Capítulo I

Introdução

Garantia de Qualidade de Serviço: um dos principais desafios a ser vencido pela tecnologia emergente dos sistemas de comunicação multimídia baseados em Redes Digitais de Serviços Integrados em Faixa Larga (RDSI-FL). A evolução tecnológica das redes de comunicação conduz a uma rede universal e flexível [BSu91, Fae94a, Pry91, SLL94] que visa suportar os mais variados serviços, tais como: distribuição domiciliar de TV de alta definição (*High Definition Television*-HDTV), educação e consultas médicas à distância, serviços meteorológicos, bem como transferência de dados a alta velocidade. Esta rede também deve ser capaz de prover serviços interativos como vídeo conferência (ou conferência multimídia), vídeo-telefonia, realidade virtual, trabalho cooperativo (*Computer Supported Cooperative Work*-CSCW), aplicações industriais e corporativas entre outros. O Modo de Transferência Assíncrona - *Asynchronous Transfer Mode* (ATM), também denominado rede ATM, deve atender esta demanda tecnológica.

Grande parte dos serviços que serão oferecidos pelas futuras redes integradas serão os serviços referentes às aplicações multimídia. Entre estes serviços, estão os chamados serviços isócronos que são caracterizados pela forte dependência temporal, onde é permitida pouca ou nenhuma variação nos atrasos de comunicação.

As aplicações multimídia, com seus fortes requisitos de alta velocidade e serviços de elevada qualidade, têm requerido o desenvolvimento de novas arquiteturas e protocolos de redes de comunicação. Como prover às aplicações uma qualidade de serviço (q*uality of service*-QoS) garantida? A solução para este problema não requer somente grandes bandas

passantes, baixos atrasos e baixa taxa de erro (ou perda), mas também limites (superiores ou inferiores) garantidos para estes índices de desempenho. É importante enfatizar que esta rede tem como grande desafio não só atender as aplicações atuais, como também as aplicações futuras que não foram ainda projetadas, ou até mesmo cogitadas.

I.2 Caracterização da Qualidade de Serviço

A qualidade do serviço de uma aplicação de tempo real é determinada pelos valores de alguns parâmetros de desempenho fim-a-fim, entre os mais comuns pode-se citar:

Vazão do Tráfego: este índice está relacionado com a necessidade de caracterizar como o tráfego do fluxo será injetado na rede. Ele expressa a quantidade de dados que será enviada através da rede por unidade de tempo.

Atraso: este índice de qualidade está relacionado com a sensibilidade das aplicações em relação ao atraso introduzido pela rede. Algumas aplicações são mais sensíveis que outras tanto em relação ao atraso quanto a variação no atraso (*jitter)*. A ITU-T denomina a característica de variação no atraso de CDV - *Cell Delay Variation*. O termo *jitter* é comumente usado para expressar variação no atraso, no entanto pode ser utilizado para expressar qualquer tipo de variação, como por exemplo *jitter* de vazão expressa variação na taxa de transmissão.

Confiabilidade: este índice está relacionado com a sensibilidade das aplicações em relação à perda de dados. Este índice de qualidade é fortemente dependente dos mecanismos de armazenamento e reserva de recursos envolvidos na transmissão da informação ao longo da rede. Existem duas fontes de perda de dados: devido à erros de transmissão (interferências ou falhas) e devido à transbordo de memória. Os recentes trabalhos apresentados na literatura consideram basicamente a perda de células devido a transbordo de memória em redes ATM.

A especificação dos parâmetros de qualidade de serviço pode ser determinística ou estatística. Por exemplo, a especificação do limite de atraso pode ser feita de forma determinística : $A_i \leq A_{max}$ para todo i , onde A_i é o atraso da i-ésima mensagem enviada pelo usuário ao destino e A_{max} é o limite máximo de atraso especificado pelo usuário. A especificação estatística seria feita da seguinte maneira: *Prob* ($A_i \leq A_{max}$) $\geq P_{min}$, onde A_i e A_{max} seguem a definição anterior e P_{min} é o limite mínimo de probabilidade do requisito de atraso ($A_i \leq A_{max}$) ser cumprido com sucesso.

O segundo problema com atrasos máximos é que eles tentam quantificar o desempenho que é aceitável para o usuário e uma aplicação, em geral, não sabe qual desempenho será aceitável para o seu usuário. Por exemplo, uma justificativa comum para especificar o máximo atraso aceitável é que considera-se humanamente difícil conversar com atrasos maiores que 100 ms. Certamente tal atraso pode tornar a comunicação menos agradável, mas é ainda possível se conversar mesmo com atrasos de vários segundos de duração, e dada a escolha entre a não conexão e um atraso longo, muitos usuários escolheriam o atraso longo.

I.3 Controle de Tráfego

Um assunto, dentre os mais complexos, que ainda hoje apresenta várias questões em aberto é o controle de tráfego em sistemas de comunicação multimídia em RDSI-FL. As dificuldades relacionadas ao controle de tráfego em ATM envolvem diversos aspectos, entre os quais destacam-se [HSu91]:

- a alta taxa de transmissão dos enlaces;
- a enorme variedade dos requisitos de serviço;
- a diversidade das características de tráfego que a rede RDSI-FL deverá suportar.

Em relação ao primeiro aspecto, a ITU-T recomenda duas velocidades de acesso à rede ATM, são elas aproximadamente 155 Mb/s e 600 Mb/s. Portanto, as células serão

transmitidas em intervalos, cuja duração será de aproximadamente 3μs ou 0,7μs. Pode-se concluir que a velocidade de processamento de células em 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. Por exemplo, em uma transmissão do Rio de Janeiro a Manaus (3000 Km de distância) considerando um atraso de propagação de 10ms e taxa de transmissão de centenas de Mb/s, o número de *bits* armazenado no meio chegaria a ordem 10Mb. 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 uma rede ATM. Em geral estes requisitos envolvem limitação (determinística ou estatística) de atraso, variação de atraso e finalmente, perda de células [Fer90]. 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. Tráfegos com alto índice de rajada deverão ser multiplexados com tráfegos cuja variação na taxa de transmissão é pequena. O índice de rajada de uma fonte de tráfego é caracterizado como a razão entre a taxa máxima e a taxa média de transmissão. Esta não é uma forma muito acurada de caracterização de uma fonte de tráfego, porém sua simplicidade a torna um parâmetro de larga utilização.

Tomando como base a pequena probabilidade de um grande número de chamadas (com alto nível de rajada) transmitir à taxa máxima simultaneamente, a técnica de multiplexação estatística aloca para cada chamada uma banda passante abaixo de sua taxa máxima.

I.3.1 Mecanismos de Controle de Tráfego

Grande parte dos mecanismos desenvolvidos se limitam a garantias de desempenho determinísticas e esquemas de alocação de recursos estáticos. Estudos mais avançados buscam mecanismos melhor adaptados a tráfegos com alto nível de rajada , onde o conceito de garantia de QoS é estatisticamente definido, permitindo uma melhor utilização dos recursos da rede. Também é objeto de estudo, a alocação dinâmica de recursos de acordo com as variações no tráfego da rede.

Diversas modalidades de controle de tráfego, para RDSI-BL, são previstas em [ITU93]. Estas modalidades classificam-se em controle de admissão de conexão, controle de prioridade, controle de congestionamento e controle de parâmetros de utilização/rede.

O controle de admissão de conexão refere-se à fase de alocação de recursos da rede, como memória e banda passante para uma nova conexão. Uma nova conexão é admitida caso a rede disponha de recursos suficientes para atender o nível de qualidade de serviço requisitado pela nova conexão, sem no entanto alterar a qualidade de serviço das conexões previamente estabelecidas.

O controle de prioridade está relacionado com o processamento dos fluxos de conexões que contêm células de diferentes prioridades, indicadas no cabeçalho através do *bit* prioridade de perda de célula. O controle dos elementos da rede, realizado para evitar que uma sobrecarga de tráfego degrade a qualidade de serviço, do sistema de comunicação, é função do controle de congestionamento.

Conexões de um sistema de comunicação baseado em RDSI-BL, compartilham recursos da rede como memória e banda passante do canal de enlace. Para que o compartilhamento de recursos ocorra de forma disciplinada, sem a degradação da qualidade de serviço é necessário que o tráfego das conexões seja devidamente monitorado e controlado. O controle de parâmetros de utilização/rede realiza esta função. Este controle também é denominado de policiamento. Seu principal objetivo é proteger os recursos da rede contra tráfegos abusivos, que violam intencionalmente, ou não, os parâmetros do contrato de admissão. Os mecanismos que realizam o controle de policiamento, além de monitorar e controlar as fontes de tráfego de forma eficaz, devem demandar baixo custo de implementação e apresentar flexibilidade de parametrização.

I.4 Motivação

O Grupo de Teleinformática e Automação (GTA) da UFRJ dedica-se atualmente 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 a demanda tecnológica dos atuais e emergentes sistemas de comunicação multimídia.

Este trabalho consiste de um estudo de mecanismos de monitoração e policiamento de tráfego, da proposta de um novo mecanismo de policiamento e da implementação de um conjunto de funcionalidades de multiplexação estatística incorporado à arquitetura de implementação GTA: a Arquitetura GTA/MM. Uma das principais características deste novo mecanismo de policiamento é sua flexibilidade de parametrização. A implementação das funcionalidades de multiplexação estatística, por sua vez, permitem que sejam desenvolvidos ambientes para o estudo de situações que envolvam compartilhamento de recursos, como por exemplo, policiamento de tráfego. As novas funções de multiplexação estatística visam ainda contribuir para o estudo e realização de experimentos relacionados a outros aspectos como: controle de congestionamento [HSu91], alocação de recursos [GAN91], controle de admissão

e sincronização de mídias [RVR92]. Estes últimos tópicos, são temas de outras teses de mestrado em desenvolvimento no GTA.

Esta tese está organizada em seis capítulos. No capítulo II são descritos os mecanismos de monitoração e policiamento de tráfego mais citados na literatura, incluindo seus modelos analíticos. A proposta do novo mecanismo de policiamento Janela Móvel Particionada é apresentada no capítulo III. O capítulo IV descreve a implementação de mecanismos de policiamento realizada sobre a Arquitetura GTA/MM. A proposta de implementação do conjunto de funções de multiplexação estatística é apresentada no capítulo V. Neste capítulo é descrito ainda um experimento que conjuga policiamento de tráfego e multiplexação estatística. Finalmente, no capítulo VI são apresentadas as conclusões.