Capítulo 1

Introdução

As Redes Digitais de Serviços Integrados de Faixa Larga (RDSI-FL) foram desenvolvidas para oferecer toda a infra-estrutura necessária para suportar uma grande variedade de classes de serviços compostos dos mais diversos tipos de tráfego com diferentes requisitos de qualidade de serviço [1, 2]. Serviços atuais, tais como telefone e telex, que requerem taxas de transmissão da ordem de algumas dezenas de kbps coexistirão com outros serviços como HDTV e teleconferência, cujas taxas de transmissão podem chegar a centenas de Mbps, taxas estas impossíveis de serem fornecidas pelas atuais redes de telecomunicações. Além disto, estima-se que uma grande parcela do tráfego seja gerado por aplicações com taxas de transmissão variáveis, como, por exemplo, vídeo e som comprimido. Nestes tipos de aplicação, o tráfego gerado pode sofrer variações bruscas onde a taxa máxima de transmissão pode ser diversas vezes superior à taxa média. Uma rede eficiente deve acomodar este tipo de tráfego sem alocar uma banda passante igual a taxa de pico.

As RDSI-FL devem, portanto, atender de forma eficiente à qualquer tipo de aplicação, ou seja, as que exigem taxas baixas, altas, fixas, variáveis e de tempo real. Tais necessidades obrigaram a uma análise dos atuais modos de transferência de dados.

O modo circuito possibilita a transferência de dados, após a fase de estabelecimento de conexão, à uma taxa constante e igual a do canal. Porém, a banda passante de uma rede comutada por circuitos consiste de canais de tamanhos fixos e estes não podem ser modificados facilmente. Como para fornecer a banda passante heterogênea, a diversidade de canais provido por uma conexão aumenta, o planejamento dos compartimentos de tempo se torna impraticável, impossibilitando o seu uso para redes que necessitem de uma grande diversidade de bandas passantes, que é o caso das RDSI-FL. Além disto, pelo fato de toda a banda passante do canal ser dedicada a uma conexão enquanto esta perdurar, mesmo que não esteja utilizando para transmitir, implica numa utilização ineficiente da rede.

Nas redes comutadas por pacote, pelo fato dos atuais protocolos terem sido desenvolvidos para uma comunicação confiável entre computadores através de canais de baixa qualidade, foram implementados procedimentos complexos de detecção e correção de erros. Toda esta complexidade torna difícil a sua implementação em comunicações de alta velocidade e não atende às necessidades de transferência em tempo real. No entanto, algumas características como a fácil implementação de bandas passantes heterogêneas e a transferência de pacotes somente quando existem dados a serem enviados são adequadas para as RDSI-FL.

Por sobrepor as deficiências dos outros modos de transferência e permitir alcançar uma alta taxa de utilização em altas velocidades de transmissão, o Modo de Transferência Assíncrona (ATM - *Asynchronous Transfer Mode*) foi proposto como a tecnologia a ser utilizada nas RDSI-FL. As redes ATM possuem as seguintes características:

- os pacotes de transmissão possuem tamanho fixo e reduzido, chamados de células. Transmitir a uma taxa variável corresponde a transmitir o número variável de células por unidade de tempo;
- o cabeçalho da célula contém campos, denominados de VCI e VPI, que identificam o endereço que são usados na multiplexação. A célula é comutada de acordo com o valor de VPI/VCI setado ordinariamente no estabelecimento da conexão. Ou seja, a célula é a unidade de transmissão e de comutação, podendo ser transmitida diretamente para uma certa rota sem requerer procedimentos especiais; e
- a transferência de informações em uma rede ATM é baseada sobre um protocolo sem controles complexos nó-a-nó tanto de fluxo quanto de erro. Apenas a verificação da correção do cabeçalho é feito nó-a-nó.

Uma das principais vantagens na utilização das redes ATM está na economia de recursos que pode ser obtida através da multiplexação estatística de suas chamadas, pois, com isto, uma banda passante menor que a taxa máxima de transmissão pode ser alocada para cada fonte e um número maior de conexões pode ser aceito. Além da rede ser eficiente, ela deve atender às expectativas do seu usuário. Uma transferência de arquivo deve ser confiável, não permitindo perda de células, e uma teleconferência não pode ser afetada por atrasos na transmissão de células. Como estas necessidades são diversas e dependem de cada aplicação em particular, denomina-se Qualidade de Serviço (QoS - *Quality of*

Service) ao conjunto de parâmetros de desempenho que a rede deve garantir à aplicação usuária. Estes parâmetros são normalmente negociados durante o estabelecimento da conexão.

Os conceitos fundamentais de multiplexação estatística e gerenciamento de recursos são descritos a seguir de forma a melhor compreender os requisitos de uma rede eficiente e do atendimento da qualidade de serviço exigida pela aplicação

1.1 Ganho de Multiplexação Estatística

Em redes ATM, várias fontes de tráfego com as mais diversas características são superpostas em um mesmo canal de transmissão, tal como um tronco de linha composto por centenas de chamadas de videofone. A esta superposição, chama-se de multiplexação estatística.

O número de células geradas por um usuário de uma rede ATM muda de acordo com a quantidade de informação a ser transferida. Portanto, a quantidade de recursos de rede para cada usuário também varia constantemente e proporcionalmente ao número de células geradas por unidade de tempo. Como estes usuários, individualmente, não alcançam seus valores máximos de taxa de transmissão simultaneamente, a quantidade total de recursos a ser fornecida pela rede será menor caso estes recursos sejam compartilhados por todos os usuários. Este fenômeno, chamado de ganho de multiplexação estatística, é uma das principais características das redes ATM.

A figura 1.1 ilustra com um exemplo, onde três conexões com diferentes requisitos de banda passante são multiplexadas, o ganho que pode ser obtido com a multiplexação estatística. Como pode se observar, a taxa máxima do tráfego gerada pelo agregado das conexões é menor que a soma das taxas máximas individuais, resultando, assim, numa boa economia de banda passante.

No entanto, o compartilhamento de recursos ou a multiplexação estatística, sem a devida monitoração e controle, podem causar situações de congestionamento, com perdas e atrasos de células, que degradam a qualidade de serviço requerida. Os mecanismos de controle de admissão de conexão, juntamente com outros mecanismos de controle de tráfego e de congestionamento, reduzem a ocorrência destas situações abusivas, por parte do usuário, no compartilhamento de recursos da rede.

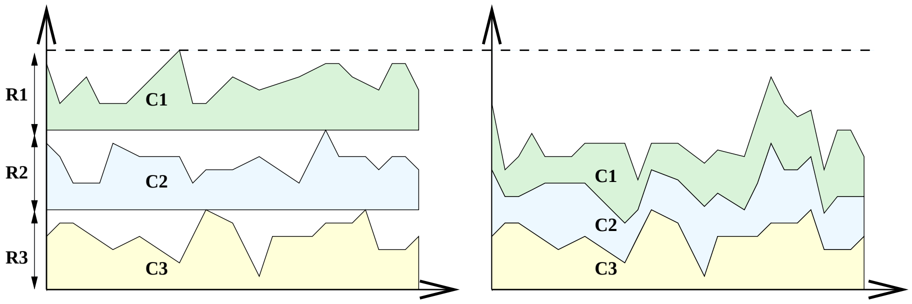

Figura 1.1: Ganho com a multiplexação estatística em redes ATM.

1.2 Gerenciamento de Recursos em Redes ATM

A flexibilidade oferecida pelas redes ATM fez com surgissem novos conceitos no gerenciamentos de recursos. Para evitar estados de congestionamento e minimizar a duração destes estados foram criados os controles de tráfego e de congestionamento.

Em RDSI-FL, o congestionamento é definido como o estado dos elementos da rede nos quais a rede não é capaz de satisfazer aos objetivos de desempenho de rede contratados para as conexões já estabelecidas e/ou para os novos pedidos de conexão [3]. Em uma rede telefônica, onde os canais são alocados exclusivamente para uma conexão, o congestionamento se dá quando não existem mais canais disponíveis para novos pedidos de conexão ao menos em parte da rede.

Geralmente um congestionamento pode ser causado por flutuações estatísticas imprevisíveis dos fluxos de tráfego ou por condições de falha dentro da rede.

O controle de tráfego é definido como o conjunto de ações tomadas pela rede para evitar condições de congestionamento, enquanto que o controle de congestionamento é definido como o conjunto de ações tomadas pela rede para minimizar a intensidade, o espalhamento e a duração dos mesmos.

Os objetivos dos controles de tráfego e de congestionamento da camada ATM são:

- suportar um conjunto de classes de Qualidade de Serviço (QoS) da camada ATM suficiente para todos os serviços previsíveis das RDSI-FL;
- não serem confiados a protocolos da camada AAL, que são específicos de serviços das RDSI-FL; e
- minimizar a complexidade da rede e do sistema terminal, enquanto maximiza a utili-

zação dos recursos da rede, através da obtenção de um conjunto ótimo de controles.

As dificuldades relacionadas ao controle de tráfego em redes ATM envolvem diversos aspectos, entre os quais destacam-se [4]: a alta taxa de transmissão dos enlaces; a enorme variedade dos requisitos de serviço; e a diversidade das características de tráfego que as redes RDSI-FL deverão suportar.

Em relação ao primeiro aspecto, a ITU-T recomenda duas velocidades de acesso às redes ATM, são elas aproximadamente 155 Mbps e 600 Mbps. Portanto, as células serão transmitidas em intervalos, cuja duração será de aproximadamente $3\mu s$ ou $0.7\mu s$. Pode-se concluir que a velocidade de processamento de células em redes ATM deverá ser comparável às altas velocidades de comutação. Para tal, a simplificação dos esquemas de processamento será fundamental.

Outro problema referente a alta taxa dos enlaces é a latência elevada nas redes ATM. O produto da taxa de transmissão pelo atraso de propagação nas fibras óticas representa um alto número de *bits* armazenados no meio. O controle reativo, onde o receptor envia mensagens de controle para regular o fluxo de transmissão quando a rede está congestionada, deixa de ser uma solução adequada, devido à dificuldade em gerenciar a enorme capacidade de armazenamento requerida.

O segundo aspecto refere-se à multiplicidade de requisitos de serviço em redes ATM. Em geral estes requisitos envolvem limitação (determinística ou estatística) de atraso, variação de atraso e finalmente, perda de células. Ao admitir uma chamada, a rede deve garantir os requisitos por esta solicitados, de forma eficiente, sem prejudicar, no entanto, a garantia dos requisitos de chamadas anteriormente admitidas. Esquemas de prioridades costumam ser utilizados para o atendimento destes requisitos.

Já o terceiro aspecto diz respeito às características do tráfego ATM. Nestas redes, aplicações com as mais diversas características são multiplexadas. Com isto, uma caracterização do tráfego gerado por estas aplicações se faz necessária para o controle efetivo da rede.

1.2.1 Mecanismos de Controle de Tráfego e de Congestionamento

Para satisfazer aos objetivos dos controles de tráfego e de congestionamento, foram propostas as seguintes funções genéricas [3]: Gerenciamento de Recursos da Rede, Controle de Admissão de Conexão, Controles de Realimentação, Controle de Parâmetros do Usuário/Rede e Controle de Prioridade.

No Gerenciamento de Recursos da Rede (NRM - *Network Resource Management*), a provisão pode ser usada para alocar recursos da rede de modo a separar os fluxos de tráfego de acordo com as características do serviço.

O Controle de Admissão de Conexão (CAC - *Connection Admission Control*) é definido como o conjunto de ações tomadas pela rede durante a fase de estabelecimento de conexão (ou durante a fase de renegociação da chamada) para estabelecer quando um pedido de conexão de canal virtual/caminho virtual pode ser aceito ou rejeitado (ou quando o pedido de realocação pode ser acomodado). O roteamento faz parte das ações do CAC.

Os Controles de Realimentação são definidos como o conjunto de ações tomadas pela rede e pelo usuário para regular o tráfego oferecido às conexões ATM de acordo com o estado dos elementos da rede.

O Controle de Parâmetros do Usuário/Rede (UPC/NPC - *Usage/Network Parameter Control*) é definido como o conjunto de ações tomadas pela rede para monitorar e controlar o tráfego, em termos de tráfego oferecido e da validade da conexão ATM, no acesso do usuário e da rede, respectivamente. Seu propósito principal é o de proteger os recursos da rede de fontes maliciosas, assim como de maus comportamentos não intencionais, que podem afetar a QoS de outras conexões já estabelecidas através da detecção da violação dos parâmetros negociados e tomando as atitudes apropriadas.

No Controle de Prioridade, o usuário pode gerar fluxos de tráfego de diferentes prioridades utilizando o *bit* de prioridade de perda de célula. Um elemento de rede congestionado pode descartar seletivamente células de baixa prioridade quando necessário para proteger o quanto possível o desempenho da rede para células de alta prioridade.

1.3 Controle de Admissão de Conexão (CAC)

Em redes ATM, a banda passante de uma conexão não é muito clara dado que toda a informação é segmentada em células e um número necessário de células são geradas e transportadas pela rede. Com o aumento do número de conexões, a QoS no nível de célula, tal como a taxa de perda de células, deteriora-se. Portanto é impossível aceitar um número ilimitado de conexões e ainda manter a QoS desejada.

Um pedido de conexão para uma determinada chamada somente deve ser aceito quando há recursos suficientes em toda a rede que permitam atender à Qualidade de Serviço (QoS) desejada desta nova conexão, assim como a manutenção da QoS contratada das demais conexões já existentes.

Chamadas que necessitam mais de uma conexão, tais como serviços multimídias (e.g., uma videoconferência), terão que realizar negociações separadas com o CAC para cada uma de suas conexões.

Dados os descritores de tráfego da fonte e a classe de QoS desejada, o CAC deve determinar a aceitação ou não da chamada, os parâmetros de tráfego necessários para o controle de parâmetros do usuário (UPC) e o roteamento e alocação de recursos da rede.

A necessidade de estimar as capacidades equivalentes de uma fonte de tráfego e do tráfego agregado para a tomada de decisão sobre a aceitação ou recusa de uma chamada leva a duas importantes considerações.

A primeira é que como não se tem o conhecimento exato da distribuição dos processos de chegadas dos tráfegos das fontes, o cálculo aproximado das capacidades equivalentes pode ser tanto superestimado como subestimado. O primeiro caso levaria a uma menor utilização da rede, enquanto que o segundo poderia acarretar em perdas de células, mesmo que todas as fontes estivessem se comportando dentro dos parâmetros especificados no estabelecimento da conexão.

A segunda é que para verificar a conformidade entre o tráfego real oferecido pela fonte e aquele especificado durante a fase de estabelecimento de conexão é introduzido o UPC. Porém, como não há um mecanismo ideal de policiamento e desde que não se queira penalizar uma parcela do tráfego bem comportado (tráfego cujo parâmetros reais estão adequados àqueles fornecidos para o CAC), sempre passará pelo UPC uma parcela de tráfego mal comportado [5]. Sendo assim, a alocação de recursos deve possuir uma margem de segurança de modo a compensar estas não idealidades.

1.4 Motivação

Hodiernamente, o desenvolvimento de meios de comunicação cada vez mais rápidos e eficientes é fundamental para suprir a demanda das atuais e emergentes aplicações multimídias. Em razão disto, o Grupo de Teleinformática e Automação (GTA) da UFRJ tem se dedicado ao estudo e desenvolvimento de protocolos de alto desempenho, aplicações, arquiteturas de implementação e mecanismos de controle e garantia de qualidade de serviço que venham suprir esta demanda. No que concerne a redes de alta velocidade, o GTA participa do projeto TRAVEL (Tecnologias em Redes de Alta Velocidade) do ProTeM-CC, financiado pelo CNPq, que visa o estudo de problemas relacionados a utilização de redes de alta velocidade e, em particular, as redes ATM. Entre os objetivos deste projeto, tem-se a incorporação de mecanismos de controle de tráfego e garantia de QoS à arquitetura de implementação do GTA, a arquitetura GTA/MM, sobre uma plataforma ATM.

Neste contexto, várias questões se colocam. Como avaliar o ganho que pode ser obtido com a multiplexação estatística? Qual o número máximo de conexões que podem ser estabelecidas sem perda na qualidade de serviço provida? Como caracterizar e classificar os tráfegos gerados pelas aplicações? As respostas para estas questões não são simples e exigem um conhecimento mais detalhado das características do tráfego gerado pelas aplicações, além de procedimentos mais eficazes para o gerenciamento dos recursos da rede.

Sendo assim, este trabalho tem como objetivos:

- o estudo de modelos de caracterização de tráfego para permitir um melhor conhecimento dos serviços existentes e uma definição do conjunto mínimo de parâmetros de tráfego que deverão ser fornecidos pelo usuário;
- a análise da multiplexação estatística para permitir uma avaliação do ganho que pode ser obtido com o uso das redes ATM. Uma das vantagens desta análise está no estudo do impacto de cada um dos parâmetros de tráfego sobre o ganho estatístico;
- o estudo dos mecanismos de CAC e a análise da eficiência de diversos mecanismos propostos na literatura.

Esta tese está organizada em cinco capítulos. No capítulo 2, é feita uma caracterização de tráfego das diversas classes de serviço das redes ATM. Vários modelos de tráfego são apresentados neste capítulo. No capítulo 3, é feita uma análise do multiplexação estatística nas redes ATM e um levamento dos parâmetros de tráfego mais importantes para esta análise. Diversos mecanismos de CAC são apresentados e analisados no capítulo 4. Neste capítulo, ainda são feitas uma classificação destes mecanismos e uma análise do impacto da variação do atraso de células sobre estes mecanismos. Finalmente, no capítulo 5, são apresentadas as conclusões.