de tempo } \Delta t / \text{buffer cheio}\}}{\Pr\{\text{chegar uma célula num intervalo de tempo } \Delta t\}} \quad (\text{B.9})$$

então,

$$P = p_K = \frac{1}{1 - \rho^{K+1}} (1 - \rho) \rho^K \quad (\text{B.10})$$

e o tráfego médio escoado é dado por $\xi = \rho(1 - p_K)$

²para $\rho = 1$, $p_0 = \frac{1}{K+1}$.

B.1.2 Tamanho Médio da Fila

Para o cálculo do *número de células presentes na fila* (incluindo a que está sendo atendida), define-se a transformada z :

$$G_n(z) \triangleq \sum_{n=0}^K z^n p_n = p_0 \frac{1 - (\rho z)^{K+1}}{1 - \rho z} \quad (\text{B.11})$$

e

$$E\{X\} = \left. \frac{d}{dz} G_n(z) \right|_{z=1} \quad (\text{B.12})$$

onde $E\{\}$ indica o operador esperança matemática. Logo,

$$E\{X\} = \frac{\rho}{(1 - \rho)} \frac{[1 - (K + 1)\rho^K + K\rho^{K+1}]}{1 - \rho^{K+1}} \quad (\text{B.13})$$

B.1.3 Tempo de Espera no *Buffer*

Seja T o *tempo de permanência no sistema* e W o *tempo de espera no buffer*. Assim,

$$E\{T\} = E\{W\} + \frac{1}{\mu} \quad (\text{B.14})$$

Segundo o teorema de Little [14], para sistemas estatisticamente estacionários em média, o número médio de elementos presentes no sistema é igual ao produto da taxa média de ingresso pelo tempo médio de permanência.

$$E\{X\} = E\{\lambda_n\} E\{T\} \quad (\text{B.15})$$

onde λ_n é a *taxa efetiva de chegadas*, e

$$E\{\lambda_n\} = \sum_{n=0}^K \lambda_n p_n = \lambda \sum_{n=0}^{K-1} p_n = \lambda (1 - p_K) \quad (\text{B.16})$$

Assim,

$$E\{W\} = \frac{1}{\mu} \frac{\rho}{1 - \rho} \frac{[1 - K\rho^{K-1} + (K - 1)\rho^K]}{1 - \rho^K} \quad (\text{B.17})$$

B.2 Fila M/D/1/K

Nesta seção são apresentados os resultados para filas M/G/1/K com base nos resultados de Baiocchi [2]. Estuda-se a distribuição de probabilidades de ocupação do *buffer* e, em particular, a probabilidade de bloqueio. A expressão para o tempo médio de espera também é encontrada. As expressões para a fila M/D/1/K, mais apropriada para a modelagem de uma rede ATM do que a fila M/M/1/K, são encontradas como caso particular da fila M/G/1/K.

Seja,

- $t_n \triangleq$ instantes de renovação, imediatamente após a n -ésima partida (fim de serviço).

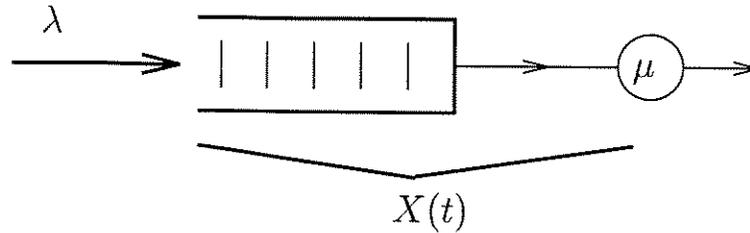


Figura B.3: Fila M/G/1/K

- $X_n \triangleq X(t_n)$ é o estado do sistema (número de usuários) nos instantes de renovação.
- $\nu_n \triangleq \nu(t_n)$ é o número de chegadas durante o serviço da n -ésima partida.
- $W_n \triangleq W(t_n)$ tempo de espera do n -ésimo usuário (elemento que partiu em t_n).
- $T_n \triangleq T(t_n)$ tempo de permanência no sistema do n -ésimo usuário.

A Figura B.3 mostra um exemplo de fila M/G/1/K.

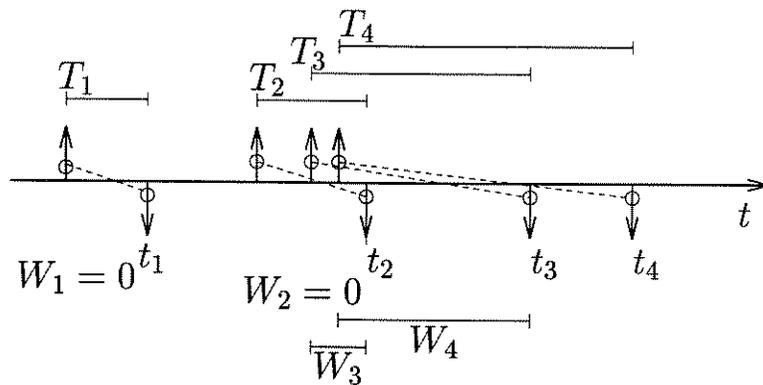


Figura B.4: Estado da fila M/G/1/K nos instantes de renovação

Observando a Figura B.4, conclui-se que:

- $X_1 = 0$, pois não há nenhum elemento no sistema no instante t_1 ;
- $X_2 = 2$, pois existem ainda dois usuários no sistema no instante t_2 ;
- $X_3 = 1$, indicando que há um usuário no sistema no instante t_3 .
- $X_4 = 0$, pois não há nenhum usuário no sistema no instante t_4 .
- $\nu_1 = 0$, indicando que não houve nenhuma chegada durante tempo de serviço $T_1 - W_1$;
- $\nu_2 = 2$, pois ocorreram duas chegadas durante o tempo de serviço $T_2 - W_2$;
- $\nu_3 = 0$, pois não houve nenhuma chegada durante o tempo de serviço $T_3 - W_3$;
- $\nu_4 = 0$, pois não houve nenhuma chegada durante o tempo de serviço $T_4 - W_4$.

Considerando a situação estacionária, seja π_k a *probabilidade de haver no sistema k células imediatamente após uma partida*, com $k = 0, \dots, K - 1$.

$$\pi_k \triangleq \Pr\{X(t_n) = k\} \quad (\text{B.18})$$

Para a determinação de π_k considera-se a probabilidade estacionária a_k assim definida:

$a_k \triangleq$ probabilidade de chegar k células durante a duração do serviço da n -ésima células (não necessariamente são aceitas). Daí,

$$a_k = \int_0^{+\infty} P(k, \theta) p_T(\theta) d\theta \quad (\text{B.19})$$

onde:

- $P(k, \theta)$ é a probabilidade de chegar k novas células durante o intervalo $(t_n - \theta, t_n)$;
- $p_T(\theta)$ é a densidade de probabilidade da duração do serviço.

O processo de chegada (poissoniano) e o de duração do serviço são supostos independentes.

$$P(k, \theta) = \exp(-\lambda\theta) \frac{(\lambda\theta)^k}{k!} \quad (\text{B.20})$$

$$G_a(z) = \int_0^{+\infty} \sum_{k=0}^{+\infty} \frac{(\lambda z \theta)^k}{k!} \exp(-\lambda\theta) p_T(\theta) d\theta = \int_0^{+\infty} \exp[-\lambda\theta(1-z)] p_T(\theta) d\theta \quad (\text{B.21})$$

Definindo-se a transformada de Laplace de $p_T(\theta)$ como

$$p_T(s) \triangleq \int_0^{+\infty} \exp(-s\theta) p_T(\theta) d\theta \quad (\text{B.22})$$

$$G_a(z) = p_T(s) \Big|_{s=\lambda(1-z)} \quad (\text{B.23})$$

então

$$a_k = \frac{1}{k!} \frac{d^k G_a(z)}{dz^k} \Big|_{z=0} (-\lambda)^k = \frac{1}{k!} \frac{d^k p_T(s)}{ds^k} \Big|_{s=\lambda} (-\lambda)^k \quad (\text{B.24})$$

Para o caso de duração de serviço exponencial:

$$p_T(\theta) = \mu \exp(-\mu\theta), \quad \theta \geq 0 \quad (\text{B.25})$$

$$p_T(s) = \mu \frac{1}{s + \mu} \Rightarrow \frac{d^k p_T(s)}{ds^k} = \frac{(-1)^k \mu k!}{(s + \mu)^{k+1}} \quad (\text{B.26})$$

Assim,

$$a_k = \frac{(-1)^k \mu}{(\lambda + \mu)^k + 1} (-\lambda)^k = \left(\frac{\rho}{1 + \rho} \right)^k \frac{1}{1 + \rho} \quad (\text{B.27})$$

Para o caso de duração de serviço determinística:

$$p_T(\theta) = \delta\left(\theta - \frac{1}{\mu}\right) \quad (\text{B.28})$$

$$p_T(s) = \mu \frac{1}{s + \mu} \Rightarrow \frac{d^k p_T(s)}{ds^k} = \left(-\frac{1}{\mu}\right)^k \exp\left(-s \frac{1}{\mu}\right) \quad (\text{B.29})$$

e

$$a_k = \frac{\rho^k}{k!} \exp(-\rho). \quad (\text{B.30})$$

O sistema de equações de equilíbrio de fluxo que determina as probabilidades π_k é dado por:

$$\begin{aligned}
 \pi_0 &= \pi_0 a_0 + \pi_1 a_0 \\
 \pi_1 &= \pi_0 a_1 + \pi_1 a_1 + \pi_2 a_0 \\
 \pi_2 &= \pi_0 a_2 + \pi_1 a_2 + \pi_2 a_1 + \pi_3 a_0 \\
 &\vdots \\
 \pi_{K-1} &= \pi_0 b_{K-1} + \pi_1 b_{K-1} + \pi_2 b_{K-2} + \cdots + \pi_{K-2} b_2 + \pi_{K-1} b_1
 \end{aligned} \tag{B.31}$$

onde

$$b_k = a_k + a_{k+1} + a_{k+2} + \cdots = 1 - \sum_{n=0}^{k-1} a_n, \quad k = 1, \dots, K-1. \tag{B.32}$$

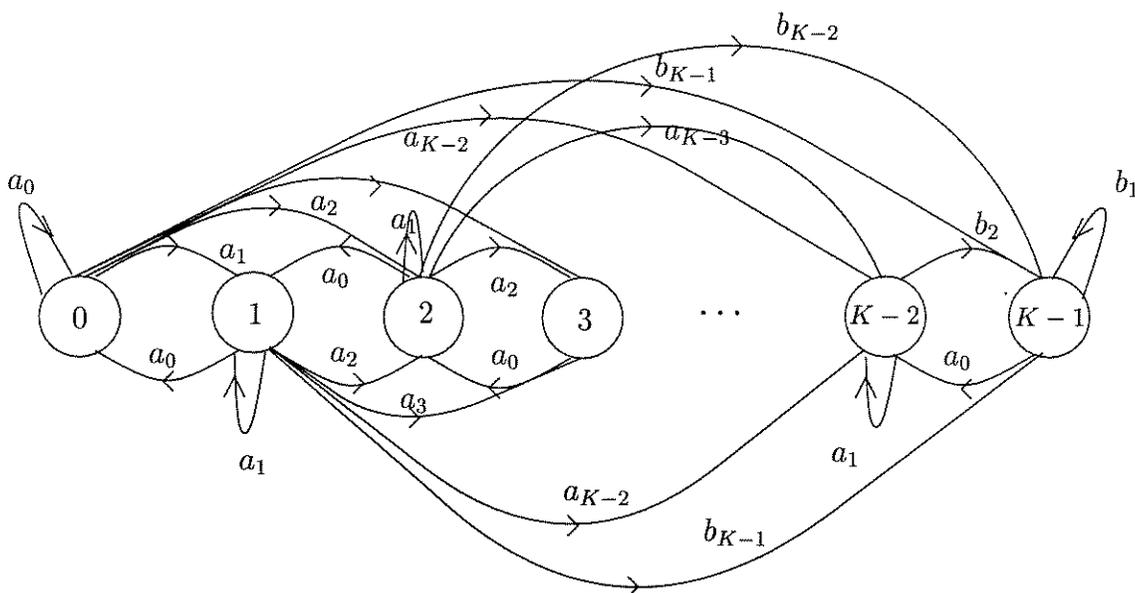


Figura B.5: Diagrama de transição da fila M/G/1/K, observado nos instantes de renovação.

O processo de ocupação da fila M/G/1/K pode ser descrito por uma cadeia de Markov discreta no tempo, mostrada na figura B.5, onde a probabilidade de transição entre os estados da fila, nos instantes de renovação, é dada pela seguinte matriz:

$$A = \begin{bmatrix}
 a_0 & a_1 & a_2 & \cdots & a_{K-2} & 1 - \sum_{k=0}^{K-2} a_k \\
 a_0 & a_1 & a_2 & \cdots & a_{K-2} & 1 - \sum_{k=0}^{K-2} a_k \\
 0 & a_0 & a_1 & \cdots & a_{K-3} & 1 - \sum_{k=0}^{K-3} a_k \\
 \vdots & \vdots & \vdots & & \vdots & \vdots \\
 0 & 0 & 0 & \cdots & a_1 & 1 - \sum_{k=0}^1 a_k \\
 0 & 0 & 0 & \cdots & a_0 & 1 - \sum_{k=0}^0 a_k
 \end{bmatrix} \tag{B.33}$$

Considerando o vetor linha $\pi = [\pi_0 \ \pi_1 \ \cdots \ \pi_{K-1}]$, tem-se:

$$\pi = \pi A \quad ; \quad \pi u = 1 \tag{B.34}$$

onde u é um vetor unitário de dimensão K .

Portanto, π pertence ao espaço nulo da matriz $(A - I)$. Métodos numéricos são disponíveis para a solução deste sistema [32].

Nota-se que, no caso particular de duração de serviço exponencial, uma solução analítica pode ser obtida como

$$\pi_k = \rho^k \pi_0, \quad k = 0, \dots, K-1; \quad \pi_0 = \frac{1-\rho}{1-\rho^K} \quad (\text{B.35})$$

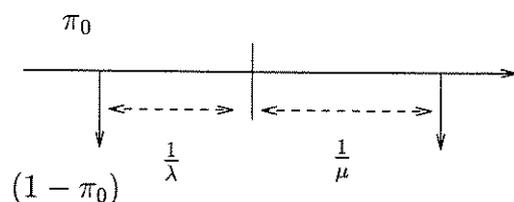


Figura B.6: Diagrama de duas partidas consecutivas para o cálculo do tráfego médio escoado.

O tráfego médio escoado ξ é calculado a partir do diagrama mostrado na Figura B.6:

$$\xi = \frac{\frac{1}{\mu}}{(1-\pi_0)\frac{1}{\mu} + \pi_0\left(\frac{1}{\lambda} + \frac{1}{\mu}\right)} = \frac{1}{1 + \frac{\pi_0}{\rho}} \quad (\text{B.36})$$

B.2.1 Probabilidade de Perda de Células

Pelo teorema de Little,

$$\xi = \rho(1 - P) \quad (\text{B.37})$$

e

$$P = 1 - \frac{1}{\rho + \pi_0} \quad (\text{B.38})$$

Observa-se que no caso de duração do serviço determinística, π_0 é obtida numericamente da resolução da equação (B.34). No caso de duração do serviço exponencial, a equação (B.35) leva à solução já encontrada na equação (B.10) confirmando o resultado da literatura [14] para a probabilidade de perda de células para fila M/M/1/K.

B.2.2 Tamanho Médio da Fila

O tamanho médio da fila pode ser encontrado através da expressão

$$E\{X\} = \sum_{n=0}^K n p_n \quad (\text{B.39})$$

com

$$p_n = \frac{\pi_n}{\rho + \pi_0}, \quad n = 0, \dots, K-1; \quad p_K = 1 - \frac{1}{\rho + \pi_0} \quad (\text{B.40})$$

B.2.3 Tempo de Espera no *Buffer*

O tempo médio de permanência no sistema é dado por:

$$E\{T\} = \frac{(\rho + \pi_0)}{\lambda} E\{X\} \quad (\text{B.41})$$

e o tempo médio de espera no *buffer*:

$$E\{W\} = E\{T\} - \frac{1}{\mu} \quad (\text{B.42})$$

Nas figuras B.7 e B.8 são mostrados, respectivamente, o logaritmo da probabilidade de bloqueio e o tempo médio de espera no *buffer* para as filas M/M/1/K e M/D/1/K.

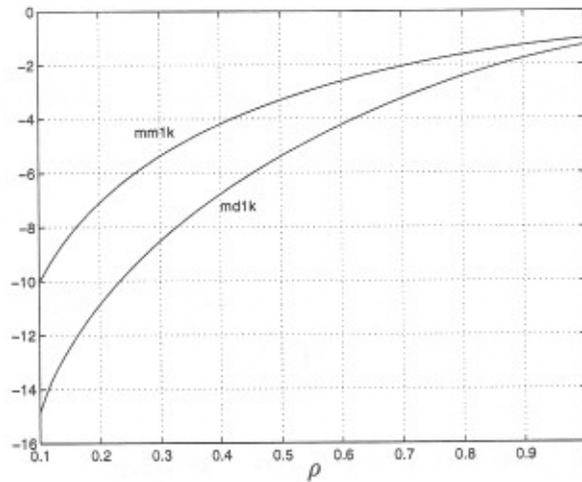


Figura B.7: Logaritmo da probabilidade de perda de células com $K = 10$ e $\mu = 1$.

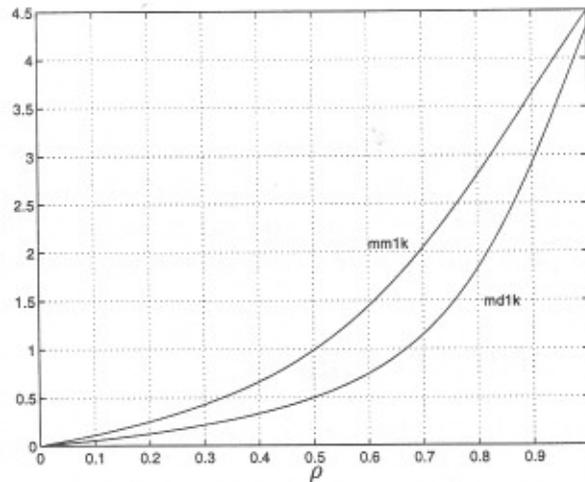


Figura B.8: Tempo médio de espera no *buffer* com $K = 10$ e $\mu = 1$.

Apêndice C

Fila MMPP/D/1/K

Fontes MMPP

Uma fonte é MMPP (*Markov-Modulated Poisson Process*) se a geração de células é poissoniana com taxa λ função de uma cadeia de Markov contínua no tempo.

O modelo MMPP consiste de uma fonte de tráfego com um certo número de fases. A fonte muda de fase conforme uma cadeia de Markov contínua no tempo. Quando a fonte se encontra em uma fase i , gera células aleatoriamente com taxa média λ_i , segundo um processo de Poisson.

A fonte MMPP é caracterizada por duas matrizes: Λ e R . A matriz Λ é uma matriz diagonal com as taxas médias de geração de células em cada fase. A matriz R apresenta as taxas de transição entre as fases.

O vetor linha q de probabilidades estacionárias de fase da fonte é dado por:

$$qR = 0; \quad qu = 1 \quad (C.1)$$

onde u é um vetor coluna unitário

A matriz de probabilidades $P(n, t)$ da fonte gerar n células no intervalo de tempo $(0, t)$, dado que no início do intervalo a fonte estava na fase i , e no final está na fase j , é dada por:

$$P(n, t) = \frac{e^{(R-\Lambda)t} (\Lambda t)^n}{n!} \quad (C.2)$$

pois pode ser obtida a partir do sistema de equações diferenciais para $t \geq 0$

$$\begin{cases} \frac{d}{dt} P(0, t) = P(0, t)(R - \Lambda) \\ \frac{d}{dt} P(n, t) = P(n, t)(R - \Lambda) + P(n-1, t)\Lambda, \quad n > 0 \end{cases} \quad (C.3)$$

com a condição inicial: $P(0, 0) = I$ e $P(n, 0) = 0, n > 0$.

Definindo-se A_n como a matriz $m \times m$, cujo elemento $a_{i,j}^n$ denota a probabilidade condicional de atingir a fase j , dado que começou na fase i , e que foram geradas n células durante o tempo de serviço de uma célula, tem-se:

$$A_n = \int_0^{+\infty} P(n, \theta) p_T(\theta) d\theta \quad (C.4)$$

onde $p_T(\theta)$ é a densidade de probabilidade da duração do serviço.

Nesta seção são apresentados os resultados da fila MMPP/D/1/K, a partir do trabalho de Baiocchi [2], apropriada ao estudo de redes ATM, uma vez que o tempo de serviço nestas redes é determinístico.

No caso do ATM, com duração de serviço constante e igual a μ^{-1} , tem-se:

$$A_n = P(n, \mu^{-1}) = \frac{e^{(R-\Lambda)\mu^{-1}} (\mu^{-1}\Lambda)^n}{n!} \quad (\text{C.5})$$

Portanto:

$$A = \sum_{n=0}^{+\infty} A_n = e^{(R-\Lambda)\mu^{-1}} e^{\mu^{-1}\Lambda} = e^{R\mu^{-1}} \quad (\text{C.6})$$

Na fila MMPP/D/1/K, a disciplina de acesso ao serviço é do tipo FIFO (*first in first out*), o processo de geração de células é do tipo MMPP, a duração do serviço é constante (determinística), existe um único servidor e a capacidade da fila, incluindo a célula que esta sendo atendida, é K . A célula é descartada quando encontra a fila lotada.

A fila MMPP/G/1/K pode ser descrita através de uma variável de estado bidimensional $S(t) = [X(t), J(t)]$, onde $X(t)$ é o número de usuários no sistema (incluindo o que está sendo atendido), enquanto $J(t)$ representa o estado do processo de Markov que modula as chegadas poissonianas. $X(t)$ pode variar de $\{0, \dots, K\}$ e $J(t)$ de $\{1, \dots, m\}$. Portanto, assume-se que o processo de Markov compreende m fases.

Os instantes de renovação $\{t_n\}$ ocorrem quando uma célula termina o serviço, sendo o estado da fila $S(t)$ completamente caracterizado por $X(t)$ e $J(t)$. O processo de ocupação da fila pode ser descrito por uma cadeia de Markov discreta no tempo, onde as probabilidades de transição entre os estados da fila, nos instantes de renovação, são dadas pela matriz:

$$P = \begin{bmatrix} UA_0 & UA_1 & UA_2 & \dots & UA_{K-2} & \sum_{n=K-1}^{\infty} UA_n \\ A_0 & A_1 & A_2 & \dots & A_{K-2} & \sum_{n=K-1}^{\infty} A_n \\ 0 & A_0 & A_1 & \dots & A_{K-3} & \sum_{n=K-2}^{\infty} A_n \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A_0 & A - A_0 \end{bmatrix} \quad (\text{C.7})$$

onde:

- P é a matriz de probabilidade de transição da cadeia de Markov, cujos elementos são:

$$p_{i,jr} = \Pr\{S(t_{n+1}) = [j, r] / S(t_n) = [i, l]\}, \quad i, j = 0, 1, \dots, K-1 \text{ e } l, r = 1, \dots, m.$$

- $U = (\Lambda - R)^{-1}\Lambda$ é uma matriz $m \times m$, onde o elemento $u_{i,j}$ representa a probabilidade condicional de atingir a fase j , dado que começou na fase i , durante um período em que o servidor esteja desocupado.

Encontrada a matriz P pode-se obter os vetores $\pi(k)$ ($k = 0, 1, \dots, K-1$) m -dimensionais, que indicam a probabilidade do sistema possuir k usuários na fila, no instante de renovação. Para cada k , a j -ésima componente do vetor indica probabilidade da fonte estar na fase j .

Então $\pi = [\pi(0) \pi(1) \dots \pi(K-1)]$, satisfaz:

$$\pi = \pi P; \quad \pi u = 1 \quad (\text{C.8})$$

onde u é um vetor unitário m -dimensional.

O tráfego médio escoado na fila é dado por:

$$\xi = [1 + \pi(0)(\Lambda - R)^{-1}\mu u]^{-1} \quad (\text{C.9})$$

A probabilidade de perda de células B é dada por:

$$B = 1 - \mu(q\Lambda u)^{-1}\xi \quad (\text{C.10})$$

O tempo médio de espera no *buffer* pode ser encontrado através do teorema de Little.

A probabilidade da fila, em qualquer instante t , possuir i células e a fonte MMPP estar na fase j é dada pelo sistema de equações:

$$\begin{cases} p(0) = \mu\xi\pi(0)(\Lambda - R)^{-1} \\ p(i) = \mu\xi[\pi(i) + \sum_{k=0}^{i-1} \pi(k)U^{i-1-k}(U - I)](\Lambda - R)^{-1}, \quad i = 1, \dots, K-1 \end{cases} \quad (\text{C.11})$$

O número médio de células na fila é:

$$E\{X\} = \sum_{n=0}^K np_n \quad (\text{C.12})$$

onde $p_n = p(n)u$.

A taxa média efetiva de chegadas que entram no *buffer* é dada por:

$$E\{\lambda_n\} = \mu\xi \quad (\text{C.13})$$

O tempo médio de permanência no sistema é encontrado através do teorema de Little:

$$E\{T\} = \frac{E\{X\}}{\mu\xi} \quad (\text{C.14})$$

e o tempo médio de espera no *buffer* é dado por:

$$E\{W\} = E\{T\} - \frac{1}{\mu} \quad (\text{C.15})$$

Referências Bibliográficas

- [1] ATM Forum, af-tm-0056.000. *Traffic Management Specification, Version 4.0*, April, 1996.
- [2] A. Baiocchi and N. Bléfari-Melazzi. Steady-state analysis of the MMPP/G/1/K queue. *IEEE Transactions on Communications*, 41(4):531–534, April 1993.
- [3] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali. *Linear Programming and Networks Flows*. John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 1990.
- [4] J.-Y. Le Boudec. An efficient solution method for Markov models of ATM links with loss priorities. *IEEE Journal on Selected Areas in Communications*, 9(3):408–417, April 1991.
- [5] J.-Y. Le Boudec. Network calculus made easy. Technical Report EPFL-DI 96/218, Laboratoire de Réseaux de Communication, Ecole Polytechnique Fédérale de Lausanne (EPFL), December 1996.
- [6] J.-Y. Le Boudec and R. Nagarajan. A CAC algorithm for VBR connections over a VBR trunk. In *Proceedings of 15th International Teletraffic Congress - ITC*, volume 1, pages 59–70, Washington, DC, USA, 1997.
- [7] C. Bruni, U. Mocci, P. Pannunzi, and C. Scoglio. Efficient capacity assignment for ATM virtual paths. In *Proceedings of ATM Hot Topics on Traffic and Performance from RACE to ACTS*, number 14, Milano, June 1995.
- [8] J. C. de Oliveira, I. S. Bonatti, A. K. Budri, and P. L. D. Peres. Encaminhamento de tráfego em redes públicas ATM. *Anais do XV Simpósio Brasileiro de Telecomunicações*, Recife, Brasil, Setembro de 1997.
- [9] J. C. de Oliveira, A. K. Budri, I. S. Bonatti, and P. L. D. Peres. Dimensionamento de enlaces ATM. *Anais do IV Congresso PETROBRÁS de Informática e Telecomunicações*, São Paulo, Brasil, Dezembro de 1997.
- [10] M. de Prycker. *Asynchronous Transfer Mode: Solution for Broadband Networks*. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1995.
- [11] 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, August 1995.

- [12] A. I. Elwalid and D. Mitra. Effective bandwidth of general markovian traffic sources and admission control of high speed networks. *IEEE Transactions on Networking*, 1(3):329–343, June 1993.
- [13] R. J. Gibbens. Traffic characterization and effective bandwidths for broadband network traces. Research Report. Submitted for publication. 1996-9, Statistical Laboratory, University of Cambridge, 1996.
- [14] D. Gross and C. M. Harris. *Fundamentals of Queueing Theory*. John Wiley & Sons, Inc., Toronto, Canada, 1974.
- [15] R. Guérin, H. Ahmadi, and M. Naghshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE Transactions on Selected Areas in Communications*, 9(7):968–981, September 1991.
- [16] ITU-T, Rec. I.150. *B-ISDN asynchronous transfer mode functional characteristics*, 1995.
- [17] ITU-T, Rec. I.311. *B-ISDN general network aspects*, 1993.
- [18] ITU-T, Rec. I.321. *B-ISDN protocol reference model and its application*, 1993.
- [19] ITU-T, Rec. I.350. *General aspects of quality of service and network performance in digital networks, including ISDNs*, 1993.
- [20] ITU-T, Rec. I.356. *B-ISDN ATM layer cell transfer performance*, 1996.
- [21] ITU-T, Rec. I.361. *B-ISDN ATM layer specification*, 1995.
- [22] ITU-T, Rec. I.362. *B-ISDN ATM adaption layer (AAL) functional description*, 1993.
- [23] ITU-T, Rec. I.363. *B-ISDN ATM adaption layer (AAL) specification*, 1993.
- [24] ITU-T, Rec. I.371. *Traffic control and congestion control in B-ISDN*, 1995.
- [25] H. Ji. An economic model for bandwidth allocation in broadband communications networks. In *Proceedings of ICC'96*, volume 1, pages 658–662, 1996.
- [26] C. A. Noronha Jr. and F. A. Tobagi. Evaluation of multicast routing algorithms for multimedia streams. In *Proceedings of IEEE International Telecommunication Symposium*, pages 390–397, Rio de Janeiro, Brazil, August 1994.
- [27] C. A. Noronha Jr. and F. A. Tobagi. Optimum routing of multicast streams. In *Proceedings of IEEE INFOCOM 94*, pages 865–873, Toronto, Canada, June 1994.
- [28] J. S. Kaufman. Blocking in a shared resource environment. *IEEE Transaction on Communications*, 29(10):1474–1481, 1981.
- [29] F. P. Kelly. Mathematical models of multiservice networks. In D. M. Titterington, editor, *Complex Stochastic Systems and Engineering*. Oxford University Press, 1994.

- [30] F. P. Kelly. Notes on effective bandwidths. In F. P. Kelly, S. Zachary, and I. B. Ziedins, editors, *Stochastic Networks: Theory and Applications*, pages 141–168. Oxford University Press, 1996.
- [31] G. Kesidis, J. Walrand, and C.-S. Chang. Effective bandwidths for multiclass Markov fluids and other ATM sources. *IEEE Transactions on Networking*, 1(4):424–428, August 1993.
- [32] The Mathworks, Inc., USA. *Matlab User's Guide*, 1991.
- [33] J. M. Pitts and J. A. Schormans. *Introduction to ATM Design and Performance*. John Wiley & Sons, Inc., Chichester, England, 1996.
- [34] M. Ritter and P. Tran-Gia. Multi-rate models for dimensioning and performance evaluation of ATM networks - COST 242. Technical report, University of Wurzburg, June 1994.
- [35] J. Roberts, U. Mocci, and J. Virtamo. *Broadband network teletraffic - performance evaluation and design of broadband multiservice networks - final report of action COST 242*. Springer-Verlag, Berlin, Heidelberg, New York, 1996.
- [36] L. F. G. Soares, G. Lemos, and S. Colcher. *Redes de Computadores: Das LANs, MANs e WANs às Redes ATM*. Editora Campus, Rio de Janeiro, 1995.
- [37] W. Stallings. *ISDN and Broadband ISDN with Frame Relay and ATM*. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1995.
- [38] TELEBRÁS. *Dimensionamento de redes de telecomunicações*, Fevereiro de 1997.
- [39] E. W. M. Wong, A. K. M. Chan, S. C. H. Chan, and K. T. Ko. Bandwidth allocation and routing in virtual path based ATM networks. In *Proceedings of ICC'96*, volume 1, pages 647–652, 1996.
- [40] S. Zhu and G. M. Huang. A new packet-loss minimization routing algorithm for ATM high-speed data networks. In *Proceedings of the 35th Conference on decision and Control*, volume 1, pages 287–292, Kobe, Japan, December 1996.