

Qualidade de Serviço em IP: nasce uma nova Internet

Carlos Frederico Marcelo da Cunha Cavalcanti¹

*Prof. do Departamento de Computação da Universidade Federal de Ouro Preto
Doutorando em Ciência da Computação na Universidade Federal de Minas Gerais
Mestre em Ciência da Computação pela Unicamp
Engenheiro Eletricista pela Pontifícia Universidade Católica de Minas Gerais*

Antônio Alfredo F. Loureiro²

*Prof. do Departamento de Computação da Universidade Federal de Minas Gerais
Doutor em Ciência da Computação pela University of British Columbia, Canadá
Mestre em Ciência da Computação pela Universidade Federal de Minas Gerais
Bacharel em Ciência da Computação pela Universidade Federal de Minas Gerais*

Antônio Otávio Fernandes³

*Prof. do Departamento de Computação da Universidade Federal de Minas Gerais
Doutor em Ciência da Computação pelo Institut National Polytechnique de Grenoble, França
Mestre em Ciência da Computação pelo Université Scientifique et Médicale de Grenoble, França
Engenheiro Eletricista pela Universidade Federal de Minas Gerais*

José Marcos S. Nogueira⁴

*Prof. do Departamento de Computação da Universidade Federal de Minas Gerais
Pós-Doutorado pela University of British Columbia, Canadá
Doutor em Engenharia Elétrica / Computação pela Unicamp
Mestre em Ciência da Computação pela Universidade Federal de Minas Gerais
Engenheiro Eletricista pela Universidade Federal de Minas Gerais*

PALAVRAS-CHAVE

Internet - Qualidade de serviço - Serviço integrado - Serviço diferenciado - Protocolo IP - DiffServ - IntServ

¹ E-mail: cfmcc@dcc.ufmg.br

² E-mail: loureiro@dcc.ufmg.br

³ E-mail: otavio@dcc.ufmg.br

⁴ E-mail: jmarcos@dcc.ufmg.br

RESUMO

Por causa do rápido crescimento do tráfego em aplicações de tempo real, as redes de computadores requerem mais funções do que anteriormente oferecidas no passado. Na presente geração da Internet, constata-se a grande demanda por aplicações comércio eletrônico, de voz e de multimídia. É essencial que a rede suporte diferentes níveis de Qualidade de Serviço (QoS) e cada nível seja associado a um usuário ou a um grupo de usuários distintos.

Em resposta ao crescimento da demanda por QoS na Internet, a *Internet Engineering Task Force* estabeleceu dois grupos de trabalho denominados DiffServ and IntServ. Este artigo apresenta os principais conceitos e soluções propostas por ambos para prover Qualidade de Serviço, estendendo a arquitetura TCP/IP

1. INTRODUÇÃO

A Internet nasceu da necessidade de interconectar redes distintas que não se comunicavam entre si. Inicialmente idealizada para utilização militar, na década de 1960, a Internet cumpriu, com sucesso, a tarefa de interconectar redes acadêmicas, particulares e públicas de forma integrada.

A rede mundial Internet está baseada no conjunto de protocolos definidos pela arquitetura TCP/IP [AR97]. Apesar de IP ser o protocolo da camada de rede que provê o encaminhamento dos pacotes IP e TCP ser o protocolo da camada de transporte que provê comunicação confiável, o termo TCP/IP é geralmente usado para denotar o conjunto de protocolos que são usados com IP, como UDP, FTP e outros.

As mensagens que são enviadas em uma rede Internet são divididas em pacotes IP que trafegam e que são encaminhados para o seu destino de uma forma mais rápida e confiável possível; por causa desta característica, o protocolo IP provê um serviço de envio de pacotes denominado “melhor esforço” (*Best Effort*). Como o protocolo IP não é orientado à conexão, os pacotes podem chegar fora de ordem, visto que a decisão de qual caminho a tomar é, geralmente, dinâmica e baseada na densidade de tráfego de cada enlace, mensurado em tempo real. Os protocolos da camada de transporte, como TCP (*Transport Control Protocol*) e RTP (*Real-Time Control Protocol*), são responsáveis em prover comunicação fim-a-fim (entre origem e destino, independentemente de quantos nós intermediários existirem) sem erros, implementando o conceito de conexão.

Devemos considerar que os protocolos da arquitetura TCP/IP foram superiores aos demais pelas seguintes razões: serem abertos, simples e tolerantes a falhas. O

fato de os protocolos TCP/IP serem abertos e simples possibilitou a sua implementação por vários fabricantes, tornando-os um padrão de *facto*. A característica de ser tolerante a falhas foi essencial para lidar com enlaces de baixa confiabilidade, garantindo uma alta disponibilidade nas redes implementadas com esses protocolos, mesmo na presença de algum tipo de falha.

Na presente geração da Internet, constata-se a grande demanda por aplicações multimídia e por aplicações de comércio eletrônico. Ambas aplicações possuem uma demanda de tráfego bem mais exigente do que uma aplicação como FTP, que provê transferência de arquivos. As aplicações multimídia, como teleconferência, devem ter o áudio e as imagens em perfeito sincronismo entre si, para uma perfeita inteligibilidade, e as aplicações de comércio eletrônico são sensíveis a longos atrasos que podem levar o consumidor a não finalizar sua compra.

O problema da atual geração da Internet deve-se ao fato de que todos os pacotes IP, independente de qual aplicativo os gerou, são processados da mesma forma. Uma arquitetura de rede, baseada nos protocolos TCP/IP, impõe sérias limitações a um tráfego com restrições de tempo real, que considera parâmetros como atraso máximo dos pacotes medidos fim-a-fim. Este problema agrava-se quando a rede está congestionada, pois protocolos de congestionamento, presentes no protocolo IP, descartam pacotes sem conhecer qual a aplicação que os gerou.

2. QUALIDADE DE SERVIÇO

Podemos observar que o aumento da velocidade dos canais de comunicação (banda de passagem) que constituem a Internet não garantirá as condições necessárias para o tráfego de voz e de aplicações multimídia. Faz-se necessário garantir que alguns parâmetros sejam limitados a valores aceitáveis que atendem ao propósito da aplicação. Consideremos os parâmetros que especificam o atraso máximo da mensagem, que é o tempo decorrido desde o início da transmissão da mensagem no dispositivo que a gerou até o seu recebimento no dispositivo de destino. O atraso máximo de uma mensagem, em uma aplicação de voz, está diretamente relacionado com a qualidade da conversação. Neste caso, os parâmetros que caracterizam o atraso máximo especificam a Qualidade de Serviço (QoS) mínima necessária que a rede deverá prover para a aplicação.

O termo Qualidade de Serviço é definido de várias maneiras na literatura [FJ99]:

Diniz [DI98] afirma que a qualidade de serviço pode ser descrita sob vários pontos de vista. Para o usuário da aplicação, ela é vista como um conjunto de características, como, por exemplo, a qualidade da imagem em termos de nitidez ou a qualidade do áudio. Para a rede, estas características são traduzidas em um conjunto de parâmetros,

como banda passante necessária, atraso máximo permitido ou taxa de erros aceitáveis.

Hafid [HA95] define qualidade de serviço como um conjunto das características qualitativas e quantitativas de um sistema multimídia distribuído, necessárias para alcançar a funcionalidade de uma aplicação.

ATM Fórum [AT96] define QoS como um conjunto de parâmetros de desempenho, tais como atraso, variação do atraso, taxa de perda de células, etc., que pode ser negociado pelo usuário, como parte de um Contrato de Tráfego, durante o estabelecimento de uma conexão.

Cada tecnologia de rede trata a QoS de uma forma específica, sendo que algumas desconsideram completamente este aspecto [LI98]. A tecnologia ATM, por exemplo, foi idealizada para atender alguns requisitos de QoS, permitindo que sejam estabelecidos circuitos virtuais fim-a-fim onde parâmetros como atraso médio da mensagem, máxima variação do atraso (*jitter*), máxima taxa de transferência, entre outros, podem ser garantidos durante o período da conexão.

3. QoS NA INTERNET

Todas as aplicações que usam a Internet devem suportar o protocolo IP, como protocolo da camada de rede. Isto implica que os vários aplicativos que hoje estão sendo amplamente usados na Internet utilizam o protocolo IP diretamente (aplicações nativas) ou indiretamente através de transladores de protocolos (denominados *gateways*)

Apesar da tecnologia ATM ser cada vez mais utilizada em *backbones*, na prática, raras aplicações utilizam ATM diretamente, trocando mensagens entre dois endereços ATM, através de um circuito virtual ATM estabelecido pela aplicação. Em outras palavras, poucas são as aplicações que usam ATM como protocolo nativo, de tal forma que acesse diretamente a rede ATM.

Por outro lado, a cada dia são geradas mais e mais aplicações que usam IP como o protocolo da camada de rede. Assim, a aplicação estabelece um fluxo de comunicação entre entidades reconhecidas univocamente, na rede, por seus endereços IP. Aplicações baseadas em IP são aplicações nativas para a Internet.

O maior desafio da Internet atualmente não é necessariamente o aumento da velocidade de comunicação, mas sim a sua utilização adequada, garantindo um

nível adequado de QoS para o correto funcionamento da aplicação mesmo em uma condição de congestionamento. Assim, pacotes IP devem se comportar de uma forma aceitável para que características funcionais da aplicação sejam preservadas, independente de congestionamentos. Como exemplo, podemos citar o parâmetro de variação do atraso de mensagens (*jitter*) em uma aplicação de voz sobre IP. Uma variação do atraso grande poderá deixar uma aplicação de voz ininteligível, apesar de, até mesmo, existir uma taxa nominal de transmissão adequada para a aplicação.

Neste sentido, a *Internet Engineering Task Force* IETF [IETF] possui dois grupos de trabalhos que objetivam estender a arquitetura atual da Internet para prover diferentes níveis de qualidade de serviço para usuários, que são os grupos de Serviço Integrado (IntServ) e Serviço Diferenciado (DiffServ), descritos a seguir.

3.1 Serviço Integrado - IntServ

A proposta do grupo de Serviço Integrado (IntServ) é estender a atual arquitetura da Internet partindo do princípio que os recursos, para garantir uma determinada qualidade de serviço, deverão ser alocados (reservados) para cada fluxo de dados. Desta forma, todos os pacotes pertencentes àquele determinado fluxo usarão os recursos previamente alocados.

Os recursos são reservados em todos os roteadores intermediários e no nó de destino usando protocolos de sinalização (controle). A alocação de recursos implica, por exemplo, em reserva de tempo de CPU, memória e banda na rede, para prover o serviço solicitado. Recursos que foram reservados devem ser atribuídos aos pacotes do seu respectivo fluxo. Finalmente, as características de tráfego negociadas no momento em que os recursos foram alocados devem ser monitoradas, para que seja verificado que as premissas assumidas quanto às características do tráfego foram respeitadas durante o período de conexão.

O protocolo RSVP (*Resource Reservation Setup Protocol*) [WH97] é uma implementação da arquitetura IntServ e foi desenvolvido como um protocolo de sinalização para reserva de banda. O RSVP estende o protocolo IP de tal forma que os dados são transmitidos sem nenhuma modificação. O protocolo troca sinais de controle (sinalização) descrevendo a qualidade de serviço dado através de um fluxo de dados. O protocolo é orientado ao nó de destino do fluxo, pois é sua função gerar as mensagens contendo os parâmetros de qualidade de serviços a serem alocados.

Propostas de implementação de arquiteturas IntServ são criticadas basicamente por sua limitada aplicação em uma rede IP com vários nós, pois os recursos são alocados para cada fluxo de dados. A implementação de IntServ em uma rede grande, como a Internet, com milhões de usuários e com vários fluxos por usuário, implica em complexos roteadores com considerável poder de processamento

e memória. Cada nó possui um estado para cada fluxo, aumentando a complexidade do sistema. O poder de processamento se fará necessário pois a tarefa de escalonamento dos pacotes para cada fluxo é complexa e fundamental para a garantia da qualidade de serviço acordada. A outra desvantagem é a perda de padrões para bilhetagem e tarifação, fazendo RSVP e IntServ recomendados apenas para redes pequenas e confinadas.

3.2 Serviço Diferenciado - DiffServ

A proposta de implementar QoS em redes IP usando serviços diferenciados tenta evitar os problemas encontrados na proposta de protocolos baseados em IntServ. A idéia do DiffServ baseia-se em reservar recursos para um conjunto de fluxos e não para um fluxo único.

Os pacotes IP são marcados pelo usuário ou pelo provedor de acesso como sendo de uma determinada classe de QoS. Os roteadores reservam recursos para lidar com pacotes de cada classe. Este conceito oferece a possibilidade de o provedor de serviço oferecer classes de QoS com custos diferenciados para seus usuários.

O *DiffServ* trata todos os pacotes dos usuários de uma mesma classe usando os recursos alocados para aquela classe de QoS. Assim, os recursos que possibilitam o envio de pacotes com uma determinada QoS são compartilhados por todos os pacotes que estão marcados como sendo da mesma classe de QoS. Este conceito não garante qualquer qualidade de serviço “absoluta” como os protocolos de reserva fim-a-fim, propostos na arquitetura IntServ, porém permite uma razoável implementação de reserva de recursos em grandes redes.

Para implementar o *DiffServ*, cada pacote é diferenciado alterando-se o estado dos bits do campo denominado DS byte do cabeçalho de IP. O DS byte nada mais é do que a redefinição do octeto TOS (*Type Of Service*) na versão 4 de IP e do octeto *Traffic Class* da versão 6.

Dos oito bits, seis são denominados PHB (*Per Hop Behavior*) e usados para definir o tratamento que será dado em cada pacote por cada roteador. Os dois bits restantes atualmente não estão usados e são denominados CU (*Currently Unused*).

3.3 Mapeando QoS em ATM

As propostas das arquiteturas para estender IP para prover QoS não descartam ATM. Tais propostas mapeam a QoS desejada - definida no nível IP - para a QoS provida pelos circuitos virtuais ou reais que estabelecem o canal de comunicação entre a origem e o destino da mensagem. No caso específico de ATM, o mapeamento de uma QoS especificada em IP implica na abertura de um canal virtual ATM com a QoS que satisfaça a especificação. A impossibilidade de atender a solicitação de

QoS implicará na rejeição da solicitação no nível ATM que irá ser transferido (mapeado) ao nível IP.

4. PROPOSTAS DE SERVIÇOS DIFERENCIADOS

Existem diferentes propostas para a implementação de DiffServ, cujas idéias principais serão discutidas abaixo. Todas elas lidam com o tratamento diferenciado dos pacotes através da informação contida no estado dos bits do DS byte.

4.1 *Premium Service*

Este tipo de serviço é similar a linhas dedicadas comercializadas por empresas de telecomunicações para conectar duas redes distintas. O usuário negocia com seu provedor a máxima banda de passagem, que não pode ser excedida. Porém, se o usuário não usar a capacidade contratada, irá pagar o mesmo valor. Neste serviço, a ausência ou não de outros usuários usando a mesma linha não influenciará na qualidade do serviço disponibilizada pelo provedor.

O roteador que está diretamente conectado ao equipamento da ponta (como a estação de trabalho do usuário final), denominado roteador de borda, possui a função de classificar os pacotes recebidos. Se estes devem ser tratados em um *Premium Service*, eles serão marcados pelo roteador de borda respeitando a máxima banda de passagem especificada para o serviço. Duas filas são contidas nos equipamentos que compõem esta rede: uma para pacotes *Premium* e a outra para os demais. Pacotes *Premium* possuem prioridade sobre os demais.

4.2 *Assured Service*

Encontramos vários casos onde o tráfego observado é de rajadas, onde na maioria do tempo o sistema está ocioso, mas quando um dado é enviado espera-se um tratamento prioritário para ele. O caso típico é o tráfego gerado por transações bancárias. Neste caso, o *Premium Service* não atenderia adequadamente pois a banda reservada seria subutilizada na maioria do tempo e, quando necessária, poderia ser insuficiente para transferir os dados com a qualidade desejada.

O *Assured Service* tenta oferecer um serviço que não pode garantir a banda de passagem mas provê uma alta probabilidade de que os pacotes marcados como de alta prioridade sejam transferidos de uma forma confiável.

A probabilidade de que pacotes sejam transportados confiavelmente depende da capacidade da rede. Um provedor, por exemplo, pode assumir que a soma de todas as bandas de passagem comercializadas como *Assured Service* deve permanecer abaixo da conexão mais lenta. A vantagem do *Assured Service* é que o usuário não necessita estabelecer a reserva por um grande tempo.

Com *Assured Service*, o usuário negocia o serviço desejado com seu provedor de serviço com a velocidade ou quantidade de serviço de alta prioridade (*Assured Service*) desejada. Para evitar modificações no sistema final, o primeiro roteador pode analisar os pacotes no que diz respeito à porta UDP/TCP usada e o endereço e atribuir a correta prioridade ao pacote (isto é, se o pacote será enviado usando um *Assured Service*).

Rajadas são suportadas através de armazenagem em *buffers*, que são enviadas de tal forma que a rede possua um comportamento estatístico.

4.3 User-Share Differentiation (USD)

Neste serviço, uma banda de passagem mínima é garantida e assim não há congestionamento nestes enlaces. O resto da banda de passagem será distribuída em uma proporção definida para cada usuário (por exemplo, 3:2:1). A proporção é atribuída pelo provedor através de configuração. Usuários podem ser distribuídos em várias classes, por exemplo, em três classes: alta, média e baixa, correspondentes à proporção 3, 2 e 1 respectivamente.

Cada roteador da rede que implementa esse serviço deve conhecer a proporção de cada classe de usuários que é distribuída pela rede. Em caso de congestionamento, o serviço do usuário deverá ser degradado até a banda mínima garantida. A implementação do serviço só é feita nos roteadores cujos enlaces experimentam congestionamento e que são considerados o gargalo do sistema

Aplicações que podem trocar sua taxa de transmissão e necessitam de uma taxa de transmissão mínima para transmitir dados, como os sistemas de tempo real e aplicações como FTP, são exemplos típicos desse tipo de serviço.

4.4 Scaleable Reservation Protocol (SRP)

O SRP representa outra proposta para uma possível implementação de DiffServ na Internet. Como sugerido pelo seu nome, o objetivo deste protocolo é fazê-lo escalável até mesmo com um grande número de fluxos de pacotes diferentes.

Cada roteador agrega todos os fluxos de dados e monitora suas características para estimar os recursos necessários (atuais e futuros) de cada nó. O processo de estimativa dos recursos necessários é de grande importância para o funcionamento do protocolo e é empregado pela origem, pelo destino e pelos roteadores entre eles. A origem possui uma avaliação otimista dos recursos requeridos para suportar o fluxo de dados. O destino calcula uma estimativa conservadora dos recursos reservados e periodicamente manda esta informação para o remetente. Sem requerer sinalização explícita de parâmetros de reserva de banda, o serviço implementado sobre SRP é feito através dos protocolos de reserva, que fazem a reserva sobre um caminho e dos protocolos de realimentação, que mandam a confirmação para a origem.

Este protocolo pode ser utilizado para uso em telefonia sobre IP. O fluxo de dados reservado pelo protocolo é usado para telefonia e a banda passante não reservada pode ser usada para as demais aplicações.

4.5 Olympic Service

Este serviço é similar ao USD, onde os usuários são classificados como de classe Ouro, Prata ou Bronze, com a proporção de 60%, 30% e 10%, respectivamente. Ao contrário do serviço USD, os pacotes são classificados já nos roteadores de borda de acordo com a sua classe. É possível que na ausência de pacotes da classe ouro e prata, os pacotes bronze ocupem toda a banda do canal. Isto significa que o canal é sempre usado na sua capacidade total e as restrições impostas por cada classe só fazem efeito no momento de congestionamento da rede. Como consequência, nenhuma garantia de largura mínima da banda de passagem é oferecida.

Provedores que desejam oferecer diferentes tipos de serviços para seus usuários são os principais interessados neste tipo de serviço.

5. CONCLUSÃO

A implantação de QoS em uma rede IP irá trazer profundas mudanças no cenário da Internet. Certamente, uma delas é a possibilidade de provedores de serviço oferecerem telefonia de boa qualidade e alta confiabilidade na Internet. Podemos inferir que o sucesso dessas propostas mudará o *backbone* das empresas de telecomunicações, pois todas as propostas de QoS implicam no uso mais racional da banda de passagem disponível e no suporte ao tráfego de tempo real.

O sucesso dessas propostas está relacionado principalmente à solução do problema da escalabilidade, isto é, se a implementação de QoS em IP poderá ser feita com baixo custo de troca de mensagens e processamento de pacotes pelos roteadores, mesmo em grandes redes.

KEYWORDS

Internet – Quality of service – Integrated service – IntServ – Differentiated service – DiffServ – IP Protocol

ABSTRACT

Current networks require many more functions than have been previously offered. It is essential for the network to be able to provide QoS (Quality of Service) that can be customized to different end users. In response to the growing demand for

QoS in the Internet, the Internet Engineering Task Force set up two Working Groups named DiffServ and IntServ which have already defined several proposals for extending the TCP/IP architecture. This paper introduces the main concepts and services proposed to provide QoS in the Internet.

REFERÊNCIAS BIBLIOGRÁFICAS

- [AR97] ARNETT, M. Inside TCP/IP; *New Riders Publishing*, 1997.
- [AT96] ATM User-Network Interface(UNI; Signaling Specification - version 4.0; *Publicação do ATM Fórum*, 1996.
- [DI98] DINIZ, A. Um Serviço de Alocação Dinâmica de Banda Passante em Redes ATM para Suporte a Aplicações Multimídia; *Dissertação (Mestrado)*, março de 1998, *Departamento de Ciência da Computação, Universidade Federal de Minas Gerais*.
- [FJ99] FIGUEIREDO JR, Themistocles. Um Serviço de Garantia Estatística do Atraso em Redes ATM para Aplicações Multimídia. *Dissertação (Mestrado)*, outubro de 1999, *Departamento de Ciência da Computação, Universidade Federal de Minas Gerais*.
- [HA95] HAFID, A. *et al.* On Quality of Service Negotiation for Distributed Multimedia Applications; *Publication Départementale 977, Université de Montreal*, 1995.
- [IETF] IETF home page, <http://www.ietf.org>
- [LI98] LIMA, F. Um Modelo de Qualidade de Serviço baseado em ATM para a Plataforma Multiware; *Dissertação (Mestrado)*, 1998, *Universidade Estadual de Campinas*.
- [MC94] MCDYSAN and SPOHN, D. ATM - Theory and Application, *McGraw-Hill on Computer Communications*, 1994.
- [WH97] WHITE, Paul. RSVP and Integrated Services in the Internet: A Tutorial, *IEEE Communications Magazine*, maio 1997, p. 100-106.