

# Uma Metodologia para Estimativa de Necessidades em Contratos de Canais de Dados com Garantia de Níveis de Serviço

Luis Felipe M. de Moraes<sup>1</sup>, Marcos Cesar Pinto<sup>1</sup>

<sup>1</sup>Laboratório de Redes de Alta Velocidade - RAVEL  
Programa de Engenharia de Sistemas e Computação - COPPE/UFRJ  
Caixa Postal: 68.511 - 21941-972 - Rio de Janeiro - RJ

{moraes, mcesar}@ravel.ufrj.br

**Abstract.** *The studies on measurement and traffic monitoring in general supply information to help the telecommunications suppliers to place the resources of its backbones in order to take care of its customers networks traffic demands. There is, however, an aspect of this issue that needs to be more accurately evaluated and represents the right definition, on the part of the customer, of these necessities. This work presents a simple methodology to predict the traffic behavior that will pass through a data channel to be rented, allowing the customer to establish the band width necessary to drain it in order to support the quality requirements that can be demanded.*

**Resumo.** *Os trabalhos sobre medição e monitoração do tráfego em geral fornecem informações para que os provedores de telecomunicações possam alocar os recursos de seus backbones de modo a atender às necessidades de escoamento do tráfego das redes de seus clientes. Há, entretanto, um aspecto desta questão que precisa ser mais bem avaliado e corresponde a adequada definição, por parte do cliente, destas necessidades. Este trabalho propõe uma metodologia simples para estimar o comportamento do tráfego que irá circular por um canal de dados a ser contratado, permitindo ao cliente decidir sobre a largura de banda necessária para escoá-lo de forma a atender os requisitos de qualidade que venham a ser exigidos.*

## 1. Introdução

Uma das principais tendências globais na gerência da tecnologia de informação (TI) consiste na terceirização de serviços e produtos através de contratos com parceiros externos [Demirkan et al. 2005]. Como o objeto de tais contratos é de natureza essencialmente técnica, seus parâmetros de performance são medidos objetivamente através de métricas conhecidas como níveis de serviço, o que os levou a serem chamados de Acordos de Nível de Serviço ou *Service Level Agreements* (SLA). Basicamente, um acordo deste tipo é composto dos seguintes itens [Verma 2004]:

- A descrição da natureza do serviço a ser fornecido;
- Os níveis de serviço exigidos;
- Os procedimentos de monitoração e geração de relatórios sobre o tráfego;
- O tempo máximo tolerável para a solução de problemas; e

- Multas e outras penalidades incidentes no caso do não cumprimento das cláusulas contratuais.

O foco deste estudo está nos acordos de nível de serviço presentes nos contratos de aluguel de canais de transmissão de dados, mais especificamente *backbones* IP, celebrados entre provedores de telecomunicações e clientes corporativos que precisam interligar unidades funcionais de suas empresas.

O tráfego de dados em redes IP agrega uma série de tipos de fluxos, oriundos de aplicações diversas. Assim sendo, é desejável classificar os fluxos, agrupando-os em classes com características semelhantes ou com o mesmo grau de prioridade de acordo com as regras de negócio da empresa. Para cada classe, diferentes níveis de serviço serão atribuídos, de acordo com a qualidade de serviço (QoS) que se julgue necessária. Os níveis de serviço característicos na contratação de um canal de dados são: a disponibilidade do canal, o atraso (tempo decorrido desde a geração de um pacote na fonte até a sua recepção no destino), o *jitter* (que representa a variação do atraso), a taxa de perda de pacotes e a vazão [Filsfil and Evans 2005]. Definidos os níveis de serviço para cada classe de aplicações, o provedor estabelece a capacidade total do canal e as reservas de largura de faixa para cada uma das classes [Filsfil and Evans 2005], de tal forma a garantir os níveis (qualidade) exigidos. Este princípio motivou uma série de estudos direcionados para a otimização dos recursos do provedor (equipamentos de conectividade e largura de banda de seus links) de forma a garantir a oferta de serviços dentro dos níveis estabelecidos nos contratos. O problema, e que representa o aspecto fundamental abordado neste artigo, diz respeito à forma como estes níveis de serviço, a serem definidos pelo cliente, influenciam a capacidade total do canal a ser contratado e as reservas de largura de faixa a serem estabelecidas. Há três questões a considerar:

- Os recursos financeiros disponíveis para arcar com as despesas do contrato;
- As exigências de qualidade intrínsecas de certos tipos de fluxos (as aplicações multimídia são o caso típico); e
- A prioridade que o cliente atribui a cada fluxo e que está relacionada com as regras de negócio específicas da empresa.

Ou seja, a definição clara do quê especificamente o cliente precisa, em termos da qualidade do serviço a ser prestado, precisa ser estudada a fundo, a exemplo do que já é feito sobre o provisionamento de recursos [Andreas Schmietendorf and Reitz 2004] [Wissam Fawaz and Du-Pont 2004]. A Figura 1 apresenta esquematicamente os passos para a elaboração e condução de SLA. A ênfase deste trabalho, portanto, está na etapa de número 1.

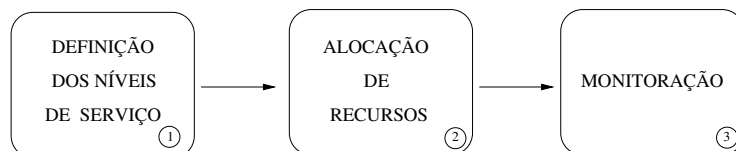


Figura 1. Etapas para a elaboração e condução de SLA

Para atender a esta necessidade, este estudo propõe uma análise estatística prévia do tráfego entre as unidades de uma empresa, de forma a estabelecer, através de uma metodologia simples, a capacidade do canal a ser contratado e as reservas de largura de

faixa, de forma a atender os níveis de serviço desejáveis e ao mesmo tempo possíveis para o fluxo total de dados e para cada categoria de aplicações. Consideraram-se apenas dois níveis de serviço: a probabilidade de perda de pacotes (ou bits) e a vazão. Os motivos são os seguintes:

- A disponibilidade não está relacionada com as características do tráfego e sim com falhas nos equipamentos dos provedores;
- O atraso e o *jitter*, ainda que influenciados diretamente pela capacidade ou velocidade de transmissão do canal, dependem de diversos outros fatores, como o tamanho das filas nos roteadores ou switches, a topologia do *backbone* do provedor e seus algoritmos de roteamento, a arquitetura de qualidade de serviço adotada (normalmente *DiffServ*), a capacidade de processamento dos nós da rede, entre outros. Todos estes fatores independem da natureza do tráfego e constituem características técnicas a serem manipuladas pelo próprio provedor; e
- A vazão e a taxa de perda de pacotes definem uma primeira estimativa da capacidade do canal a ser alugado. Numa etapa seguinte, se os valores toleráveis de atraso e *jitter* não forem alcançados, pequenos ajustes ainda poderão ser feitos, não necessariamente associados a aumento da capacidade do canal ou mudança nas políticas de reserva de largura de faixa [Aniruddha Diwan and Kumar 2002] [Van der Wal 1997].

O restante deste artigo está estruturado, como segue. A seção 2 apresenta as etapas do método proposto, acrescidas das observações pertinentes. Um exemplo de aplicação da técnica é abordado na seção 3, na forma de um estudo de caso. A quarta parte do trabalho valida os resultados encontrados e a última seção expõe as conclusões e propostas de trabalhos futuros.

## 2. Metodologia

### 2.1. Princípios

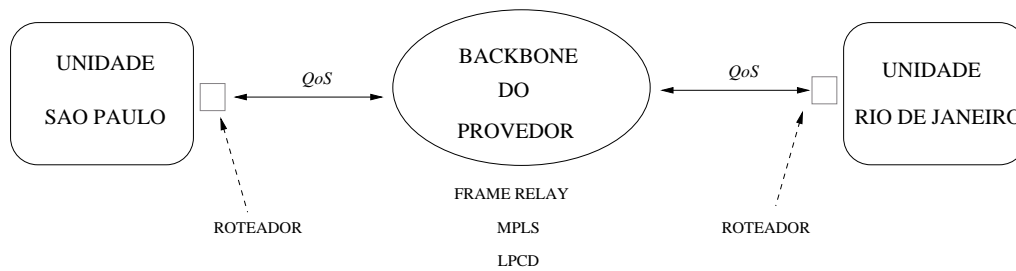
Os seguintes princípios orientaram a elaboração da metodologia: simplicidade, redução do desperdício e opção pelo resultado mais conservador na ocorrência de qualquer incerteza sobre o valor estimado para a capacidade do enlace.

### 2.2. Cenário

A metodologia proposta neste estudo se aplica a um único enlace ponto-a-ponto (na visão do cliente), interligando duas unidades funcionais de uma empresa (Figura 2). A técnica independe da estrutura do *backbone* do prestador de serviços (*Frame Relay*, *MPLS*, *LPCD*, etc.), uma vez que não cabe ao cliente, em princípio, avaliar as vantagens técnicas de cada uma delas. O trabalho considera que a empresa contratante negocia um acordo com o provedor para que um canal de dados provisório seja instalado com capacidade muito acima das necessidades de largura de banda oferecidas em situações semelhantes. O canal deverá permanecer ativo durante um período de tempo previamente acordado, suficiente para que sejam feitas as coletas e análises necessárias. O adequado dimensionamento deste período será objeto de estudo de seções subsequentes.

### 2.3. Etapas

A metodologia proposta pode ser dividida em quatro etapas fundamentais:



**Figura 2. Enlace entre duas unidades de uma corporação**

1. Definição das classes de tráfego;
2. Coleta de dados;
3. Filtragem dos dados coletados;
4. Análise estatística dos dados; e
5. Definição da capacidade do canal a ser alocada para cada categoria de tráfego.

### 2.3.1. Definição das classes de tráfego

Esta etapa é a que exige a maior interação com o cliente. É preciso conhecer todas as aplicações utilizadas no ambiente corporativo, listá-las e identificar os números IP e portas correspondentes. Em seguida, é preciso que o cliente estabeleça o grau de importância de cada aplicação, de acordo com seu valor estratégico. As prioridades podem ser representadas por números ou pesos, num escala que vai de 1 a 5, por exemplo. Esta avaliação pode ser muito subjetiva ao ponderar critérios como segurança, custos, imposições contratuais, entre outros e pode apresentar diferenças significativas entre empresas. O último passo consiste em agrupar estas aplicações em classes de tráfego, de acordo com as prioridades já definidas pelo cliente juntamente com as restrições técnicas inerentes às mesmas. Neste estudo, por exemplo, foram consideradas três classes de aplicações, como será apresentado no estudo de caso da Rede Rio.

### 2.3.2. Coleta de dados

Uma vez definidas as classes ou categorias de aplicações, a etapa seguinte consiste em extrair as informações relevantes sobre o tráfego no enlace, que neste estudo correspondem aos valores observados da vazão. Um dos softwares utilizados para processar informações sobre o tráfego é o NetFlow, da Cisco [Netflow], que efetua amostras periódicas do tráfego, compila os dados correspondentes no próprio roteador e os exporta de tempos em tempos para uma unidade de armazenamento e processamento (estação de trabalho ou servidor) que recebe o nome de coletor. No coletor, outros aplicativos são utilizados com o objetivo de formatar os dados acumulados referentes às amostras para posterior tratamento na etapa de análise. As amostras devem ser realizadas por meio de coleta passiva, em intervalos de tempo constantes, cuja duração deve levar em conta alguns fatores, abaixo listados:

- Volume de tráfego esperado;
- Recursos computacionais disponíveis; e

- Período de tempo acordado com o provedor para a realização da coleta.

Intervalos muito pequenos podem levar a um volume de dados muito grande, impossível de ser manipulado se o coletor não for suficientemente robusto. Além disso, é preciso evitar a amostragem destrutiva [Töpke 2001] que ocorre quando o fluxo de dados entre o roteador e o coletor tornar-se significativo se comparado com os fluxos efetivos, podendo ter influência nos resultados das estimativas.

### 2.3.3. Filtragem dos dados coletados

A filtragem das amostras tem por objetivo selecionar os dados que efetivamente representam a demanda esperada no enlace. Dois aspectos precisam ser considerados: o período (instante de tempo) a que se referem as amostras (hora, dia do mês ou da semana) e a identificação de valores inconsistentes, resultantes de problemas como falta de energia ou interrupção dos serviços no roteador ou coletor. Com relação ao instante de leitura dos dados, não podem ser consideradas, por exemplo, amostras obtidas fora do horário do expediente, referentes a aplicações utilizadas somente neste período. Da mesma forma, finais de semana e feriados devem ser analisados separadamente. Podem existir, entretanto, programas que funcionem ininterruptamente, 24 horas por dia, todos os dias da semana, para os quais todo o universo das amostras precisa ser considerado.

### 2.3.4. Análise estatística dos dados

O passo seguinte consiste em definir o comportamento estatístico do tráfego no enlace e com base nessa informação, estabelecer a capacidade do canal ( $C$ ) que garanta uma probabilidade de perda ( $P_{perda}$ ) inferior a um limite pré-estabelecido (nível de serviço a ser acordado). Este procedimento deve ser realizado para o tráfego como um todo e repetido, separadamente, para cada classe de fluxos. É importante mencionar que o conceito de probabilidade de perda não pode ser confundido com taxa média de perda. No primeiro caso, considera-se a fração de tempo em que uma parte (não especificada) dos dados foi perdida, enquanto no segundo, mede-se a proporção entre o volume efetivo de pacotes enviados e descartados, denominado *workload loss ratio* (WLR) [BÍRÓ 2004]. Assim:

$$P_{perda} = P(\text{ocorrer perda}), \quad WLR = \frac{E[(X - C)^+]}{E[X]}$$

onde o operador  $E[ ]$  representa a média estatística da variável aleatória considerada.

Para calcular  $C$ , alguns estudos [Cao et al. 2001] sugerem que a probabilidade de saturação do link, definida como  $P_{sat} \triangleq P(X > C)$  pode ser considerada uma boa aproximação para  $P_{perda}$ , desconsiderando o efeito das filas nos nós da rede. Assim sendo, o método consiste em estimar a função de distribuição de probabilidade  $F_X(x)$  que caracteriza adequadamente o tráfego, e com base em um valor de  $P_{perda}$  previamente definido, calcular a correspondente capacidade  $C$  do canal:

$$F_X(x) \triangleq P(X \leq x) \tag{1}$$

$$P_{perda} \approx P_{sat} = 1 - F_X(C) \tag{2}$$

Para estimar  $F_X(x)$ , com base nas amostras coletadas, cada valor único de vazão observado ( $x_i$ ) deve ter a ele associada uma frequência relativa ( $f_i$ ) que corresponde ao número total de ocorrências de  $x_i$  dividido pelo número total de amostras  $n$ . Em seguida, calcula-se a função de distribuição cumulativa empírica dos dados observados ou *empirical cumulative distribution function* (ecdf), uma função escada, com saltos de altura  $f_i$ , em cada um dos pontos  $x_i$  (Figura 3). Quando  $n$  tende para infinito, a ecdf converge para função de distribuição cumulativa da vazão. Finalmente, define-se um valor para  $C$ , através de um nível de confiança a ser determinado (95%, por exemplo), de tal forma que a probabilidade de perda seja menor ou igual ao percentual máximo estabelecido (Figura 4).

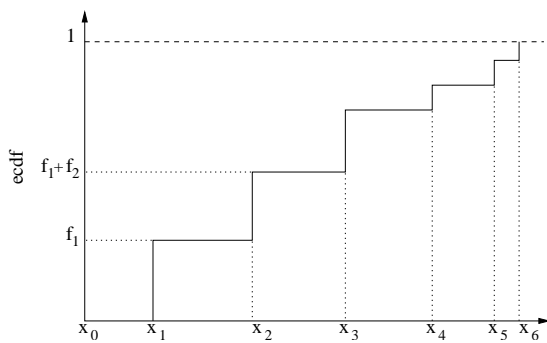


Figura 3. Função de distribuição cumulativa empírica

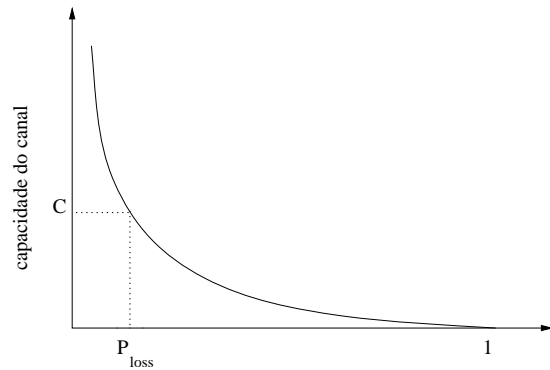


Figura 4. Capacidade x Probabilidade de perda

Considerações e Hipóteses Simplificadoras:

- A aplicação da técnica sugerida deveria partir do pressuposto de que as amostras são independentes e identicamente distribuídas (iid), o que não é o caso. Porém, assumir esta hipótese simplificadora reduz significativamente a complexidade do tratamento dos dados. Concessão semelhante é aplicada na modelagem do intervalo de chegada entre pacotes na porta de entrada de um roteador; e

Cálculo do intervalo de confiança:

O intervalo de confiança dos valores estimados para a capacidade do canal pode ser definido como  $(C - \varepsilon, C + \varepsilon)$ , sendo  $\varepsilon$  a margem de erro. A fórmula matemática do valor de  $\varepsilon$ , com base no desvio padrão amostral ( $V_n$ ), no número de amostras ( $n$ ) e em um nível de confiança  $(1 - \alpha) \times 100\%$ , pode ser obtida através da equação 3, considerando uma distribuição normal.

$$\varepsilon = \pm \frac{z_{\alpha/2, n-1} * V_n}{\sqrt{n}} \text{ onde } z \sim N(0, 1) \tag{3}$$

Como a variável aleatória em questão (vazão no canal) em princípio não é Gaussiana, uma possível solução para o cálculo do intervalo de confiança consiste em reunir os dados coletados em  $n_g$  grupos, cada qual com um número suficientemente grande de amostras. Para cada grupo, equivalente, por exemplo, à coleta de um dia, obtém-se a ecdf correspondente. A partir daí, a variável aleatória a ser considerada passa a ser a função

*quantile*,  $Q(q)$ , referente a cada grupo (ou dia), assim definida:

$$\text{(dado } q, \text{ o menor valor de } x \text{ tal que)} \quad P(X \leq x) > q \quad (4)$$

onde  $q$  representa o complemento das possíveis probabilidades de perda consideradas (10%, 1% e 0,1%, por exemplo)

O valor da função *quantile*, em cada ponto  $q$  considerado, pode ser tratado como uma variável gaussiana, uma vez que representa o valor médio de um número grande de amostras (teorema do limite central), o que permite a aplicação da equação 3, considerando o desvio padrão amostral do valor de  $Q(q)$  entre os  $n_g$  grupos criados.

Será necessário aplicar o método a um número inicial de amostras e calcular o valor proporcional do erro para cada valor de perda considerado, ou seja,  $\varepsilon/Q(q)$ . Se o erro for muito grande, o número de amostras terá sido insuficiente e mais dias de coleta serão necessários.

### 2.3.5. Definição da capacidade do canal a ser alocada para cada categoria de tráfego

A decisão do cliente sobre a capacidade total do link a ser contratado deve ser assessorada pelos técnicos que conduziram os testes. A escolha estará condicionada aos custos de aluguel do link, contrapostos às probabilidades de perda aceitáveis. Assim sendo, monta-se uma tabela com os valores sugeridos para  $P_{perda}$  e as correspondentes capacidades de transmissão exigidas, acrescida dos custos associados. Como não há um critério absoluto para tomada desta decisão, uma possível alternativa seria garantir os menores níveis de perda, alocando a banda correspondente, para as categorias de fluxos consideradas prioritárias. Quanto à capacidade total do enlace a ser contratado, sugere-se um nível de serviço mais frouxo, mas que ainda assim preserve um mínimo de qualidade para as classes de menor importância. O estudo de caso da seção seguinte apresenta um exemplo concreto que facilita o entendimento da idéia aqui proposta.

Um último detalhe precisa ser destacado. Aparentemente, a demanda total calculada para o enlace  $C_T$ , considerando um determinado valor de  $P_{perda}$ , poderia ser obtida através do somatório das demandas de cada classe ( $\Sigma C$ ). Entretanto, estes valores são distintos ( $C_T < \Sigma C$ ), isto porque a adição pura e simples elimina o efeito de suavização observado quando variáveis aleatórias são sobrepostas. Ou seja, quando um único canal atende a um número grande de fontes de tráfego, a tendência é a maior probabilidade de suportar as eventuais rajadas, mesmo quando a capacidade total alocada corresponde apenas a média do somatório das fontes [Kleinrock 1976]. As vantagens da multiplexação estatística do uso de um canal de dados, quando comparada ao somatório simples da largura de faixa efetiva de cada uma das fontes (ou conjunto de fontes, neste estudo), também é apresentada em [Choudhury 1996].

## 2.4. Fluxograma

A metodologia proposta neste estudo está representada esquematicamente na Figura 5, que apresenta de forma seqüencial, todas as etapas descritas nos itens anteriores.

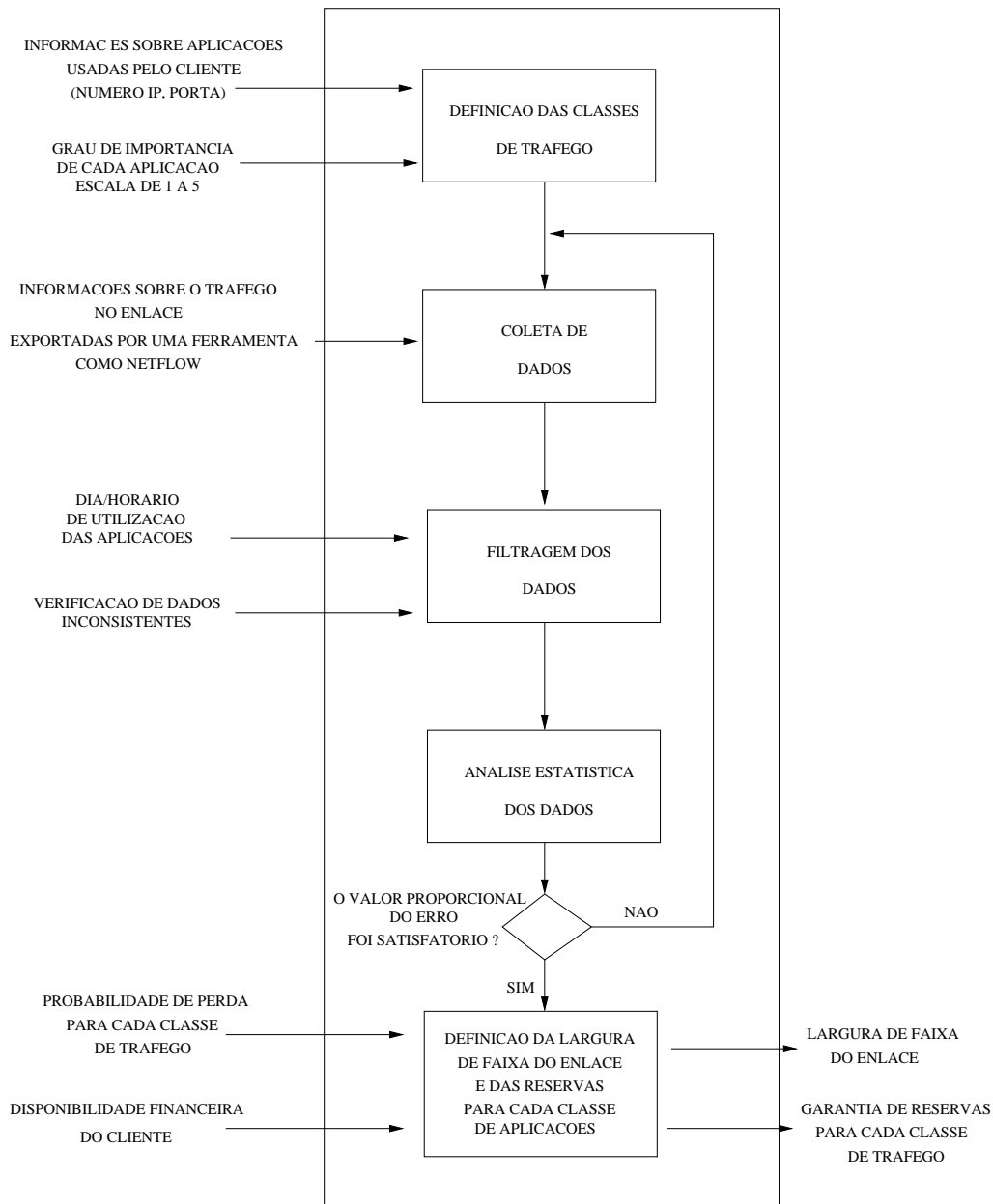


Figura 5. Fluxograma

## 3. Estudo de Caso: Rede Rio (RR)

O estudo de caso que será apresentado não corresponde fielmente ao cenário proposto para a aplicação da metodologia. Entretanto, por tratar-se de um enlace digital de dados, por onde passam diferentes tipos de fluxo, é razoável utilizá-lo como exemplo, imaginando que os pacotes estariam trafegando em um enlace corporativo ponto-a-ponto



entre a RR e a Internet. Este estudo se aproveitou da estrutura já existente para coleta do tráfego na Rede Rio (RR) e que representa uma das linhas de pesquisa do Laboratório de Redes de Alta Velocidade (Ravel) da COPPE-UFRJ, conhecido como IPTRAF [IPTRAF]. A seguir, serão apresentados todos os passos da metodologia, aplicados ao caso:

### 3.1. Definição das classes de tráfego

Como não havia a figura de um cliente que pudesse priorizar os fluxos de dados de acordo com as regras de negócio da empresa, as classes de tráfego foram estabelecidas levando-se em conta somente suas características intrínsecas. Foram considerados quatro tipos de aplicações, identificados como o auxílio da relação de portas conhecidas e registradas no *Internet Assigned Numbers Authority* [IANA ]:

- **Classe 1 (aplicações adaptativas ou elásticas):** são as aplicações TCP genéricas, que possuem controle de fluxo e congestionamento.
- **Classe 2 (aplicações não adaptativas):** correspondem, no caso geral, aos fluxos UDP. Destacam-se os serviços *Peer-to-Peer* (P2P) e de mensagens instantâneas.
- **Classe 3 (aplicações em Tempo Real):** aplicações de voz e vídeo em tempo real. Programas deste tipo são mais exigentes com relação aos valores máximos tolerados de atraso e *jitter*.

*Problemas Encontrados:* A principal dificuldade encontrada na etapa de classificação dos fluxos na RR foi o fato de que um percentual muito grande do tráfego (cerca de 30%) não faz parte de nenhuma das classes pré-definidas. Ou seja, considerando-se o tráfego referente às portas das aplicações selecionadas, ainda restava uma parcela significativa dos fluxos não identificada. Além disso, a distribuição dos pacotes entre portas, a partir da porta 8000 (aproximadamente) é praticamente uniforme, como mostra a Figura 6. Ou seja, o tráfego não identificado estava espalhado em mais de 50000 portas diferentes, não sendo possível identificar que tipos de aplicações seriam responsáveis pela geração destes fluxos. A solução, pelo critério de simplicidade, foi considerá-los apenas como "outros tipos de fluxo", alocando-os na categoria 4.

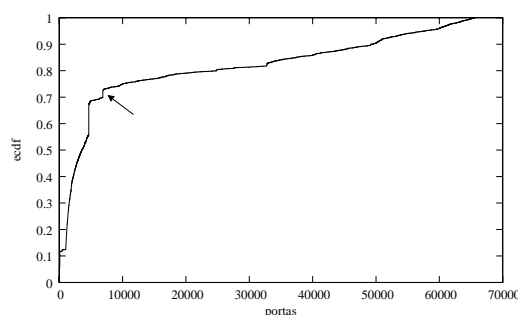


Figura 6. Distribuição do tráfego por portas

### 3.2. Coleta de dados no enlace a ser contratado

Foram recolhidas 10868 amostras para cada classe de aplicações e para o tráfego total. A Figura 7 exhibe um gráfico relativo a um único dia de coleta, contendo as curvas do tráfego total e de cada uma das classes que permite chegar a algumas conclusões:

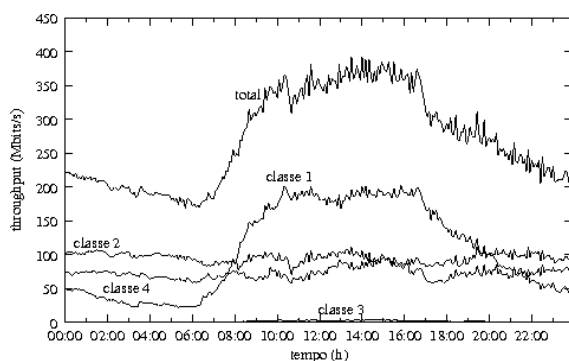


Figura 7. Tráfego por classes em um dia de coleta

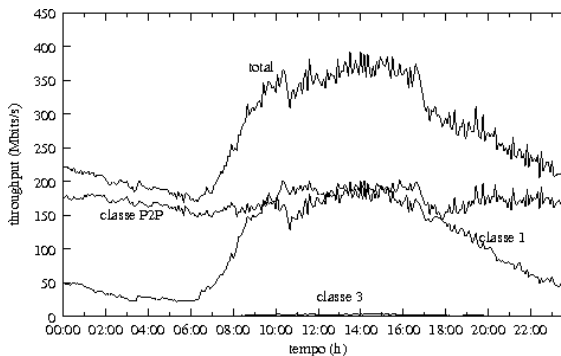


Figura 8. Tráfego por classes em um dia de coleta (P2P agregado)

- A semelhança entre as curvas das classes 2 e 4 sugere que há alguma relação entre os dois tipos de fluxo, especialmente fora do horário do expediente. Para verificar a existência deste vínculo, calculou-se o coeficiente de correlação de *Pearson* (equação 5) entre estes fluxos, num período de seis dias. Tomando-se as amostras ao longo de todo o dia, o valor obtido foi de 0,47. Entretanto, ao desconsiderar as leituras realizadas entre às 10:00h e às 16:00h, este valor chegou a 0,80. O mesmo coeficiente se calculado entre os fluxos das categorias 1 e 2, e entre as classes 1 e 4, produz resultados da ordem de  $10^{-5}$ . Portanto, o tráfego que não havia sido enquadrado nas demais classes foi considerado como decorrente (ainda que parcialmente) de aplicações P2P e as classes 2 e 4, reunidas numa mesma categoria (Figura 8);

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (5)$$

- O tráfego gerado pelos fluxos P2P apresenta um comportamento uniforme ao longo de todo o dia, com pouca alteração entre o início e o término do expediente. A razão para tal fato reside no uso ininterrupto de aplicações *peer-to-peer*, uma vez que não é necessária a presença de um usuário diante do computador para fazer o *download* de músicas e vídeos. Ou seja, as máquinas permanecem ligadas 24 horas por dia, baixando arquivos.

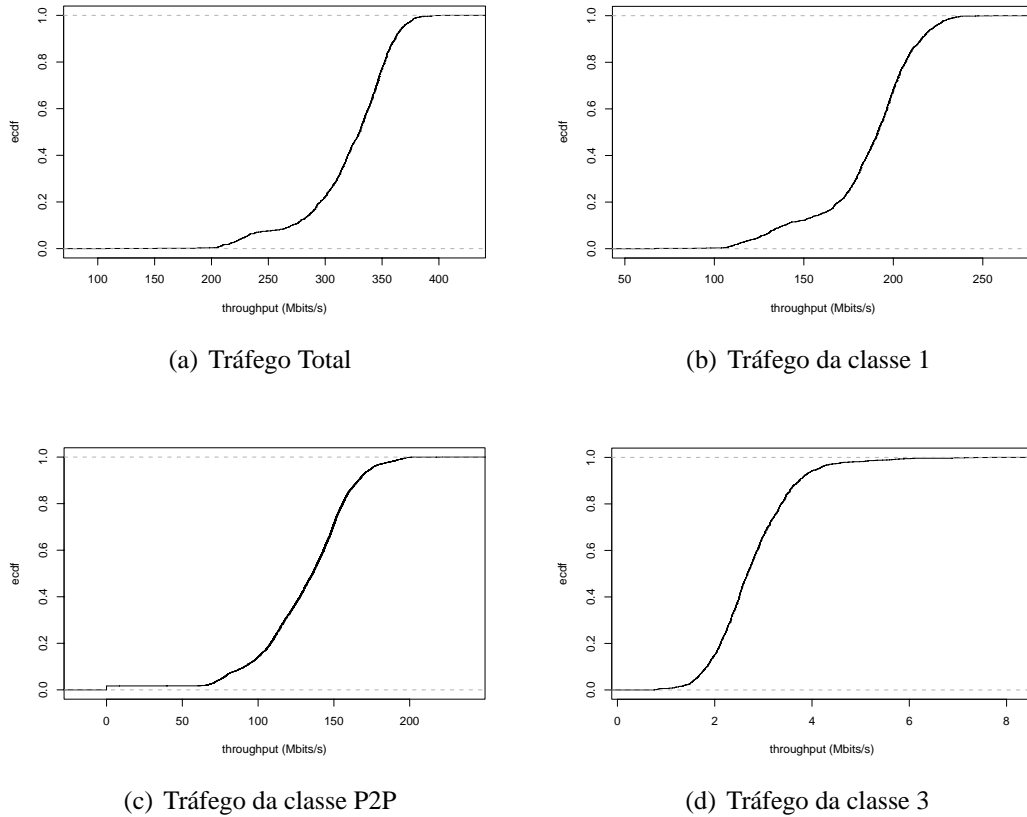
### 3.3. Filtragem dos dados coletados

Levando-se em conta as observações realizadas na etapa de coleta, os seguintes critérios foram estabelecidos para filtragem das amostras:

1. Para os fluxos das classes 1 e 3, somente as amostras coletadas no horário normal de trabalho foram consideradas, mais especificamente entre as 10:00h e as 16:00h, período em que as curvas do tráfego apresentam-se mais estáveis. Feriados e finais de semana também foram descartados; e
2. Todas as amostras da categoria P2P foram aproveitadas.

### 3.4. Análise estatística dos dados

Seguindo o procedimento definido pela metodologia, foram traçadas as curvas da ecdf relativas ao tráfego total e às três categorias de fluxo (Figura 9), com o auxílio da ferramenta R [R Development Core Team 2005].



**Figura 9. Ecdf das classes de tráfego**

A capacidade do canal e a margem de erro  $\varepsilon$  foram calculadas para cada classe de tráfego e para cada probabilidade de perda assumida. O número de grupos  $n_g$ , foi de 20 para as classes 1 e 3, e 38 para a categoria P2P, de acordo com os dias considerados válidos na etapa de filtragem. Os resultados estão nas Tabelas 1 e 2. A última etapa da metodologia representa a definição da capacidade necessária do canal, para cada classe de tráfego, de acordo com a máxima probabilidade de perda desejada. A Tabela 1 apresenta as capacidades calculadas, considerando os *quantiles* correspondentes nas curvas de distribuição. O valor proporcional da margem de erro, obtida em cada caso, encontra-se na Tabela 2.

**Tabela 1.**

CLASSES	CAPACIDADE DO CANAL (MBITS/S)		
	PROBABILIDADE DE PERDA		
	10%	1%	0,1%
total	360,44 ± 4,61	378,01 ± 9,03	384,28 ± 29,26
classe 1	215,10 ± 3,45	230,26 ± 8,64	234,77 ± 23,69
classe P2P	165,01 ± 0,64	192,15 ± 4,23	198,21 ± 15,56
classe 3	3,73 ± 0,01	5,14 ± 0,07	6,00 ± 0,24

Como já fora observado, a simples adição das demandas por classe, considerando, por exemplo, perda inferior a 0,1%, excede o valor calculado quando se considera o trá-

Tabela 2.

CLASSES	$\varepsilon$ / (capacidade estimada do canal) PROBABILIDADE DE PERDA		
	10%	1%	0,1%
total	1,27%	2,33%	7,61%
classe 1	1,6%	3,75%	10,01%
classe P2P	0,3%	2,2%	7,84%
classe 3	0,22%	1,36%	4,10%

fego como um todo. No primeiro caso, obtém-se uma demanda de 478,47 Mbits/s, enquanto para o tráfego total, o valor calculado seria de 413,54 Mbits/s.

Assim sendo, considerando que os fluxos adaptativos e em tempo real são prioritários se comparados com o tráfego P2P, as classes 1 e 3 poderiam ter garantidas para si:

classe 1: 258,46 Mbits/s  
 classe 3: 6,24 Mbits/s  
 total prioritário: 264,70 Mbits/s

Quanto a capacidade total do enlace a ser contratado, considerando uma probabilidade de saturação de até 10%, o valor de 365,05 Mbits/s seria suficiente. Este valor, se comparado com os 413,54 Mbits/s necessários para garantir 0,1% de perda para o tráfego total, representa uma redução de aproximadamente 12%.

#### Observações:

- Os valores calculados são fracionários. Na prática, os enlaces são disponibilizados em valores inteiros, múltiplos de uma unidade básica que corresponde a menor capacidade de uma das portas de um roteador. Desta forma, os resultados precisam ser adaptados para um contexto real. Um cenário possível seria o da contratação de um link de 400 Mbits/s, ao invés de 500Mbits/s, ou mesmo 1Gbit/s;
- Os erros encontrados nos cálculos da demanda, associados à perda não superior a 0,1% chegaram, no pior caso, a 10% (Tabela 2). Este percentual tende a diminuir, considerando um número maior de dias para coleta dos dados na rede;
- Os fluxos da categoria P2P serão os maiores prejudicados com a implementação das reservas de banda para as classes 1 e 3. Entretanto, tais fluxos não são prioritários e a consequência das restrições impostas não passa de um tempo maior para o *download* de arquivos;

## 4. Validação

Para validar os números sugeridos pela metodologia, foram analisados os 18 dias seguintes aos 20 primeiros tratados na coleta inicial. A comparação entre os valores sugeridos e os resultados obtidos encontra-se na Tabela 3:

Os números da Tabela 3 indicam que as estimativas foram coerentes e flutuaram dentro da margem de erro considerada, à exceção dos fluxos em tempo real. No caso do tráfego total, por exemplo, considerando o limite inferior do intervalo de confiança (355,83 Mbits/s), as amostras que excederam este valor não passaram de 8,88%.

**Tabela 3. Comparação entre a previsão de saturação e os resultados observados**

	capacidade sugerida Mbits/s	máx. probabilidade saturação prevista	saturação observada
total	365,05	10%	3,02%
classe 1	258,46	0,1%	0,0%
classe 3	6,24	0,1%	0,22%

Para a classe 1, entretanto, a opção pelo valor médio calculado para uma perda de 0,1% (234,77 Mbits/s), desconsiderando a margem de erro, resultaria num percentual de perda de 0,38%, acima portanto do estabelecido. A categoria 3, por sua vez, ficou fora dos limites calculados. Para as aplicações multimídia, apenas uma reserva de 6,82 Mbits/s garantiria uma perda não superior a 0,1%. O erro da previsão dos fluxos de voz e vídeo tem duas possíveis causas. A primeira delas está no fato de possuírem um valor médio bem inferior às demais categorias, o que sugere um intervalo de amostragem menor do que os 5 minutos considerados. Além disso, a cauda da distribuição empírica da classe 3 é a mais longa de todas (Figura 9). Ainda que o erro proporcional do *quantile* correspondente a 0,1% tenha sido de apenas 4,10% (Tabela 2), possivelmente um número maior de amostras seja necessário.

## 5. Conclusão e Trabalhos Futuros

O principal objetivo deste estudo foi demonstrar a importância da adequada estimativa de necessidades em SLA. Não faz sentido desenvolver mecanismos sofisticados para garantir qualidade de serviço se os clientes não conhecem as características do tráfego gerado por suas aplicações e nem tampouco tem a oportunidade de priorizar os fluxos mais relevantes segundo as regras de negócio da empresa. A abordagem proposta é simples, mas os resultados apresentados demonstram que pode vir a ser o ponto de partida para estudos mais avançados. A adoção do critério da probabilidade de saturação (ou perda), juntamente com o traçado das curvas de distribuição cumulativa empírica (ecdf) mostraram-se adequados, ainda que sejam necessários complementos relativos aos valores tolerados de atraso e *jitter*. A opção por uma abordagem estatística ao invés da tentativa de definição de modelos analíticos é coerente com o princípio adotado da simplicidade, uma vez que não é necessária uma precisão rígida dos resultados encontrados. Propostas de trabalhos futuros correspondem a análises semelhantes feitas em períodos mais longos, a avaliação do comportamento do atraso e do *jitter*, a tentativa de identificar a tendência de crescimento do tráfego para o planejamento de contratos futuros e a análise dos fluxos no contexto da camada de aplicação e não mais considerando apenas o par IP/porta como critério suficiente para classificação dos tráfegos.

## Referências

- Andreas Schmietendorf, R. D. and Reitz, D. (2004). Sla Management - Challenges in the Context of Web-Service-Based Infrastructures. In *IWCS*, page 606. IEEE Computer Society.
- Aniruddha Diwan, J. K. and Kumar, A. (2002). Optimal per-node rate allocation to provide per-flow end-to-end delay guarantees in a network of routers supporting guaranteed service class. In *Proceeding of IEEE ICC*.

- BÍRÓ, J. J. (2004). Equivalent capacity estimators for bufferless fluid flow multiplexing. In *GLOBECOM*, volume 3, pages 1406–1410.
- Cao, J., Cleveland, W. S., Lin, D., and Sun, D. X. (2001). On the nonstationarity of internet traffic. In *SIGMETRICS '01: Proceedings of the 2001 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 102–112, New York, NY, USA. ACM Press.
- Choudhury, G. L. (1996). Squeezing the most out of atm. In *IEEE TRANSACTIONS ON COMMUNICATIONS*, volume 44, pages 203–217.
- Demirkan, H., Goul, M., and Soper, D. S. (2005). Service level agreement negotiation: A theory-based exploratory study as a starting point for identifying negotiation support system requirements. In *HICSS*, volume 01, page 37b.
- Filsfils, C. and Evans, J. (2005). Deploying diffserv in backbone networks for tight sla control. In *IEEE Internet Computing*, volume 9, pages 58–65, Piscataway, NJ, USA. IEEE Educational Activities Department.
- IANA. Internet Assigned Numbers Authority. <http://www.iana.org/assignments/port-numbers>.
- IPTRAF. Portal de Análise de Tráfego do Laboratório Ravel. <http://iptraf.ravel.ufrj.br/>.
- Kleinrock, L. (1976). *Queueing systems*. volume 2, New York, NY, USA. John Wiley and Sons.
- Netflow. Netflow services and applications. Cisco White Paper. [http://www.cisco.com/warp/public/cc/cisco/mkt/ios/netflow/tech/napps\\_wp.pdf](http://www.cisco.com/warp/public/cc/cisco/mkt/ios/netflow/tech/napps_wp.pdf).
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Töpke, C. R. (2001). Uma metodologia para caracterização de tráfego e medidas de desempenho em backbones ip. Master's thesis, COPPE/UFRJ.
- Van der Wal, K., M. M. B. H. (1997). Delay performance analysis of the new internet services with guaranteed qos. In *Proceedings of the IEEE*.
- Verma, D. C. (2004). Service level agreement on ip networks. In *Proceedings of the IEEE*, volume 92, pages 1382–1388.
- Wissam Fawaz, Belkacem Daheb, O. A. and Du-Pont, M. (2004). Service Level Agreement and Provisioning in Optical Networks. *IEEE Communications Magazine*.