

UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO

Edemilson Bay

Uma Metodologia para a Identificação e a
Definição do Perfil de Usuário para QoS com
Serviços Diferenciados

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação.

Prof. Roberto Willrich, Dr.
Orientador

Florianópolis, Março de 2003.

Uma Metodologia para a Identificação e a Definição do Perfil de Usuário para QoS com Serviços Diferenciados

Edemilson Bay

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, Área de Concentração Sistemas de Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Prof. Fernando Ostuni Gauthier, Dr
Coordenador do Curso

Banca Examinadora

Prof. Roberto Willrich, Dr.
Orientador

Prof. Vitório Bruno Mazzola, Dr.
UFSC

Prof. Mário Antônio Ribeiro Dantas, Dr.
UFSC

*“Para realizar grandes conquistas
devemos não apenas agir mas também sonhar;
não apenas planejar mas também acreditar”*

Anatole France

Ofereço este trabalho aos meus familiares, principalmente
à minha mãe e minha avó, como demonstração
de que não foram em vão todos os seus esforços, desde o início,
para me proporcionar este momento.

Agradecimentos

Primeiramente ao meu orientador, professor Roberto Willrich, pelo seu empenho a fim de enriquecer este trabalho e sempre pronta disposição para atendimento quando precisei.

À empresa Brasil Telecom, especialmente ao Luiz Antônio da Costa Silva, por proporcionar a liberação em vários momentos visando a realização do Mestrado.

Aos colegas de trabalho e colegas do curso pelo incentivo que me deram com vistas ao atingimento deste objetivo final.

Aos mestres das disciplinas cursadas, que com seus conhecimentos, experiências e amizade tornaram o aprendizado melhor e mais fácil.

A todos os demais, não nomeados aqui, mas que de forma direta ou indireta contribuíram para a realização deste trabalho.

E sobretudo à DEUS, por me permitir esta oportunidade maravilhosa, além de me dar força e ânimo nos momentos de fraqueza.

SUMÁRIO

LISTA DE FIGURAS.....	VIII
LISTA DE TABELAS.....	IX
LISTA DE ACRÔNIMOS	X
RESUMO.....	XIII
ABSTRACT	XIV
1. INTRODUÇÃO	15
1.1 OBJETIVOS DA DISSERTAÇÃO	18
1.2 ESTRUTURA DA DISSERTAÇÃO	18
2. A INTERNET	20
2.1 EVOLUÇÃO E ADMINISTRAÇÃO DA INTERNET.....	21
2.2 A ARQUITETURA DA INTERNET	24
2.2.1 <i>Protocolo TCP</i>	27
2.2.2 <i>Protocolo UDP</i>	29
2.2.3 <i>Protocolo IP</i>	31
2.3 APLICAÇÕES E NEGÓCIOS NA INTERNET	35
2.4 CONCLUSÃO	40
3 QUALIDADE DE SERVIÇO NA INTERNET	41
3.1 PARÂMETROS DE QUALIDADE DE SERVIÇO.....	42
3.2 REQUISITOS DE QOS EM APLICAÇÕES IP	44
3.2.1 <i>Requisitos para Aplicações de Voz</i>	46
3.2.2 <i>Requisitos para Aplicações de Vídeo</i>	49
3.2.3 <i>Requisitos para Aplicações Web</i>	51
3.3 SERVIÇOS DIFERENCIADOS.....	53
3.3.1 <i>Arquitetura DiffServ</i>	55
3.3.2 <i>Mecanismos para Qualidade de Serviços</i>	57
3.3.3 <i>Campo DS</i>	63
3.3.4 <i>PHB</i>	64
3.3.5 <i>Classes de Serviço</i>	67
3.3.6 <i>Acordo de Nível de Serviço (SLA)</i>	70
3.3.7 <i>Visão Geral do Funcionamento</i>	70
3.4 CONCLUSÃO	72
4. ACORDO DE NÍVEL DE SERVIÇO	73
4.1 CONHECENDO O SLA	74
4.2 ESTRUTURA DO SLA/SLS	77
4.2.1 <i>Parte A: Parâmetros Administrativos, Legais e Econômicos</i>	77
4.2.2 <i>Parte B: Parâmetros Técnicos</i>	79
4.3 CONCLUSÃO	83

5. PROPOSTA DE IDENTIFICAÇÃO E DEFINIÇÃO DO PERFIL DO USUÁRIO.....	85
5.1 VISÃO GERAL DA METODOLOGIA.....	86
5.2 FASE I – IDENTIFICANDO O PERFIL	87
5.2.1 <i>Etapa 1 – Identificar aplicações, características e prioridades</i>	88
5.2.2 <i>Etapa 2 – Identificar a abrangência e limitações</i>	91
5.2.3 <i>Etapa 3 – Identificar o perfil de tráfego</i>	93
5.2.4 <i>Etapa 4 – Identificar a escala de atividade e confiabilidade</i>	96
5.3 FASE II – COMPONDO O SLS	97
5.4 CONCLUSÃO	101
6. ESTUDO DE CASO	103
6.1 DEFINIÇÃO DO CONTEXTO	104
6.2 APLICAÇÃO DA METODOLOGIA PROPOSTA.....	106
6.3 PROGRAMAÇÃO DOS SLS’S NA REDE DIFFSERV	118
6.4 CONCLUSÃO	119
7. CONCLUSÕES.....	120
7.1 TRABALHOS FUTUROS	122
REFERÊNCIAS	123
ANEXO A - PROGRAMAÇÃO DO ROTEADOR DE BORDA	127

Lista de Figuras

FIGURA 2.1 – ARQUITETURA DO TCP/IP	26
FIGURA 2.2 – SEGMENTO TCP	29
FIGURA 2.3 – DATAGRAMA UDP	30
FIGURA 2.4 – ENDEREÇOS IP NAS REDES CLASSES A, B E C	32
FIGURA 2.5 – DATAGRAMA IP	32
FIGURA 2.6 – RELAÇÕES DE <i>E-COMMERCE</i>	38
FIGURA 2.7 – VENDAS VIA <i>E-COMMERCE</i> NO MUNDO	39
FIGURA 3.1 – DOMÍNIO DS	56
FIGURA 3.2 – ARQUITETURA GERAL DE DOMÍNIOS DS	57
FIGURA 3.3 – MECANISMOS BÁSICOS PARA IMPLEMENTAR <i>DIFFSERV</i>	58
FIGURA 3.4 – <i>LEAKY BUCKET</i> OU LIMITADOR DE TRÁFEGO	60
FIGURA 3.5 – <i>TOKEN BUCKET</i> OU CONFORMADOR DE RAJADAS	60
FIGURA 3.1 – ESTRUTURA DO CAMPO DS	63
FIGURA 6.1 – TOPOLOGIA DA REDE: OPERADORA E CLIENTE	104

Lista de Tabelas

TABELA 3.1 – REQUISITOS POR TIPO DE APLICAÇÃO	45
TABELA 3.2 – MOS	46
TABELA 3.3 – VALORES DE ATRASO PARA VOIP	48
TABELA 3.4 – VALORES DE VARIAÇÃO DO ATRASO PARA VOIP	48
TABELA 3.5 – VALORES DO DSCP PARA CADA PHB AF	66
TABELA 5.1 - IDENTIFICAÇÃO DAS APLICAÇÕES, CARACTERÍSTICAS E PRIORIDADES.	90
TABELA 5.2 – IDENTIFICAÇÃO DA ABRANGÊNCIA DAS APLICAÇÕES	92
TABELA 5.3 – IDENTIFICAÇÃO DE LIMITAÇÕES LÓGICAS.....	92
TABELA 5.4 – IDENTIFICAÇÃO DO PERFIL DE TRÁFEGO DESEJADO PELO CLIENTE	95
TABELA 5.5 – IDENTIFICAÇÃO DA ESCALA DE ATIVIDADE E CONFIABILIDADE	96
TABELA 5.6 – FORMA DE IDENTIFICAÇÃO DE PACOTES PARA ALGUMAS APLICAÇÕES	98
TABELA 5.7 – TAXA DE TRANSMISSÃO TOTAL GERADA POR APLICAÇÃO	99
TABELA 6.1 – CLASSES DE SERVIÇO E APLICAÇÕES	105
TABELA 6.2 – PARÂMETROS DE QoS PARA AS CLASSES DE SERVIÇO	105
TABELA I - IDENTIFICAÇÃO DAS APLICAÇÕES, CARACTERÍSTICAS E PRIORIDADES.	107
TABELA II – IDENTIFICAÇÃO DA ABRANGÊNCIA DAS APLICAÇÕES.....	108
TABELA III – IDENTIFICAÇÃO DE LIMITAÇÕES LÓGICAS.....	109
TABELA IV – IDENTIFICAÇÃO DO PERFIL DE TRÁFEGO DESEJADO PELO CLIENTE	110
TABELA V – IDENTIFICAÇÃO DA ESCALA DE ATIVIDADE E CONFIABILIDADE	111

Lista de Acrônimos

TCP	Transmission Control Protocol
ADSL	Assimetric Digital Subscriber Line
AF	Assured Forwarding
ANS	Advanced Network and Services
AQUILA	Adaptive Resource Control for QoS Using na IP based Layered Architecture
ARIN	American Registry of Internet Numbers
ARPA	Advanced Research Projects Agency
ATM	Assynchronous Transfer Protocol
B2B	Business-to-Business
B2C	Business-to-Consumer
B2G	Business-to-Government
BA	Behavior Aggregate
BE	Best Effort
C2B	Consumer-to-Business
C2C	Consumer-to-Consumer
C2G	Consumer-to-Government
CADENUS	Creation and Deployment of End-User Services in Premium IP Networks
CERT/CC	Computer Emergency Response Team - Coordinate Center)
CRM	Customer Relationship Management
CRTP	Compressed Real Time Protocol
DiffServ	Differenciaded Services
DRI	Defense Research Internet
DSCP	Differentiated Service CodePoint
EBONE	European Backbone
EDI	Electronic Data Interchange
EF	Expedited Forwarding
ERP	Enterprise Resource Planning

FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FTP	File Transfer Protocol
G2B	Governement-toBusiness
G2C	Government-to-Consumer
G2G	Government-to-Government
HTTP	Hiper Text Transfer Protocol
HDTV	High Definition TV
IAB	The Internet Architecture Board
IANA	Internet Assigned Numbers Authority
IETF	The Internet Engineering Task Force
InterNIC	The Internet Network Information Center
IntServ	Integrated Services
IP	Internet Protocol
IRTF	The Internet Research Task Force
ISOC	The Internet Society
ITU	International Telecommunications Union
LAN	Local Area Network
LDP	Label Distribution Protocol
LNCC	Laboratório Nacional de Computação Científica
MF	Multi Field
MOS	Mean Option Score
MPEG	Moving Picture Experts Group
MPLS	MultiProtocol Label Switching
NSF	National Science Foundation
NBAR	Network Based Application Recognition
OSI/ISO	Open System Interconnection/International Standard Organization
PDU	Protocol Data Unit
PHB	Per-Hop Behavior
PoP	Point of Presence
PPP	Point-to-Point Protocol
QoS	Quality of Service
RFCs	Request For Comments

RNP	Red Nacional de Pesquisas
RSVP	ReSerVation Protocol
RTP	Real Time Protocol
SDTV	Standard Definition TV
SEQUIN	<i>S</i> ervice <i>Q</i> uality across <i>I</i> ndependently managed <i>N</i> etworks
SLA	Service Level Agreement
SLS	Service Level Specification
SMTP	Simple Mail Transfer Protocol
TEQUILA	Traffic Engineering for Quality of Service in the Internet at Large Scale
UDP	User Datagram Protocol
UIT	União Internacional de Telecomunicações
VAD	Voice Activity Detection
VCR	Vídeo Cassete Recorder
VoIP	Voz sobre IP
WAN	Wide Area Network
WWW	World Wide Web

Resumo

Com o grande crescimento do número de aplicações e usuários na Internet, o tráfego na rede tem aumentado rapidamente. Mesmo com o aumento da capacidade de transmissão nas redes dos provedores de serviços, o alto tráfego pode causar congestionamentos que impeçam o bom funcionamento de várias aplicações. Assim, os provedores começam a oferecer Qualidade de Serviço (QoS - *Quality of Service*) em suas redes, visando proporcionar serviços com garantias de desempenho a seus usuários. Uma das soluções mais aceitas para QoS atualmente é o modelo Serviços Diferenciados (DiffServ) proposto pela IETF.

Neste contexto, a qualidade de serviço oferecida por um provedor a um cliente deve estar acordada na forma de um contrato chamado Acordo de Nível de Serviço (SLA - *Service Level Agreement*). Várias informações estão contidas no SLA, entre elas a Especificação do Nível de Serviço (SLS - *Service Level Specification*), que define os detalhes técnicos dos serviços. A especificação do SLS é uma tarefa difícil e requer um conhecimento das necessidades do cliente e das capacidades do provedor.

Esta dissertação propõe uma metodologia de identificação das necessidades do usuário e a definição do seu perfil, para utilização de serviços com QoS, baseados em DiffServ. O objetivo é auxiliar na elaboração do SLS adequado aos serviços desejados pelo usuário, aqui denominado cliente.

Abstract

With the large expansion of users and applications for the Internet, the network traffic has been growing fast. Even with the transmission capacity growing at the services provider networks, the high traffic can degrade the quality for some applications due to network traffic congestion. Thus, the providers start offering QoS (Quality of Service) in their networks, aiming at delivering services with performance guarantees to their users. One of the most accepted solutions for QoS nowadays is the DiffServ (Differentiated Services Model) proposed by the IETF.

In this context, the quality of service offered by a provider to a client must be assigned according to a contract called SLA (Service level Agreement). Much information is described in the SLA, such as the SLS (Service level Specification), that defines the technical details for the services. The SLS specification is an hard task and requires knowledge about the customer needs and the provider capabilities.

This dissertation proposes a methodology to identify the user needs and the definition of his profile for use of services with QoS based on Diffserv. The objective is assist to elaborate the right SLS for the services expected by the user, here named client.

1. INTRODUÇÃO

A Internet representa um dos mais bem sucedidos exemplos de benefícios que podem ser obtidos com a manutenção do investimento em pesquisas e compromisso com o desenvolvimento tecnológico. Neste caso, o resultado está sendo a implantação de uma infra-estrutura adequada para o tráfego da informação, ou seja, uma rede de comunicação de dados com abrangência globalizada e relação custo-benefício muito atrativa. Depois das primeiras pesquisas sobre transmissão de dados baseada em pacotes, há várias décadas, hoje o governo, a indústria e o meio acadêmico continuam sendo parceiros no desenvolvimento dessa tecnologia. Termos como nome@nomedeempresa.com (ou nome@nomedeempresa.com.br, no caso do Brasil) e <http://www.nomedeempresa.com> (ou <http://www.nomedeempresa.com.br>, no caso do Brasil) são usados diariamente por milhões de pessoas ao redor do mundo.

Novas aplicações e serviços variados têm sido desenvolvidos nos últimos anos, utilizando a Internet como meio de comunicação e facilitando a realização de tarefas como comunicação (interativa ou não), negócios, disseminação de informação, serviços de utilidade pública, entre outras. Tudo isso, independente da localização geográfica dos usuários e computadores. Com isso, o número de conexões à Internet tem crescido muito rapidamente.

O número de usuários da Internet no Brasil dobra a cada ano, e um fato importante é que os usuários estão se tornando cada vez mais exigentes com relação ao desempenho dos serviços, quer estejam usando aplicações de texto, áudio, animações ou vídeo. Como exemplo de serviços que podem usar a Internet temos a vídeo-conferência, a telefonia, entretenimento e educação à distância, a tele-medicina, o tele-notícias, os

serviços bancários, entre outros. Na área da saúde, por exemplo, hospitais de referência podem estar interligados a outros hospitais em qualquer lugar do mundo, tornando possível a troca de informações em tempo real através da Internet.

No entanto, a Internet é baseada na arquitetura TCP/IP (*Transmission Control Protocol/Internet Protocol*), também usando o protocolo UDP (*User datagram Protocol*) no lugar do TCP para algumas aplicações. Os equipamentos da rede estão configurados para utilizar o protocolo com a política do melhor esforço de entrega (*Best Effort*). Isto significa que um usuário da rede Internet, enviando seus dados (pacotes), compartilha igualmente a taxa de transmissão disponível na rede com os dados dos outros usuários, que estejam cursando pela rede no mesmo instante. Com a política do melhor esforço, a rede tenta entregar os dados ao destino da melhor maneira possível, baseada nas rotas melhores disponíveis no momento. Não há garantia de que os pacotes sejam entregues e nem da taxa de transmissão obtida. Com o aumento do uso da Internet, existe um aumento na probabilidade de ocorrer congestionamentos, mesmo aumentando a banda disponível. Em casos de congestionamento na rede, alguns pacotes podem ser descartados sem qualquer distinção, fazendo com que a aplicação deva retransmiti-los para que cheguem ao destino, e isso pode ser indesejável para algumas aplicações. Dessa forma, aplicações como voz e vídeo ficam praticamente inviabilizadas na Internet, já que precisam de garantia de entrega e desempenho.

A Internet é cada vez mais usada para alavancar negócios e otimizar processos, seja em empresas privadas (grandes corporações, médias e pequenas) ou instituições governamentais. Estes usuários (Clientes) contratam serviços de Internet através de prestadores de serviço (Operadoras de Telecomunicações, Provedores de Internet, etc.). Neste trabalho, será definido que o prestador de serviços de Internet será uma Operadora de Telecomunicações. Quando o serviço a ser prestado deve ter garantia de qualidade, a Operadora e o Cliente firmam um contrato, chamado Acordo de Nível de Serviço (SLA - *Service Level Agreement*). No SLA existem várias características e parâmetros de qualidade do serviço que devem ser cumpridas pelas partes, para a correta prestação do serviço. Dependendo do SLA contratado, pode existir a necessidade da Operadora trabalhar com sistemas tecnológicos que forneçam garantia de qualidade na rede.

Tecnologias que garantem a qualidade do tráfego de dados na rede Internet já estão em desenvolvimento, e o IETF (*Internet Engineering Task Force*) é uma entidade que coordena este trabalho, além de outros na área da Internet. Ao conjunto de soluções para garantia de qualidade na Internet, convencionou-se chamar de Qualidade de Serviço ou QoS (*Quality of Service*). Com QoS é possível oferecer maior confiabilidade de entrega e garantia de desempenho, através de esquemas como priorização ou alocação de banda para determinados pacotes selecionados. Assim, aplicações avançadas como voz sobre IP (VoIP) e vídeo-conferência, por exemplo, passam a ter maior prioridade na rede, enquanto usuários de aplicações tradicionais continuam utilizando apenas o melhor esforço.

Atualmente, existem dois modelos para implementar QoS na Internet: Serviços Integrados (*IntServ*) [BRA94] e Serviços Diferenciados (*DiffServ*) [BLA98]. *IntServ* é um modelo baseado em reserva de recursos na rede, enquanto que *DiffServ* é uma proposta onde os pacotes são marcados de acordo com classes de serviços pré-determinadas e que serão tratadas na rede de forma diferenciada. A Operadora deve trabalhar com o modelo que atenda as necessidades de seus clientes, sem contudo inviabilizar a rentabilidade do negócio da Operadora.

Nesta dissertação será focado QoS baseado no modelo *DiffServ*, sendo esta a proposta de QoS mais aceita para a Internet. No modelo *DiffServ*, QoS é obtido através da marcação dos pacotes de dados do Cliente objetivando o tratamento diferenciado destes dentro da rede. A rede tratará os pacotes de acordo com classes de serviços previamente definidas pela Operadora, e ajustadas de acordo com as necessidades da aplicação do Cliente.

Visando definir o tipo de serviço a ser prestado ao Cliente e sua correta configuração (marcação de pacotes e tratamento na rede), a Operadora deverá obter várias informações, como necessidades do Cliente e características das suas aplicações. Com isto é possível definir o perfil do usuário, e a partir deste, configurar as especificações do nível de serviço, ou SLS (*Service Level Specifications*). O SLS é moldado conforme o perfil do usuário, mas dentro dos limites das classes de serviço oferecidas pela Operadora.

Dessa forma, o SLS também faz parte do SLA, e define os detalhes técnicos necessários para a implantação do SLA acordado. Trata-se, portanto, de um assunto com importância crescente, principalmente no cenário da concorrência em prestação de serviços por Operadoras de Telecomunicações.

1.1 Objetivos da Dissertação

Esta dissertação propõe uma metodologia de identificação e definição do perfil do usuário de serviços com QoS, baseados em DiffServ, visando auxiliar na elaboração do SLS adequado aos serviços desejados pelo cliente do provedor. O SLS é composto por vários parâmetros técnicos que vão definir o QoS do serviço prestado, dependendo de cada aplicação do cliente. Assim, é interessante que exista um procedimento para obtenção dos dados que vão compor esse SLS

A importância desta proposta advém da dificuldade em obter todas as informações necessárias do Cliente, com relação aos requisitos técnicos e parâmetros de suas aplicações mais avançadas. Dessa forma, pretende-se sugerir uma metodologia que facilite a obtenção de informações e necessidades do Cliente, visando a elaboração do SLS de cada aplicação, baseado na definição das classes de serviço da Operadora.

1.2 Estrutura da Dissertação

Esta dissertação, dividida em capítulos, está organizada na forma que segue.

No capítulo 2 são apresentados conceitos relacionados à Internet, contendo um histórico da evolução e crescimento, além da arquitetura de rede utilizada. Também é comentado sobre a utilização atual da Internet.

O capítulo 3 é feito um estudo sobre algumas aplicações que podem usar a internet como meio de transporte e seus requisitos para bom funcionamento. Este capítulo também aborda a necessidade da migração para uma rede que forneça Qualidade de Serviços (QoS) e apresenta um estudo sobre os modelos de QoS, em especial o modelo *DiffServ*.

Um estudo sobre acordo de nível de serviço (SLA) e especificações de nível de serviço (SLS) é feito no capítulo 4, apresentando uma possível estrutura para estes documentos.

O capítulo 5 apresenta a proposta de uma metodologia de identificação e definição do perfil do usuário, mostrando como podem ser obtidas e tratadas todas as informações sobre aplicações e necessidades do Cliente, com o objetivo de compor o SLS das aplicações, mapeando a solução nas classes de serviço fornecidas pelo provedor.

No capítulo 6 é feito um estudo de caso, tomando por base uma empresa fictícia que tem necessidades de comunicação através da Internet, e que deseja certa garantia na prestação desses serviços pelo provedor.

O capítulo 7 apresenta as conclusões e sugestões para a continuidade deste trabalho.

2. A INTERNET

A Internet tem revolucionado o mundo dos computadores, das comunicações e, indiretamente, o rumo da humanidade como um todo. Nenhuma invenção anterior foi capaz de mudar comportamentos de forma tão rápida. O telégrafo, o telefone, o rádio e o computador prepararam o terreno para essa integração de funcionalidades que a Internet está viabilizando. Ela tornou-se um mecanismo de disseminação e divulgação da informação a nível mundial, além de um meio para colaboração e interação tanto de pessoas físicas quanto jurídicas, independente de suas localizações geográficas.

A Internet vem a ser um conjunto de redes de computadores interligadas pelo mundo inteiro, que têm em comum um conjunto de protocolos e serviços de forma que os usuários a ela conectados podem usufruir de serviços de informação e comunicação de alcance mundial. Esse conjunto de redes tem em comum o protocolo TCP/IP, que permite a comunicação entre elas. Dessa forma, o TCP/IP é a linguagem comum dos computadores que integram a Internet, desde os grandes computadores até os micro-computadores pessoais.

Uma analogia para entender a Internet seria defini-la como sendo similar a uma malha de rodovias (federais, estaduais e vicinais) por onde trafegam bytes sob a forma de pacotes TCP/IP. Nestes pacotes está contida a informação a ser trocada entre os computadores. A informação, quer seja texto, som ou imagem, trafega em alta velocidade entre qualquer computador conectado nessa rede. Por esta razão que a Internet é algumas vezes chamada da "Super Rodovia da Informação".

Este capítulo tem o objetivo de estudar a Internet de uma forma abrangente, porém simplificada. Na seção 2.1 é feito um relato sobre a evolução da Internet e os

órgãos que operam e administram sua utilização e desenvolvimento. A arquitetura da rede e seus protocolos são estudados na seção 2.2. Na seção 2.3 são apresentadas algumas formas de utilização da Internet, que simplificam e melhoram a maneira de viver e ajudam no crescimento econômico do mercado.

2.1 Evolução e Administração da Internet

A Internet, em seu início, era bem diferente de como é conhecida nos dias de hoje. Surgiu a partir de um projeto da agência norte-americana ARPA (*Advanced Research and Projects Agency*) com o objetivo de conectar os computadores dos seus departamentos de pesquisa [CYC99]. Essa conexão iniciou-se em 1969, entre 4 localidades (Universidades de Los Angeles e Santa Bárbara, na Califórnia, Universidade de Utah e o Instituto de Pesquisa de Stanford), e passou a ser conhecida como ARPANET.

Esse projeto inicial foi colocado à disposição de pesquisadores, o que resultou em uma intensa atividade de pesquisa durante a década de 70, cujo principal resultado foi a concepção do conjunto de protocolos conhecido como arquitetura TCP/IP, e que até hoje é a base da Internet,.

No início da década de 80 a ARPA iniciou a integração das redes de computadores dos outros centros de pesquisas à ARPANET. Em 1985, a entidade americana NSF (*National Science Foundation*) interligou os supercomputadores de seus centros de pesquisa, o que resultou na rede conhecida como NSFNET, que em 1986 foi conectada à ARPANET. O conjunto de todos os computadores e redes ligados a esses dois backbones (espinhas dorsais de uma rede) passou a ser conhecido oficialmente como INTERNET. Em 1988 a NSFNET passou a ser mantida com o apoio das organizações IBM, MCI (operadora de telecomunicações) e MERIT (instituição responsável por uma rede de computadores de instituições educacionais de Michigan), que formaram uma associação conhecida como ANS (*Advanced Network and Services*).

Em 1990 o backbone ARPANET foi desativado, criando-se em seu lugar o backbone DRI (Defense Research Internet). Em 1991/1992 a ANS desenvolveu um novo backbone, conhecido como ANSNET, que passou a ser o backbone principal da Internet. Nessa mesma época iniciou-se o desenvolvimento de um backbone europeu (EBONE), interligando alguns países da Europa à Internet.

A partir de 1993 a Internet deixou de ser uma instituição de natureza apenas acadêmica e passou a ser explorada comercialmente, com a construção de novos backbones por empresas privadas e fornecimento de serviços diversos [FONSE].

No Brasil, a Internet chegou em 1988 por iniciativa da comunidade acadêmica de São Paulo (FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo) e Rio de Janeiro (UFRJ - Universidade Federal do Rio de Janeiro e LNCC - Laboratório Nacional de Computação Científica).

O Ministério da Ciência e Tecnologia criou a Rede Nacional de Pesquisas (RNP) em 1989, uma instituição com os objetivos de iniciar e coordenar o oferecimento de serviços de acesso à Internet no Brasil. Como ponto de partida foi criado um backbone conhecido como o backbone RNP, interligando instituições educacionais à Internet com pontos de presença - PoP (*Point of Presence*) - em 11 estados.

A exploração comercial da Internet foi iniciada em dezembro de 1994, onde foram permitidos acessos à Internet inicialmente através de linhas discadas, e posteriormente, em abril de 1995, através de acessos dedicados via rede de pacotes ou via circuitos PPP (*Point-to-Point Protocol*).

De 1995 até os dias de hoje o mercado de Internet no Brasil popularizou-se e amadureceu. Hoje é composto por provedores de backbone, provedores de acesso e de conteúdo, com diversas origens e tamanhos. A Internet no Brasil dobra o número de usuários a cada ano. Hoje são cerca de 150 milhões de usuários conectados no mundo todo, com uma evolução de 70% ao ano, taxa nunca antes atingida por qualquer outro meio de comunicação. Hoje a Internet conecta aproximadamente 220 países.

Apesar de todo esse crescimento, tanto a administração quanto a operação da Internet são descentralizadas, ou seja, a Internet não é controlada por nenhuma organização ou pessoa. Apenas algumas tarefas são centralizadas, tais como a coordenação das pesquisas e padrões para funcionamento da rede, e a distribuição de

endereços e registros de domínios para interligação na rede. As principais instituições responsáveis por essas tarefas são:

- ISOC (*The Internet Society*) [ISOC] - através de fóruns, debates e publicações, procura orientar a pesquisa e utilização da Internet.
- IAB (*The Internet Architecture Board*) [IAB] - integrado à Internet Society em 1992, coordena toda a pesquisa e desenvolvimento envolvido no funcionamento da Internet, coordenando duas frentes de trabalho, que são os grupos de pesquisadores voluntários IRTF e IETF.
- IRTF (*The Internet Research Task Force*) [IRTF] - grupo com o objetivo de desenvolver pesquisas a longo prazo sobre funcionamento da Internet.
- IETF (*The Internet Engineering Task Force*) [IETF] - grupos responsáveis pelo desenvolvimento de padrões para funcionamento da Internet. Elaboram os documentos RFCs (*Request For Comments*), que embora tenham sido criados apenas como propostas para padronização, na prática tornaram-se os padrões oficiais da Internet.
- IANA (*Internet Assigned Numbers Authority*) [IANA] - entidade responsável pelo gerenciamento de todos os registros da Internet, de forma direta ou via terceiros. Mantém a administração dos números de protocolos e portas, mas delegou o registro de nomes de domínio ao InterNIC e os registros de endereços a órgãos regionais, como o ARIN (*American Registry of Internet Numbers*) para as Américas.
- InterNIC (*The Internet Network Information Center*) [INNIC] - composto por três empresas - AT&T, PSI e General Atomics - esta instituição centraliza a distribuição de informações da Internet Society (RFCs,...), além de coordenar a distribuição de endereços e registros de domínio para provedores a nível mundial.
- CERT/CC (*Computer Emergency Response Team/Coordinate Center*) [CERT] - grupo ligado a um departamento da Carnegie Mellon University, cuja atividade principal é a de coordenar respostas e ações a incidentes de segurança na Internet.

No Brasil, a instância máxima consultiva é o Comitê Gestor da Internet [CGINT], criado em junho de 1995. Tem como objetivo a coordenação da implantação do acesso à Internet no Brasil. A FAPESP é responsável pelos registros de domínio e endereços de redes cadastradas no Brasil, através do serviço conhecido como Registro BR [REGBR] e permite consulta e manutenção de registros.

Mesmo com a rápida evolução da Internet, o panorama de acesso continua sofrendo grandes mudanças no Brasil, principalmente a partir de 2001, quando as operadoras de telecomunicações começaram a fazer grandes investimentos em implantação de redes e tecnologias para comunicação de dados e Internet. Hoje está sendo priorizado o acesso dedicado à Internet em banda larga, com tecnologias de acesso como Wireless, Cable-Modem e ADSL (Assimetric Digital Subscriber Line). Devido ao baixo custo para o usuário final, o ADSL está despontando como preferido, e o número de usuários está crescendo rápido. Com a banda larga podem surgir aplicações distribuídas cada vez mais complexas, possibilitando novos serviços e demandando uma grande quantidade de conteúdo multimídia. Isto transformará totalmente o perfil de tráfego nas redes de telecomunicações baseadas no TCP/IP, que suportam a Internet.

2.2 A Arquitetura da Internet

Conforme já apresentado, a Internet não possui nenhum “centro” que controle a rede. Ela é um conjunto de mais de 70 mil redes no mundo inteiro, e mais de 150 milhões de usuários conectados, obtendo uma evolução de 70% ao ano. O que faz com que todas essas redes e usuários consigam realmente se comunicar ?

A resposta começa no início do desenvolvimento da rede do Departamento de Defesa Norte-Americano em 1969. Estava no auge a Guerra Fria e os cientistas queriam uma rede que continuasse funcionando mesmo em caso de um bombardeio. Surgiu então o conceito principal da Internet: uma rede em que todos os pontos se equivalem e não existe um comando central. Assim, se B deixa de funcionar, A e C

continuam podendo se comunicar. Esse projeto teve como resultado a concepção do conjunto de protocolos conhecido como TCP/IP Protocol Suíte, ou Protocolo Internet TCP/IP, ou ,simplesmente, TCP/IP devido a seus dois protocolos mais importantes:

- TCP = Transmition Control Protocol;
- IP = Internet Protocol.

Todas as redes e computadores que compõem a Internet possuem em comum o TCP/IP. Ele executa a conectividade em nível de rede, o que permite a comunicação entre aplicações de computadores em redes distintas, sem a necessidade de conhecimento da topologia envolvida nesse processo [SOA95].

Uma outra característica importante do TCP/IP é a flexibilidade de adaptação às tecnologias de redes existentes e futuras. Isso é possível porque o TCP/IP foi concebido de forma independente das tecnologias de redes, e é adotado comercialmente por muitos ambientes como UNIX, Novell, Windws NT e OS/2.

A arquitetura básica do TCP/IP é mostrada na figura 2.1 e é composta por quatro níveis [MUR00]:

- **Aplicação** – oferece ao usuário um conjunto de protocolos e serviços padronizados para acesso à rede. Qualquer usuário pode criar suas próprias aplicações, pois o TCP/IP é uma arquitetura aberta. Uma aplicação é um processo de usuário cooperando com outro processo, no mesmo servidor ou em outro. Exemplos de protocolos de aplicação: correio eletrônico (SMTP = *Simple Mail Transfer Protocol*), conexão remota (TELNET), transferência de arquivo (FTP = *File Transfer Protocol*), etc.
- **Transporte** – fornece serviços de transferência de dados fim-a-fim entre aplicações. Os serviços neste nível devem oferecer transparência com relação aos níveis inferiores e dar suporte para o controle de vários canais de comunicação entre as aplicações, simultaneamente. Existem dois protocolos neste nível: TCP (Transmition Control Protocol), que é orientado a conexão e garante a entrega dos dados na ordem correta; e UDP (User Datagram Protocol), que não é orientado a conexão e não garante a entrega de dados, mas tem menos overhead e

funcionalidades mais simplificadas, sendo útil para aplicações que precisam de transporte rápido. O acesso das aplicações à camada de transporte é feito através de portas que recebem um número inteiro para cada tipo de aplicação, podendo também tais portas serem criadas ao passo em que necessidades vão surgindo com o desenvolvimento de novas aplicações.;

- **Rede** – responsável pelo endereçamento dos equipamentos e pelo roteamento dos dados na rede, permitindo que aplicações troquem informações, mesmo estando em estações conectadas a redes completamente diferentes. . O IP (Internet Protocol) é o protocolo dessa camada e não é orientado a conexão. Em conjunto com o TCP pode oferecer um serviço de alta confiabilidade. Outros protocolos deste nível são ARP, ICMP, RARP, além de protocolos de roteamento como RIP, OSPF e EGP;
- **Enlace e Físico** – este nível não define um padrão próprio já que o objetivo é acomodar os diversos tipos de rede existentes (X.25, ATM, FDDI, Ethernet, Token Ring, Frame Relay, SLIP, etc.). Consiste de rotinas de acesso à rede física, interagindo com o hardware e permitindo que as demais camadas sejam independentes do hardware da rede.

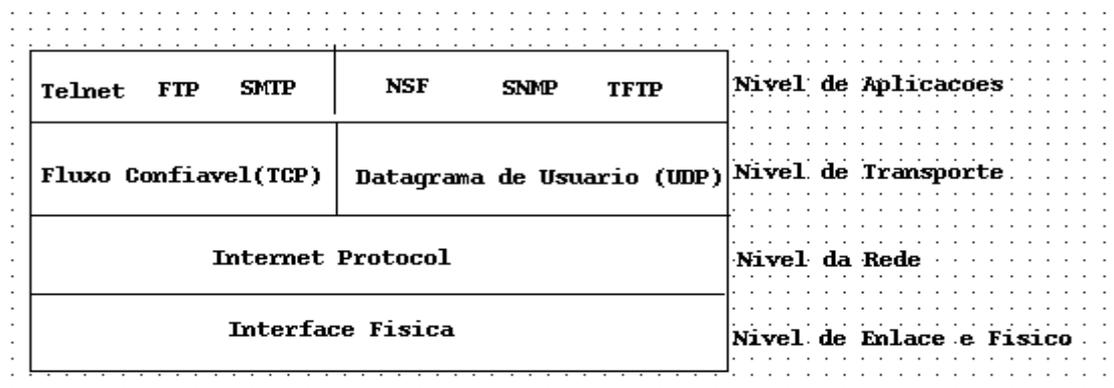


FIGURA 2.1 – ARQUITETURA DO TCP/IP

2.2.1 Protocolo TCP

O TCP é um dos protocolos do nível de transporte e é orientado à conexão, o que indica que neste nível vão ser solucionados todos os problemas de erros que não forem solucionados na camada inferior (nível da rede-IP). Ele trabalha com mensagens de reconhecimento, especificação do formato da informação e mecanismos de segurança. Garante que todas as PDU's (Protocol Data Unit) serão enviadas com sucesso, já que realiza transmissões orientadas à conexão, ou seja, independente da qualidade das sub-redes, o TCP garante que as PDU's chegarão ao destino. Quando executado, utiliza o protocolo IP, não orientado à conexão. O TCP então fica responsável pelo controle dos procedimentos da transferência segura de dados. Cabe salientar que o IP não é o único protocolo não orientado à conexão que pode ser utilizado pelo TCP.

A seguir as principais funções implementadas pelo TCP:

- Estabelecimento e liberação de conexão - antes e depois das transferências de dados, através de um mecanismo chamado *three-way-handshake* no qual os TCP's origem e destino trocam as mensagens necessárias para a função;
- Transferência de dados confiável- através de mensagens de tamanho variável em full-duplex, com espera de reconhecimento (ACK);
- Multiplexação - as mensagens de cada aplicação simultânea são multiplexadas para repasse ao IP, através do uso de portas. Ao chegar ao destino, o TCP demultiplexa as mensagens para as aplicações destinatárias;
- Segmentação - quando o tamanho do pacote IP não suporta o tamanho do dado a ser transmitido, o TCP segmenta (mantendo a ordem) para posterior remontagem na máquina destinatária, sem prejudicar a aplicação ;
- Controle do fluxo - através de um sistema de bufferização denominado janelas deslizantes, o TCP envia uma série de pacotes sem aguardar o reconhecimento de cada um deles, otimizando a transmissão. Na medida em que recebe o reconhecimento de cada bloco enviado, atualiza o buffer (caso reconhecimento

positivo) ou reenvia (caso reconhecimento negativo ou não reconhecimento após um timeout), obtendo controle do fluxo de dados;

- Precedência e segurança - os níveis de segurança e prioridade de dados durante a transmissão podem ser controlados pelo administrador.

O TCP permite que múltiplos programas de aplicação numa determinada máquina se comuniquem simultaneamente, e se encarrega de demultiplexar o tráfego TCP entrante para os programas de aplicação. É usado o número de porta para identificar o último destino numa máquina. A cada porta é associado um número inteiro pequeno para identificá-lo.

Quando se diz que o TCP é orientado à conexão, significa conexão de um circuito virtual e não de uma porta individual. Então, uma conexão consiste de um circuito virtual entre dois programas de aplicações, e é identificada por um par de "endpoints". Um *endpoint* é definido como um par de inteiros (*host,port*), onde *host* é o endereço IP para um computador e *port* é uma porta TCP nesse computador. A seguir são apresentados exemplos sobre *endpoints*:

- o *endpoint* 128.10.2.3,25 especifica a porta TCP número 25 na máquina com o endereço IP 128.10.2.3.
- uma conexão entre as máquinas 192.107.104.12 e 143.54.2.99 usando as portas 1069 e 25, respectivamente, seria definida pelos dois *endpoints* : (192.107.104.12,1069) e (143.54.2.99,25).

Já que o TCP identifica uma conexão por um par de *endpoints*, uma mesma porta pode ser compartilhada por múltiplas conexões na mesma máquina, possibilitando a multiplexação

Na figura 2.2 é mostrado o formato do segmento TCP.

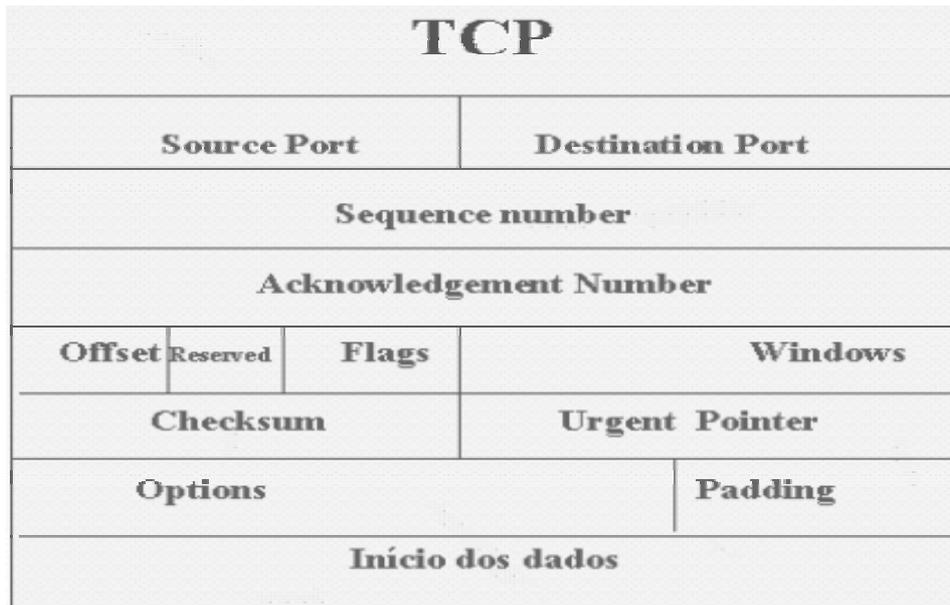


FIGURA 2.2 – SEGMENTO TCP

Considerando a quantidade de campos do segmento TCP, será apresentado apenas o significado dos campos de interesse para o escopo deste trabalho:

- **Source Port** (16 bits) - este campo no cabeçalho TCP contém o número da porta TCP de origem. É usado pelo receptor para responder.
- **Destination Port** (16 bits) - este campo no cabeçalho TCP contém o número da porta TCP de destino da mensagem.

2.2.2 Protocolo UDP

Este protocolo do nível de transporte provê um serviço não orientado à conexão, usando IP para transportar mensagens entre duas máquinas. O UDP é um protocolo mais rápido que o TCP pelo fato de ter menos overhead, e por não verificar o reconhecimento das mensagens enviadas. Por este mesmo motivo, não é confiável como o TCP. Não controla o fluxo podendo os datagramas IP chegarem fora de seqüência ou

até mesmo não chegarem ao destino. Opcionalmente pode conter um campo checksum. Também existe um mecanismo de portas, como no TCP, que o transmissor usa para distinguir entre múltiplos receptores numa mesma máquina.

Cada datagrama UDP, conforme mostra a figura 2.3 , é formado por um cabeçalho UDP e uma área de dados. O formato do cabeçalho UDP está dividido em quatro campos de 16 bits.

O significado dos campos do segmento UDP é apresentado a seguir:

- **Source Port** (16 bits) - este campo contém o número da porta UDP de origem da mensagem. É um campo opcional, e quando é usado especifica a porta para a qual uma resposta poderia ser enviada. Se não é usado contém zeros.
- **Destination Port** (16 bits) - este campo contém o número da porta UDP de destino da mensagem.
- **Length** - contém um contador de bytes do datagrama UDP. O valor mínimo é oito, sendo este só o comprimento do cabeçalho.
- **Checksum** - este campo é opcional. Um valor zero indica que o checksum não é computado.

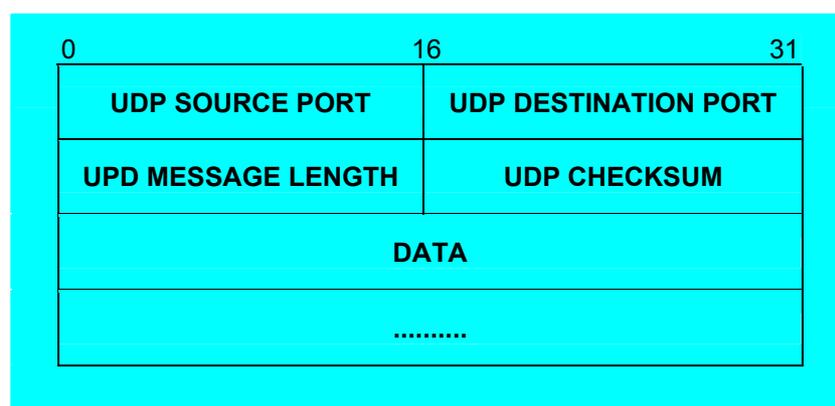


FIGURA 2.3 – DATAGRAMA UDP

2.2.3 Protocolo IP

A principal função do protocolo IP é transportar os datagramas de uma rede à outra na Internet. Ele é um protocolo de transmissão não orientado à conexão, e por ser mais básico, não apresenta muitas características do TCP. Não possui mecanismos de retransmissão, não dando garantia de uma transmissão íntegra ou ordenada. Dessa forma, todas as funções de consistência para a íntegridade dos dados transmitidos fica por conta do TCP.

A função de roteamento, escolhendo um caminho sobre o qual os dados serão enviados é executada pelo IP. Ele também inclui um conjunto de regras que envolvem a idéia da expedição de pacotes não confiáveis. Estas regras indicam como os hosts ou gateways poderiam processar os pacotes; como e quando as mensagens de erros poderiam ser geradas; e as condições em que os pacotes podem ser descartados.

O endereçamento na Internet é executado associando um endereço íntegro de 32 bits, chamado endereço IP, para cada computador da rede. Especificamente um endereço IP define o identificador da rede ao qual o *host* está conectado e também a identificação de um único computador nessa rede. Cada endereço é um par (*netid*, *hostid*), onde *netid* identifica a rede, e *hostid* identifica um computador nesta rede.

As redes são divididas hierarquicamente, existindo redes maiores e menores, com relação à possibilidade de um número maior ou menor de *hosts* conectados. De maneira geral, existem as redes classes A, B e C. O formato dos endereços IP é mostrado na figura 2.4.

A seguir são apresentadas as quantidades de redes e endereços de hosts par cada uma das três classes:

	Redes	Endereços
Classe A	126	16.777.214
Classe B	16.382	65.534
Classe C	2.097.150	254

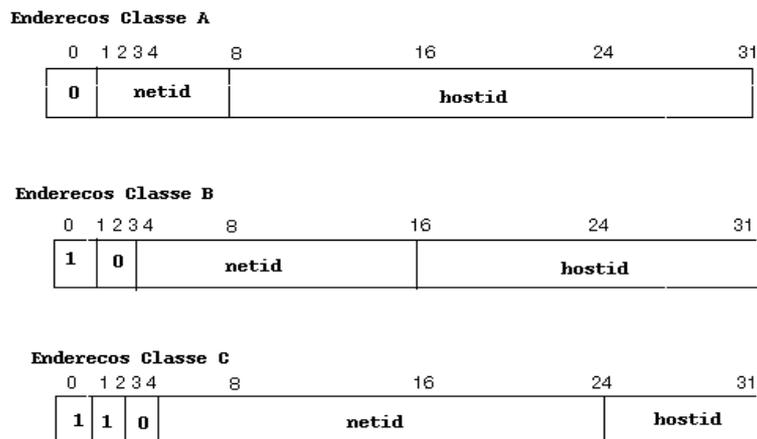


FIGURA 2.4 – ENDEREÇOS IP NAS REDES CLASSES A, B E C

O datagrama IP é a unidade básica de dados no nível IP, e está dividido em duas áreas, uma área de cabeçalho e outra de dados. O cabeçalho contém toda a informação necessária que identifica o conteúdo do datagrama. Na área de dados está encapsulado o pacote do nível superior, ou seja um pacote TCP ou UDP. Existem várias versões do protocolo IP, mas a Internet atual utiliza a versão IPv4. No futuro a Internet deve migrar para a versão IPv6 (evolução do IPv4). Na figura 2.5 é mostrado o formato do datagrama IPv4.

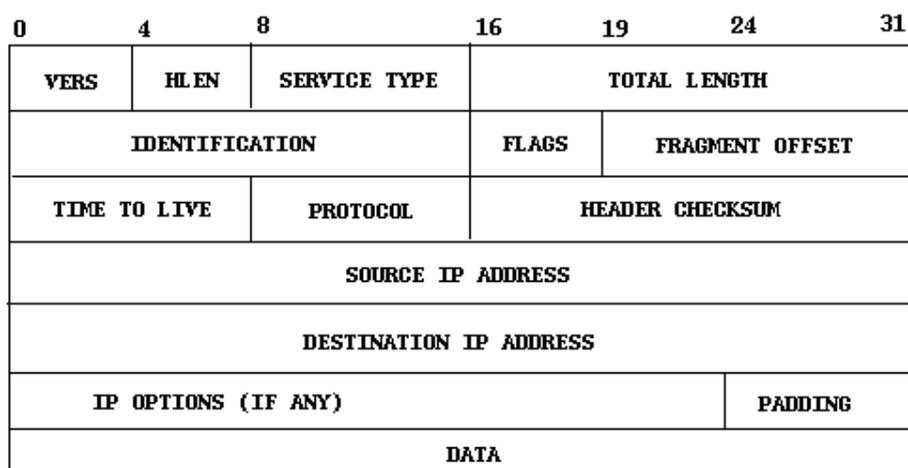


FIGURA 2.5 – DATAGRAMA IP

O significado dos campos que interessam ao escopo deste trabalho, são apresentados a seguir:

- **Vers** (4 bits) - versão do protocolo IP que foi usada para criar o datagrama.
- **Service-Type** - este campo especifica como o datagrama poderia ser manejado pela rede. Este campo é dividido da seguinte forma:
 - **Precedence** (3 bits): indica precedência de datagramas com valores desde 0 (precedência normal) até 7 (controle da rede). Com estes bits permite-se ao transmissor indicar a importância de cada datagrama que ele está enviando.
 - **TOS-Type of Service** (4 bits): indicam o tipo de transporte que o datagrama deseja, como baixo retardo, alto throughput, alta confiabilidade, baixo custo ou normal. É possível que estes tipos de serviços não possam ser oferecidos, já que dependem das condições físicas da rede.
- **Total-Length** - este campo proporciona o comprimento do datagrama medido em bytes, incluindo cabeçalho e dados.
- **Time To Live** - especifica o tempo em segundos que o datagrama tem permissão para permanecer no sistema Internet. Gateways e hosts que processam o datagrama devem decrementar o campo TTL cada vez que um datagrama passa por eles e devem removê-lo quando seu tempo expirar.
- **Protocol** - especifica qual protocolo de nível imediatamente superior foi usado para criar a mensagem que está sendo transportada na área de dados do datagrama.
- **Header- Checksum** - assegura integridade dos valores do cabeçalho.
- **Source e Destination IP Address** - especifica o endereço IP de 32 bits do remetente e do receptor deste datagrama.

O datagrama IP deve ser enviado ao destino correto através do roteamento, que é feito baseando-se no endereço IP destino contido no datagrama. A Internet é composta de múltiplas redes físicas interconetadas por computadores chamados *gateways*, que possuem conexão direta para duas ou mais redes. A diferença entre um

gateway e um *host* é que este último está ligado a somente uma rede física. No roteamento IP participam *hosts* e *gateways*, e o roteamento pode ser direto ou indireto.

No roteamento direto a transmissão do datagrama é de uma máquina à outra, sendo que as duas devem estar na mesma rede. O transmissor encapsula o datagrama num quadro físico (frame: nível de enlace), liga o endereço IP destino ao endereço físico (de hardware) correspondente, e envia o quadro resultante diretamente ao destino sem precisar passar por um gateway..

No roteamento indireto o remetente deve identificar um gateway ao qual o datagrama pode ser enviado, depois o gateway deve enviar o datagrama à rede destino. Isso é executado por uma rotina de roteamento IP, que vai selecionar o próximo *gateway* que formará parte do caminho que levará o datagrama ao *host* destino. A rotina ou algoritmo de roteamento usa uma tabela de roteamento IP em cada máquina que armazena informações acerca de possíveis destinos e as maneiras de chegar a eles.

Tanto os hosts como gateways têm tabelas de roteamento. Tipicamente, a informação armazenada nas tabelas de rotas é o par (N,G), onde N é o endereço IP da rede destino e G é o endereço IP do próximo gateway no caminho da rede N. Assim, uma tabela de roteamento num gateway G somente especifica um passo do caminho de G a uma rede destino. O gateway não sabe o caminho completo ao destino. A tabela sempre aponta aos gateways que podem ser alcançados através da rede à qual esse gateway está conectado. Todos os gateways listados na tabela de roteamento de uma máquina M devem conectar-se em redes às quais M está conectada diretamente. Nas tabelas de roteamento não é possível armazenar as informações de cada máquina destino, seria impossível manter as tabelas atualizadas, além de que as máquinas teriam problemas com armazenamento para toda a informação.

Como o IP provê um serviço de expedição de datagramas sem conexão e não confiável, e além disso um datagrama viaja de um gateway a outro até alcançar um gateway que possa expedí-lo diretamente ao host destino; é necessário um mecanismo que emita informações de controle e de erros quando acontecerem problemas na rede. Alguns dos problemas típicos que podem acontecer são:

- Um gateway não pode expedir ou rotear um datagrama.
- Um gateway detecta uma condição não usual tal como congestionamento.

O mecanismo de controle que emite mensagens quando acontece algum erro é executado pelo protocolo ICMP, permitindo aos gateways enviar mensagens de erros ou de controle a outros gateways ou hosts. O ICMP somente reporta condições de erros à fonte original. A fonte deve relatar os erros aos programas de aplicação que deve tomar ação para corrigir o problema. Uma das mensagens que o ICMP pode enviar é: Destination Unreachable (destino não alcançável), o qual, por sua vez pode ser dos seguintes tipos:

- Network Unreachable (rede não alcançável)
- Host Unreachable (host não alcançável)
- Port Unreachable (porta não alcançável)
- Destination Host Unknown (Host destino desconhecido)
- Destination Network Unknown (rede destino desconhecida)

Assim, o protocolo ICMP também tem funções de importância dentro de uma rede IP.

2.3 Aplicações e Negócios na Internet

A Internet permitiu a disseminação de informações e conhecimento, antes distantes e difíceis de serem obtidos, através de uma forma rápida e simples como o simples clicar de um mouse. É como se barreiras de tempo tivessem sido rompidas, permitindo o acesso e disponibilizando informação num ambiente de alcance mundial.

A explosão no uso da Internet também foi impulsionado por aplicações tais como o correio eletrônico (e-mail), que possibilitou a comunicação quase que instantânea entre pessoas, e o uso de navegadores (browsers), que trouxeram aos olhos do usuário uma informação visual de boa qualidade e com facilidade de acesso, inclusive encaminhando o usuário entre diversos computadores da rede, inserindo o conceito de WWW(World Wide Web), ou, a teia de aranha eletrônica disseminadora de informações.

Com estas facilidades, um grande número de empresas passou a fazer uso da Internet para se comunicar com clientes e fornecedores através de e-mail, e disponibilizar informações através de suas páginas eletrônicas (home-pages). Com o mercado potencial, houve um crescimento no número de provedores de Internet, empresas dedicadas a prover acesso à rede, tanto para pessoas físicas como jurídicas. Hoje, a grande maioria de servidores conectados à Internet é de natureza comercial.

No ramo bancário, por exemplo, a Internet revolucionou o atendimento e retirou milhares de pessoas das filas de bancos. O uso de técnicas avançadas de segurança trouxe ao meio da Internet um aval tecnológico gerando confiança ao usuário. Isto motivou outros segmentos a acreditarem no uso eficaz da Internet para prover serviços seguros. A Internet também está sendo usada para desburocratização em vários segmentos públicos.

No ramo empresarial, incluindo indústrias e comércio, a Internet está começando a ser usada para prover comunicação entre matriz e filiais, substituindo aos poucos as redes dedicadas e Frame Relay. Para ter maior segurança podem ser usados os conceitos de VPN-IP (Virtual Private Network), que garantem um rede privativa através da Internet. Esta comunicação também pode ser feita com fornecedores, distribuidores, etc.

Dessa forma, as empresas conseguem reduzir custos e continuar usando os sistemas corporativos como ERP (Enterprise Resource Planning – Planejamento de Recursos Empresarias) e CRM (Customer Relationship Management – Gerenciamento de Relacionamento com o Cliente); sistemas contábeis, de estoque e de vendas por exemplo; através de modelo cliente-servidor (que está na matriz). Pela rede também pode ser usado e-mail e acesso www, através do servidor que está na matriz.

Com a versatilidade e abrangência proporcionada pela Internet, e com o desenvolvimento de novas tecnologias de acesso à rede, mais rápidas e com menor custo como o ADSL, novas aplicações, avançadas e distribuídas, começam a ser utilizadas, e novos serviços podem ser oferecidos. Serviços estes principalmente baseados em sistemas multimídia. Conforme [WIL00], sistema multimídia é aquele capaz de manipular ao menos um tipo de mídia discreta e um tipo de mídia contínua, sendo as duas numa forma digital.

Alguns exemplos de novas aplicações avançadas são:

- **Telefonia e fax sobre IP (VoIP)** – possibilitando redução de custos ao usuário final.
- **Comércio eletrônico (*e-commerce*)** – possibilita que as empresas vendam seus produtos de maneira rápida e simples, além de ter abrangência mundial. Também é usado para relacionamento com fornecedores e distribuidores, entre outros.
- **Vídeo sobre IP** – possibilita transmissão de vídeo através da Internet, com aplicações diversas.
- **Educação à distância** – é uma forma de distribuir conhecimento rapidamente e a um custo relativamente baixo
- **Vídeo-conferência** – pode ser usada para realizar reuniões entre executivos e gerentes em uma empresa, de forma que não seja preciso fazer deslocamentos.
- **Aplicações colaborativas e de grupo** – sistemas onde um grupo de pessoas compartilha um espaço de trabalho, como por exemplo um monitor de vídeo. A idéia é facilitar trabalhos em grupo, onde os usuários tem objetivos em comum.
- **Aplicações tempo real** – alguns sistemas como os de automação industrial, ou telemedicina, necessitam de respostas e ações em tempo real, e já existem aplicações via internet que se propõem a disponibilizar isto.

Dos exemplos acima, a aplicação de *e-commerce* é a que tem crescimento mais acelerado atualmente. Trata-se do uso de canais eletrônicos, incluindo a Internet, para efetuar uma transação comercial. O processo pode se caracterizar como EDI (*Electronic Data Interchange*), quando é realizado entre sistemas, ou como B2C (*Business-to-Consumer*) ou B2B (*Business-to-Business*), por exemplo, quando existe interação humana em uma das pontas. Os principais agentes envolvidos nessas transações são os consumidores, as empresas e o governo. As relações são descritas a seguir e mostradas na figura 2.6.

- **B2B (*business-to-business*)**: transações entre empresas. Ex: EDI, portais verticais de negócios, como e-procurement.
- **B2C/C2B (*business-to-consumer / consumer-to-business*)**: transações entre empresas e consumidores. Ex: lojas e *shoppings* virtuais.

- **B2G/G2B** (*business-to-government* / *government-to-business*): transações envolvendo empresas e governo. Ex: EDI, portais de compras.
- **C2C** (*consumer-to-consumer*): transações entre consumidores finais. Ex: *sites* de leilões, classificados *on-line*.
- **G2C/C2G** (*government-to-consumer* / *consumer-to-government*): transações envolvendo governo e consumidores finais. Ex: pagamento de impostos, serviços de comunicação.
- **G2G** (*government-to-government*): transações entre governo e governo.

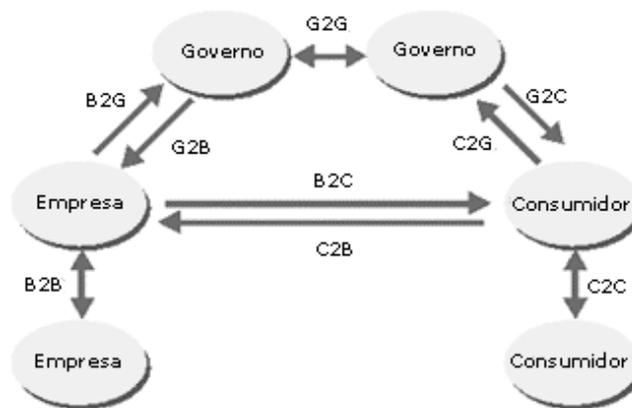


FIGURA 2.6 – RELAÇÕES DE *E-COMMERCE*

Comprar e vender pela Internet tornou-se uma necessidade para empresas que procuram uma forma de crescer ou manter-se no mercado, e também para entidades que buscam uma alternativa para melhorar a prestação de seus serviços. Os números saltam a cada dia e o comércio na Web movimentava bilhões de dólares. Estar na Internet, portanto, é como manter uma loja aberta 24 horas por dia, e no mundo.

O *e-commerce* abre inúmeras possibilidades de compra e venda de produtos e serviços num mercado potencial de milhões de usuários espalhados pelo mundo todo. Conforme [MCT01], pode-se observar a explosão do comércio eletrônico, verificando que em 1998 o comércio total no mundo foi em torno de US\$ 80 bilhões, e a previsão para 2003 é de mais de US\$ 3 trilhões. Um crescimento de 109% ao ano, num período de 5 anos. Isto pode ser observado na figura 2.7.

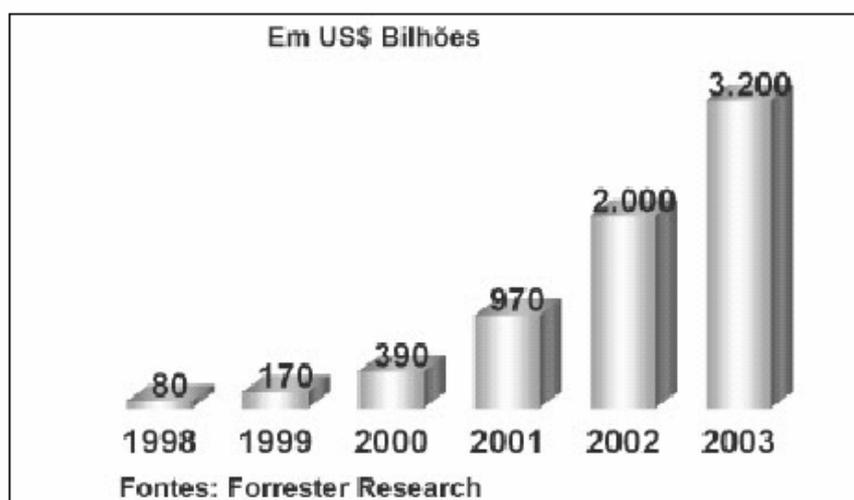


FIGURA 2.7 – VENDAS VIA *E-COMMERCE* NO MUNDO

Assim, a Internet está alavancando negócios para muitas empresas, principalmente para aquelas que aprenderam a utilizar as facilidades da rede mundial a seu favor. Mas a Internet não para, e as novas aplicações avançadas, citadas anteriormente, serão motivo de grandes mudanças econômicas e tecnológicas no futuro próximo.

No entanto, aplicações avançadas possuem requisitos técnicos que devem ser oferecidos pela rede de transporte para o perfeito funcionamento, ou seja, requisitos de qualidade devem ser atingidos pela Internet. Hoje, ela não está preparada para isso, e torna-se de fundamental importância a adequação da rede para operação com qualidade de serviço (QoS).

2.4 Conclusão

Pode-se dizer que está sendo muito rápido o desenvolvimento da Internet desde sua criação e utilização para objetivos comerciais. Desde 1993, quando começou a ser explorada comercialmente, até agora passaram-se apenas 10 anos e muitas aplicações foram desenvolvidas. O fato da Internet não ser controlada por um órgão único, também abriu espaço para a globalização deste instrumento.

A confiabilidade e rapidez proporcionada pela arquitetura utilizada, através dos protocolos TCP, UDP e IP, além da abrangência da Internet, permitiram a utilização dessa rede para implementar diversos serviços de grande utilidade. A comunidade mundial se beneficiou disso, e principalmente quem está sabendo usar a Internet a seu favor está tendo um grande retorno, como no caso de empresas que utilizam o e-commerce.

Como o desenvolvimento de novas aplicações e serviços não para, e estes estão cada vez mais complexos, é natural que seus requisitos com relação à rede também o sejam. Dessa forma, as novas aplicações, principalmente baseadas em multimídia, exigem características de qualidade que a Internet atual não proporciona. Torna-se necessário, portanto, que sejam desenvolvidas e implantadas metodologias ou facilidades na rede Internet, que forneçam garantias de qualidade de serviço (QoS), sendo este o tema abordado no próximo capítulo.

3 QUALIDADE DE SERVIÇO NA INTERNET

Cerca de 150 milhões de usuários estão conectados na Internet em 217 países e com taxa de crescimento de 70% ao ano. A quantidade de novas aplicações cresce rapidamente, sendo cada vez mais complexas e demandando uma grande quantidade de conteúdo multimídia, transformando totalmente o perfil de tráfego nas redes de telecomunicações.

A Internet, hoje, fornece um serviço do tipo melhor esforço, ou seja, o tráfego é tratado tão rápido quanto possível, mas não há garantias de tempo nem limites de erro ou perdas de pacotes. Dessa forma, várias aplicações sensíveis a tempo de transmissão e/ou perda de pacotes, como voz e vídeo, acabam tendo seu desempenho prejudicado, ou até mesmo se tornam inviáveis. Assim, é necessário o desenvolvimento de meios que permitam à tecnologia IP garantir a Qualidade de Serviço (QoS) dentro da rede de transporte, viabilizando o uso de aplicações sensíveis ao desempenho da rede.

Com a rápida transformação da Internet em uma infra-estrutura comercial, o fornecimento de qualidade de serviço está sendo considerado cada vez mais um requisito essencial. Os termos qualidade e serviço podem ser vistos como a capacidade do sistema em diferenciar tráfegos, ou tipos de serviços, de forma que as classes de tráfego possam ser tratadas diferentemente umas das outras. Neste sentido, é interessante que uma Operadora, ou Provedor de Serviços Internet, possa fornecer a seus clientes diversas alternativas em termos de qualidade de serviço, garantindo de um certo modo a taxa de bits e atraso, por exemplo. Um serviço com mais qualidade terá um custo maior para o cliente, do que um serviço do tipo melhor esforço.

Segundo [FER98], Qualidade de Serviço pode ser definida como o fornecimento de serviços de entrega de dados de forma consistente e previsível. Assim,

os usuários da rede podem ter certeza de que as necessidades de tráfego de suas aplicações poderão ser satisfeitas, tomando por base os parâmetros de qualidade de serviço.

Como visto no capítulo anterior, cabe ao protocolo IP o envio e recebimento dos pacotes desde a estação origem até a estação destino. No entanto, essa arquitetura apresenta alguns inconvenientes, como nenhuma garantia de vazão e entrega dos pacotes (que podem ser simplesmente descartados ou perdidos), e nenhuma garantia de tempo para entrega dos pacotes. Ou seja, o IP não oferece nenhuma garantia de qualidade de serviço de rede.

Assim, o desempenho e a confiabilidade dos serviços de comunicação da Internet atual podem não ser suficientes para garantir o perfeito funcionamento de certas aplicações, principalmente aplicações de tempo-real ou multimídia (vídeo sob demanda, vídeo-conferência, voz, tele-medicina, etc.). Existem alguns parâmetros técnicos definidos para especificar certas características da rede, e a partir deles pode-se dizer qual a qualidade de serviço que esta rede poderá fornecer.

Este capítulo tem o objetivo de apresentar os principais conceitos associados a QoS. A seção 3.1 apresenta os parâmetros de qualidade de serviço. Em seguida, a seção 3.2 mostra como algumas aplicações são sensíveis aos parâmetros de QoS e relaciona os seus requisitos. Métodos para obter QoS, detalhando o modelo Serviços Diferenciados é o tema da seção 3.3.

3.1 Parâmetros de Qualidade de Serviço

A especificação da qualidade do serviço a ser oferecido a uma aplicação é feita a partir da indicação de um conjunto de parâmetros de qualidade de serviço. Estes parâmetros são, portanto, quantificações de características que a rede pode oferecer ao fluxo de pacotes.

Segundo [MAR00] e [SEM01], os parâmetros mais utilizados são:

- **Vazão (taxa de transmissão):** é a taxa efetiva de bits trafegada na rede, medida em bits/s. Pode-se dizer que é a capacidade do sistema em transferir dados. A vazão da maioria das redes varia com o tempo, já que os recursos da rede são utilizados de forma estatística. Também podem ocorrer congestionamentos em determinadas rotas, ocasionando menor vazão durante esses períodos.
- **Atraso (latência):** é o tempo total entre o envio de um pacote, ou mensagem, a partir da origem até sua recepção no destino. É a somatória dos atrasos gerados pelos meios de transmissão e pelos equipamentos envolvidos na rota fim-a fim da transmissão, incluindo roteadores, switches, etc. Do ponto de vista da aplicação, a latência (atraso) resulta em um tempo de resposta ou tempo de entrega da informação. Os principais fatores que influenciam na latência são atraso de propagação do sinal no meio de transmissão, velocidade da transmissão e tempo de processamento nos equipamentos da rede. Em redes WAN o efeito da latência é muito mais observado que em redes LAN. Os próprios *hosts* e servidores também geram atrasos que devem ser considerados se a medição for fim-a fim na aplicação.
- **Variação do atraso:** pode ser visto como a variação do atraso de entrega dos pacotes de um mesmo fluxo. A variação de atraso é gerada principalmente pela variação nos tempos de processamento dos pacotes pelos equipamentos de rede, como roteadores. Isto ocorre, em geral, devido às variações na quantidade de tráfego que cursa pela rede. Aplicações interativas de voz e vídeo, ou aplicações de tempo-real podem ser afetadas pela variação do atraso.
- **Taxa de perda de pacotes:** é a razão entre a quantidade de pacotes perdidos na rede e a quantidade de pacotes que foram transmitidos. Esta perda ocorre principalmente devido ao descarte de pacotes nos roteadores, durante períodos de congestionamentos, onde os *buffers* internos não comportam a quantidade de pacotes a armazenar. O processo de descarte depende de fatores como condicionamento do tráfego, enfileiramento e tratamento dos pacotes nos *buffers* internos. Também podem ocorrer perda de pacotes no meio de transmissão, mas nos dias de hoje esses sistemas já são muito confiáveis (a maioria dos grandes

sistemas usa Fibra Ótica). Algumas aplicações multimídia são muito sensíveis a perda de pacotes, principalmente aplicações de voz.

3.2 Requisitos de QoS em Aplicações IP

Os parâmetros de qualidade definem características técnicas da rede, e as aplicações são sensíveis a eles de formas diversas. Uma são mais sensíveis ao atraso e variação de atraso, outras mais sensíveis à vazão e perda de pacotes, por exemplo. Nesta seção são estudados os requisitos de algumas aplicações IP com relação aos parâmetros de qualidade de serviço, para que funcionem de forma adequada.

Deve-se considerar que nem todas as aplicações necessitam realmente de garantias fortes e rígidas quanto a QoS, pois funcionam satisfatoriamente com o serviço de melhor esforço da rede atual. Essas são as aplicações tradicionais. No entanto, as novas aplicações, multimídia em geral, exigem maior desempenho e confiabilidade da rede.

Existem diversos tipos de aplicações e com requisitos distintos, portanto para cada aplicação o nível de QoS necessário é diferente. Uma boa alternativa é juntar aplicações com comportamentos e requisitos semelhantes para obter grupos com necessidades de QoS parecidas, visando facilitar o processo de negociação entre o provedor de serviços (Operadora) e os clientes. A classificação pode ser feita usando um critério de diferenciação em três grupos, segundo [SAL00]:

- **Aplicações Multimídia Interativas** – onde os dados são trocados e apresentados imediatamente (Ex: vídeo-conferências e VoIP). Essas aplicações dependem principalmente da vazão, atraso e variação do atraso.
- **Aplicações Multimídia Streaming** – onde os dados são recebidos pelo destino e bufferizados antes da apresentação (Ex: áudio e vídeo on demand, educação à distância e internet TV). Essa aplicações dependem principalmente da vazão.

- **Aplicações de Missão Crítica** – onde os dados devem ser transferidos com integridade (Ex: B2B e on-line banking). Essas aplicações dependem principalmente da vazão e da taxa de perda de pacotes.

Dentro de um mesmo grupo as necessidades das aplicações podem não ser exatamente iguais, e considerando que o número de aplicações aumenta cada vez mais, é difícil trabalhar apenas com o critério apresentado. [WAA99] propõe uma classificação mais adequada, usando um número maior de grupos por tipo de aplicação. Esta classificação é apresentada na tabela 3.1.

TABELA 3.1 – REQUISITOS POR TIPO DE APLICAÇÃO

Tipos de Aplicações	Exemplos	Requisitos		
		Largura de Banda	Atraso	Tolerância a perdas
Aplicações interativas confiáveis	Telnet, database accesss, transações HTTP curtas	Baixa	Baixo	Sim
Aplicações multimídia interativas	Voz sobre IP, Vídeo sobre IP	Baixa/alta	Baixo	Mínima
Aplicações multimídia não interativas	Áudio e vídeo broadcast, educação à distância	Alta	Variável, mas dentro de faixa pré-determinada	Mínima
Aplicações tipo requisição e resposta	Cliente-servidor, NFS	Baixa/alta	Baixo para transações curtas	Sim
Aplicações em batch	Backup, ftp, transações HTTP longas	Baixa/alta	Alto	Sim
Controle de rede	SNMP, BGP, OSPF	Baixa	Baixo	Alguma (não para SNMP)

É importante notar que é difícil fazer uma classificação exata e que possa ser tomada como definitiva, pois existem aplicações que podem se encaixar em mais de um tipo, como acontece com o HTTP, ou um tipo de aplicação pode ter requisitos variados como no caso de aplicações multimídia interativas onde a largura de banda pode ser baixa (para voz), mas deve ser alta (para vídeo).

Para a prestação de serviços profissionais, como aqueles providos pelas operadoras de telecomunicações, é importante que seja feito um trabalho de

reconhecimento das necessidades do cliente, visando fornecer o serviço correto, que possa atendê-lo com a qualidade desejada. Assim, a operadora também precisa conhecer as aplicações do cliente e seus requisitos. A seguir, serão apresentados mais detalhadamente os requisitos de algumas aplicações. No escopo deste trabalho, serão consideradas especialmente as aplicações web, vídeo e VoIP.

3.2.1 Requisitos para Aplicações de Voz

Sem dúvida, atualmente a aplicação de voz sobre IP (VoIP) é a que mais desperta interesse nos usuários, principalmente empresas. Esse interesse é devido principalmente à redução de custos que proporciona com relação às ligações telefônicas interurbanas. Para as operadoras, VoIP também é importante pois aquela que não puder disponibilizar este serviço, perderá mercado para a concorrente que possui o serviço.

A qualidade de voz percebida pelos usuários pode ser medida subjetivamente através do MOS (*Mean Opinion Score*) [ITU96]. O MOS varia numa escala de 1 a 5, conforme a tabela 3.2, e é interessante que um sistema de voz possua MOS acima de 3,9.

TABELA 3.2 – MOS

MOS	SATISFAÇÃO DO USUÁRIO
4,4 a 5,0	Muito satisfeito
4,0 a 4,3	Satisfeito
3,6 a 3,9	Alguns usuários insatisfeitos
3,1 a 3,5	Muitos usuários insatisfeitos
2,6 a 3,1	Quase todos insatisfeitos
1,0 a 2,5	Não recomendado

Para que uma aplicação VoIP tenha um funcionamento adequado, e considerando que a interatividade é fundamental, os parâmetros de qualidade importantes são a vazão, atraso, variação de atraso e taxa de perda de pacotes.

Vazão

A vazão necessária para uma aplicação de voz vai depender basicamente do codec utilizado. Existem vários padrões para codec de voz [VOI02], sendo que os mais utilizados e suas características são:

- G.711 – gera voz codificada numa taxa constante de 64 Kbit/s e faz a conversão do sinal do sinal analógico para digital em 1ms. O MOS fica em 4,1.
- G.729 – taxa de 8 Kbit/s e leva em média 25 ms para fazer a conversão e o processo de compressão do sinal. O MOS fica em 3,92 para G.729 e 3,7 para G.729 a (que utiliza menos poder de processamento da CPU).
- G.723 – possui dois algoritmos de compressão gerando taxas de 6,3 ou 5,3 Kbit/s. Leva em média 67 ms para a conversão. O MOS fica em 3,9 e 3,65 respectivamente.

Esses padrões podem usar a técnica de supressão de silêncio (para reduzir ou zerar a taxa de envio durante momentos de silêncio na fala), e por isso geram uma taxa não constante. Os codecs trabalham a nível de aplicação e a taxa de bits efetivamente gerada no nível de rede é superior ao apresentado acima, pois deve ser adicionado o overhead gerado pelos protocolos inferiores.

As aplicações que usam a recomendação H.323 devem encapsular o fluxo de bits gerados no protocolo RTP (Real Time Protocol). Em seguida este é encapsulado pelo UDP e este pelo IP, gerando um overhead de 40 bytes, não considerando o protocolo da camada 2 que pode ser Ethernet, Frame Relay, PPP, ATM, etc.

É muito importante notar que as taxas geradas pelos codecs, apresentadas acima, representam a vazão vista pela aplicação. A nível de rede, esta taxa aumenta devido ao encapsulamento nos protocolos RTP, UDP e IP. A taxa efetiva de um fluxo de voz gerado sobre uma rede PPP, por um codec G.729, é de 26,4 Kbit/s conforme [VOI02]. Para G.711 a taxa ficará em 82,4 Kbit/s e para G.723 a taxa será de 16 Kbit/s ou 13,4 Kbit/s. Estes são valores que devem ser usados para projetos de VoIP.

Visando reduzir mais ainda a taxa de bits para um fluxo de voz, foi proposto o CRTP (Compressed Real Time Protocol), conforme [CON02], que realiza a compressão do cabeçalho IP/UDP/RTP de 40 bytes para aproximadamente 2 a 5 bytes. Com isso se consegue uma redução muito grande no overhead, já que os pacotes de voz

são pequenos (de 20 a 160 bytes). Para G.729 a taxa efetiva passa de 26,4 para 11,2 Kbit/s e para G.711 passa de 82,4 para 67,2 Kbit/s.

Atraso

O atraso deve ser muito baixo, pois a VoIP é uma aplicação multimídia interativa. Caso o atraso seja muito grande, a comunicação interativa de voz é inviabilizada.

Conforme [MIR02], o atraso deve ter um valor menor do que 100 ou 150 ms. Caso exista um atraso acima dos valores sugeridos, a qualidade da voz ficará afetada, segunda a tabela 3.3 que mostra recomendações de valores para o atraso.

TABELA 3.3 – VALORES DE ATRASO PARA VOIP

Atraso	Efeito na qualidade da voz
< 150 ms	Não afeta a qualidade
150-250 ms	Qualidade ainda aceitável. Atraso um pouco perceptível.
> 250-300 ms	Qualidade inaceitável

Variação de atraso

O valor da variação de atraso deve ser baixo, apesar de seus efeitos poderem ser minimizados através da técnica de bufferização. Esta técnica tem o objetivo de mascarar a variação de atraso, liberando os pacotes para a aplicação sem a variação gerada na rede. Para pacotes que cheguem com variação de atraso muito grande, não será possível a correção. A tabela 3.4 mostra recomendações de valores para a variação do atraso, segundo [MIR02].

TABELA 3.4 – VALORES DE VARIAÇÃO DO ATRASO PARA VOIP

Varição do Atraso	Efeito na qualidade da voz
< 40 ms	Não afeta a qualidade
40-75 ms	Qualidade ainda aceitável. Voz pode ficar confusa/misturada
> 75 ms	Qualidade inaceitável

Taxa de Perda de pacotes

A perda de pacotes será um problema caso ela ocorra em rajadas, pois não será possível recuperar a qualidade de voz. O efeito da perda de pacotes também depende do tamanho dos pacotes: pacotes maiores geram maior problema em caso de perda. Segundo [MIR02], levando em conta o MOS, a perda de pacotes ideal é de até 1%, podendo ser aceitável até 2% caso o atraso seja menor que 150 ms.

3.2.2 Requisitos para Aplicações de Vídeo

A transmissão de vídeo oferece serviços avançados de comunicação, colaboração (entre grupos de trabalho, por exemplo) e entretenimento, mas demanda recursos consideráveis da rede quando comparado a outras mídias. Os requisitos de um fluxo de vídeo dependem da aplicação, sendo que existem duas grandes modalidades de aplicação: o vídeo interativo e o vídeo streaming.

Em aplicações de vídeo interativo, a interatividade entre os usuários é a característica principal e deve ser mantida para a satisfação destes. Alguns exemplos de aplicações de vídeo interativo incluem videoconferência, ambientes de colaboração virtual, aplicações médicas como cirurgia remota, etc.

Em aplicações de vídeo *streaming*, a interatividade não é a principal característica, pois se trata da transmissão de vídeo ao vivo ou armazenado. Esta transmissão é unidirecional para um grupo de usuários, ou para um único usuário, sendo que não existe canal de retorno. Alguns exemplos de streaming de vídeo incluem a transmissão de notícias gravadas, transmissão de vídeo ao vivo, TV paga, vídeo on demand, etc.

Vazão

Para aplicações de vídeo streaming, os requisitos de vazão dependem do codec e da qualidade de vídeo desejada pelo usuário. A vazão necessária pode variar de

dezenas de Kbit/s a 1 Mbit/s para aplicações mais modestas, como Real Vídeo, ou até vários Mbit/s ou Gbit/s para aplicações que exijam alta qualidade de vídeo.

Segundo [MIR02], para aplicações com maior qualidade são utilizados os codesc MPEG (*Moving Picture Experts Group*). Para qualidade próxima ao VCR (Vídeo Cassete Recorder) o codec MPEG-1 com taxas entre 1,5 e 3 Mbit/s é utilizado. O MPEG2 pode ser usado para obter qualidade de TV (SDTV – Standard Definition TV) na taxa de 6 Mbit/s. Também pode ser usado para obter qualidade de broadcast HDTV (High Definition TV) na taxa de 19,4 Mbit/s.

Como os custos de transmissão aumentam de forma proporcional à vazão, foi desenvolvido o padrão MPEG-4 que gera taxas bem menores, possibilitando a redução dos custos de transmissão. MPEG-4 obtém a mesma qualidade do MPEG-1, mas com taxas de 300 a 400 Kbit/s. O foco principal é a aplicação de vídeo on demand.

Para aplicações de vídeo interativas, considerando a recomendação H.323, os requisitos de vazão variam, segundo [MIR02], de 28,8 Kbit/s até $N \times 64$ Kbit/s usando os padrões de codec para vídeo H.261 e H.263. Este último pode ser usado desde taxas mais baixas (< 64 Kbit/s) até taxas de 2 Mbit/s. Para vídeos com muito movimento, é importante utilizar taxas de 384 Kbit/s até 768 Kbit/s, para obter qualidade de vídeo aceitável. Também podem ser usados codecs MPEG, para maior qualidade, mas estes são mais usados para aplicações de vídeo streaming.

Para obter a qualidade desejada, deve-se tomar cuidado ao definir as taxas de vazão necessárias. Como exemplo, no caso da videoconferência multi-ponto, para definir a taxa de vazão do ponto de concentração, é necessário considerar a soma das taxas referentes a todos os pontos remotos.

Atraso e Variação de Atraso

Para as aplicações interativas, como videoconferência, o baixo atraso e a baixa variação de atraso são essenciais, já que a sincronização entre áudio e vídeo é importante. Segundo [NBA02], o atraso deve ser menor que 200 ms e a variação de atraso pode chegar numa média máxima de 110 ms, mas devendo ser minimizada.

Nas aplicações de vídeo, também é utilizada a técnica de bufferização, para aliviar problemas de variação de atraso. Esta técnica implica no aumento do atraso, devendo ser utilizada com critérios para que não venha gerar muito aumento no atraso.

Em aplicações de vídeo *streaming*, onde não se tem a interatividade, o atraso não é importante podendo chegar a até 8 segundos e a variação de atraso pode chegar a 500ms, conforme [MIR02]. Com o uso da bufferização, tanto o atraso quanto a variação de atraso acabam não sendo parâmetros importantes.

Taxa de Perda de Pacotes

Em aplicações de vídeo streaming, como a tolerância a atrasos é alta, é possível implementar técnicas de correção de erros, segundo [MIR02], fazendo com que possa existir também uma pequena tolerância a perda de pacotes. Para sistemas de alta qualidade de vídeo, onde a aplicação exige alta vazão, a perda de pacotes não é bem aceita, e deverá ser no máximo 1%.

Nas aplicações de vídeo interativo, a perda de pacotes é mais bem aceita do que no vídeo streaming, conforme [MIR02], mas até certo limite. Este limite não é bem definido pois depende da tolerância dos usuários com relação à distorção do vídeo, mas um valor típico fica em 2%.

3.2.3 Requisitos para Aplicações Web

As aplicações WEB utilizando o protocolo http são baseadas em requisições e respostas. O cliente http (software browser) faz a requisição para o servidor http, este envia a resposta com o recurso solicitado (que pode ser um arquivo por exemplo), e depois da resposta, o servidor fecha a conexão.

Vazão

Como os acessos WEB são interativos, é importante que as respostas não sejam demoradas. Assim, a vazão deve ser boa, principalmente em transações WEB rápidas. Conforme [MAR00], uma vazão típica fica entre 10 e 500 Kbit/s.

A vazão necessária vai depender do tamanho dos recursos disponibilizados no servidor http, e do tempo desejado para efetuar a transferência dos recursos. Caso os recursos (páginas, arquivos) sejam grandes (vários Mbytes), é indicado que a vazão seja relativamente alta. Como exemplo, pode-se fazer o seguinte cálculo: para transferir um recurso de 400 Kbytes, em 20 segundos, a vazão deve ser de 160 Kbit/s. Assim, é importante que os recursos sejam tão menores quanto possível, pois dessa forma a transação será mais rápida e a vazão necessária será menor.

Atraso e Variação de Atraso

Como os acessos web devem ter respostas rápidas, devido à característica de interatividade, é importante que o atraso seja baixo para permitir um bom tempo de resposta. O atraso pode chegar até a 2 segundos, sem causar nenhuma insatisfação nos usuários.

No caso de transações http mais longas, onde a interatividade não é a característica principal, o atraso pode ser maior. A variação de atraso não tem influência sobre a aplicação de acesso web.

Taxa de perda de pacotes

A perda de pacotes pode ser tolerada, já que as técnicas de correção de erros resolvem o problema. Porém, no caso de altas perdas, a consequência será o aumento do tempo de resposta pois muitas retransmissões serão efetuadas pelo sistema de correção de erros, até que os dados corretos cheguem ao destino.

3.3 Serviços Diferenciados

Para implementar qualidade de serviço numa rede deve-se considerar que é um mecanismo fim-a-fim de garantia de entrega de informações. Assim, a implementação de QoS na rede implica em atuar nos equipamentos envolvidos na comunicação fim-a-fim visando o controle dos parâmetros afim.

O IETF tem proposto vários modelos de serviço e mecanismos para satisfazer a necessidade de QoS na Internet, proporcionando um melhor controle sobre o tráfego, na forma de priorização de certas aplicações (com certas restrições temporais) em detrimento do restante (tráfego essencialmente melhor esforço). Entre estes trabalhos estão o modelo Serviços Integrados/RSVP [BRA94], o modelo Serviços Diferenciados [BLA98] e o MPLS (*MultiProtocol Label Switching*) [ROS01].

A solução MPLS (*MultiProtocol Label Switching*) tem uma certa relação com a questão da qualidade de serviço em redes, pois um dos ganhos mais importantes com sua utilização é a simplificação da função de roteamento, reduzindo o overhead e as latências nos roteadores. É uma solução mais orientada para uma engenharia de tráfego de pacotes na rede do que para qualidade de serviço.

O funcionamento é baseado na marcação dos pacotes com um rótulo (label MPLS) de 20 bits, nos roteadores de entrada da rede. Os roteadores internos à rede verificam esses rótulos que indicam o próximo salto, e assim, sucessivamente encaminham o pacote até o destino de uma forma bastante rápida. Na saída da rede, o rótulo é retirado. Com os rótulos é possível estabelecer um caminho determinado de forma a utilizar as melhores rotas, menos congestionadas, e obter um desempenho melhor para o fluxo de dados.

Para a operação efetiva do MPLS faz-se necessário a distribuição dos rótulos entre os roteadores e a gerência dos mesmos. O protocolo LDP (*Label Distribution Protocol*) é um protocolo de sinalização desenvolvido com esta finalidade.

O MPLS consegue melhorar as condições de operação na rede e isso leva a uma melhor qualidade de serviço. Entretanto, o MPLS não provê controles específicos quanto à garantia de QoS na rede e não é controlável pela aplicação.

O modelo Serviços Integrados (IntServ - Integrated Services) visa a implantação de uma infra-estrutura para a Internet que possa suportar o transporte de áudio, vídeo e dados em tempo real além do tráfego de dados atual. QoS na arquitetura IntServ é garantida através de mecanismos de reserva de recursos na rede, onde a aplicação deve reservar os recursos que deseja utilizar antes de iniciar o envio dos dados, usando o protocolo RSVP (*ReSerVation Protocol*).

O RSVP é um protocolo de sinalização que atua sobre o tráfego de pacotes IP numa rede (Internet, redes privadas, etc). É um protocolo eficiente do ponto de vista da qualidade de serviço (QoS) na medida em que provê granularidade e controle fino das solicitações feitas pelas aplicações. Sua maior desvantagem é a complexidade inerente à sua operação nos roteadores que, eventualmente, pode gerar problemas nos backbones de grandes redes, como das Operadoras.

Uma questão ser observada para definir mecanismos de qualidade de serviço, é que a necessidade de garantir QoS se mostra mais importante nos períodos de pico de tráfego, quando a rede enfrenta uma situação de congestionamento ou de carga muito elevada. Nesta situação os mecanismos de QoS devem buscar soluções para decisões do tipo:

- como alocar os escassos recursos, como banda de transmissão ;
- como selecionar o tráfego de pacotes ;
- como priorizar os pacotes ;
- como descartar pacotes, ou seja, quais e quando descartá-los.

Aderente a essas necessidades, o modelo Serviços Diferenciados (*DiffServ - Differentiated Services*) foi introduzido pelo IETF devido às limitações no modelo *IntServ*, e se mostra como uma excelente alternativa já sendo utilizado para implementar QoS na Internet2 [MIR02]. No DiffServ os pacotes são marcados diferentemente, na entrada da rede, para criar classes de pacotes. As classes recebem serviços, ou

tratamentos diferenciados na rede, conforme definições pré-estabelecidas. Serviços Diferenciados é essencialmente um esquema de priorização de pacotes.

Um serviço pode ser caracterizado por alguns aspectos significativos relacionados à transmissão dos pacotes, de acordo com parâmetros como vazão, atraso, variação de atraso ou perda de pacotes. Além de DiffServ atender aos requisitos de QoS de aplicações heterogêneas, e diferentes expectativas de usuários, também permite uma tarifação diferenciada para os serviços providos pelas Operadoras.

Nas próximas seções será mais detalhado o DiffServ em sua arquitetura, funcionamento e componentes.

3.3.1 Arquitetura DiffServ

A arquitetura dos Serviços Diferenciados é apresentada em [BLA98], propondo diferenciação de serviços com base na especificação e marcação do campo DS. Os Serviços Diferenciados são oferecidos no interior de um domínio DS (Differentiated Services), que é composto por um conjunto de nós compatíveis com a proposta DiffServ, compartilham uma mesma política de provisão de serviços e implementam um conjunto de comportamentos de encaminhamento.

Os serviços oferecidos por um domínio DS são todos para tráfego unidirecional e para tráfegos agregados, não fluxos individuais. Os tráfegos agregados também são chamados de comportamentos agregados (Behavior Aggregate – BA), e são definidos como conjuntos de pacotes que tem a mesma marcação no campo DS, e que atravessam um caminho na mesma direção. Esse pacotes podem vir de origens ou aplicações distintas.

A figura 3.1 representa um domínio DS, com nós de fronteira ou borda (roteadores 1,4 e 5) onde os fluxos são agregados (BA), e estes são encaminhados pelos nós internos (roteadores 2 e 3) para o destino. Os nós (ou nodos) de borda assumem, dependendo do sentido do tráfego tratado, o papel de nós de ingresso ou egresso do domínio. Portanto, na figura 3.1 o nó 1 representa um roteador de ingresso enquanto o nó 4 representa um roteador de egresso para os fluxos simbolizados. Na prática um

domínio DS pode ser uma rede de uma empresa, universidade, operadora de telecomunicações, etc.

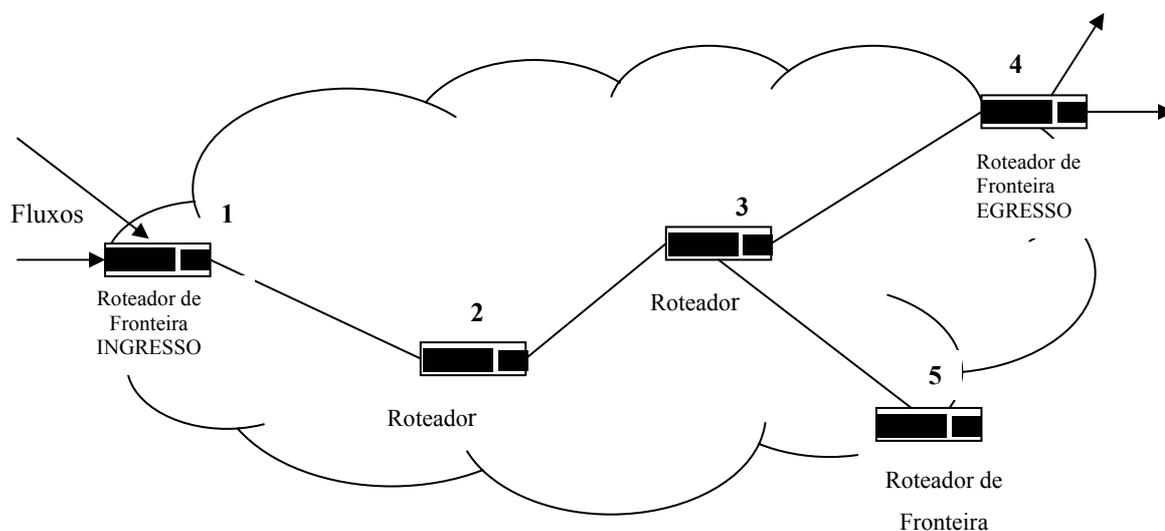


FIGURA 3.1 – DOMÍNIO DS

Nos roteadores de borda, que realizam a conexão da rede do usuário com o domínio DS, reside a maior complexidade do modelo DiffServ. Eles devem realizar as funções como classificação e condicionamento dos pacotes de cada fluxo entrante no domínio. Esses pacotes são transmitidos aos roteadores internos, na forma de fluxos agregados, ou BA's. O tráfego e processamento nos roteadores internos normalmente é maior, e visando não carregá-los ainda mais, no modelo DiffServ eles devem apenas encaminhar o BA conforme o PHB definido para o serviço.

Domínios DS também podem ser concatenados, ou seja, domínios diferentes podem estar ligados entre si. Também podem estar conectados a domínios não DS, tomando como exemplo uma conexão de uma rede DS de uma Operadora com uma rede não DS de um Cliente.

Os serviços entre os domínios são definidos por um contrato, chamado SLA (Service Level Agreement), que especifica a forma de encaminhamento e garantias de prestação do serviço através do domínio DS. O SLS (Service Level Agreement), que faz parte do SLA, define garantias de parâmetros de QoS e inclui os perfis e as regras de condicionamento de tráfego que devem ser aplicadas aos fluxos de dados.

Na figura 3.2 é apresentada a arquitetura geral de domínios DS, onde podemos ver nós de borda (ingresso e egresso), nós internos (interiores), domínios DS, e onde é aplicado o SLA e SLS.

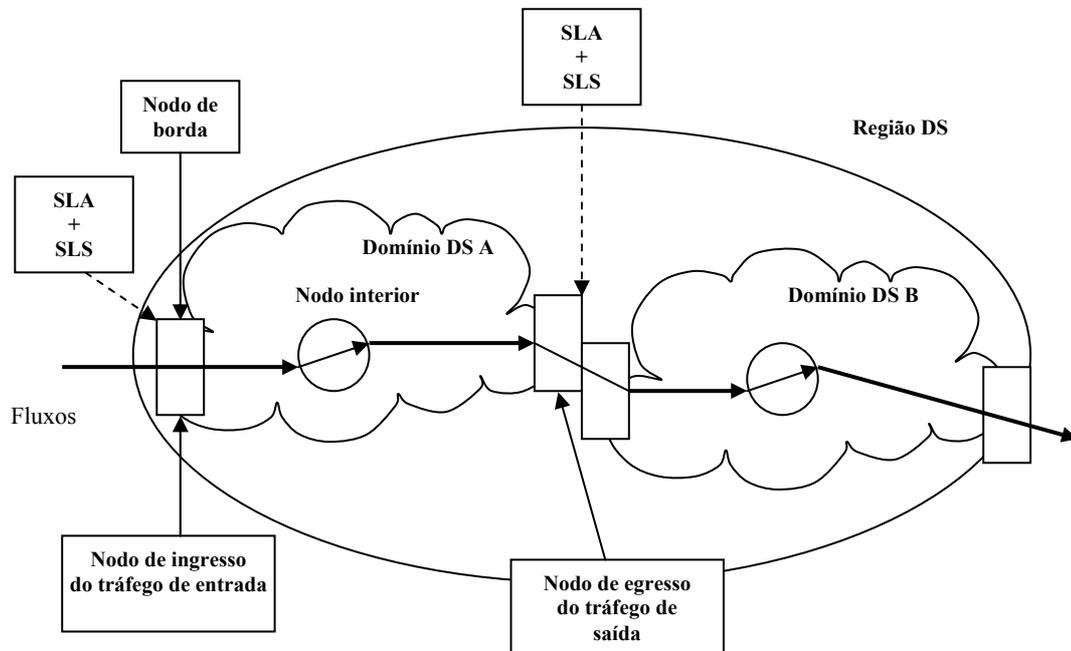


FIGURA 3.2 – ARQUITETURA GERAL DE DOMÍNIOS DS

3.3.2 Mecanismos para Qualidade de Serviços

Para que seja possível implementar QoS nas redes IP é necessário o uso de diversos mecanismos nos roteadores. Estes mecanismos podem e devem ser usados para prover as facilidades de tratamento para dos pacotes conforme sugerido no modelo DiffServ. Também podem ser usados para objetivos mais abrangentes.

A figura 3.3 mostra os mecanismos básicos que devem estar presentes num roteador compatível com DiffServ. Em seguida são apresentados esses mecanismos e também os mecanismos de diferenciação de serviços e prevenção de congestionamento.

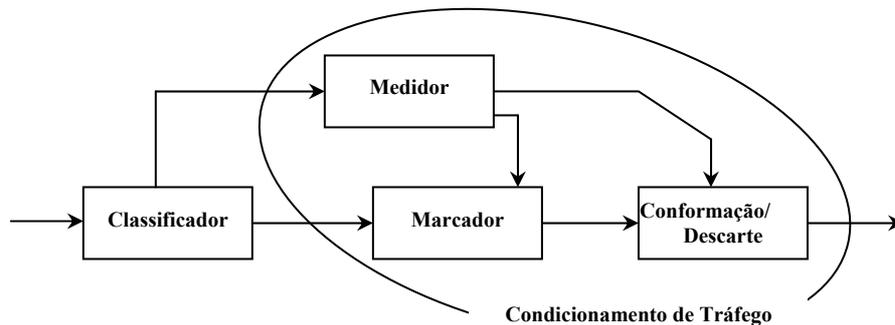


FIGURA 3.3 – MECANISMOS BÁSICOS PARA IMPLEMENTAR DIFFSERV

Classificação de pacotes

É necessário classificar os pacotes e depois condicioná-los para poder realizar o tratamento de QoS. O classificador de pacotes seleciona os pacotes de um fluxo de tráfego baseado no conteúdo de campos do cabeçalho destes pacotes. DiffServ define dois tipos de classificadores de pacotes:

- **Comportamento Agregado (BA)** - seleciona pacotes através do valor do DSCP.
- **Multi-campo (Multi Field – MF)** - seleciona pacotes baseado na combinação de valores de um ou mais campos do cabeçalho, conforme [NIC98]. Estes campos podem ser os endereços IP de origem e/ou destino, campo DS, protocolo e porta de origem e/ou destino do segmento TCP ou UDP. Também podem ser usadas outras informações para fazer a classificação como a interface física de entrada do fluxo.

A realização da classificação de pacotes de forma eficiente e consistente é o primeiro passo para uma implementação correta de DiffServ, e é uma questão ainda bastante discutida e pesquisada. Mais recentemente, com o advento das aplicações de voz e vídeo que usam o protocolo RTP com alocação dinâmica de portas, tornou-se necessário o desenvolvimento de técnicas mais avançadas para definir se o pacote RTP é de áudio ou vídeo. Alguns fabricantes já disponibilizam essa facilidade em seus roteadores. O fabricante CISCO Systems fornece o NBAR (*Network Based Application*

Recognition) que permite fazer a classificação de pacotes baseado no protocolo que está sendo usado pela aplicação, podendo verificar informações das camadas 4 até 7.

Condicionamento de Tráfego

O condicionamento de tráfego está geralmente localizado nos nodos de borda do domínio DS, mas também podem estar em nós interiores. O condicionador de tráfego pode conter todos os blocos descritos a seguir, ou apenas alguns:

- **Medidor:** faz a medida do fluxo para determinar se um pacote está dentro ou fora do perfil de tráfego contratado no SLS. Esta informação é passada aos outros blocos para que ações distintas de condicionamento sejam aplicadas aos pacotes dentro do perfil e pacotes fora do perfil;
- **Marcador:** marca ou remarca o campo DS do cabeçalho do pacote para um valor DSCP específico, tornando-o pertencente a um determinado BA;
- **Conformador:** também chamado de suavizador de tráfego (*traffic shaping*) pois sua função é atrasar um ou mais pacotes, de um fluxo fora do perfil, até que estes estejam em conformidade com o perfil contratado, descrito no SLS, e possam ser encaminhados pela rede
- **Descartador:** também chamado de policiador pois sua função é descartar pacotes fora do perfil, tomando por base regras descritas no SLS,

Os pacotes fora de perfil podem ser tratados de formas distintas, podendo ser descartados ou remarcados para uma classe menor através da função de policiamento do tráfego, ou conformados/suavizados até que fiquem dentro do perfil. Na prática, as funções de medição e conformação ou descarte são implementadas, principalmente, através do método *Token Bucket* (Balde de Fichas) e sua variante *Leaky Bucket* (Balde Furado ou Gotejador). Esse método presume a existência de um recipiente, ou balde, com capacidade para “b” tokens (fichas), e que será esvaziado a uma taxa constante “r”. Essa taxa é implementada através de um gerador de tokens.

O método *Leaky Bucket* é usado para limitar em determinado valor um tráfego entrante qualquer. Neste caso, a capacidade do balde comporta somente um

único token, que será usado para a passagem de um pacote pelo regulador. Como a taxa de geração de tokens é “ r ”, a taxa de saída será limitada a esse valor. Isto pode ser visto na figura 3.4.

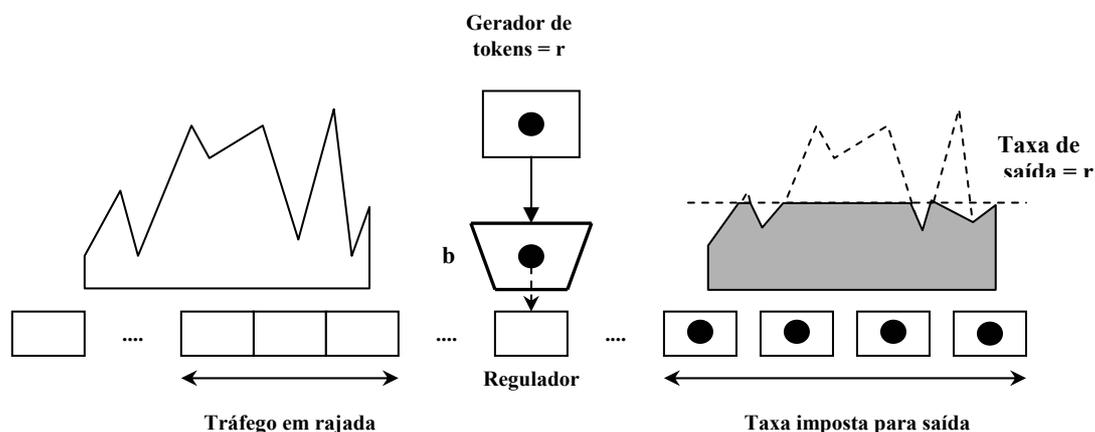


FIGURA 3.4 – LEAKY BUCKET OU LIMITADOR DE TRÁFEGO

O método *Token Bucket*, mostrado na figura 3.5, é usado para permitir tráfegos em rajada, mas de forma controlada. O controle é feito através da capacidade “ b ” estabelecida para o balde, em bytes. Este método acumula créditos (tokens) no balde para admitir os pacotes, e enquanto houverem créditos o tráfego entrante será aceito, mesmo que seja acima da taxa média “ r ” (rajada). Mais detalhes sobre esse método podem ser consultados em [COM02] e [CON02].

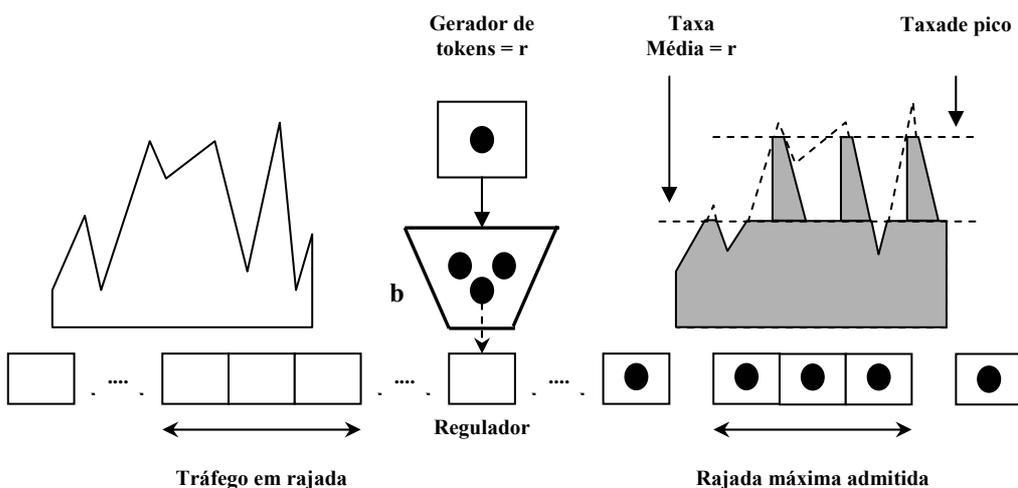


FIGURA 3.5 – TOKEN BUCKET OU CONFORMADOR DE RAJADAS

A utilização de dois Token Buckets, em série, proporciona uma configuração de duas taxas de rajada, uma considerada normal e uma considerada como excesso, que também poderá ser transmitida caso existam recursos. Acima deste limite os pacotes serão descartados ou remarcados (função de policiamento). Com a aplicação destes mecanismos pode-se condicionar o comportamento do tráfego para uma grande variedade de perfis.

Quando o token bucket é usado para função de conformação de tráfego, ele faz uso de um buffer de dados, onde os pacotes são armazenados, e atrasados, até que estejam dentro do perfil. Dessa forma, pacotes em rajada poderão ser transmitidos, gerando um tráfego mais constante e conformado dentro da taxa contratada. Para a função de policiamento de tráfego não existe o buffer, pois os pacotes fora de perfil são descartados os remarcados para classes inferiores.

Mecanismos de diferenciação de serviços

Estes mecanismos proporcionam facilidades sobre as quais pode ser implementada a arquitetura necessária para obter o comportamento desejado das classes de serviços, identificadas pelos PHB's. Também durante momentos de congestionamentos na rede, estes mecanismos tentam garantir os parâmetros de QoS das classes. Esses mecanismos são compostos por dois elementos fundamentais:

- **Estrutura de filas** – é a utilização de várias filas para armazenamento dos pacotes e posterior reenvio. As filas são tratadas com prioridades distintas, possibilitando que os pacotes armazenados sejam encaminhados conforme o PHB desejado. A política de prioridades é a base do modelo DiffServ para obter a diferenciação nos serviços.
- **Escalonamento de pacotes** – mecanismo responsável, juntamente com a estrutura de filas, em colocar os pacotes na fila de saída e garantir que cada BA obtenha a parte dos recursos que foi prometida (contratada). A capacidade do canal de transmissão é dividida entre os BA's conforme definido pelos serviços contratados. Qualquer recurso disponível deverá ser distribuído de forma justa.

Na prática, é muito usado o método CBWFQ (Class-Based Weighted Fair Queueing) para obter filas com prioridades diferenciadas conforme a necessidade de tráfego das classes de serviço, segundo [CON02]. Para obter uma fila com prioridade total sobre as outras, é usado o método PQ (Priority Queueing). A utilização desses elementos possibilita a implementação dos PHB's EF e AF do modelo DiffServ que serão apresentados na seção 3.3.4.

Prevenção de congestionamento

Essa técnica monitora as cargas de tráfego da rede com o objetivo de prever e prevenir congestionamentos. Baseia-se no controle dos fluxos de pacotes nas filas dos roteadores. Normalmente, sem a prevenção as filas tendem a encher até o limite, e os roteadores começam a descartar esses pacotes. O protocolo TCP verifica isso, e reduz a taxa de transmissão para diminuir o congestionamento, mas nem sempre é suficiente.

O mecanismo RED (Random Early Detection) previne congestionamentos através do descarte aleatório de pacotes antes da fila transbordar. Com estes descartes, o TCP reduz a taxa de transmissão e o congestionamento é evitado antes de acontecer. O início do descarte aleatório é programável no equipamento que está implementando o RED. Conforme [CON02], a Cisco Systems adota uma variante chamada WRED (Weighted Random Early Detection), onde o descarte é iniciado em classes de tráfego com menor prioridade, contribuindo para o objetivo da arquitetura DiffServ.

Aplicação dos mecanismos para qualidade de serviços

Todos os mecanismos apresentados possuem grande importância para a implementação de DiffServ, e são usados nos roteadores da seguinte forma:

- **Roteadores de ingresso** – efetuam classificação de pacotes tipo MF, em geral, e condicionamento de tráfego nos fluxos entrantes. Também podem aplicar o PHB

apropriado, baseado no processo de classificação, através de mecanismos de diferenciação de serviços.

- **Roteadores internos** – efetuam classificação tipo BA e aplicam o PHB apropriado através dos mecanismos de diferenciação de serviços. Também devem fazer a prevenção de congestionamentos.
- **Roteadores de egresso** – normalmente efetuam conformação do tráfego nos pacotes de saída . Também pode fazer a classificação MF ou BA e remarcação de pacotes caso exista um SLA com o domínio de saída.

3.3.3 Campo DS

No DiffServ, os pacotes são marcados diferentemente para criar várias classes de pacotes, que serão tratadas de forma diferenciada através dos serviços disponíveis na rede. Os serviços são oferecidos conforme o BA, que é identificado através do DSCP (Differentiated Services CodePoint).

O DSCP está contido no campo DS (Differentiated Service) definido na RFC-2474 [NIC98], conforme a figura 3.1. Esta definição substitui a definição do IPv4, onde o campo se chama *Type of Service*, ou TOS, e substitui a definição do IPv6, onde o campo se chama *Traffic Class*.

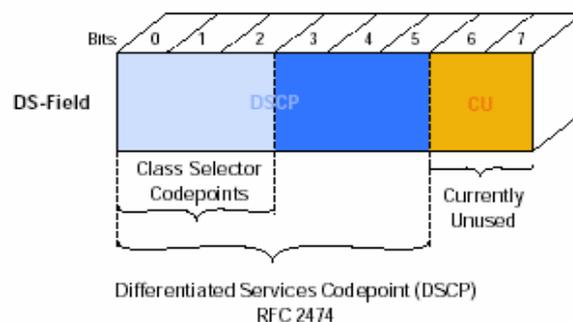


FIGURA 3.1 – ESTRUTURA DO CAMPO DS

O campo DS é formado por 6 bits (DSCP) usados para definir o comportamento por salto, ou PHB (*Per-Hop Behavior*), que o pacote sofrerá em cada

nó. Este campo é tratado como um índice de uma tabela que é usada para selecionar um mecanismo de manipulação de pacotes implementado em cada dispositivo. Este campo é definido como um campo não estruturado para facilitar a definição de futuros PHBs, suportando até 64 valores diferentes, ou *codepoints*. O campo DS também possui dois bits CU (Currently Unused) reservados para o futuro.

Um nodo da rede (roteador) que seja compatível com *DiffServ* deverá usar todos os 6 bits do campo DSCP, e nenhum outro, para selecionar um PHB, fornecendo um serviço apropriado ao BA.

Com o objetivo de permitir o gerenciamento de *codepoints*, a RFC 2474 [NIC98] sugere três conjuntos, com a seguinte definição:

- “xxxxx0” – conjunto de 32 valores que será usado com o propósito de padronização;
- “xxxx11” – conjunto de 16 valores com propósitos experimentais e uso local;
- “xxxx01” – conjunto de 16 valores que serão inicialmente usados para uso experimental e local, e poderão no futuro vir a ser usados para padronização.

3.3.4 PHB

PHB é o comportamento de encaminhamento observável externamente aplicado a um BA por um nó (roteador) suportando *DiffServ*. Num mesmo BA podem existir pacotes vindos de várias origens ou aplicações distintas.

Em outras palavras, um PHB refere-se ao comportamento de um nó com relação à marcação, enfileiramento, policiamento e condicionamento de pacotes pertencentes a um BA, de acordo com a configuração prévia do serviço a ser oferecido pelo domínio DS. É comum que vários PHB's sejam implementados num nó, permitindo a utilização eficaz dos recursos do nó e de suas conexões, sendo extensível a todo o domínio DS.

Existem quatro padrões de PHB disponíveis, e são apresentados a seguir.

PHB Default (DSCP = 000000)

Definido na RFC 2474 [NIC98], especifica que um pacote marcado com este valor de DSCP, receberá do nó *DiffServ* o serviço tradicional de melhor esforço (Best Effort – BE). Se um pacote chegar a um nó *DiffServ* com seu DSCP não mapeado em nenhum PHB, este pacote será mapeado para o PHB Default.

PHB Class-Selector (DSCP=xxx000)

Definido na RFC 2474 [NIC98], este PHB foi criado para preservar compatibilidade com redes que ainda empregam o esquema de precedência IP, definida na RFC 791. O valor de DSCP na forma xxx000, pode ter valores de x como 1 ou 0. Estes valores são chamados *Class-Selector CodePoints*. O *codepoint* 0000 do PHB Default também é chamado de *Class-Selector CodePoint*.

Estes PHB's mantêm quase o mesmo comportamento que nós implementando classificação e encaminhamento baseados em precedência IP. Por exemplo, pacotes com DSCP=100000 (precedência IP 100 = primeiros três bits) devem ter tratamento de encaminhamento preferencial quando comparados a pacotes com DSCP=011000 (precedência IP 011).

PHB EF – Expedited Forwarding (DSCP=101110)

Definido na RFC 2598 [JAC99], o PHB EF é um ingrediente chave para *DiffServ*, e tem o objetivo de habilitar serviços fim-a-fim com baixo atraso, baixa variação de atraso, baixa perda de pacotes e vazão garantida. Aplicações como VoIP e vídeo requerem da rede um tratamento robusto, e o EF é indicado para estes casos. Este PHB pode ser implementado através de algumas técnicas distintas, segundo [JAC99]

Para eficiência máxima, o PHB EF deve ser reservado para as aplicações mais críticas, pois em momentos de congestionamento não é possível tratar o tráfego todo, ou na maior parte, com alta prioridade. O BA referente ao tráfego EF deve receber

a vazão estipulada independentemente da intensidade de qualquer outro tráfego existente, objetivando diminuir a permanência dos pacotes em filas no interior dos nós, diminuindo o atraso e sua variação. Os roteadores de entrada na rede DiffServ, devem policiar e condicionar o tráfego EF de forma a assegurar que a taxa de chegada aos nós interiores esteja em conformidade com a taxa contratada, não permitindo que taxas maiores entrem na rede.

PHB AF – Assured Forwarding

Definido na RFC 2597 [HEI99], o PHB AF foi sugerido para aplicações que necessitam de confiabilidade e prioridade maior no tráfego de dados com relação ao que é oferecido pelo serviço melhor esforço (BE). O encaminhamento assegurado (AF) é composto por um grupo de PHB's que criam uma hierarquia de classes de serviço. A expectativa é obter o perfil das classes mesmo durante períodos de congestionamento.

Cada classe deverá exibir um comportamento diferente quanto ao método de encaminhamento, independente do comportamento de outras classes, sendo que cada uma deverá apresentar níveis de precedência quanto à decisão de descarte de pacotes. Atualmente existem quatro classe definidas, com três níveis de precedência caracterizada em baixa, média e alta (maior probabilidade de descarte). A tabela 3.5 mostra os 12 valores de DSCP para as classes AF. Em caso de congestionamento no nó, a precedência determina a importância de cada pacote e influi no comportamento de descarte. Para manter compatibilidade com a precedência IP (RFC 791), é aconselhável que o serviço mais prioritário seja mapeado na classe AF4, e o menos prioritário na classe AF1.

TABELA 3.5 – VALORES DO DSCP PARA CADA PHB AF

<i>Descarte</i>	<i>Classe 1</i>	<i>Classe 2</i>	<i>Classe 3</i>	<i>Classe 4</i>
<i>Precedência Baixa</i>	001010 (AF11)	010010 (AF21)	011010 (AF31)	100010 (AF41)
<i>Precedência Média</i>	001100 (AF12)	010100 (AF22)	011100 (AF32)	100100 (AF42)
<i>Precedência Alta</i>	001110 (AF13)	010110 (AF23)	011110 (AF33)	100110 (AF43)

Através do AF, a Operadora pode oferecer a seus clientes serviços com níveis diferenciados de QoS, porém, com menor garantia do que o serviço *Premium* (PHB EF). Como os serviços são diferenciados, os preços também podem ser (mais garantia, implica em maior preço). O serviço prestado depende da quantidade de recursos alocados para a classe AF do serviço, da carga atual desta classe e, em caso de congestionamento, do valor de precedência de descarte. Para cada classe AF deve estar alocado um mínimo configurável de recursos (buffer e largura de banda).

A garantia no modelo AF é que os pacotes marcados serão entregues com alta probabilidade enquanto o tráfego BA não exceder a taxa contratada definida no perfil de tráfego do usuário. No entanto, é permitido ao usuário exceder as taxas contratadas, mas ele deve estar ciente de que o tráfego excedente não terá a mesma probabilidade de entrega.

Um perfil associado a cada tráfego define o serviço esperado. Assim, um mecanismo de condicionamento de tráfego atua nas fronteiras do domínio DS, de forma a marcar os pacotes de acordo com sua conformidade ao perfil. Os roteadores de fronteira são responsáveis por manter a granularidade dos serviços.

Não é necessário implementar todos os níveis de precedência de descarte em cada classe AF, mas pelo menos dois. Isso fica a critério de cada empresa que venha a implementar o modelo AF, dependendo da quantidade de congestionamento esperada.

O PHB AF detecta e reage a congestionamentos de longa duração dentro das classes através do descarte de pacotes e também da remarcação de precedência. Também aceita em sua fila pacotes que causam congestionamento de curta duração. Para realizar essas tarefas deve ser utilizado um mecanismo para prevenção de congestionamento, como RED visto anteriormente.

3.3.5 Classes de Serviço

O modelo DiffServ não especifica os serviços ou classe de serviços a serem prestados por um provedor. É de responsabilidade deste (uma operadora de

telecomunicações, por exemplo) decidir que classes de serviços vai disponibilizar a seus clientes. A seguir, alguns exemplos de serviços que poderiam ser oferecidos:

- **Serviço Premium** - para aplicações que necessitam de baixo atraso e baixa variação de atraso. Neste caso, o usuário pode negociar com o provedor a máxima largura de banda para enviar pacotes através da rede e as alocações são feitas em termos de taxa de pico. Uma desvantagem é o fraco suporte a tráfegos em rajada e o fato de que o usuário paga mesmo quando não usa completamente a largura de banda. Este serviço é implementado usando o PHB EF, em geral.
- **Serviço Assegurado** - para aplicações que necessitam mais confiabilidade que o serviço de melhor esforço. Este serviço não garante a largura de banda como o Serviço Premium, mas fornece uma alta probabilidade de que o provedor transfira os pacotes marcados de forma bastante confiável. É implementado usando PHB AF.
- **Serviço Olympic** – este serviço pode fornecer algumas classes de serviços. Por exemplo, três classes chamadas de Ouro, Prata e Bronze. Os pacotes da classe Ouro têm um encaminhamento mais rápido e garantido do que os pacotes Prata, e estes do que os Bronze. As necessidades dos clientes são moldadas dentro destas classes de serviço, obtendo a diferenciação necessária. Dentro de cada classe ainda podem existir níveis de precedência para descarte de pacotes, em casos de congestionamentos. É implementado através do PHB AF.

Os serviços prestados podem ser categorizados como qualitativos ou quantitativos, dependendo de como são definidos e apresentados os parâmetros de qualidade. Os serviços qualitativos oferecem garantias relativas que somente podem ser avaliadas por comparação. Exemplo:

- o tráfego oferecido no nível de serviço “A” será entregue com baixo atraso;
- o tráfego oferecido no nível de serviço “B” terá baixa perda de pacotes.

Os serviços quantitativos oferecem garantias concretas que podem ser avaliadas por medições convenientes, independentes de outros serviços, por exemplo:

- 90 % do tráfego dentro do perfil para o serviço “C”, terá menos que 50 ms de atraso;

- 95 % do tráfego dentro do perfil para o serviço “D”, será efetivamente entregue no destino (perda de pacotes igual a zero).

Também pode ser feita uma quantificação relativa, por exemplo:

- o tráfego oferecido no nível de serviço “E” terá três vezes a largura de banda que o tráfego oferecido no nível de serviço “F”.

A definição dos serviços, ou classes de serviços, que serão prestados pelo provedor ficam a seu critério, mas de uma forma geral pode-se dizer que um portfólio completo inclui serviços para voz, vídeo e dados com alguns níveis de prioridade. Segundo [RET02], além do melhor esforço poderiam ser disponibilizadas cinco classes de serviço como:

- Classe Voz – serviço otimizado para aplicações em tempo real, muito sensíveis ao tempo, como VoIP e sistemas de controle. Classe implementada com PHB EF
- Classe Vídeo Interativo – serviço para aplicações em tempo real de vídeo, como videoconferência. Classe implementada com PHB AF41.
- Classe Interativa – serviço para aplicações de negócios que exijam rápido tempo de resposta, consideradas críticas. Classe implementada com PHB AF31.
- Classe Negócios – serviço adequado para aplicações transacionais, comuns de negócios, como tráfego cliente/servidor e web corporativa. Classe implementada com PHB AF21.
- Classe Geral – serviço de mais baixo custo, para tráfego de negócios não críticos como e-mail e streaming de vídeo, como vídeo on demand ou e-learning por exemplo. Classe implementada com PHB AF11.

É certo que a prioridade desejada para as aplicações varia conforme as necessidades de cada cliente, portanto os exemplos acima não são fixos.

3.3.6 Acordo de Nível de Serviço (SLA)

Afim de que os clientes recebam serviços com garantias de QoS de seus provedores, ou operadoras de telecomunicações, deve ser firmado um Acordo de Nível de Serviço (SLA – Service Level Agreement). Este acordo é formado por vários itens administrativos e comerciais, além dos detalhes técnicos que aparecem na Especificação do Nível de Serviço (SLS – Service Level Specification). O SLS é que especifica as classes de serviço suportados e o perfil de tráfego permitido para cada classe..Alguns dos parâmetros mais comumente utilizados na definição do SLS são a vazão (banda), atraso (latência), variação do atraso, taxa de perda de pacotes e disponibilidade.

A disponibilidade é uma medida da garantia de execução da aplicação ao longo do tempo. As empresas dependem cada vez mais da Internet para a viabilização de seus negócios (comércio eletrônico, home-banking, atendimento online , etc) e, neste sentido, a disponibilidade é um requisito importante. A título de exemplo, requisitos de disponibilidade acima de 99% do tempo são comuns para QoS em aplicações WEB, aplicações cliente/ servidor e aplicações de forte interação com o cliente.

Mas os parâmetros que mais chamam atenção são os de qualidade de serviço como vazão, atraso, variação do atraso e taxa de perdas que vão definir quais aplicações o cliente poderá utilizar com garantias de funcionamento. Mais detalhes sobre SLA e SLS serão vistos no capítulo 4.

3.3.7 Visão Geral do Funcionamento

No modelo DiffServ a base de funcionamento é a marcação do valor DSCP no campo DS do cabeçalho IP, indicando o tipo de serviço que deverá ser provido para este pacote. Essa marcação pode ser feita pelo roteador que liga o cliente à rede do ISP (roteador de borda) baseado na classificação MF. Outra alternativa, não considerada neste trabalho, é a marcação feita diretamente pela aplicação do cliente

No ingresso à rede DS os pacotes são classificados e é feito um condicionamento de tráfego para torna-los conformes ao perfil de tráfego previamente

contratado. As regras de classificação e condicionamento, assim como montante de espaço de bufferização e largura de banda nos roteadores são derivadas do SLA, mais especificamente do SLS, de onde se constrói todo o perfil de tráfego do cliente. Um exemplo simples de perfil de tráfego poderia ser:

- medir o fluxo de pacotes do endereço IP a.b.c.d., e se sua taxa ficar abaixo de 384 Kbit/s, setar o campo DS para o valor X, senão setar o campo DS para o valor Y. Se a taxa exceder 512 Kbit/s, descartar os pacotes excedentes.

Os perfis são configurados pelo provedor de forma manual ou via sistema de sinalização, e são baseados no SLA/SLS. Como os perfis são dependentes das aplicações e necessidades do cliente, é fora do escopo da padronização DiffServ.

Dentro da rede DS, nos roteadores internos ao domínio, o campo DS é usado para determinar como os pacotes são tratados. O tratamento, também chamado de PHB ou comportamento por salto, pode incluir diferentes prioridades envolvendo atraso de enfileiramento (escalonamento), diferentes prioridades na decisão de descarte na sobrecarga de filas (gerenciamento de filas), seleção de rota, liberação de largura de banda, etc.

De sessenta e quatro possíveis significados, dados pelos bits do DSCP, o grupo de trabalho do DiffServ padronizou apenas alguns PHBs globalmente aplicáveis, e deixou o resto para uso experimental e futuro. Se os experimentos indicarem que um certo PHB não padronizado é claramente útil, ele pode ser padronizado posteriormente, pois o modelo está em constantes debates e atualizações.

No modelo DiffServ, um serviço assegurado também pode ser fornecido por um sistema que suporta parcialmente os Serviços Diferenciados. Desde que pacotes serviço assegurado têm menos probabilidade de serem cortados em roteadores com DS, e roteadores sem DS simplesmente ignoram o campo DS, o desempenho total do tráfego serviço assegurado será melhor que o tráfego melhor esforço, possibilitando implementações gradativas de DiffServ nas redes dos provedores.

3.4 Conclusão

A qualidade de serviços na internet proporciona aos clientes confiabilidade com relação aos serviços prestados por seus provedores. Para implementar QoS existem alguns modelos, sendo que Serviços Diferenciados está despontando como uma ótima alternativa, devido a sua escalabilidade e flexibilidade.

No entanto, existem várias técnicas e mecanismos que devem ser observados não só para a implantação do domínio DS pelos provedores, mas também para o correto mapeamento das aplicações dos clientes nos serviços disponibilizados. Devem ser observados os requisitos técnicos das aplicações como vazão e atraso, entre outros, para garantir o perfeito funcionamento da aplicação do cliente.

O provedor ou operadora de telecomunicações pode desenvolver um conjunto de serviços que abranja a grande maioria, ou a totalidade das aplicações de clientes, como VoIP, videoconferência, web confiável e com desempenho, entre outras. Mas para essas aplicações, onde os clientes utilizam com foco cada vez mais de negócio, é muito importante que exista um acordo que garanta a qualidade do serviço e defina como este serviço será prestado. Este acordo é o SLA e será abordado mais detalhadamente no capítulo 4.

4. ACORDO DE NÍVEL DE SERVIÇO

Com a privatização das Telecomunicações no Brasil houve um grande desenvolvimento no setor e, conseqüentemente, o surgimento de novos serviços para o mercado. O maior crescimento é o de serviços baseados em redes IP, principalmente pela redução de custos que propicia ao cliente, comparando com outras tecnologias. Enquanto antes a oferta de serviços mal atendia a demanda, e a preocupação das operadoras de Telecomunicações era apenas de criar opções de serviços, agora o cenário é bastante diferente. Grande parcela da demanda está atendida, e torna-se de vital importância para as operadoras a manutenção de seus clientes, ou seja, não perdê-los para a concorrência, e prospectar novas oportunidades de negócios.

A manutenção dos clientes passa não somente pelo provimento de serviços tecnologicamente atualizados, mas também pela excelência no serviço prestado. O mercado está exigindo mais confiabilidade dos serviços e melhor atendimento por parte do provedor. Assim, a qualidade torna-se o diferencial na competição entre os provedores, sejam operadoras de telecomunicações ou outros fornecedores de serviços .

Os provedores já perceberam a importância de possuir qualidade em seus serviços, e por isso correm contra o tempo para disponibilizar em suas redes os recursos necessários para obtenção de QoS. Mas é necessário que exista uma forma de explicitar ao cliente o que ele está comprando, definindo o tipo de serviço, quais suas características e garantias de fornecimento, além de outras informações. Com este objetivo está sendo cada vez mais utilizado um documento chamado Acordo de Nível de Serviço (SLA-*Service Level Agreement*). O SLA pode ser visto como um contrato

efetuado entre provedor e cliente, para oferecer as garantias necessárias ao cliente quanto a prestação do serviço adquirido.

Como não existe um padrão para o SLA, mas existem vários trabalhos sendo efetuados com o objetivo de obter uma forma mais unificada para este instrumento, este capítulo pretende apresentá-lo de uma forma geral. A seção 4.1 fala sobre alguns conceitos envolvendo o SLA, enquanto que a seção 4.2 mostra uma possível estrutura e conteúdo para este instrumento

4.1 Conhecendo o SLA

Com o desenvolvimento de aplicações que necessitam de QoS para seu bom funcionamento, e o maior nível de exigência dos clientes, os provedores precisam especificar seus serviços de forma mais clara e formal. Isto é realizado através do SLA, que deve definir claramente parâmetros de qualidade oferecidos pelo serviço, visando fornecer qualidade para as aplicações do cliente, além de esclarecer os direitos e deveres das partes envolvidas.

Para a prestação de um serviço com qualidade garantida, o compromisso entre custo/benefício deve ser avaliado pelos clientes, pois quanto maior o benefício (ou garantias oferecidas pelo provedor) maior também será o preço do serviço.

Não existe uma definição única para SLA, sendo que abaixo são apresentadas duas possíveis definições:

- SLA é um contrato de serviço entre um provedor de serviços e seu cliente, que define as responsabilidades do provedor em termos de parâmetros da rede e tempos de disponibilidade, métodos de medidas, penalidades caso os níveis de serviço não sejam alcançados ou se os níveis de tráfego são ultrapassados pelo cliente, e todos os custos envolvidos. Esta é a definição conforme o QoS Fórum [STA99].
- SLA é um contrato de serviço entre um cliente e um provedor de serviço que especifica o serviço de encaminhamento que o cliente deve receber. O cliente pode

ser uma organização usuária (domínio fonte) ou outro domínio DS. Esta é a definição conforme [BLA98].

Existem dois tipos de SLA, o interno e externo. O interno é usado entre departamentos de uma mesma empresa, ou dentro de um mesmo departamento, sendo que tem como objetivos criar e estimular a cultura de cliente interno (foco no cliente) e estabelecer metas da área visando aumento de qualidade ou produtividade. Ele não precisa ser um documento independente, mas pode fazer parte de planos estratégicos e de desenvolvimento da empresa. Já o SLA externo é um acordo entre empresas distintas, sendo negociado entre as partes, e constitui um documento legal (contrato) que pode e deve ser utilizado em casos de não cumprimento de cláusulas.

Alguns objetivos do SLA externo são:

- oficializar as expectativas com relação ao serviço contratado;
- facilitar a negociação do padrão de funcionamento versus condições de funcionamento do serviço;
- propiciar segurança e conforto aos clientes e também ao provedor.

A adoção de SLA oferece várias vantagens para as partes. Para os provedores pode-se citar algumas vantagens:

- é uma forma para diferenciar seus serviços da concorrência;
- assegura que o cliente tenham uma clara compreensão sobre a qualidade do serviço, evitando expectativas errôneas;
- assegura que o cliente tenha conhecimento formal dos custos e penalidades;
- definição sobre utilização de recursos pelos clientes;
- facilita o diagnóstico e resolução de possíveis problemas;
- clientes mais satisfeitos e fidelizados.

A seguir algumas vantagens para os clientes:

- é um mecanismo para comparar a qualidade dos serviços entre provedores e avaliar preço/benefício;
- maiores garantias que contribuem para viabilizar o *outsourcing*, ou seja, terceirização de operações, e conseqüente redução de custos;
- possibilita auditoria de desempenho e atribuição de responsabilidades;
- entendimento único das garantias envolvidas;
- garantia da prestação do serviço adquirido, ou ressarcimento em caso contrário.

Assim, além da definição de informações administrativas, legais, e econômicas, o documento SLA também é composto por definições de parâmetros técnicos para obter o desempenho desejado, e esta parte técnica do acordo é chamada Especificação do Nível de Serviço (SLS – *Service Level Specification*). Segundo [GRO02], SLS pode ser definido como sendo um conjunto de parâmetros e seus valores que juntos definem o serviço oferecido a um fluxo de tráfego por um domínio DS.

Para que o provedor e o cliente possam comprovar se o SLA está sendo realmente cumprido, são necessárias ferramentas e metodologias para medir o desempenho do serviço prestado. Este é o objetivo da Gerência de Nível de Serviço (SLM – *Service Level Management*). [STU00] define SLM como uma metodologia e procedimentos pró-ativos, e disciplinados, usados para garantir que níveis adequados de serviços sejam entregues aos usuários de acordo com prioridades comerciais e a um custo acessível.

Visando maior aprofundamento na definição e utilização do SLA e do SLS, será apresentada no próximo item uma estrutura para esses documentos, com os parâmetros mais utilizados e seus significados.

4.2 Estrutura do SLA/SLS

Conforme comentado anteriormente, não existe um padrão para a estrutura de um SLA, mas existem vários esforços de entidades e grupos a nível mundial como TEQUILA em [GOD01] e [TRI01], AQUILA em [SAL00], CADENUS em [CADEN], e SEQUIN em [SEV02], com o objetivo de definir um modelo mais formalizado.

De uma forma geral, pode-se dizer que a estrutura do Acordo de Nível de Serviço (SLA) é formada pela uma parte administrativa, legal e econômica, e por uma parte técnica, que constitui o SLS. Dentro do SLS existe a Especificação de Condicionamento de Tráfego (TCS – *Traffic Conditioning Specification*). TCS é um conjunto de parâmetros e seus valores que juntos especificam um grupo de regras de classificação e um perfil de tráfego [GRO02]. Um SLA pode conter mais de um SLS, sendo que cada SLS refere-se exclusivamente a um determinado serviço.

A estrutura de um SLS pode conter parâmetros que variam ou não no tempo, definindo um SLS estático ou dinâmico. O dinâmico possui parâmetros com aspectos que podem ser alterados, pelo cliente ou pelo provedor, sem a aprovação da outra parte, dentro de determinados limites pré-acordados. No SLS estático os parâmetros não podem ser alterados sem a aprovação de ambas as partes. Caso seja necessária alguma alteração, configura-se um novo SLS

Tomando por base [GOD01] e [SEV02], pode-se definir uma possível estrutura modelo para um SLA. Ela é formado por duas partes: a parte A que contém parâmetros administrativos, legais e econômicos; e a parte B que contém o SLS com parâmetros técnicos, incluindo o TCS. A seguir é apresentada essa estrutura para SLA.

4.2.1 Parte A: Parâmetros Administrativos, Legais e Econômicos

Esta primeira parte do SLA compreende uma série de itens que definem detalhes comerciais e procedimento para o fornecimento do serviço em questão. A seguir são apresentados os principais itens da parte A:

Contato técnico e administrativo – identifica pelo menos uma pessoa e seus dados para contato técnico e administrativo em cada uma das partes envolvidas no SLA.

Tempo de duração – especifica o período pelo qual o acordo é válido.

Garantias de disponibilidade – define o modelo para o cálculo da disponibilidade do serviço e como os dados necessários serão obtidos. Uma forma de obter os dados seria através do histórico de chamados técnicos no sistema do provedor. Neste item também podem ser definidas uma média de disponibilidade do serviço a ser atingida dentro de um período especificado, um limite de tempo para indisponibilidade e fórmulas para ressarcimento em caso de não atingimento do estabelecido.

Monitoramento do SLA – deve especificar como e quando será monitorado o SLA, além de definir os parâmetros e métricas visíveis ao cliente. Também deve ser especificado como ele terá acesso ao dados de monitoramento.

Tempos de resposta – apresenta os tempos de resposta que o provedor deve garantir para casos onde o cliente solicita alterações no SLA ou SLS, e para configurações ou alterações necessárias em dispositivos ou circuitos de acesso.

Tratamento de falhas – são especificadas as ações que o provedor deve tomar quando ocorrem falhas na prestação do serviço definido no SLS, e também os tempos de reação para estes casos.

Qualidade e performance do suporte – especificação detalhada da infra-estrutura de suporte ao serviço contratado.

Preço do serviço – definição detalhada dos preços, levando em conta aspectos de monitoramento e infra-estrutura utilizada.

Descrição do serviço – o serviço prestado deve ser descrito de maneira geral, informando de forma qualitativa suas características (em termos de atraso, variação do atraso, vazão e perdas) e também a sua operação.

4.2.2 Parte B: Parâmetros Técnicos

Esta segunda parte do SLA é efetivamente o SLS, incluindo o TCS, e compreende parâmetros técnicos essenciais para o fornecimento de qualidade de serviço ao cliente, de maneira unidirecional, definindo exatamente o serviço que deverá ser prestado. Podem existir vários SLS's dentro de um mesmo SLA, sendo que cada SLS refere-se a um determinado serviço. Segundo [GOD01], nem todos os parâmetros do SLS precisam ser definidos para determinados serviços. A seguir são apresentados os principais parâmetros da parte B:

Escopo – define a região geográfica/topológica na qual o serviço ser prestado. Este campo deve definir onde os pacotes entram e saem do domínio DS, sempre de maneira unidirecional (isto não quer dizer que não possam existir serviços bidirecionais, mas estes devem ser feitos através da combinação de dois ou mais SLS's). Segundo [GOD01], o escopo é definido por um par entrada e saída de interfaces, ou seja:

Escopo = (entrada, saída) ; onde entrada e saída podem ser uma interface, ou um grupo de interfaces, ou quaisquer interfaces.

Dessa forma podem ser habilitados serviços 1/1 (um-para-um), 1/N (um-para-alguns), 1/quaisquer, N/1(alguns-para-1) e quaisquer/1, dependendo das interfaces definidas.

Identificador de fluxo – sua função é definir para quais pacotes IP o SLS em questão deve ser cumprido, ou seja, indica quais pacotes IP devem receber o PHB definido para esse serviço. A identificação do fluxo de pacotes IP pode ser efetuada através do campo DSCP, e/ou endereços IP origem/destino, e/ou informação referente à aplicação (como protocolo e/ou porta TCP/UDP de origem/destino do pacote). O identificador pode ter os seguintes atributos:

Identificador de fluxo = (valor do DSCP, IP origem, IP destino, aplicação)

Esses atributos podem ser definidos por um valor único, ou um grupo de valores ou quaisquer valores. Podem ser combinados entre si, e devem conter a informação necessária para a classificação MF, ou BA, dos pacotes na borda do domínio DS. Deve-se observar que esse DSCP não tem nada a ver com o valor do DSCP usado

internamente no domínio para efetuar a diferenciação por PHB. No nó de entrada os pacotes podem ser classificados usando o valor do DSCP (identificador), mas depois ele poderá ser remarcado para uso interno ao domínio.

A informação referente à aplicação pode ser muito importante para a classificação MF. Pode ser informado o protocolo, ou as portas TCP/UDP usadas pela aplicação do cliente. Várias portas já estão definidas para certas aplicações e não podem ser usadas por outras. A relação de portas está disponível no IANA, que é o órgão responsável pela administração de números, conforme visto na seção 2.1.

Mas, nem sempre a aplicação possui portas bem definidas. Por exemplo, o padrão H.323 (usado para videoconferência) aloca dinamicamente as portas UDP, através do protocolo RTP, para as *streams* de voz e vídeo. A identificação dos pacotes de aplicações como esta, não pode ser feita com certeza usando apenas uma informação de portas UDP. Para resolver essa situação, alguns fabricantes como a CISCO Systems desenvolveram ferramentas que analisam toda a pilha do pacote, do nível 4 ao 7, definindo com certeza absoluta o tipo de aplicação. Isto pode ser verificado em [NBA02] e [NBARE].

Descritor do tráfego e Conformidade do Tráfego – o descritor do tráfego descreve o perfil de tráfego desejável para o fluxo IP (identificado pelo identificador de fluxo), através de um conjunto de parâmetros de conformidade de tráfego. O perfil de tráfego especifica características às quais o fluxo deve estar conforme para que possa obter as garantias do serviço contratado.

Existe um algoritmo de conformidade de tráfego que usa os parâmetros de conformidade para definir quais pacotes, dentro do fluxo identificado, estão em conformidade com o perfil de tráfego definido no descritor. Os pacotes conformes ao perfil são chamados “dentro de perfil” (“*in-profile*”), e os pacotes não conformes são chamados “fora de perfil” (“*out-of-profile*”). Estes são considerados como excesso. O teste de conformidade é aplicado, geralmente, nos nós de borda do domínio DS.

Os parâmetros de conformidade descrevem valores de referência para o perfil de tráfego desejável. Um algoritmo de conformidade muito utilizado é o Token Bucket usando os

parâmetros “b” e “r”, conforme apresentado na seção 3.3.2 deste trabalho. Alguns parâmetros de conformidade usados são:

taxa de pico = p (bit/s) ;

taxa média = r (bit/s) ;

tamanho da rajada = b (bytes)

Existem casos onde o tamanho da rajada poderá ser definido por até dois valores, sendo que o primeiro definirá uma rajada normal e o segundo definirá uma rajada máxima. Isto possibilita mais opções para tratamento de excesso, como remarcação do pacote e descarte apenas para pacotes que ultrapassam a rajada máxima.

Tratamento de excesso – descreve qual o tratamento para os pacotes em excesso, ou seja, pacotes fora de perfil encontrados no teste de conformidade. Esses pacotes poderão ser descartados, conformados/suavizados (atrasados até que se situem dentro do perfil desejado), ou remarcados para outro DSCP. Dependendo do tratamento a ser realizado, alguns parâmetros serão necessários, como novo valor do DSCP no caso da remarcação de pacotes ou o tamanho do buffer de conformação para pacotes a serem conformados. Caso o parâmetro Tratamento de excesso não seja especificado, o tráfego em excesso será descartado.

Garantias de desempenho – descreve as garantias de transporte do serviço que a rede oferece para o fluxo IP, já identificado, e dentro do escopo definido . Existem quatro parâmetros de desempenho:

- Atraso – indica o atraso máximo na transferência do pacote desde a entrada até a saída, para tráfego *in-profile*. Exemplos: atraso < 150 ms ; ou atraso < 100ms em 99% dos pacotes num período especificado.
- Variação de atraso – indica a máxima variação de atraso na transferência do pacote desde a entrada até a saída; para tráfego *in-profile*. Exemplos: variação < 25ms ; ou variação < 20ms em 98% dos pacotes.

- Perda de pacotes – “p” indica a razão entre a perda de pacotes, dentro do Escopo, e o total de pacotes do fluxo que entraram na rede, para os pacotes *in-profile*. Exemplos: perda $< 1 \times 10^{-5}$; ou perda máxima = 1% dos pacotes

$$p = \text{pacotes perdidos no escopo} / \text{pacotes que entraram}$$

- Vazão – é a taxa de velocidade de todos os pacotes do fluxo, medida na saída do Escopo, incluindo pacotes *in-profile* e *out-of-profile*. Exemplo: taxa = 512 Kbit/s.

Segundo [GOD01], os parâmetros de desempenho podem ser quantitativos ou qualitativos. Um parâmetro é quantitativo quando seu valor é especificado por um valor numérico, e as garantias de desempenho são ditas quantitativas se ao menos um dos quatro parâmetros for quantificado.

As garantias de desempenho são ditas qualificadas, se nenhum dos parâmetros é quantificado. Dessa forma, os parâmetros atraso e a perda de pacotes podem ser qualificados com valores ditos alto, médio ou baixo. O provedor do serviço é quem define a diferença relativa, por exemplo: alto = 2 x médio e médio = 2 x baixo.

Se as garantias de desempenho não são quantificadas nem qualificadas, então o serviço oferecido pelo SLS é apenas o de melhor esforço (best-effort).

Escala do serviço – indica o momento de início e o momento final da prestação do serviço. Pode ser especificado por um período de horas, dias, etc., ou de forma infinita. Exemplo: início as 08:00h e término as 18:00h todos os dias, ou somente nos dias úteis.

Confiabilidade – especifica o tempo médio de indisponibilidade (MDT – *Mean Down Time*) aceitável por ano, que pode ser expresso em minutos. Também especifica o máximo tempo de reparo (TTR-*Time To Repair*) em caso de interrupção do serviço, e pode ser expresso em horas ou minutos.

Outros parâmetros podem ser incluídos num SLA/SLS, mas os parâmetros citados acima seriam os mais indicados para um modelo formal básico. Vários estudos continuam neste sentido e novidades vão surgir com o tempo.

Depois que o SLA, e o(s) SLS(s), estão definidos para o serviço a ser prestado ao Cliente, o provedor precisa programar nos nós/roteadores da rede todas as funções e parâmetros necessários para fornecer o serviço contratado. Essa programação pode ser manual, sendo feita pela equipe técnica do provedor, ou automatizada através de softwares desenvolvidos para essa função. Este trabalho considera a forma manual de programação dos parâmetros SLS. Em [PRI01] é sugerida uma forma para implementar um sistema de mapeamento e configuração de parâmetros na rede, a partir do SLS.

A confirmação de que o serviço está, ou não, sendo prestado da forma adequada ao que foi contratado, deve ser apresentada ao Cliente através de ferramentas que monitorem a qualidade do serviço, principalmente os parâmetros de desempenho como atraso, variação de atraso, perda de pacotes, vazão e também a confiabilidade (MDT e TTR), segundo [STU00]. Isso é implementado pela Gerência de Nível de Serviço (SLM), cujo procedimento deve estar definido no SLA – parte A. O SLM não faz parte do escopo deste trabalho.

4.3 Conclusão

Com a maior quantidade de aplicações que necessitam de garantias de qualidade da rede de transporte, e com a necessidade dos provedores em fornecer serviços que se adequem ao mercado, foi necessário a adoção do SLA entre provedor e cliente. É o documento que define o tipo de serviço contratado, suas características técnicas e comerciais, além das responsabilidades do provedor e do cliente.

A estrutura do SLA, incluindo o SLS, não é padronizada, mas existem vários grupos de trabalho estudando a melhor forma de obter um modelo que inclua todas as informações pertinentes. Neste capítulo foi apresentada uma estrutura que pode ser adotada por um provedor para elaborar um SLA, definindo os vários itens que devem ser levados em conta para especificar o serviço contratado pelo cliente.

No entanto, a elaboração do SLS depende de várias informações que dizem respeito às aplicações de cada cliente, e como este irá utilizá-las. Assim, é necessário identificar e definir o perfil das aplicações do cliente, mas em várias situações isso não é tarefa simples, e o capítulo 5 tem por objetivo propor uma metodologia para identificação e definição do perfil do cliente, para chegar ao SLS das aplicações.

5. PROPOSTA DE IDENTIFICAÇÃO E DEFINIÇÃO DO PERFIL DO USUÁRIO

Atualmente, o Acordo de Nível de Serviço (SLA) passa a ser adotado cada vez com maior frequência entre provedores e clientes, visando definir de forma mais detalhadamente o serviço contratado. Como parte desse acordo, o SLS especifica parâmetros técnicos como garantias de desempenho, perfil de tráfego e regras para identificação de cada aplicação, definindo a forma como o serviço deverá ser prestado. O SLA é peça fundamental na arquitetura de Qualidade de Serviço DiffServ [BLA98], pois a partir dele são definidos o tipo de serviço a ser prestado, as características técnicas do serviço, as responsabilidades e também as penalidades a serem aplicadas caso o serviço não atinja o nível de qualidade desejado.

A estrutura do SLS não é padronizada, mas existem vários estudos neste sentido. No capítulo 4 foi apresentada uma possível formatação para os parâmetros a serem usados no SLS, tomando por base [GOD01] e [SEV02]. Dependendo do serviço a ser prestado, os parâmetros podem ser quantitativos ou qualitativos e em certos casos apenas alguns parâmetros ser definidos.

Os provedores que oferecem QoS através de DiffServ em suas redes, disponibilizam os serviços, ou classes de serviços, conforme seu interesse em atender o mercado. Uma vez definidas quais classes de serviço serão providas, o provedor especifica para cada uma delas os parâmetros de QoS que serão fornecidos por sua rede. Os clientes contratam classes de serviço para transportar suas aplicações, entretanto, a definição de qual classe contratar passa pela definição do SLS, que depende de informações sobre a aplicação e necessidades do cliente. O conjunto dessas informações pode ser chamado de perfil do usuário, ou cliente.

O perfil do cliente é usado para definir o SLS de cada tipo de aplicação, e essas informações precisam ser obtidas junto ao cliente. No entanto, existem muitas dificuldades em obtê-las, principalmente pelo desconhecimento do cliente com relação a detalhes de suas aplicações. Um cliente pode não saber especificar a vazão necessária para sua aplicação de voz, mas pode saber que seu equipamento implementa o padrão G.729, por exemplo. Também é difícil traduzir as necessidades do cliente nos parâmetros do SLS, já que são necessárias informações detalhadas como vazão a ser utilizada, limites de tráfego e rajadas aceitas, tratamento do excesso de tráfego, entre outras. Assim, torna-se importante a existência de uma ferramenta ou metodologia que facilite o processo de obtenção das informações, e que seja relativamente padronizada.

Para o provedor é muito interessante que sua própria equipe de vendas possa obter todos os dados necessários do cliente, identificando seu perfil, sem demandar a presença de um especialista técnico do provedor. Com o uso de uma ferramenta que facilite esse processo, torna-se mais ágil e econômica, para o provedor, a tarefa de formatação do SLS e o cliente não precisa se envolver em detalhes, podendo permanecer focado no seu próprio negócio.

O objetivo deste capítulo é propor uma metodologia para identificar e definir o perfil do usuário e compor o SLS para cada aplicação desejada pelo cliente. Na seção 5.1 é apresentada uma visão geral da metodologia proposta, enquanto que nas seções 5.2 e 5.3 são descritas as duas fases que a compõem.

5.1 Visão Geral da Metodologia

A metodologia proposta tem o objetivo de possibilitar que a equipe de vendas do provedor possa identificar e definir o perfil do cliente, visando compor o SLS. Este poderá ser verificado pela equipe técnica do provedor antes de ser validado, inserido no SLA e assinado o contrato com o cliente. Como QoS é um assunto bastante recente no meio comercial, é necessário que a equipe de vendas tenha um treinamento básico sobre qualidade de serviço na Internet, mais especificamente sobre DiffServ.

Duas fases distintas compõem a metodologia. Na Fase I o objetivo é levantar todas as informações referentes às aplicações que o cliente pretende usar, além de suas necessidades e desejos com relação ao desempenho dessas aplicações, identificando seu perfil. Na Fase II o perfil do cliente deve ser tratado com o objetivo de compor o SLS das aplicações desejadas, levando em conta a estrutura apresentada no capítulo 4.

A Fase I é composta por quatro etapas, sendo que na Etapa 1 são identificadas as aplicações do cliente, suas características e as prioridades desejadas pelo cliente para cada aplicação. Em seguida, na Etapa 2 é feita a identificação da abrangência física que devem ter as aplicações e possíveis limitações lógicas. Na sequência, a Etapa 3 é responsável por identificar o perfil de tráfego desejado pelo cliente. Finalmente, a escala de atividade e a confiabilidade desejada para as aplicações são identificados na Etapa 4.

A fase seguinte, Fase II, é responsável pela composição do SLS para cada aplicação identificada na Fase I. Os valores dos parâmetros do SLS são definidos a partir das informações do perfil do cliente. Também são definidas as classes de serviço nas quais cada aplicação deverá ser mapeada. Alguns parâmetros de SLS não precisam ser definidos para algumas classes de serviços, conforme visto na seção 4.2.1.

Deve-se lembrar que Diffserv tem seus serviços definidos para tráfego unidirecional, portanto o nível de serviço definido será para o tráfego de *upload* da empresa contratada. Ou seja, o foco do SLS são os pacotes IP cuja fonte é a rede do cliente.

5.2 Fase I – Identificando o Perfil

Nesta fase são levantadas todas as informações como necessidades e desejos do cliente, além das aplicações que vai utilizar e suas características. A equipe de vendas vai obter as informações através do preenchimento de tabelas, em conjunto com o cliente. O integrante da equipe de vendas do provedor deve auxiliar o cliente no

preenchimento das tabelas, esclarecendo eventuais dúvidas. Nestas tabelas são inseridas as informações conforme as quatro etapas apresentadas a seguir. O número de linhas nas tabelas está limitado mas na prática não tem limite.

5.2.1 Etapa 1 – Identificar aplicações, características e prioridades

Esta etapa visa realizar o levantamento das aplicações do cliente, suas características e as prioridades desejadas pelo cliente. Para viabilizar esse processo, definiu-se a Tabela 5.1 (ver adiante), que deve ser preenchida objetivando o levantamento dos dados necessários. A seguir tem-se a descrição de cada coluna desta tabela:

- **Aplicação ou Tipo de fluxo** – relacionar nesta coluna as aplicações ou tipo de fluxo de pacotes que o cliente deseja utilizar através das classes de serviço com QoS do provedor, como voz, vídeo, acesso WEB, sistemas de controle, etc.

Para o caso de aplicações multimídia, que contenham mais de um tipo de fluxo, deverá ser considerado e relacionado cada tipo de fluxo. Por exemplo, uma aplicação de videoconferência deverá ser especificada através de dois tipos de fluxo, como vídeo interativo e voz. Caso o cliente tenha outras aplicações de voz, todas poderão ser consideradas como o mesmo tipo de fluxo, inserindo apenas uma linha na Tabela 5.1 para tipo de fluxo = voz, se o Codec usado (especificado na coluna respectiva) for o mesmo para todas. Caso o Codec seja diferente, deverá ser inserida uma linha para tipo de fluxo de voz, informando os dados solicitados nas colunas seguintes como Codec, Num. conexões, etc.

- **Aplicativo** – preencher com o nome do aplicativo e/ou fabricante, como Outlook (para e-mail) ou Polycom (para videoconferência), ou informação de uso. Essa informação poderá ser útil, mas seu preenchimento não é obrigatório.
- **Protocolo ou padrão** – informar o protocolo da aplicação ou padrão utilizado.

- **Codec** – para o caso de aplicações de voz ou vídeo, informar o Codec utilizado ou selecionado pelo cliente na aplicação.
- **Taxa gerada** – nesta coluna deve ser informada a taxa de bits, conhecida ou estimada, que vai ser gerada pela aplicação. A taxa a ser informada nesta coluna é por conexão. A taxa para aplicações de voz não precisa ser informada, pois será definida pelo tipo de Codec, já informado na coluna respectiva.
- **Número conexões** – informar o número máximo de conexões simultâneas a serem aceitas pela aplicação. Caso não seja conhecido, estimar um valor aproximado. Deve-se ter atenção para o fato de que determinadas aplicações acabam gerando mais de um tipo de conexão, como a videoconferência que além de gerar conexões de vídeo também gera conexões de voz. Cada conexão deve ser contabilizada para não gerar erros no processo.
- **Prioridade** – nesta coluna, o cliente deve indicar o nível de prioridade que cada aplicação deverá ter na rede. A notação usada é numérica, sendo 1 a mais prioritária. Esses níveis de prioridades serão implementados pelas classes de serviço. Segundo [NBA02] e [RET02], e considerando as seções 3.2 e 3.3.5 desta dissertação, recomenda-se que as aplicações de voz sejam de prioridade 1 e as de vídeo interativo sejam de prioridade 2, caso este tipo de tráfego seja de interesse do cliente. Também é possível repetir prioridades para aplicações diferentes. A seguir são sugeridos níveis de prioridade para algumas aplicações ou tipos de fluxo:
 - voz: nível de prioridade deve ser máximo, ou seja, nível 1.
 - vídeo interativo: nível 1 ou 2. Sugere-se nível 2 caso exista aplicação de voz.
 - outras aplicações: o cliente pode optar pelo nível de prioridade conforme seu desejo, sem nenhuma restrição, sempre levando em conta que quanto maiores as garantias oferecidas pelas classes de serviço, maior será o preço do serviço.
Para aplicações interativas, que exijam rápida resposta, sugere-se o nível 3, ou menor. Caso não exista vídeo interativo, pode-se optar pelo nível 2 também.
Para aplicações transacionais, incluindo vídeo streaming, sugere-se o nível 3 ou 4 para baixo.

TABELA 5.1 - IDENTIFICAÇÃO DAS APLICAÇÕES, CARACTERÍSTICAS E PRIORIDADES.

Aplicação ou Tipo de fluxo	Aplicativo (informação de uso)	Protocolo ou Padrão	Codec (para voz e vídeo)	Taxa gerada (por conexão)	Num. Conexões (máx. simultaneas)	Prioridade desejada 1 a 5 (1 para maior)
- Voz						
- Vídeo interativo						
- Vídeo streaming						
- Web/http						
- e-mail						
- telnet						
- FTP						
-						
-						

5.2.2 Etapa 2 – Identificar a abrangência e limitações

Nesta etapa são obtidas informações quanto à abrangência de cada aplicação, ou seja, qual a região em que deve estar sendo prestado o serviço conforme desejo do cliente. Pode ser uma região geograficamente limitada ou não. Também podem existir limitações lógicas, como apenas alguns endereços IP tendo permissão para utilizar o serviço.

Os dados de abrangência são colocados na Tabela 5.2, e os dados de limitações lógicas na Tabela 5.3. Deve-se observar que para cada aplicação ou tipo de fluxo, deve ser gerada uma tabela 5.2 e uma tabela 5.3. A seguir tem-se a descrição de cada coluna da Tabela 5.2:

- **Número circuito de entrada** – neste caso deve ser especificado o número do circuito físico por onde os pacotes entram na rede do provedor. Este número identifica uma interface física da rede DS do provedor. Caso seja um novo circuito, a ser instalado ainda, o provedor é quem vai definir este número. Podem ser definidos mais de um número nesta coluna.
- **Número circuito de saída** – neste caso deve ser especificado o número do circuito físico por onde os pacotes saem da rede do provedor. Este número identifica uma interface física da rede DS do provedor. Caso seja um novo circuito, a ser instalado ainda, o provedor é quem vai definir este número. Podem ser definidos mais de um número nesta coluna.

Estas informações servirão para definir o escopo do SLS, e conforme apresentado na seção 4.2.1 este parâmetro pode definir serviços 1/1 (um-para-um), 1/N (um-para-alguns), 1/qualquer, N/1 (alguns-para-um) e qualquer/1. Dessa forma, o número de circuito poderá ser um único número, alguns números ou a indicação de qualquer, definindo que não existe limitação geográfica.

5.2.3 Etapa 3 – Identificar o perfil de tráfego

Nesta etapa devem ser obtidas informações com relação ao perfil de tráfego desejado pelo cliente. Ele poderá definir um perfil de tráfego com limitações para o tráfego entrante, aceitando ou não tráfego em rajadas, e definindo ações a serem tomadas sobre pacotes que estejam fora do perfil. O tráfego também pode ser conformado segundo o desejo do cliente. Todas essas informações devem ser inseridas na Tabela 5.4.

Não existem valores de rajadas ou conformação sugeridos, pois o cliente é quem deverá decidir se deseja ou não aceitar rajadas e conformar o tráfego, e quais esses valores. Na prática, o provedor do serviço terá preços variando conforme o percentual de rajadas e/ou conformação do tráfego desejados, e isso será um fator limitante para o cliente. No entanto, o escopo deste trabalho não considera o fator preço. A seguir tem-se a descrição de cada coluna da Tabela 5.4.

- **Aplicação ou Tipo de fluxo** – relação das aplicações, a partir das listadas na Tabela 5.1, nas quais deverá ser efetuado algum tratamento de perfil de tráfego, conforme desejo do cliente. Em adição às aplicações listadas explicitamente pelo cliente, uma aplicação chamada “melhor esforço” está incluída na Tabela 5.4 para especificar um limite de tráfego, se desejado, com relação a fluxos de aplicações existentes, mas que não foram listadas na Tabela 5.1, ou seja, fluxos que serão tratados apenas com o serviço melhor esforço. Aplicações de voz e vídeo não devem ser listadas nesta tabela pois serão tratadas com policiamento de limite de tráfego na taxa de pico, e essa taxa será definida através dos dados da Tabela 5.1.
- **Conformar tráfego** – nesta coluna pode ser informado o valor da taxa média ou de pico para efetuar a conformação do tráfego, conforme desejado pelo cliente para cada aplicação. A conformação de tráfego foi apresentada na seção 3.3.2 deste trabalho, sendo que é um método através do qual pacotes que estejam fora do perfil especificado podem ser atrasados até que estejam em conformidade ao perfil, e possam então ser transmitidos adiante. A taxa de pico é informada quando se deseja

aproveitar melhor a banda disponível, para aplicações que tenham tráfego irregular, ou em rajadas, consistindo na maioria das aplicações de dados.

- **Limitar tráfego** – pode ser definido um valor no qual o tráfego da aplicação deverá ser policiado ou limitado. Podem ser aceitas rajadas ou não, conforme definido nas colunas seguintes. Pacotes que estão dentro do limite de tráfego serão transmitidos normalmente, e pacotes acima do limite poderão ser descartados ou remarcados para nível de prioridade menor dentro da mesma classe de serviço, conforme apresentado na seção 3.3.2 deste trabalho. As ações a serem tomadas com pacotes fora de perfil, ou acima do limite, são especificadas nas colunas Ação pacotes.
- **Rajada normal** – o cliente pode definir que deseja aceitar uma rajada, em bytes, acima do valor indicado na coluna limitar tráfego. Esta rajada será considerada normal, e o tráfego até este limite será transmitido normalmente também. Na prática, o provedor terá preços diferenciados conforme o montante de tráfego em rajadas desejado, e isto será um fator limitante para o cliente.
- **Rajada máxima** – poderá ser definido um segundo valor de rajada, em bytes, que também poderá ser transmitido pela rede, mas apenas se houver recursos disponíveis, segundo [CON02]. Este valor define o máximo tráfego em rajada aceitável e é o limite para as ações definidas nas próximas colunas.
- **Ação pacotes entre rajada normal e máxima** – define a ação que deverá ser tomada para tráfego que se situa entre os valores rajada normal e rajada máxima, sendo que o pacote poderá ser descartado, ou então remarcado para um nível de prioridade menor dentro da mesma classe de serviço. Por exemplo: remarcado da classe AF21 para a classe AF22, sendo que neste caso a probabilidade de descarte deste pacote dentro da rede será maior pois a classe AF22 tem maior precedência de descarte do que a classe AF21, conforme visto na seção 3.3.4 deste trabalho.
- **Ação pacotes acima rajada máxima** – define a ação para tráfego que esteja acima do valor definido na coluna Rajada máxima. É sugerido que a ação seja o descarte dos pacotes tráfego, mas a remarcação também é possível, conforme [CON02].

5.3 Fase II – Compendo o SLS

Com todas as informações de identificação do perfil do cliente obtidas durante a Fase I, o próximo passo é a Fase II. Esta fase tem por objetivo compor o SLS de cada aplicação, definindo valores para os seus parâmetros, e também definir em qual classe de serviço cada aplicação deve ser mapeada.

O SLS é formado por uma série de parâmetros, sendo que na seção 4.2.2 deste trabalho foi apresentada uma estrutura segundo [GOD01] e [SEV02]. A seguir é apresentada a forma pela qual cada um dos parâmetros do SLS será definido usando as informações obtidas na Fase I.

- **Escopo** – as interfaces físicas de entrada e saída do tráfego são identificadas pela Tabela 5.2 através do número de circuito, que é associado a uma interface na rede do provedor, definindo o escopo da aplicação. Caso seja um novo circuito, ainda não instalado, o provedor definirá o número do circuito. As configurações possíveis para o escopo são aquelas apresentadas na seção 4.2.1 deste trabalho.
- **Identificador de fluxo** – a identificação de fluxo ou pacotes IP, no escopo deste trabalho, não será feita através de DSCP vindo do cliente, mas via regras de classificação. Caso a Tabela 5.3 esteja preenchida, a aplicação deverá ser limitada aos endereços IP definidos nessa tabela, e o fluxo deverá ser identificado através dos endereços IP de origem e destino relacionados nessa tabela.

Caso não exista limitação definida por endereços IP, a classificação será feita a partir do tipo da aplicação. Neste caso, conforme visto na seção 4.2.2 deste trabalho, geralmente a classificação é feita através dos campos que definem as portas de origem e destino nos segmentos TCP e UDP.

No entanto, algumas aplicações como voz e vídeo utilizam portas UDP alocadas de forma dinâmica através do protocolo RTP, e neste caso não é seguro fazer a classificação usando os números das portas UDP. A classificação deve ser feita verificando o cabeçalho do protocolo RTP, mais precisamente o campo *payload*, que indica se o pacote é de áudio, vídeo ou de outra aplicação.

Na tabela 5.6 são apresentadas as formas de identificação de pacotes para alguns tipos de aplicação.

TABELA 5.6 – FORMA DE IDENTIFICAÇÃO DE PACOTES PARA ALGUMAS APLICAÇÕES

Aplicação :	Forma de identificação :
Voz	payload do protocolo RTP, sobre UDP
vídeo	payload do protocolo RTP, sobre UDP
http/web	porta TCP - 80
telnet	porta TCP - 23
smtp/e-mail	porta TCP - 25
ftp	porta TCP - 20

Conforme pode ser visto na tabela 5.6, grande parte dos fluxos podem ser identificados através das portas TCP/UDP (portas TCP ou UDP) definidas para protocolos padrão ou não. A relação de portas está disponível no IANA, que é o órgão responsável pela administração de números, visto na seção 2.1. Portanto, o identificador de fluxo para classificar o pacote através da aplicação será a porta TCP/UDP usada pelo protocolo da aplicação, que será obtido na Tabela 1, coluna Protocolo. Para facilitar a tarefa de programação dos roteadores, alguns trabalham com o nome dos protocolos também e não somente com o número da porta.

Quando a aplicação for voz ou vídeo, ou outra aplicação que utilize o protocolo RTP, a identificação deve ser feita analisando o cabeçalho do protocolo RTP, e não as portas TCP/UDP. Isto pode ser feito via software, e conforme [NBA02] e [NBARE] a CISCO Systems desenvolveu uma facilidade chamada NBAR (Network Based Application Recognition), que identifica dentro do RTP se a stream é de voz ou vídeo. O NBAR pode ser usado para identificar qualquer tipo de pacote, não somente para pacotes que usam RTP, podendo ser usado para identificar pacotes telnet, por exemplo.

Neste caso, o identificador de fluxo será o próprio tipo de fluxo identificado na Tabela 1, coluna Aplicação ou Protocolo, que será usado como parâmetro para o NBAR. A tendência é aumentar o uso de facilidades como NBAR pois conseguem identificar qualquer tipo de pacote de forma bastante simples, como poderá ser visto no exemplo do capítulo 6.

- **Descritor do tráfego e conformidade do tráfego** – neste item serão obtidos parâmetros que especificam o perfil de tráfego da aplicação, como taxa de transmissão gerada e definições de limites para policiamento/conformação do tráfego. Para obter a taxa total gerada pela aplicação, é utilizada a tabela 5.7.

TABELA 5.7 – TAXA DE TRANSMISSÃO TOTAL GERADA POR APLICAÇÃO

Aplicação	Taxa considerada	Como obter a taxa total gerada
voz	Máxima	Taxa real gerada pelo Codec x Num. Conexões
vídeo	Máxima	Taxa gerada x Num. conexões
outras	Média	Taxa gerada x Num. conexões

Os valores de Taxa gerada e Num. conexões são obtidos através das colunas respectivas na Tabela 5.1. Para aplicações de voz, a taxa total gerada é obtida a partir da taxa real gerada pelo Codec, sendo este especificado na Tabela 5.1. Essa taxa real foi apresentada na seção 3.2.1 deste trabalho para os codecs de voz mais utilizados.

Para aplicações de voz e vídeo a taxa considerada será o valor máximo, e o tráfego será policiado descartando pacotes que estejam acima desse valor de pico, não sendo admitidas rajadas.

Para outras aplicações é considerada a taxa média, e o cliente pode especificar também o perfil de tráfego, sendo essas informações obtidas na Tabela 4. Essas aplicações podem ter um tráfego máximo policiado (coluna Limitar tráfego), com aceitação de rajadas ou não (colunas Rajada normal e Rajada Máxima), ou podem ter tráfego conformado para um valor médio ou de pico (coluna Conformar tráfego).

Tendo os valores de conformação e/ou policiamento (com rajadas ou não), o algoritmo de conformidade mais utilizado na prática é o *Token Bucket*, segundo [CON02], usando os parâmetros “r” (tráfego desejado) e “b” (rajada). Caso existam duas rajadas desejadas, o Token Bucket será duplo conforme apresentado na seção 3.3.2.

- **Tratamento de excesso** – os pacotes “fora de perfil”, ou seja, que não estão em conformidade ao perfil definido no item anterior, podem ser descartados ou

remarcados para nível menor, através do policiamento, ou conformados através da conformação. Na Tabela 4, o tratamento de excesso é definido pela coluna Conformar tráfego ou para casos de policiamento, pelas colunas Ação pacotes entre rajada normal e máxima e Ação pacotes acima rajada máxima. O cliente poderá definir a ação desejada para cada aplicação. Para voz e vídeo os pacotes acima da taxa máxima são descartados.

- **Garantias de desempenho** – três dos quatro parâmetros deste item são estabelecidos pelas classes de serviço disponibilizadas pelo provedor. O atraso, a variação de atraso, e a perda de pacotes possuem valores definidos pelo provedor, variando conforme as classes de serviço.

Segundo [NBA02] e [RET02], atualmente os provedores estão oferecendo, em geral, cinco classes de serviço como apresentado na seção 3.3.5 deste trabalho. As classes para voz e vídeo interativo devem possuir garantias mais rígidas e quantitativas, enquanto as outras classes devem oferecer menos garantias e de forma qualitativa, sendo que a prioridade destas na rede vai decrescendo. Um exemplo será visto no capítulo 6.

Conforme apresentado na Etapa 1 desta metodologia, na seção 5.2.1, a prioridade sugerida para aplicações de voz e vídeo interativo são 1 e 2 respectivamente. Caso seja seguida essa sugestão pelo cliente, voz será mapeada na classe de serviço de nível mais alto e vídeo na classe seguinte.

Através da coluna Prioridade desejada, da Tabela 5.1, pode-se identificar em quais classes de serviço o cliente deseja inserir suas aplicações. Aplicações com prioridade 3 terão tratamento preferencial na rede com relação à aplicações com prioridade 4, pois serão mapeadas numa classe de serviço de nível mais alto (AF31 por exemplo), enquanto as aplicações com prioridade 4 serão mapeadas numa classe de nível mais baixo (AF21, por exemplo).

A definição da vazão depende do perfil de tráfego desejado pelo cliente, sendo que o processo para sua definição é o mesmo apresentado no item Descritor do tráfego e Conformidade do tráfego. Para voz e vídeo a vazão considerada será a taxa máxima, e para outras aplicações será a taxa média. Como na prática, todas as vazões podem

estar compartilhando um mesmo acesso físico (circuito), o dimensionamento deste será visto no exemplo do capítulo 6.

- **Escala do serviço** – o período de prestação do serviço desejado pelo cliente pode ser obtido na Tabela 5.5, nas colunas Período para o serviço e Observações. Esse tipo de facilidade ainda não é comum ser ofertada pelos provedores, mas com o aumento da demanda visando redução de custos nos clientes, a facilidade deverá ser mais comum em breve.
- **Confiabilidade** – os provedores podem oferecer níveis diferenciados de confiabilidade, tanto para o Tempo Médio de Indisponibilidade (MDT) como para o Máximo Tempo de Reparo (TTR). Considerando estes níveis, o desejo do cliente será identificado a partir da Tabela 5.5, na coluna Confiabilidade.

Com os parâmetros do SLS de cada aplicação definidos através da Fase II, o provedor pode apresentar ao cliente a solução técnica e comercial para negociações finais, visando concluir e assinar o acordo SLA. Nestas negociações podem ser reavaliadas questões técnicas, como tráfego desejado ou classes de serviço mapeadas por exemplo, caso o cliente necessite reduzir o custo da solução ou tenha condições de contratar uma solução ainda melhor.

Após o fechamento do contrato e assinatura do SLA, o provedor vai programar seu roteador, ou roteadores, de borda no ambiente do cliente, conforme o SLS estabelecido para cada aplicação, e iniciará a prestação dos serviços.

5.4 Conclusão

A utilização crescente da rede internet para aplicações avançadas e de negócios, com o objetivo de redução de custos, motivou o desenvolvimento dos mecanismos de QoS. Eles melhoram a utilização dos recursos disponíveis na rede e proporcionam uma certa garantia de desempenho no transporte da informação.

As garantias que o provedor de serviço oferece ao cliente são estabelecidas através do SLA, e os parâmetros técnicos são definidos pelo SLS. Não existe uma estrutura padronizada para o SLS, mas existem propostas. No capítulo 4, foi apresentada uma estrutura composta por alguns parâmetros que definem o serviço a ser prestado.

A elaboração do SLS não é uma tarefa simples, pois depende de cada aplicação a ser utilizada, além das necessidades e desejos individuais de cada cliente. Clientes podem querer usar a mesma aplicação, mas com necessidades de tráfego distintas, por exemplo.

Com a metodologia apresentada neste capítulo é possível obter as informações sobre o perfil do cliente, e compor o SLS para cada aplicação desejada. No capítulo 6 é apresentado um estudo de caso, com o objetivo de comprovar que a metodologia facilita e, de certa forma, padroniza a identificação do perfil do cliente e definição do SLS. Essa metodologia poderá e deverá ser aperfeiçoada, mesmo porque os estudos com relação a QoS e SLA continuam sendo efetuados por vários grupos de trabalho a nível mundial.

6. ESTUDO DE CASO

As empresas estão usando a Internet cada vez mais para fazer negócios com clientes e parceiros, além de realizar a comunicação de dados entre suas unidades, com o objetivo de reduzir custos. O surgimento de aplicações avançadas que podem utilizar a Internet , como voz e vídeo, tornam essa rede ainda mais atrativa. No entanto, para o perfeito funcionamento de certas aplicações e para obter a satisfação dos usuários, é necessário que a rede ofereça alguma garantia de desempenho para o tráfego, e isso é obtido com os mecanismos de QoS.

Os provedores de serviços começam a oferecer QoS em suas redes, disponibilizando pacotes de serviços diferenciados, chamados classes de serviço. Elas possuem garantias diferenciadas, e assim, podem ser usadas para viabilizar aplicações que tenham diferentes requisitos. Como as classes oferecem serviços diferenciados, os preços para utilização das classes também são diferenciados.

O cliente contrata com o provedor os serviços desejados através do SLA, sendo que as características técnicas que são definidos através do SLS para cada aplicação. No capítulo 5 foi apresentada uma metodologia para obter todas as informações necessárias do cliente, com relação às suas aplicações, necessidades e desejos, definindo seu perfil e finalizando com a composição do SLS adequado.

Neste capítulo será apresentado um estudo de caso, com o objetivo de utilizar a metodologia proposta verificando sua funcionalidade e aplicabilidade prática.

6.1 Definição do Contexto

Para este estudo de caso será considerado um exemplo de uma empresa fictícia, que está instalada em três pontos, sendo uma matriz e duas filiais em cidades distintas. A empresa é cliente de uma operadora de telecomunicações e deseja utilizar a internet como meio de transporte para várias aplicações, algumas restritas aos três pontos e outras não.

A operadora oferece serviços via internet com garantias de desempenho, através de sua rede IP com DiffServ. A topologia da rede para o estudo de caso é apresentada na figura 6.1.

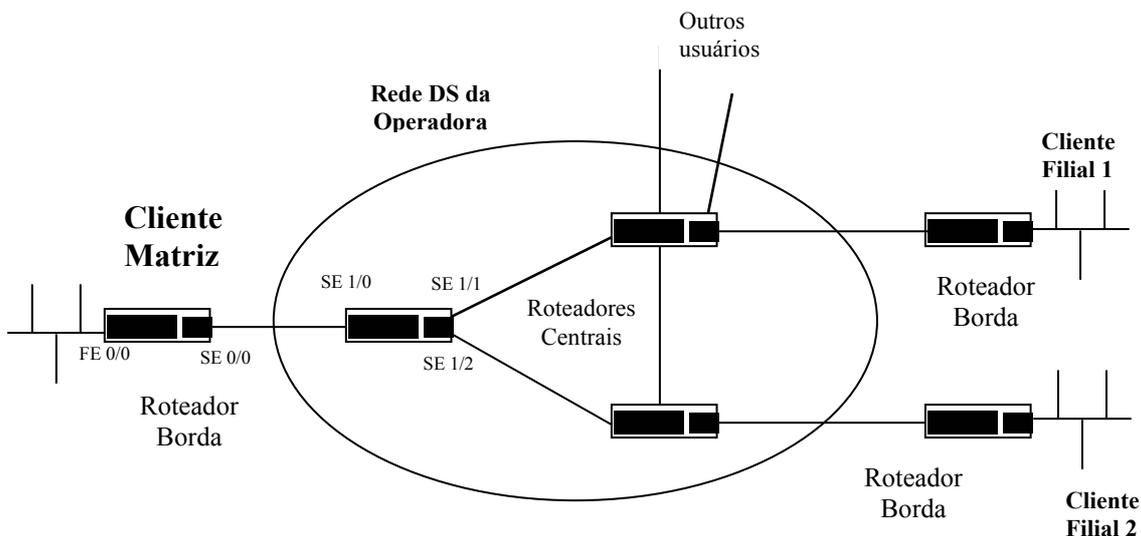


FIGURA 6.1 – TOPOLOGIA DA REDE: OPERADORA E CLIENTE

Além do serviço de melhor esforço, a operadora disponibiliza mais cinco classes de serviços com QoS, segundo sugerido em [NBA02], e apresentado na seção 3.3.5 deste trabalho. As classes de serviço, o PHB utilizado, assim como os tipos de aplicações às quais as classes de serviço se adequam são relacionadas na Tabela 6.1, e estão em conformidade com [NBA02] e [RET02]. Os nomes das classes de serviço foram criados apenas para servir como exemplo neste estudo

TABELA 6.1 – CLASSES DE SERVIÇO E APLICAÇÕES

Classe de Serviço	Valor DSCP/PHB	Tipos de aplicações
Super VIP	46 (101110) / EF	Tempo real (VoIP, sistemas de controle)
VIP	34 (100010) / AF41	Tempo real vídeo interativo (vídeoconferência)
Ouro	26 (011010) / AF31	Aplicações dados interativas; rápida resposta (ERP, telnet)
Prata	18 (010010) / AF21	Aplicações dados transacionais, normais (cliente/servidor, web)
Bronze	10 (001010) / AF11	Aplicações não interativas (e-mail, ftp, streaming de vídeo)

A classe de serviço na qual deve ser mapeada determinada aplicação do cliente vai depender da própria aplicação, das necessidades e desejos do cliente com relação à prioridade desta aplicação em relação às outras. O preço que a operadora cobra por cada classe de serviço também pode influenciar na decisão do cliente. Por exemplo, “telnet” pode ser bastante prioritário para determinado cliente, mas para outro não, e dessa forma a aplicação pode ser mapeada em classes distintas para cada cliente.

A operadora deve indicar os parâmetros de QoS fornecidos por seus serviços, ou seja, as garantias de desempenho que cada classe de serviço pode disponibilizar ao cliente. Alguns parâmetros de garantias podem ser quantitativos e outros qualitativos. Será considerado que a operadora deste exemplo fornece serviços com as garantias de QoS mostradas na tabela 6.2. Estes valores foram obtidos levando em consideração [GOD01], [NBA02] e a seção 3.2.2 deste trabalho. Reforça-se o conceito de que cada provedor define as garantias de desempenho dos seus serviços.

TABELA 6.2 – PARÂMETROS DE QOS PARA AS CLASSES DE SERVIÇO

Classe de Serviço	Atraso	Varição Atraso	Perda Pacotes	Vazão
Super VIP	< 100 ms	< 40 ms	máx. 1 %	conforme cliente
VIP	< 150 ms	<100 ms	máx. 1 %	conforme cliente
Ouro	< 200 ms	ND	máx. 2 %	conforme cliente
Prata	baixo	ND	máx. 3 %	conforme cliente
Bronze	médio	ND	máx. 3 %	conforme cliente

A operadora pretende atender o cliente em suas necessidades, oferecendo seus serviços com QoS, mas para isso precisa conhecer todas as aplicações, necessidades e desejos do cliente visando definir o SLS necessário. Para isso será aplicada a metodologia proposta no capítulo 5. Com o SLS definido, a operadora poderá fazer uma proposta comercial apresentando os serviços que atenderão ao cliente e, possivelmente, fechar o contrato.

6.2 Aplicação da Metodologia Proposta

Todas as necessidades, desejos e prioridades do cliente com relação às suas aplicações devem ser determinadas pela operadora, com o objetivo de definir o SLS e a partir daí mapear as aplicações do cliente nas classes de serviço adequadas. Para realizar essa tarefa, o vendedor ou consultor de vendas da operadora vai utilizar a metodologia proposta no capítulo 5.

Como DiffServ trabalha de forma unidirecional, por motivos de simplificação para este estudo, será analisada apenas a necessidade da matriz da empresa no sentido de saída do fluxo dos dados (matriz → rede da operadora), pois já será suficiente para validar a metodologia. Na prática devem ser analisadas todas as necessidades de fluxos desejados pelo cliente, como por exemplo os fluxos de saída das filiais em direção à matriz já que o SLS será diferente provavelmente.

Estabelecido o acima, o consultor de vendas aplica junto ao cliente as quatro etapas da Fase I, preenchendo as tabelas que fazem parte da metodologia. Para este estudo de caso, considera-se que as informações obtidas são as apresentadas nas Tabelas I, II, III, IV e V, mostradas a seguir.

TABELA I - IDENTIFICAÇÃO DAS APLICAÇÕES, CARACTERÍSTICAS E PRIORIDADES.

Aplicação ou Tipo de fluxo	Aplicativo (informação de uso)	Protocolo ou Padrão	Codec (para voz e vídeo)	Taxa gerada (por conexão)	Num. Conexões (máx. simultaneas)	Prioridade desejada 1 a 5 (1 para maior)
- Voz	Ramais e Videoconferência	H.323	G.729	-----	4 + 2	1
- Vídeo interativo	Videoconferência	H.323	H.263	256 Kbit/s	2	2
- Vídeo streaming	Não					
- Web/http	Servidor web	http		30 Kbit/s estimado	10 estimado	3
- e-mail	Servidor e-mail	smtp		20 Kbit/s estimado	5 estimado	5
- telnet	Emulação de terminal TN 3270	telnet		20 Kbit/s	5 estimado	4
- FTP	Não					
-						
-						

TABELA II – IDENTIFICAÇÃO DA ABRANGÊNCIA DAS APLICAÇÕES

Aplicação ou tipo de fluxo = VOZ	
Número(s) circuito(s) entrada (caso já exista)	Número(s) circuito(s) saída (caso já exista)
043-1415	043-4041
	043-5051

Aplicação ou tipo de fluxo = Vídeo Interativo	
Número(s) circuito(s) entrada (caso já exista)	Número(s) circuito(s) saída (caso já exista)
043-1415	043-4041
	043-5051

Aplicação ou tipo de fluxo = Web/http	
Número(s) circuito(s) entrada (caso já exista)	Número(s) circuito(s) saída (caso já exista)
043-1415	quaisquer

Aplicação ou tipo de fluxo = telnet	
Número(s) circuito(s) entrada (caso já exista)	Número(s) circuito(s) saída (caso já exista)
043-1415	043-4041
	043-5051

Aplicação ou tipo de fluxo = e-mail	
Número(s) circuito(s) entrada (caso já exista)	Número(s) circuito(s) saída (caso já exista)
043-1415	quaisquer

TABELA III – IDENTIFICAÇÃO DE LIMITAÇÕES LÓGICAS

Aplicação ou tipo de fluxo = <u>VOZ</u>	
IP origem	IP destino
Sem limitação	Sem limitação

Aplicação ou tipo de fluxo = <u>Video interativo</u>	
IP origem	IP destino
Sem limitação	Sem limitação

Aplicação ou tipo de fluxo = <u>Web/http</u>	
IP origem	IP destino
Sem limitação	Sem limitação

Aplicação ou tipo de fluxo = <u>telnet</u>	
IP origem	IP destino
Sem limitação	Sem limitação

Aplicação ou tipo de fluxo = <u>e-mail</u>	
IP origem	IP destino
Sem limitação	Sem limitação

TABELA IV – IDENTIFICAÇÃO DO PERFIL DE TRÁFEGO DESEJADO PELO CLIENTE

(para aplicações que não sejam voz ou vídeo, informar a conformação de tráfego ou policiamento desejado)

Aplicação ou	Conformar tráfego	Limitar tráfego	Rajada normal	Rajada máxima	Ação para pacotes	Ação para pacotes
Tipo de fluxo	(médio ou pico) bits/s	(policiar) bit/s	bytes	bytes	entre rajada normal e máxima	acima rajada máxima
Melhor esforço	XXXXXXXXXXXXX	64 Kbit/s	400	XXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX
Web/http	400 Kbit/s pico					
e-mail		100 Kbit/s	0 (zero)	0 (zero)		
telnet		100 Kbit/s	1200	1000	Remarcar	Descartar

TABELA V – IDENTIFICAÇÃO DA ESCALA DE ATIVIDADE E CONFIABILIDADE

Aplicação ou Tipo de fluxo	Escala para o serviço (horas)	Observações (sobre período)	Confiabilidade (Super alta, alta, normal)
VOZ	08:00 as 22:00 h	dias úteis	alta
Vídeo interativo	08:00 as 22:00 h	dias úteis	normal
web	24 horas	todos os dias	super alta
e-mail	24 horas	todos os dias	normal
telnet	24 horas	dias úteis	alta

Com as tabelas preenchidas a Fase I está concluída, e pode-se passar para a etapa de composição do SLS das aplicações desejadas utilizando a Fase II da metodologia, descrita na seção 5.3 deste trabalho.

Conforme informação da Tabela 1, este cliente deseja utilizar cinco aplicações, com características e necessidades distintas de QoS. São elas: voz, vídeo interativo (para uso em videoconferência), servidor web, servidor de e-mail e emulação de terminal através de telnet.

Como na Tabela 4 existe informação sobre policiamento no serviço melhor esforço, isso deve ser considerado e, portanto, devem ser compostos seis SLS's distintos nesta Fase II.

Nas páginas seguintes são apresentados os SLS's obtidos através da aplicação do processo descrito na seção 5.3 deste trabalho, para cada um dos seis serviços a serem prestados ao cliente exemplo deste estudo de caso.

1) SLS da aplicação = VOZ

- **Escopo** – obtido a partir da Tabela II.

Escopo = (int 0431415, int 0434041)
(int 0431514, int 0435051)

- **Identificador de Fluxo** –

Como a aplicação é voz, e não existe limitação conforme tabela III, os pacotes serão identificados através da aplicação, usando a facilidade NBAR, como descrito na seção 5.3. Portanto, ID Fluxo = (NBAR,audio).

- **Descritor do tráfego e Conformidade do Tráfego** – obtido da tabela I.

Codec é G.729, portanto taxa efetiva gerada é 26,4 Kbit/s. O número máx. de conexões simultâneas é 4, mas tem 2 conexões possíveis para a aplicação de Vídeoconferência também, assim, são 6 conexões de voz no total. Portanto :

Tráfego = 26,4 K x 6 = 158,4 K. Arredondando → 159 Kbit/s (tráfego máximo)

Algoritmo de Conformidade = Token Bucket

“r” = 159000 bits/s ; “b” = 0 bytes

Obs: não existem rajadas pois já é considerado o tráfego máximo.

- **Tratamento de excesso** –

Não especificado, pois na prática é estabelecido o descarte de pacotes.

- **Garantias de Desempenho** – a prioridade é definida na Tabela I.

Como voz possui prioridade 1 (mais alta), a aplicação é mapeada na classe Super VIP da operadora, obtendo os parâmetros definidos na tabela 6.2. A vazão é aquela definida no descritor de tráfego. Então temos os seguintes valores de garantias de desempenho:

Atraso : < 100 ms
Variação de atraso: < 40 ms
Perda de Pacotes: máx. 1 %
Vazão : 159 Kbit/s

- **Escala do serviço** – conforme tabela V, das 08:00 as 22:00h (em dias úteis)
- **Confiabilidade** – conforme tabela V, Confiabilidade = alta

2) SLS da aplicação = Vídeo interativo

- **Escopo** - obtido a partir da Tabela II.

Escopo = (int 0431415, int 0434041)
(int 0431514, int 0435051)

- **Identificador de Fluxo** –

Como a aplicação é vídeo interativo, e não existe limitação conforme tabela III, os pacotes serão identificados através da aplicação, usando a facilidade NBAR, como descrito na seção 5.3. Portanto, ID Fluxo = (NBAR,video).

- **Descritor do tráfego e Conformidade do Tráfego** –

Codec de vídeo é H.263, e a taxa é definida pela qualidade desejada de usuário. Neste caso, o usuário vai utilizar uma qualidade que gera 256 Kbit/s. Como o número máx. de conexões simultâneas é 2, temos :

Tráfego = 256 K x 2 = 512 Kbit/s (tráfego máximo)

Algoritmo de Conformidade = Token Bucket

“r” = 512000 bits/s ; “b” = 0 bytes

Obs: não existem rajadas pois já é considerado o tráfego máximo.

- **Tratamento de excesso** –

Não especificado, pois na prática é estabelecido o descarte de pacotes.

- **Garantias de Desempenho** – a prioridade é definida na Tabela I.

Como vídeo interativo possui prioridade 2, a aplicação é mapeada na classe VIP da operadora, obtendo os parâmetros definidos na tabela 6.2. A Vazão é aquela definida no descritor de tráfego. Então, temos os seguintes valores de garantias de desempenho:

Atraso : < 150 ms
Variação de atraso: < 100 ms
Perda de Pacotes: máx. 1 %
Vazão : 512 Kbit/s

- **Escala do serviço** – conforme tabela V, das 08:00 as 22:00h (em dias úteis)
- **Confiabilidade** - conforme tabela V, Confiabilidade = normal

3) SLS da aplicação = acesso web

- **Escopo** – obtido a partir da Tabela II.

Escopo = (int 0431415, int quaisquer)

- **Identificador de Fluxo** – obtido a partir da Tabela I.

Como a aplicação não é voz nem vídeo, e não existe limitação conforme tabela III, será identificada pela porta TCP/UDP utilizada. Na prática pode ser especificado o número da porta ou o nome do protocolo, portanto:

ID Fluxo = (porta TCP 80) ou ID Fluxo = (http).

- **Descritor do tráfego e Conformidade do Tráfego** – obtido das Tabelas I e IV.

Tráfego médio = 30 K x 10 = 300 Kbit/s

Tráfego deve ser conformado em 400 Kbit/s de pico, ou seja, 100 Kbit/s acima do tráfego médio, podendo ser considerada uma rajada de 12800 bytes no Token Bucket.

Algoritmo de Conformidade = Token Bucket

“r” = 300000 bits/s

“b” = 12800 bytes

- **Tratamento de excesso** – obtido na Tabela IV.

Conformar pacotes para 400 Kbit/s de pico.

- **Garantias de Desempenho** – a prioridade é definida na Tabela I.

Como a prioridade desejada pelo cliente é 3, a aplicação será mapeada na classe de serviço Ouro, que é um nível abaixo da prioridade 2 (serviço VIP). Os valores para classe Ouro estão definidos na tabela 6.2. A vazão é o tráfego médio definido no descritor de tráfego. Então temos os seguintes valores de garantias de desempenho:

Atraso :	< 200 ms
Varição de atraso:	não definida
Perda de Pacotes:	máx. 2 %
Vazão :	300 Kbit/s

- **Escala do serviço** – conforme tabela V, 24 horas (todos os dias)

- **Confiabilidade** - conforme tabela V, Confiabilidade = super alta

4) SLS da aplicação = telnet

- **Escopo** – obtido a partir da Tabela II

Escopo = (int 0431415, int 0434041)
(int 0431415, int 0435051)

- **Identificador de Fluxo** –

Como a aplicação não é voz nem vídeo, e não existe limitação conforme tabela III, será identificada pela porta TCP/UDP utilizada. Na prática pode ser especificado o número da porta ou o nome do protocolo, portanto:

ID Fluxo = (porta TCP 23) ou ID Fluxo = (telnet).

- **Descritor do tráfego e Conformidade do Tráfego** – obtido das Tabelas I e IV.

Tráfego médio = 20 K x 5 = 100 Kbit/s

Policado em 100 Kbit/s. Rajada normal = 1200 bytes. Rajada máxima = 1000bytes

Algoritmo de Conformidade = Token Bucket

“r” = 100000 bits/s

“b” = 1200 bytes e “b1” = 1000 bytes

- **Tratamento de excesso** – obtido na Tabela IV.

Pacotes entre rajada normal e máxima = remarcar (precedência maior) e transmitir

Pacotes acima rajada máxima = descartar

- **Garantias de Desempenho** – a prioridade é definida na Tabela I.

Como a prioridade desejada pelo cliente é 4, a aplicação será mapeada na classe de serviço Prata, que é um nível abaixo da prioridade 3 (serviço Ouro). Os valores para classe Prata estão definidos na tabela 6.2. A vazão é o tráfego médio definido no descritor de tráfego. Então temos os seguintes valores de garantias de desempenho:

Atraso : baixo, porém sem valor definido
Variação de atraso: não definida
Perda de Pacotes: máx. 3%
Vazão : 100 Kbit/s

- **Escala do serviço** – conforme tabela V, 24 horas (dias úteis)

- **Confiabilidade** - conforme tabela V, Confiabilidade = alta

5) SLS da aplicação = e-mail

- **Escopo** - obtido a partir da Tabela II.

Escopo = (int 0431415, int quaisquer)

- **Identificador de Fluxo** –

Como a aplicação não é voz nem vídeo, e não existe limitação conforme tabela III, será identificada pela porta TCP/UDP utilizada. Na prática pode ser especificado o número da porta ou o nome do protocolo, portanto:

ID Fluxo = (porta TCP 25) ou ID Fluxo = (smtp).

- **Descritor do tráfego e Conformidade do Tráfego** – obtido das Tabelas I e IV.

Tráfego médio = 20 K x 5 = 100 Kbit/s

Policiado em 100 Kbit/s. Rajada normal = 0 bytes. Rajada máxima = 0 bytes

Algoritmo de Conformidade = Token Bucket

“r” = 100000 bits/s

“b” = 0 bytes

- **Tratamento de excesso** – obtido na Tabela IV.

Descartar pacotes pois rajadas devem ser igual a zero.

- **Garantias de Desempenho** – a prioridade é definida na Tabela I.

Como a prioridade desejada pelo cliente é 5, a aplicação será mapeada na classe de serviço Bronze, que é um nível abaixo da prioridade 4 (serviço Prata). Os valores para classe Bronze estão definidos na tabela 6.2. A vazão é o tráfego médio definido no descritor de tráfego. Então temos os seguintes valores de garantias de desempenho:

Atraso : médio, porém sem valor definido

Varição de atraso: não definida

Perda de Pacotes: máx. 3%

Vazão : 100 Kbit/s

- **Escala do serviço** – conforme tabela V, 24 horas (todos os dias)

- **Confiabilidade** - conforme tabela V, Confiabilidade = normal

6) SLS da aplicação = serviço melhor esforço

- **Escopo** – não definido, portanto será o *default*. Para este estudo:

Escopo = (int 0431415, int quaisquer)

- **Identificador de Fluxo** – nenhum

- **Descritor do tráfego e Conformidade do Tráfego** –

Tráfego médio = não definido

Policiado em 64000 Kbit/s. Rajadas não permitidas

Algoritmo de Conformidade = Token Bucket

“r” = 64000 bits/s

“b” = 0 bytes

- **Tratamento de excesso** – não especificado, portanto será descarte
- **Garantias de Desempenho** – não existem
- **Escala do serviço** – não definida, portanto 24 horas (todos os dias).
- **Confiabilidade** - não definida, portanto é normal

Definidos os SLS's de todas as aplicações e considerando a topologia da rede, vista na figura 6.1, pode-se observar que os fluxos das diferentes aplicações passam através do mesmo circuito físico entre matriz do cliente e operadora. Qual deve ser a taxa ou vazão deste circuito? A taxa ideal seria a somatória da vazão de todas as aplicações com QoS, mais a vazão permitida para o melhor esforço (64K), conforme descrito nos SLS's anteriormente definidos. Então:

Taxa ideal = 159 K + 512 K + 300 K + 100 K + 100 K + 64 K = 1.235 Kbit/s

Portanto, a taxa ideal para o circuito de acesso entre a matriz do cliente e a operadora seria 1.235 Kbit/s. Este não é um valor de taxa de transmissão comercialmente disponibilizado pelas operadoras. Comercialmente existem as taxas de 1 Mbit/s ou 1,5 Mbit/s, e o cliente deverá optar entre as duas. Caso opte pela taxa de 1,5

Mbit/s pagará um preço maior e talvez nunca venha a utilizar toda a vazão disponível. O cliente pode optar também pela taxa de 1 Mbit/s, com o objetivo de reduzir custos. Mesmo com taxa de 1 Mbit/s o sistema funcionará perfeitamente já que as aplicações trafegam dados de forma estatística, sendo pouco provável um congestionamento por tráfego máximo de todas as aplicações num mesmo instante.

Em caso de congestionamento, pouco provável com taxa de 1Mbit/s, a política de QoS baseada em DiffServ e implantada no roteador de borda, vai priorizar as aplicações na seqüência Super VIP, VIP, Ouro e assim por diante, garantindo as prioridades definidas pelo cliente. Dentro da rede DS da operadora o funcionamento é semelhante porém com mecanismos que detectam e previnem congestionamentos com antecedência, conforme apresentado na seção 3.3.2. Esses mecanismos são usados dentro da rede da operadora pois o tráfego é muito maior e os recursos devem ser utilizados de forma otimizada para atender o tráfego de todos os seus clientes.

Com todos os SLS's definidos, e as condições de preço da solução fechadas com o cliente, é assinado o acordo SLA, e a fase de implementação da solução é iniciada. Um exemplo de implementação de DiffServ em roteador de borda, com os SLS's obtidos para este estudo de caso, é apresentado na seção 6.3, considerando alguns comandos de programação para roteadores da CISCO Systems.

6.3 Programação dos SLS's na rede DiffServ

A implementação da solução baseada em DiffServ para este estudo de caso consiste na instalação, pela operadora, de um roteador de borda na matriz do cliente conectado com sua rede DS, conforme a figura 6.1. Neste roteador serão programadas funções de classificação de pacotes e condicionamento de tráfego, além de implementar também os PHB's para as classes de serviço contratadas.

Fisicamente, isso é realizado efetuando a programação das funções e parâmetros no roteador de acordo com os SLS's já definidos, visando obter da rede o comportamento desejado, pelo cliente, para suas aplicações.

No anexo A é apresentada a programação do roteador de borda para este estudo de caso, tomando como base a plataforma CISCO, segundo [CON02], [NBARE] e [NBA02]. É considerando apenas o tráfego de ingresso pela matriz, ou seja, tráfego da matriz para a rede DS da operadora. Por motivo de simplificação, não são consideradas as informações de Escopo, Escala de serviço e Confiabilidade constantes no SLS. A programação dos roteadores centrais e de egresso de tráfego também não são apresentadas por não fazerem parte do escopo deste trabalho.

6.4 Conclusão

Para definição dos serviços a serem prestados ao cliente, é necessário que o provedor obtenha todas as informações sobre aplicações utilizadas, necessidades e prioridades desejadas pelo cliente. De posse das informações pode-se elaborar os SLS's de cada aplicação. Para realizar todo esse processo, foi proposta a metodologia do capítulo 5.

Como o número de aplicações é muito grande, com seus requisitos de desempenho variáveis, além de que as necessidades e desejos dos clientes são as mais diversas, é importante que os provedores utilizem alguma forma para facilitar e padronizar, tanto quanto possível, o processo de identificação do perfil do cliente com o objetivo de definir quais os serviços que serão usados para atender o cliente.

Conforme o estudo de caso apresentado neste capítulo, onde a metodologia proposta foi usada para definir os serviços a serem prestados por uma operadora a uma empresa fictícia, pode-se concluir que a aplicação da metodologia facilita a obtenção e organização de todas as informações necessárias, permitindo a composição do SLS para cada aplicação através do auxílio na definição dos valores de seus parâmetros. Dessa forma podem ser definidos os serviços adequados ao cliente.

7. CONCLUSÕES

Com o crescimento acelerado do número de usuários da internet, e com a utilização cada vez maior por parte de empresas para aplicações de negócios, o tráfego nessa rede tem aumentado muito. Aplicações mais avançadas, como voz e vídeo, também têm sido desenvolvidas para funcionar sobre a internet, e para seu bom funcionamento é necessário que a rede ofereça garantias de desempenho, e não apenas o melhor esforço como faz hoje.

Os mecanismos de qualidade de serviço (QoS) foram desenvolvidos para oferecer garantias de desempenho na rede, e podem ser oferecidos pelos provedores na forma de classes de serviços aos clientes, viabilizando aplicações avançadas e críticas através da rede internet.

A revisão bibliográfica efetuada sobre Qualidade de Serviço (QoS) na internet, e sobre o modelo DiffServ de forma mais específica, foi de grande importância para o entendimento das características gerais e específicas que envolvem esse vasto e fascinante assunto, além dos mecanismos utilizados para implementar QoS.

Pode-se concluir que o modelo DiffServ é bem aceito para implementar QoS pois possui uma flexibilidade muito grande, além da escalabilidade que proporciona. Trata-se basicamente um esquema de prioridades conferido aos pacotes que trafegam pela rede. O tratamento dos pacotes é feito conforme a prioridade estabelecida a eles através do DSCP, e seu encaminhamento é definido pelo respectivo

PHB. Os PHB's do DiffServ são implementados pelos provedores através das classes de serviços. A classe de serviço é contratada pelo cliente através da assinatura do acordo SLA com o provedor, sendo que a parte técnica do acordo é o SLS.

Chega-se à conclusão que a correta elaboração do SLS é fundamental para atender as expectativas do cliente, sendo necessário conhecer todas as informações sobre suas aplicações, além das necessidades e desejos individuais. Dessa forma, a tarefa de elaborar o SLS não é simples, pois depende de informações detalhadas e corretas, sendo que nem sempre o cliente tem conhecimento suficiente para repassá-las.

O objetivo deste trabalho foi propor uma metodologia para identificar e definir o perfil do usuário, ou cliente, visando compor o SLS. A utilização da metodologia é importante e vantajosa pois define um processo de obtenção das informações e o tratamento destas, possibilitando a elaboração do SLS de cada aplicação do cliente e a definição das classes de serviço corretas a serem utilizadas.

É importante notar que a metodologia poderá ser aplicada pela força de vendas do provedor de serviços, sem a necessidade de um especialista técnico, o que é extremamente interessante e desejável. Por outro lado, a força de vendas deve ser treinada para aplicar a metodologia, e ter um mínimo de conhecimento técnico, para que possa obter os resultados esperados, de forma detalhada e precisa. A metodologia também pode ser aplicada com o objetivo de pesquisa de mercado, proporcionando ao provedor conhecer a demanda e as necessidades do mercado potencial.

No capítulo 6, através de um estudo de caso bastante próximo de um caso real, foram verificadas a aplicabilidade e funcionalidade da metodologia proposta, além de demonstrar, através do Anexo A, a versatilidade da programação de roteadores que implementam QoS, e que são fornecidos por diversos fabricantes.

7.1 Trabalhos Futuros

A Qualidade de Serviço (QoS) na Internet é um tema bastante amplo e muitos grupos de trabalho a nível mundial como TEQUILA, AQUILA, CADENUS, SEQUIN, QoS Fórum entre outros, continuam a discutí-lo. Dessa forma, este trabalho não tem a pretensão de esgotar o assunto acerca deste tema, mesmo porque sempre existirão aperfeiçoamentos, e no decorrer do seu desenvolvimento pôde-se vislumbrar alguns trabalhos futuros como:

- implementação da metodologia proposta em software, visando automatizar o processo;
- desenvolvimento de interface para programação de roteadores de borda, considerando determinado fabricante, visando inserir na rede da operadora o SLS estabelecido para o cliente;
- adaptação da metodologia para utilização como pesquisa de mercado e demanda por parte de operadoras/provedores, para desenvolvimento de serviços com QoS adaptados à realidade do mercado;
- estudo visando aperfeiçoar a metodologia levando em conta não só o SLS, mas também o SLA, e adaptando conforme novas sugestões dadas pelos grupos de trabalho citados.

REFERÊNCIAS

- [BLA98] BLAKE, S., et al. **An Architecture for Differentiated Services**. RFC 2475, Internet Engineering Task Force, Dezembro 1998.
- [BRA94] BRADEN, R.; CLARK, D.; SHENKER, S. **Integrated Services in the Internet Architecture: An Overview**. RFC 1633, Internet Engineering Task Force, Junho 1994.
- [CADEN] IST project, **Creation and Deployment of End-User Services in Premium IP Networks**, CADENUS. Disponível em <<http://www.cadenus.org>>. Acesso em: 14 novembro 2002.
- [CERT] Computer Emergency Response Team/Coordinate Center. Disponível em <<http://www.cert.org>> . Acesso em: 07 fevereiro 2002.
- [CGINT] Comitê Gestor da Internet. Disponível em <<http://www.cg.org.br>> . Acesso em: 12 julho 2002.
- [COM02] CISCO Systems. **Cisco IOS Quality of Service Solutions Command Reference**. Release 12.2., Dezembro 2002. Disponível em <http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/122cgr/fqos_r/qrfbook.pdf>. Acesso em: 11 janeiro 2003.
- [CON02] CISCO Systems. **Cisco IOS Quality of Service Solutions Configuration Guide**. Release 12.2., Dezembro 2002. Disponível em <http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/122cgr/fqos_c/qcfbook.pdf> . Acesso em: 11 janeiro 2003.
- [CYC99] CYCLADES Brasil. **Guia Internet de Conectividade**. Cyclades Brasil, São Paulo, 5^a. edição, 1999.
- [FONSE] FONSECA, E.; SAMPAIO, E. **História da Internet e Órgãos de Coordenação**. Disponível em <<http://www.dcc.ufmg.br/~mlbc/cursos/internet/historia>> . Acesso em: 09 fevereiro 2002.
- [FER98] FERGUSON, P., HUSTON, G. **Quality of Service: Delivering QoS on the Internet and in Corporate Networks**. John Wiley & Sons, Janeiro 1998.

- [GOD01] GODERIS, D., et al. **Service Level Specifications Semantics, Parameters and negotiation requirements**. TEQUILA Consortium, draft-tequila-sls-01.txt, Junho 2001. Disponível em <<http://www.ist-tequila.org/>>. Acesso em: 10 novembro 2002.
- [GRO02] GROSSMAN, D. **New Terminology and Clarification for Diffserv**. RFC 3260, Internet Engineering Task Force Abril 2002.
- [HEI99] HEINANEN, J. et al. **Assured Forwarding PHB Group**. RFC 2597, Internet Engineering Task Force, Junho 1999.
- [IAB] The Internet Architecture Board. Disponível em <<http://www.iab.org>>. Acesso em: 07 fevereiro 2002.
- [IANA] Internet Assigned Numbers Authority. Disponível em <<http://www.iana.org>>. Acesso em: 08 fevereiro 2002.
- [IETF] The Internet Engineering Task Force. Disponível em <<http://www.ietf.org>>. Acesso em: 10 fevereiro 2002.
- [INNIC] The Internet Network Information Center. Disponível em <<http://www.internic.net>>. Acesso em: 07 fevereiro 2002.
- [IRTF] The Internet Research Task Force. Disponível em <<http://www.irtf.org>>. Acesso em: 08 fevereiro 2002.
- [ISOC] The Internet Society. Disponível em <<http://www.isoc.org>>. Acesso em: 10 fevereiro 2002.
- [ITU96] ITU-T. **Methods for subjective determination of transmission quality**. ITU-T Recommendation P.800. Agosto 1996
- [JAC99] JACOBSON, V., NICHOLS, K., PODURI, K. **An Expedited Forwarding PHB**. RFC 2598, Internet Engineering Task Force, Junho 1999.
- [MAN02] CISCO Systems. **Managing Real-Time and Peer-to-Peer Traffic**. CISCO Systems, Novembro 2002. Disponível em <http://www.cisco.com/warp/public/779/servpro/promotions/bbip/pdfs/bip_v3.23.pdf>. Acesso em: 25 janeiro 2003.
- [MAR00] MARTINS, J. **Qualidade de Serviço (QoS) em redes IP: Princípios Básicos, Parâmetros e Mecanismos**. UNIFACS, 2000
- [MCT01] MCT. **Internet Comercial – Conceitos, estatísticas e aspectos legais**. Ministério da Ciência e Tecnologia. Abril 2001. Disponível em <<http://www.mct.gov.br>>. Acesso em: 12 julho 2002.

- [MIR02] MIRAS, D., et al. **Network QoS Needs of Advanced Internet Applications-A Survey**. Internet2 - QoS Working Group, Novembro 2002. Disponível em <<http://www.internet2.edu/>>. Acesso em: 25 janeiro 2003.
- [MUR00] MURHAMMER, M.W. et al. **TCP/IP Tutorial e Técnico**. Makron Books. 2000
- [NBA02] CISCO Systems. **Network Based Application Recognition RTP Payload Classification**. CISCO Systems, White Paper, Dezembro 2002. Disponível em <http://www.cisco.com/warp/public/cc/pd/iosw/prodlit/nbarw_wp.htm>. Acesso em: 28 janeiro2003.
- [NBARE] CISCO Systems. **Network-Based Application Recognition and Distributed Network-Based Application Recognition**. CISCO Systems. Disponível em <<http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/122newft/122t/122t8/dtnbarad.pdf>>. Acesso em: 28 janeiro 2003.
- [NIC98] NICHOLS, K. et al. **Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers**. RFC 2474, Internet Engineering Task Force, Dezembro 1998.
- [PRI01] PRIETO, A. G., BRUNNER, M. **SLS to Diffserv configuration mappings**. 12th International Workshop on Distributed Systems: Operations and Management, Outubro 2001.
- [REGBR] Registro BR; FAPESP. Disponível em <<http://www.registro.fapesp.br> >. Acesso em: 07 fevereiro 2002.
- [RET02] RETIERE, C. **Quality of Service – MPLS**. DESS Telecom 2002/2003. Equant, Novembro 2002.
- [ROM00] ROMANO, S., et al. **Service Level Agreements for Premium IP Networks**, draft-cadenus-sla-00.txt, Internet Draft, November 2000.
- [ROS01] ROSEN, E., VISWANATHAN, A., FELDMAN, N. **Multiprotocol Label Switching Architecture**. RFC 3031, Internet Engineering Task Force, Janeiro 2001.
- [SAL00] SALSANO, S., et al. **Definition and usage of SLS in the AQUILA consortium**, draft-salsano-aquila-sls-00.txt, Internet Draft, Novembro 2000.

- [SEM01] SEMERIA, C., STEWART III, J. W. **Supporting Differentiated Service Classes in Large IP Networks**. White Paper - Juniper Networks, Dezembro 2001.
- [SEV01] SEVASTI, A., CAMPANELLA, M. **Service Level Agreements specification for IP Premium Service**, Outubro 2001.
- [SEV02] SEVASTI, A., et al. **SLA Definition for the provision of an EF-based service**, 16th International Workshop on Communications Quality & Reliability, Maio 2002.
- [SOA95] SOARES, L.F.G.; LEMOS, G.; COLCHER, S. **Redes de Computadores das LANS MANS e WANS às Redes ATM**. Editora Campus. 1995.
- [STA99] STARDUST.com. QoSFORUM.com, **Introduction to QoS policies**. White paper, version 3, Setembro 1999
- [STU00] STURM, R., MORRIS, W., JANDER, M., **Foundations of Service Level Management**, Sams, Abril 2000.
- [TRI01] TRIMINTZIOS, P., et al. **An Architectural Framework for Providing QoS in IP Differentiated Services Networks**, 7th IFIP/IEEE International Symposium on Integrated Network management, 2001.
- [VOI02] CISCO Systems. **Voice/Data Integration Technologies**. CISCO Systems, Fevereiro 2002. Disponível em <http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/voicdata.htm>. Acesso em: 23 novembro 2002.
- [XIA99] XIAO, X., NI, L. M. **Internet QoS: A Big Picture**, IEEE Network Magazine, Março 1999.
- [WAA99] WAAIJ, B. D., GAIDUKOV, A.V., JANSEN, J.A. **Internet Next Generation Management: Quality-Based Service Management**. KPN Research, Dezembro 1999.
- [WIL00] WILLRICH, R. **Sistemas Multimídia Distribuídos**. Universidade Federal de Santa Catarina. Agosto 2000.

Anexo A - Programação do roteador de borda

No roteador de borda (ingresso dos pacotes), a programação para classificação de pacotes será feita de duas formas distintas, a título de exemplo. O mecanismo Access-lists pode considerar informações como protocolo de nível 4, portas TCP/UDP utilizadas e os endereços IP de origem/destino. O mecanismo NBAR (Network Based Application Recognition) analisa o protocolo das camadas 4 a 7 para identificar os tipos de pacotes, inclusive de aplicações mais avançadas.

A seguir são apresentados os comandos básicos para implementar os SLS's definidos no estudo de caso do capítulo 6.

```

/*definição de três listas de acesso (Access-lists). www substitui o número da porta = tcp 80*/
access-list 101 permit tcp any any eq www
access-list 102 permit tcp any any eq smtp
access-list 103 permit tcp any any eq telnet

/*pacotes identificados são relacionados à classes. Pacotes de áudio à classe EF, por exemplo*/
class-map match-any EF
    match protocol rtp audio          /*NBAR - identifica áudio dentro do protocolo RTP*/
class-map match-any AF41
    match protocol rtp video          /*NBAR - identifica vídeo dentro do protocolo RTP*/
class-map match-any AF31
    match access-group 101
class-map match-any AF21
    match access-group 102
class-map match-any AF11
    match access-group 103

/*policiamento de tráfego e marcação de pacotes*/
/*tráfego EF até 159Kbit/s tem os pacotes marcados com DSCP 46. Acima de 159 Kbit/s são descartados*/
/*tráfego AF21 até 100Kbit/s com burst de 1200 bytes é marcado com DSCP 18. Se burst até 2400 (1200+1200) bytes, é marcado com nível menor=DSCP 20. Tráfego acima será descartado*/
policy-map CONDICIONA-E-MARCA-DSCP
    class EF
        police 159000 0 0 conform-action set-dscp-transmit 46 exceed-action drop
    class AF41
        police 512000 0 0 conform-action set-dscp-transmit 34 exceed-action drop
    class AF31
        set ip dscp 26          /*não policia, apenas marca DSCP com 26*/
    class AF21

```

```

        police 100000 1200 1000 conform-action set-dscp-transmit 18 exceed-
        action set-dscp-transmit 20 violate-action drop
    class AF11
        police 100000 0 0 conform-action set-dscp-transmit 10 exceed-action
        drop
    class class-default
        set ip dscp 0      /*melhor esforço para pacotes de nenhuma classe anterior*/

/*ativa a regra CONDICIONA-E-MARCA-DSCP na direção de entrada da interface ethernet, ou seja,
para os pacotes que estão entrando no roteador de borda*/
interface FastEthernet0/0
    service-policy input CONDICIONA-E-MARCA-DSCP

/*associa o fluxo BA ao nome da classe de serviço*/
class-map match-all supervip
    match ip dscp 46
class-map match-all vip
    match ip dscp 34 36 38
class-map match-all ouro
    match ip dscp 26 28 30
class-map match-all prata
    match ip dscp 18 20 22      /*define os PHB's AF21, AF22 e AF23*/
class-map match-all bronze
    match ip dscp 10 12 14
class-map match-all melhor-esforço
    match ip dscp 0

/*regras que definem o PHB de cada serviço, conforme as necessidades do cliente*/
/*uma fila prioritária de até 159 Kbit/s é reservada para a classe supervip - EF*/
/*outras classes de serviço(AF), são definidas em termos de percentual da banda disponível na interface.
30% da banda disponível será reservada para a classe vip, em caso de congestionamento*/
policy-map TRATA-TRAFEGO
    class supervip
        priority 159      /*define uma fila prioritária com banda garantida até 159Kbit/s*/
    class vip
        bandwidth percent 30      /*define maior prioridade=30%, entre as classes AF*/
    class ouro
        shape peak 400000      /*conforma o tráfego Ouro em 400Kbit/s de pico*/
        bandwidth percent 20
    class prata
        bandwidth percent 15
    class bronze
        bandwidth percent 10
    class melhor-esforço      /*limita o tráfego de melhor esforço em 64 Kbit/s*/
        police 64000 400 0 conform-action transmit exceed-action drop

/*ativa a regra TRATA-TRÁFEGO na direção de saída da interface serial. Pacotes saintes*/
interface Serial0/0
    service-policy output TRATA-TRAFEGO

```