

2 Qualidade de serviço

A partir dos anos 90, as redes comutadas por pacotes começaram a substituir as redes comutadas por circuitos devido à ineficiente utilização das redes utilizando tal comutação e a possibilidade de se obter ganhos com a multiplexação estatística derivada da comutação em pacotes. Além do que, uma rede com um único tipo de comutação oferece ganhos em questões de interoperabilidade e simplicidade.

Anteriormente a voz trafegava em redes telefônicas e dados eram trafegados em redes IP. A partir desse momento, ocorreu a chamada “convergência IP” onde dados e voz começaram a compartilhar as mesmas redes.

Há várias razões e benefícios para tal migração de redes comutadas por circuito para redes comutadas por pacotes (IP), como por exemplo:

- Excelente suporte para aplicações multimídia, pois com uma conectividade unificada, o número de dispositivos necessários se torna menores, reduzindo os esforços de instalação, operação e manutenção;
- A integração dos dispositivos de rede tem o potencial de simplificar o gerenciamento da segurança fim-a-fim e ao mesmo tempo tornar a rede mais robusta;
- Redes IP convergentes se beneficiam dos desenvolvimentos no roteamento baseado em “QoS” e da diferenciação de classes de serviço, resultando em uma melhor utilização dos recursos da rede;
- A arquitetura IP é capaz de lidar com as diferentes tolerâncias a atrasos, perda de pacotes e taxas de erros das diferentes tipos de aplicações

- Redes IP são redes altamente escaláveis fazendo com que suas aplicações possam ser trafegadas em qualquer parte do mundo;
- Redes IP convergentes oferecem redução de custos em termos de hardware e utilização de espaços físicos.

Na rede IP original o tráfego de dados era denominado de “store and forward”, ou seja, quando ocorria um erro ou perda na transmissão de pacotes, uma resposta negativa era enviada ao transmissor e o pacote errôneo era então retransmitido. Este tráfego também denominado de “melhor esforço” (do Inglês, “Best-effort”) se encaixava muito bem a essa necessidade não oferecendo nenhuma garantia de serviço. O objetivo então era assegurar que os terminais teriam os protocolos e inteligência suficientes para garantir uma transmissão sem erros ou perdas.

Quando as redes de voz e dados se fundiram, novos desafios tecnológicos surgiram fazendo com que o tráfego “melhor esforço” não fosse adequado o suficiente para atender aos diversos requisitos de desempenho, algumas vezes conflitantes dos vários tipos de informação (voz e vídeo) que começaram a ser trafegados nas redes IP convergentes.

E a tecnologia criada para solucionar esses problemas técnicos foi denominada de “Qualidade de Serviço” (QoS).

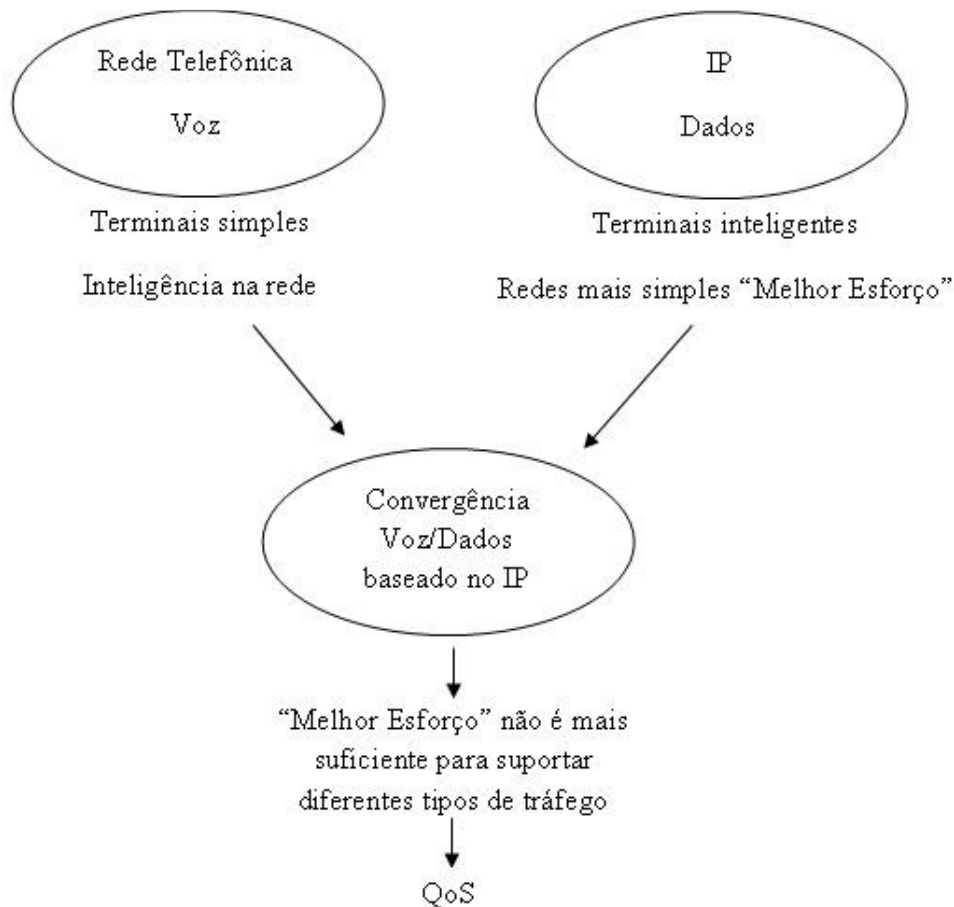


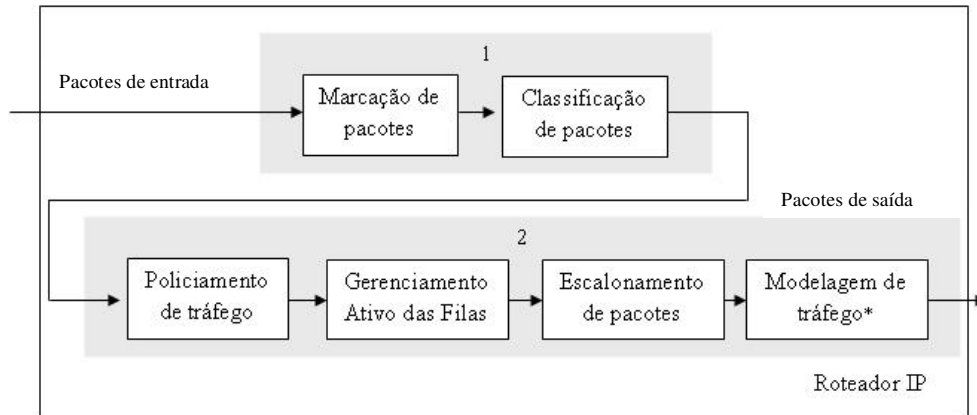
Figura 1 - Evolução das redes de telecomunicações

Qualidade de serviço, então, pode ser definida por dois pontos de vista: o do usuário e o da rede. Sob o ponto de vista do usuário, é a qualidade percebida do tipo de serviço/aplicação que contratado da provedora de serviço. Já sob o ponto de vista da rede, a qualidade de serviço depende da alocação dos recursos de comunicação disponíveis para manter certos parâmetros técnicos da rede em patamares necessários à oferta adequada de cada tipo de serviço requisitado.

Logo para prover qualidade de serviço, a rede precisa:

1. Diferenciar entre os tipos de fluxos de tráfego e de serviço para que sejam tratados diferentemente entre si;
2. Tratar as diversas classes de serviço para que estas recebam tratamento diferenciado no que concerne a garantias na alocação dos recursos da rede.

A figura seguinte sumariza os requerimentos necessários para prover qualidade de serviço, em blocos funcionais de um roteador IP, de acordo com os itens acima.



* Do inglês "Traffic Shaping"

Figura 2 – Requerimentos funcionais de QoS em um roteador IP

A forma como os usuários expressam sua satisfação com o serviço oferecido está relacionada com parâmetros técnicos da rede tais como atraso, jitter taxa de transmissão de dados e outros, cujos valores devem estar localizados em patamares característicos de cada tipo de serviço requisitado.

Como a rede atende a um conjunto de diferentes tipos de serviços com diferentes requisitos de desempenho, nem sempre um parâmetro técnico da rede tem um impacto significativo em um tipo de serviço. Assim sendo o mecanismo de QoS procura reduzir o impacto negativo dos parâmetros da rede significativos para cada tipo de serviço, visando assim a melhoria da satisfação dos usuários com os serviços recebidos.

Os parâmetros técnicos que são levados em consideração neste trabalho estão restritos a atraso, jitter (variância do atraso) e taxa de transmissão de dados, devido à importância dos mesmos no bom funcionamento das classes de serviço implementadas. Uma métrica importante que não está sendo considerada devido a forma como o sistema em questão é projetado é a probabilidade de perda de pacotes. Como será mostrado adiante, o sistema se baseia no funcionamento do roteador de borda da rede de uma provedora de serviços logo não há informações sobre quantos nós da rede o pacote percorreu nem há o conhecimento do tempo

máximo de vida de um pacote na rede para calcular a probabilidade de perda de pacotes.

2.1. Atraso

Atraso é o tempo medido entre o momento que um pacote de dados ingressa no sistema e o momento em que o destinatário recebe o pacote. Há várias fontes de atraso, tais como: codificação de fonte, empacotamento, codificação de canal, bufferização dos pacotes causando jitter, enfileiramento de pacotes e atraso de propagação. Neste trabalho é considerado apenas o atraso referente ao enfileiramento de pacotes, pois o sistema implementado considera filas de pacotes para serem entregues ao destino levando em conta os efeitos do processamento dos pacotes em roteadores de borda de uma provedora de serviços de telecomunicações, não há implementação de buffers para minimizar efeitos de jitter e os efeitos de propagação e codificação não estão sendo considerados.

Em relação aos tipos de aplicações consideradas, o atraso deve obedecer às seguintes relações:

Aplicações de áudio (VoIP) : para manutenção de um serviço satisfatório, é necessário baixo atraso máximo e médio. O limite máximo suportável de atraso em aplicações de áudio é da ordem de 400 ms [14].

Aplicações de dados (FTP e HTTP): para suportar aplicações interativas de dados com baixas taxas de transmissão de dados, é necessário um baixo atraso médio.

2.2. Jitter

O Jitter é definido como uma grandeza que expressa a flutuação temporal do atraso. Pode ser caracterizada por meio da variância estatística da variável aleatória que define o atraso dos pacotes, ou pode ser simplesmente definida como o valor máximo desta mesma variável. Seu efeito é diferenciado em aplicações de vídeo e áudio, pois estas aplicações são típicas de tempo real. É causado pela característica de rajada do tráfego de pacotes e é combatido através da inserção de buffers que os armazenam e retransmitem a uma taxa aproximadamente constante.

Há um compromisso natural entre tamanho do buffer e atraso inserido no sistema. Se o tamanho temporal deste buffer exceder 100 ms [15], há comprometimento na aplicação trafegada, principalmente nas aplicações em tempo real.

2.3.

Taxa de transmissão de dados

A taxa efetiva de transmissão de dados (para distinguir da taxa nominal de transmissão de dados) é a velocidade com que os pacotes de cada tipo de aplicação percorrem o caminho entre origem e destino no que concerne à transmissão da informação. Sendo assim esta taxa é sensível, a atrasos, retransmissões, e tudo aquilo mais que temporalmente afeta a transmissão dos pacotes.

No sistema proposto, a taxa efetiva de transmissão é medida para mensurar a sua adesão a taxa de transmissão e com isso avaliar se a alocação dos recursos da rede está sendo feita de forma justa entre usuários e tipos de aplicação.

Com a evolução das redes IP convergentes, um novo conceito vem ganhando cada vez mais espaço na literatura relativa a este assunto: trata-se do que são chamadas de “Redes da Próxima Geração (do inglês “NGN” - Next Generation Networks) [16]. A idéia geral desse conceito, é que uma mesma rede transporte todas as todas as informações e serviços (voz, dados e todos os tipos de mídias como, por exemplo, vídeo), encapsulando-os em pacotes tal como é feito o tráfego de dados na Internet. NGN’s são geralmente construídas com base no protocolo IP.

Com a integração de todas as informações e serviços, novos paradigmas e desafios vêm à tona levantando questões referentes à qualidade de serviço dessas informações e serviços convergentes [17]. Isso simboliza que independente da arquitetura usada, qualidade de serviço é um tema crucial para o sucesso tanto das redes IP convergentes quanto das redes NGN’s.