

# **Estratégias de Controlo de Admissão no contexto dos modelos de QoS propostos pelo IETF**

Davide Lourenço, Paulo Loureiro, Gonçalo Quadros, Edmundo Monteiro  
pdavide@dei.uc.pt, loureiro@estg.iplei.pt, {quadros, edmundo}@dei.uc.pt

Laboratório de Comunicações e Telemática da Universidade de Coimbra (LCT-UC)  
DEI, Polo II, Pinhal de Marrocos, 3030-290 Coimbra  
<<http://lct.dei.uc.pt>>

**Palavras chave:** QoS, Controlo de Admissão.

## **Resumo**

*Qualquer estratégia de fornecimento de QoS, além dos mecanismos necessários à diferenciação de tráfego – nomeadamente ao nível do escalonamento e de gestão de filas de espera – tem igualmente de contemplar mecanismos que confirmam ao sistema de comunicação a capacidade de controlar e regular o volume de carga suportado. Neste artigo são apresentadas várias estratégias de Controlo de Admissão que dotam os sistemas desta funcionalidade.*

## **I. Introdução**

Num sistema de comunicação onde se pretenda disponibilizar serviços com qualidade é fundamental que a carga suportada não condicione o serviço fornecido a cada fluxo, apesar do carácter finito dos recursos de que é constituído. Desta forma, o utilizador não deverá perceber qualquer perturbação no funcionamento da sua aplicação devido ao aumento do volume de tráfego concorrente aos mesmos recursos do sistema de comunicação.

Neste contexto, a capacidade de controlar e regular o volume de tráfego admitido – gerindo, desta forma, os recursos do sistema em função da política de fornecimento de serviços e dos compromissos estabelecidos – é fundamental para que os sistemas de comunicação em geral e as redes IP, em particular, possam fornecer o serviço adequado (e esperado) às aplicações suportadas.

A capacidade de controlo e regulação do tráfego admitido num sistema de comunicação é fornecida pelos mecanismos de Policiamento de Tráfego (PT) e de Controlo de Admissão (CA). Enquanto que o mecanismo de PT tem como objectivo verificar a conformidade do tráfego em relação ao perfil definido, quando estabelecido o acordo de fornecimento de serviço, o mecanismo de CA tem como principal objectivo condicionar a admissão de um novo fluxo, tendo subjacente a política de aceitação adoptada pelo sistema em causa.

Em termos gerais, um mecanismo de CA deverá condicionar o suporte de uma aplicação com base na política global de aceitação de tráfego<sup>1</sup> definida (estratégia ou critério de admissão), nos compromissos de serviço já estabelecidos e nos recursos disponíveis *versus* requisitos da aplicação.

Em concreto, uma aplicação só deverá ser admitida se a rede dispuser da quantidade de recursos necessária para poder fornecer o serviço solicitado e se a admissão de um novo fluxo não condicionar, de forma perceptível aos utilizadores, a QoS fornecida aos fluxos gerados pelas aplicações já suportadas. Por outro lado, só poderão ser admitidas as aplicações e os utilizadores que se apresentem em conformidade com a política de aceitação definida no sistema de comunicação.

A estratégia de CA a adoptar em determinado sistema de comunicação deve ser definida tendo em consideração os seguintes factores:

- o posicionamento do mecanismo – centralizado num agente ou distribuído por todos os encaminhadores da rede ou apenas pelo encaminhadores de fronteira;
- a informação utilizada para suportar a decisão – valor indicativo do nível de congestão ou quantidade de recursos não reservados;
- a utilização, ou não, de protocolos de sinalização;
- o carácter associado ao pedido de admissão – implícito ou explícito<sup>2</sup>, e
- a necessidade de antecipadamente conhecer os parâmetros técnicos que caracterizam o fluxo.

Na secção II deste artigo são apresentadas diferentes abordagens de CA, sendo enfatizada, nas secções III e IV, a influência do conjunto de factores que acaba de ser referido na definição da estratégia de CA mais adequada no contexto dos modelos de QoS propostos pelo *Internet Engineering Task Force*<sup>3</sup> (IETF). Por fim, na secção V, as diferentes estratégias são avaliadas.

## II. Estratégias de CA

A estratégia de CA a adoptar em redes não orientadas à conexão é uma questão complexa, estando longe de ser evidente como realizar este procedimento de forma eficiente. Não obstante a complexidade, desde logo parece claro que a estratégia a adoptar está intimamente relacionada quer com o nível de qualidade dos serviços fornecidos, quer com o grau de compromisso assumido pela rede (garantias de serviço) no fornecimento dos mesmos serviços. Desta forma, a definição da estratégia de CA é directamente influenciada pelas características do sistema de comunicação alvo, nomeadamente pelas características do modelo de QoS adoptado.

---

<sup>1</sup> A política global de aceitação de tráfego define quais as aplicações e os utilizadores que podem utilizar a infra estrutura de rede, sendo igualmente definida a estratégia de atribuição de recursos, nomeadamente a distribuição de largura de banda pelos diversos serviços disponibilizados.

<sup>2</sup> Um pedido de admissão explícito é desencadeado por um procedimento formal suportado por um protocolo de sinalização. Se não estiver definido um procedimento com tal objectivo o pedido de admissão é desencadeado pelos pacotes de dados, daí o carácter implícito.

<sup>3</sup> <http://www.ietf.org>

Neste sentido, em termos de CA, um modelo que fornece garantias de desempenho quantitativas (débito garantido e/ou atraso máximo para pacotes em trânsito) apresenta requisitos bem diferentes dos apresentados por um modelo que apenas estabelece garantias qualitativas (“atraso médio de pacotes em trânsito reduzido”) ou, ainda, das de um modelo que apenas garante a diferenciação de tratamento de fluxos, agrupados num conjunto de classes servidas “o melhor possível”.

Por outro lado, a eficiência que se pretende obter na utilização dos recursos do sistema de comunicação é igualmente factor condicionante na definição da estratégia de CA<sup>4</sup>.

Desta forma, o mecanismo de CA deverá implementar a estratégia mais adequada ao modelo de serviço em causa, tendo em consideração designadamente o tipo de garantias fornecidas e a eficiência pretendida na utilização dos recursos do sistema de comunicação.

Em termos gerais, para suportar o critério de admissão um mecanismo de CA adopta uma das seguintes abordagens:

- (1) controlo com base na disponibilidade de recursos - recursos não atribuídos;
- (2) controlo com base num valor indicativo do nível de congestão.

A primeira abordagem requer que, através de um protocolo de sinalização, a aplicação especifique antecipadamente à rede os parâmetros técnicos que caracterizam o fluxo que irá gerar. Desta forma, o sistema de comunicação pode avaliar se os recursos que tem disponíveis – recursos não atribuídos ao processamento de tráfego – são suficientes para fazer face aos requisitos do serviço solicitado.

Esta abordagem é a mais adequada no âmbito de fornecimento de serviços cuja performance tem de ser estritamente controlada. O *Guaranteed Quality of Service*<sup>5</sup> (GQS) [1] do modelo *Integrated Services (IntServ)* [2] é um exemplo da sua aplicação.

No contexto do modelo *Differentiated Services (DiffServ)* [3], a mesma abordagem é implementada através de um agente centralizador das actividades de gestão designado por *Banwidth Broker* (BB). Neste cenário, um nó que pretenda desencadear um serviço de comunicação tem de o solicitar explicitamente ao BB através de um protocolo de sinalização como seja o RSVP [4] ou o *Boomerang* [5].

A segunda possibilidade de abordar o CA passa por suportar o critério de admissão num valor indicativo do nível de congestão experimentado pelo sistema de comunicação, podendo este valor ser obtido através de medidas de congestão efectuadas ao sistema. As estratégias de medição são tradicionalmente classificadas em *passive measurement*, quando obtidas pela monitorização directa do tráfego de dados, ou *active measurement*, se realizadas de forma indirecta através de tráfego de prova (*probing traffic*).

Embora mais adequada no âmbito do modelo *DiffServ* a realização do procedimento de CA com base em medidas de congestão foi inicialmente proposta no contexto do modelo *IntServ* [6].

---

<sup>4</sup> Neste contexto, por eficiência entende-se o nível de utilização efectiva dos recursos por tráfego gerado pelas aplicações suportadas.

<sup>5</sup> Também conhecido por “Serviço Garantido”.

Ainda de acordo com a segunda possibilidade de abordar o CA, a proposta apresentada em [7] sugere, como alternativa à avaliação do estado de congestão do sistema de comunicação por medição, a utilização da informação associada ao cálculo de caminhos por um mecanismo de encaminhamento que considera parâmetros de QoS (QoS).

Esta estratégia contorna a necessidade do procedimento inicial de aferição do estado de congestão por tráfego de prova, evitando, assim, a sobrecarga e a intrusão associado a esse processo preliminar.

Nas duas secções seguintes são descritas com maior detalhe as abordagens ao mecanismo de CA apresentadas, sendo analisados os pontos fortes e as respectivas limitações.

### III. CA com base na disponibilidade de recursos

Por mais bem apetrechado que seja um sistema de comunicação, os seus recursos terão sempre um carácter finito. Assim sendo, a largura de banda das ligações, a memória e a capacidade de processamento do equipamento activo são recursos limitados que é necessário gerir, designadamente por intermédio de mecanismos de CA de tráfego.

#### A. Modelo *IntServ*

De acordo com a abordagem tradicional, de que é exemplo o modelo *IntServ*, o CA de tráfego é realizado confrontando a disponibilidade de recursos<sup>6</sup> de rede face aos requisitos de qualidade do serviço solicitado. Neste contexto, uma aplicação só poderá ser suportada se estiver disponível – em todos os encaminhadores intervenientes – a quantidade de recursos necessária para fornecer o serviço solicitado, sendo, de outra forma, recusada pelo mecanismo de CA.

A definição de um mecanismo de CA, segundo esta estratégia, apresenta-se incontornável no âmbito do modelo *IntServ* onde é pretendido fornecer, nomeadamente, um serviço garantido no que se refere ao atraso máximo de pacotes em trânsito. Desta forma, o modelo *IntServ* preconiza a utilização de um mecanismo de reserva de recursos que lhe permite fornecer garantias de qualidade de serviço aferidas de forma quantitativa. O carácter “tudo ou nada” do GQS especificado no âmbito deste modelo depende precisamente do mecanismo de reserva de recursos.

O sistema de comunicação em função dos recursos disponíveis, por um lado, e das características de QoS do serviço requerido pela aplicação, por outro, toma a decisão de aceitação ou de rejeição do serviço solicitado. Desta forma, antes de ser estabelecido qualquer compromisso entre a rede e a aplicação há lugar a uma fase inicial de negociação onde a aplicação solicita determinado serviço e a rede verifica se tem capacidade para fazer face à solicitação de que é alvo sem, no entanto, condicionar qualquer compromisso já estabelecido. Ambas as actividades são realizadas por intermédio do protocolo RSVP [4].

---

<sup>6</sup> No âmbito do modelo *IntServ*, quando se refere disponibilidade de recursos deve entender-se recursos não reservados através do protocolo RSVP.

No caso da aplicação ser aceite é legítimo esperar que ao tráfego gerado, desde que em conformidade com o perfil sinalizado, seja fornecido um serviço de acordo com o solicitado e posteriormente acordado.

### B. Modelo *DiffServ*

A capacidade de escalabilidade evidenciada pelo *DiffServ* é a principal vantagem que o modelo apresenta em relação ao *IntServ*. Esta característica deve-se ao facto de os recursos, num domínio *DiffServ*, serem atribuídos por agregado de tráfego e não ao nível do fluxo individual. Assim, no âmbito deste modelo, a realização de reserva de recursos apenas poderá ter sentido se definida ao nível do agregado de tráfego (ou seja, da Classe de Serviço - CoS). Neste caso, o volume de informação de estado a manter por cada encaminhador é proporcional ao número de classes de serviço implementadas nesse domínio e não ao número individual de fluxos suportados.

No contexto do modelo *DiffServ*, a motivação subjacente à utilização de protocolos de sinalização e de reserva de recursos pode ser de alguma forma considerada fora dos princípios que fundamentaram o aparecimento do modelo. Contudo, desde o princípio do seu desenvolvimento foram definidos cenários onde as aplicações fazem uso de protocolos de sinalização para poderem caracterizar o serviço requerido em domínios *DiffServ*. Um dos primeiros cenários definidos neste contexto – para muitos uma referência na área dos serviços diferenciados – foi apresentado por Nichols [8]. Esta abordagem preconiza a utilização de um agente em cada domínio *DiffServ* – *Bandwidth Broker*<sup>7</sup> (BB) – para efectuar a atribuição de recursos.

O principal objectivo de um BB é centralizar o conjunto de funções relacionadas com a gestão dos serviços disponibilizados pelo domínio, designadamente as referentes à gestão de recursos. Este conjunto de funções de gestão tem subjacente a política global de serviço praticada no domínio.

Neste sentido, o BB mantém informação do nível de utilização dos recursos do domínio, implementando o procedimento de CA de um novo fluxo com base na disponibilidade de recursos e nos SLA's estabelecidos com os domínios adjacentes.

Em termos conceptuais, como referido em [9] e em [10], o BB introduz na arquitectura do modelo *DiffServ* um plano de controlo e gestão de QoS, desnivelando as funções relacionadas com estas actividades das funções tradicionais de *forwarding*<sup>8</sup>. Em concreto, a implementação de um BB permite libertar os encaminhadores das funções de controlo e de gestão de parâmetros de QoS, sendo esta uma das vantagens mais evidentes da sua adopção. Esta separação possibilita a implementação de mecanismos de CA eficientes sem introduzir perda de desempenho ao nível do encaminhamento.

---

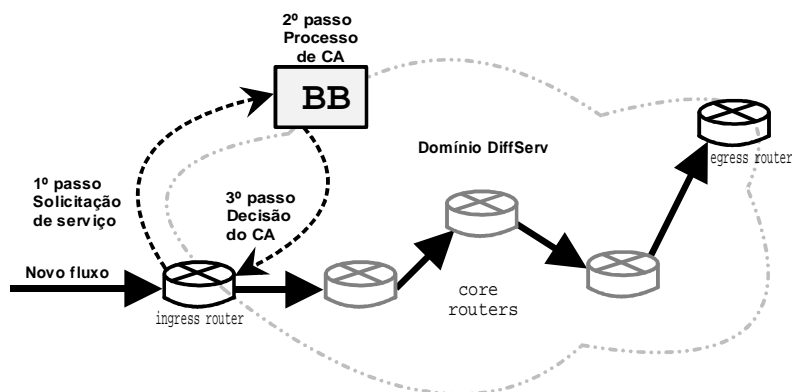
<sup>7</sup> De forma a introduzir alguma redundância, a abordagem de Nichols prevê a possibilidade de distribuir, de forma hierárquica, algumas das funções atribuídas ao BB de um domínio. No entanto, apenas um BB em cada domínio tem a seu cargo as funções de negociação com os BB dos domínios adjacentes.

<sup>8</sup> O termo *forwarding* é utilizado para designar de uma forma global os mecanismos, envolvidos no processamento realizado pela camada IP de um encaminhador, que permitem que um pacote seja retirado do interface de entrada e seja colocado no interface de saída associado ao caminho seleccionado pelo mecanismo de encaminhamento.

Contudo, a adopção de um BB acarreta alterações sensíveis nos procedimentos subjacentes ao estabelecimento de um serviço de comunicação através de um ou mais domínios *DiffServ*.

Desta forma, para que uma aplicação possa ser suportada é necessário que o serviço requerido seja caracterizado e solicitado ao BB do domínio a que o nó está ligado. Este pedido é realizado directamente pela aplicação, via encaminhador de ingresso (*ingress router*) no domínio, através de um protocolo de sinalização. Como já referido, o RSVP é o protocolo de sinalização mais utilizado neste contexto pelas aplicações.

Ao receber uma mensagem de sinalização RSVP, referente à caracterização de um novo fluxo, o encaminhador de ingresso questiona o BB acerca da capacidade do domínio para suportar essa nova solicitação. Este procedimento de interacção entre o encaminhador de ingresso e o BB é suportado pelo protocolo RSVP – neste caso o encaminhador limita-se a encaminhar a mensagem RSVP proveniente da aplicação até ao BB – ou pelo protocolo *Common Open Policy Service* (COPS) [11] que faz passar a informação relevante contida na mensagem RSVP para o BB. A Figura 1 ilustra este procedimento de interacção entre encaminhador de ingresso e o BB do domínio local.



**Figura 1** – Interação entre encaminhador de ingresso e o BB [10]

Depois de determinar se a aplicação e o utilizador são potencialmente admissíveis à luz da política de serviço praticada pelo domínio, o BB verifica se entre o encaminhador de ingresso (*ingress router*) e o encaminhador de saída (*egress router*) do domínio estão disponíveis os recursos necessários para fornecer o serviço requerido.

Seguidamente, no caso de o serviço solicitado ser aceite pelo BB local e no caso da negociação inter-domínios ser dinâmica, este BB negocia com o BB do domínio vizinho a quantidade de recursos necessária para que a aplicação seja alvo da QoS requerida. Este, por sua vez, realiza a mesma negociação com o domínio seguinte.

Estas negociações bilaterais repetem-se até que seja atingido o domínio a que está ligado o sistema destino do serviço de comunicação<sup>9</sup>.

<sup>9</sup> A abordagem de Nichols contempla a possibilidade da negociação entre domínios ser estática, reduzindo as funções do BB apenas à gestão interna de recursos. Neste caso é verificado se na ligação entre os dois domínios estão disponíveis os recursos requeridos pela aplicação. Esta verificação repete-se entre todos os domínios adjacentes até que seja atingido o domínio destino.

Segundo Nichols [8] este conjunto de acordos bilaterais e uma adequada gestão de recursos intra-domínio permitirão fornecer à aplicação o serviço *end-to-end* com a qualidade esperada.

Apesar do apresentado nos parágrafos anteriores, referente ao enquadramento de um BB num domínio *DiffServ*, não foi até agora definido qualquer RFC que o normalize. Desta forma, esta abordagem apresenta diversas particularidades no seu modo de operação, fundamentalmente devido:

- ao conjunto de funções de gestão atribuídas ao BB;
- à forma como são atribuídos os recursos inter-domínios e intra-domínio: dinâmica em função das necessidades ou estática/pré-estabelecida pelos serviços disponibilizados;
- à configuração dos encaminhadores fronteira: parametrização “reactiva” em função das solicitações de serviço ou “proactiva”, isto é, pré-estabelecida pelo BB.

Por outro lado, falta igualmente definir como combinar o BB e os diversos aspectos relacionados com a arquitectura do modelo *DiffServ*, designadamente ao nível da sinalização das aplicações, da configuração de PHB's e da atribuição estática de recursos entre outros, de forma a fornecer um serviço às aplicações suportadas com a qualidade esperada e fazendo uma utilização eficiente dos recursos do domínio.

#### **IV. CA com base num valor indicativo do nível de congestão**

Em sistemas onde não sejam realizadas reservas explícitas de recursos o critério de admissão a adoptar pelo mecanismo de CA pode ser suportado por um valor indicativo do seu estado de congestão, sendo este valor obtido por medidas de congestão efectuadas de forma directa ou indirecta ao sistema de comunicação.

Por outro lado, quando o sistema de comunicação realiza o encaminhamento de pacotes com base em parâmetros de QoS, a informação associada ao cálculo de caminhos pode ser utilizada pelo mecanismo de CA como estimativa do nível de congestão. Esta alternativa baseia-se em medições indirectas do nível de congestão, fornecidas pelo mecanismo de encaminhamento com QoS para suportar o critério de admissão.

##### *A. Nível de congestão obtido por medição*

Independente do método utilizado, *passive* ou *active measurement*, o valor obtido reflecte o nível de congestão experimentado durante um período de tempo, que pode ser o que decorre durante a realização da medição. Se num instante seguinte se repetir o procedimento de medição ou se o intervalo de tempo em que esta se efectua for alterado, o valor médio de congestão obtido será, muito provavelmente, diferente.

Por outro lado, em redes *connectionless* é provável que os pacotes referentes a um fluxo sejam encaminhados por diferentes percursos entre os *endpoints* envolvidos no serviço de comunicação. Este facto introduz determinada imprecisão no nível de congestão obtido entre o emissor e o receptor, já que nem todos os equipamentos envolvidos apresentarão o mesmo estado de congestão. Importa ainda acrescentar que o intervalo de tempo em que se efectua a medição pode não ser suficientemente grande para fazer

reflectir correctamente no valor obtido o estado de congestão do sistema de comunicação.

Pelo exposto, pode-se afirmar que a avaliação do estado de congestão do sistema de comunicação por medição apresenta alguma imprecisão. Esta característica, intrínseca à estratégia de medição, faz com que esta abordagem não seja a mais adequada no âmbito de serviços com garantias estritas dos parâmetros de QoS acordados, como seja o “Serviço Garantido”. Contudo, esta abordagem apresenta-se como uma alternativa interessante no contexto de serviços *soft real-time* onde não são estabelecidos compromissos rígidos referentes à performance fornecida pelo sistema de comunicação.

#### 1) *Passive measurement*

O procedimento de CA realizado com base em medidas de congestão – *Measurement-Based Admission Control* (MBAC) – foi inicialmente proposto por Jamin [6] no contexto do modelo *IntServ*, com o objectivo de fornecer um serviço designado por *Predictive* a aplicações de tempo real. Este serviço pretende oferecer um limite máximo para o atraso de pacotes em trânsito sem, no entanto, estabelecer a esse respeito um rigoroso compromisso.

Esta abordagem tira partido da capacidade evidenciada por algumas aplicações em se adaptarem a situações ocasionais de perda de qualidade no serviço de que são alvo. Assim, para poderem funcionar correctamente, estas aplicações, designadas genericamente por “elásticas” ou adaptativas, não requerem um limite rigoroso para o atraso de pacotes em trânsito, adaptando-se dinamicamente a variações moderadas no atraso da sua chegada.

Segundo Jamin, o serviço *Predictive* apresenta-se como uma evolução do “Serviço Garantido”, tirando partido da capacidade adaptativa das aplicações para aumentar a eficiência da utilização dos recursos da rede.

O procedimento de CA proposto por Jamin utiliza a caracterização do novo fluxo – feita através do protocolo de sinalização RSVP – e a estimativa do volume de carga suportado pelo encaminhador – obtido através de medições realizadas ao tráfego de dados processado – para tomar a decisão de admissão.

Segundo a abordagem de Jamin todos os encaminhadores envolvidos são intervenientes activos no processo de CA, no qual utilizam de forma independente os valores obtidos relativamente à largura de banda utilizada e ao atraso médio de pacotes em fila de espera.

O objectivo é controlar e delimitar o atraso de pacotes em fila de espera ao nível de cada encaminhador. Os requisitos da aplicação no que refere a atraso *end-to-end* são salvaguardados pelos sistemas terminais envolvidos. Assim sendo, para que uma aplicação possa ser suportada, todos os encaminhadores envolvidos têm de verificar se a admissão de um novo fluxo não conduz o atraso em fila de espera para valores não adequados, quer ao serviço requerido pela nova aplicação, como ao fornecido ao conjunto de aplicações já suportadas.



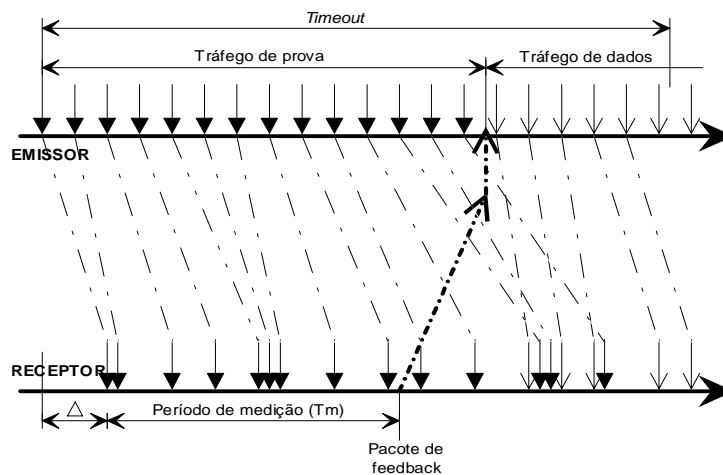
## 2) Active measurement

Mais recentemente foram propostos vários trabalhos preconizando a utilização de tráfego de prova para obter o nível de congestão entre os *endpoints*<sup>10</sup> envolvidos no serviço de comunicação que se pretende estabelecer. Esta estratégia, designada genericamente por *End-to-end Measurement-Based Admission Control*<sup>11</sup> (EMBAC) difere da tradicional abordagem MBAC fundamentalmente em duas vertentes. Primeiro, os encaminhadores do interior do domínio são completamente alheios a solicitações de estabelecimento de novos serviços, sendo a decisão de admissão tomada unicamente pelos *endpoints* intervenientes. Segundo, a medição do nível de congestão é realizada por cada “ligação” que se pretende estabelecer e não obtida e mantida por todos os encaminhadores do domínio.

Enquadrados no contexto do modelo *DiffServ* destacam-se os trabalhos referenciados em [12], [13] e [14]. De uma forma geral, a ideia básica subjacente à estratégia de CA definida em todas estas referências é a realização do procedimento tendo por base o nível de perdas ou de marcas explícitas de congestão<sup>12</sup> verificado no tráfego de prova.

Antes de uma aplicação poder ser admitida, os *endpoints* intervenientes verificam o nível de congestão do caminho que os interliga, quantificando-o em termos de perdas ou de marcas ECN. O objectivo desta fase de medição é verificar se a aplicação pode ser suportada, sendo alvo de um serviço com determinado nível de qualidade e não comprometendo, ao mesmo tempo, a QoS fornecida às aplicações já suportadas.

A Figura 2 apresenta o processo subjacente à admissão de um novo fluxo com base em medidas de congestão. Como é possível verificar, a aceitação de um novo fluxo é precedida de uma fase de aferição (*probing phase*) onde a aplicação gera um fluxo de prova (*probing traffic*) destinado ao receptor do serviço de comunicação.



**Figura 2** – Admissão de um novo fluxo com base em medidas de congestão [13]

<sup>10</sup> Neste contexto, são considerados *endpoints* tanto os *edge routers* como os nós terminais.

<sup>11</sup> Esta estratégia é também conhecida na literatura da especialidade por *Endpoint Admission Control* assim como por *Probe-Based Admission Control*.

<sup>12</sup> Com base numa estratégia dinâmica de gestão de filas como o RED [16], é possível adoptar conforme sugerido em [15] a utilização de marcas explícitas de congestão (*Explicit Congestion Notification* – ECN [17, 18]) como indicador do estado de congestão do sistema de comunicação.

Ao receber o primeiro pacote do fluxo de prova (isto é, após o intervalo tempo representado na figura por  $\Delta$ ), o receptor dá início à contabilização estatística de todos os pacotes recebidos durante o período de medição ( $T_m$ ). Este processo de contabilização pode ser constituído pela simples quantificação do número de pacotes recebidos (ou não recebidos) no intervalo de medição ou por algoritmos mais complexos que permitam aferir o atraso e a variação de atraso sofrida pelos pacotes recebidos.

Ao fim do período de medição o receptor decide, com base no valor obtido e no limiar de admissão especificado para o fluxo em causa, se a aplicação pode ou não ser admitida. O emissor é então notificado da decisão através de um ou mais pacotes (pacotes de *feedback*).

Ao receber a notificação o emissor pára de gerar pacotes de prova e, em conformidade com conteúdo do pacote de *feedback*, passa a gerar o tráfego de dados referente ao serviço de comunicação estabelecido (aplicação suportada) ou interrompe a tentativa de estabelecimento do mesmo serviço (aplicação rejeitada). No caso de o nível de congestão da rede ser tal que nenhum dos pacotes do fluxo de prova consegue chegar ao destino, a tentativa de estabelecimento do serviço de comunicação é anulada por *timeout*.

#### B. Nível de congestão obtido através de informação de routing

Os mecanismos de encaminhamento utilizados nas redes IP actuais não consideram critérios de QoS no cálculo de caminhos. Os caminhos são seleccionados não fazendo distinção do tráfego a encaminhar, isto é, o melhor caminho seleccionado para o tráfego *best-effort* é assumido que o seja para o tráfego com requisitos de QoS.

A adopção de um mecanismo de encaminhamento onde o cálculo de caminhos é realizado tendo em consideração os requisitos de QoS do tráfego a encaminhar e o estado de congestão dos vários caminhos alternativos<sup>13</sup> é fundamental na definição de uma plataforma coerente e integrada que tenha como objectivo fornecer serviços *end-to-end* diferenciados por parâmetros de qualidade.

Desta forma, o caminho seleccionado não será obrigatoriamente o mais curto – estratégia tradicional dos algoritmos de encaminhamento como sejam o RIP e o IS-IS – mas o mais adequado face aos requisitos de QoS. Esta característica dos protocolos de encaminhamento com QoS (QoS<sup>14</sup>) é essencial para que a rede possa fornecer o tratamento diferenciado e adequado às necessidades de desempenho do tráfego e, ao mesmo tempo, promova uma eficiente utilização de recursos.

Assim, os encaminhadores da rede têm de ter conhecimento do estado de congestão quer dos encaminhadores vizinhos quer das ligações entre estes. O protocolo de encaminhamento deve promover a divulgação desta informação por todos os

---

<sup>13</sup> No contexto do modelo *DiffServ*, os requisitos do tráfego a encaminhar são obtidos através da classe a que está associado, enquanto o valor indicativo do estado de congestão é obtido através de uma métrica de QoS.

<sup>14</sup> As estratégias de encaminhamento que consideram os requisitos de QoS do tráfego designam-se genericamente por *Quality of Service Routing* (QoSR).

encaminhadores da rede – incluindo os encaminhadores de ingresso onde são realizados os procedimentos de controlo de admissão e de regulação de tráfego.

Ao ter disponível nestes encaminhadores um valor indicativo do nível de congestão associado ao caminho seleccionado, o mecanismo de CA pode ser realizado com base nesta informação como sugerido em [7].

Segundo esta abordagem de CA, uma aplicação só será admitida se o valor de congestão, associado ao caminho seleccionado pelo mecanismo de QoS como o mais adequado às características da classe de serviço em que o fluxo será agregado em caso de admissão da aplicação, apresentar um valor inferior a determinado limiar. A definição do valor de limiar de não admissão tem em consideração a política de serviços praticada no sistema de comunicação, designadamente no que respeita ao tipo de compromissos estabelecidos e ao nível de eficiência pretendida na utilização de recursos.

## V. Avaliação das estratégias

De forma a salientar os pontos fortes e fracos das estratégias de CA apresentadas, pretende-se agora promover a sua avaliação (ver Tabela 1), tendo por base o seguinte conjunto de critérios:

- Garantia: grau de compromisso face ao nível de qualidade associado ao serviço;
- Eficiência: nível de utilização efectiva dos recursos;
- Latência: espera subjacente ao processo de admissão;
- Intrusão: utilização ou não de tráfego de prova.

|                                          | Garantia      | Eficiência    | Latência | Intrusão    | Modelo/serviço alvo       |
|------------------------------------------|---------------|---------------|----------|-------------|---------------------------|
| Com base num protocolo de sinalização    | elevada       | reduzida      | média    | não utiliza | <i>IntServ</i>            |
| Com base num BB                          | média/elevada | elevada/média | elevada  | não utiliza | <i>DiffServ</i>           |
| Com base em <i>passive measurements</i>  | média         | elevada       | média    | não utiliza | <i>IntServ/Predictive</i> |
| Com base em <i>active measurements</i>   | média         | média         | elevada  | utiliza     | <i>DiffServ</i>           |
| Com base na informação de <i>routing</i> | média         | elevada       | reduzida | não utiliza | <i>DiffServ</i>           |

**Tabela 1** – Avaliação de estratégias de CA

A avaliação apresentada pela Tabela 1 é detalhadamente fundamentada nos parágrafos seguintes.

Em redes *IntServ* os recursos são atribuídos aos fluxos suportados. O mecanismo de CA é realizado com base no protocolo de sinalização RSVP, sendo a aplicação aceite apenas se todos os encaminhadores intervenientes puderem atribuir a quantidade de recursos necessária para fornecer o serviço com a qualidade requerida e anunciada explicitamente pela aplicação.

Esta estratégia permite que às aplicações seja garantido o nível de qualidade que caracteriza o serviço requerido. Contudo, no caso das aplicações suportadas apresentarem um comportamento pouco constante esta abordagem pode apresentar-se pouco eficiente na utilização dos recursos de rede. Esta característica deve-se ao facto de os recursos quando estão atribuídos a determinado fluxo não serem partilhados mesmo que temporariamente disponíveis.

De forma a aumentar a eficiência de utilização dos recursos, a estratégia de CA proposta por Jamin para o modelo *IntServ* preconiza a utilização de medidas de congestão efectuadas pelos encaminhadores da rede (*passive measurement*) para suportar a decisão de admissão/rejeição de um fluxo caracterizado através do protocolo RSVP. O objectivo é, claramente, aumentar a utilização efectiva dos recursos da rede à custa de um compromisso menos rigoroso relativamente ao nível de qualidade associado ao serviço fornecido.

A definição de uma estratégia de gestão de recursos adequada à arquitectura do modelo *DiffServ* é, ao contrário do que acontece com o modelo *IntServ*, uma questão em aberto não estando ainda definida a forma de implementar um mecanismo de CA. Contudo, várias propostas têm vindo a ser apresentadas com o intuito de dotar os domínios *DiffServ* de um mecanismo de CA adequado.

Uma das primeiras abordagens neste sentido preconiza a utilização em cada domínio *DiffServ* de um BB centralizador das tarefas de gestão [8], incluindo as relacionadas com a gestão dos recursos. Conceptualmente, o BB introduz na arquitectura do modelo *DiffServ* um plano superior – plano de gestão – que na prática permite retirar dos encaminhadores as funções de controlo e gestão. Esta é, reconhecidamente, a principal vantagem desta abordagem já que possibilita que sejam desenvolvidos algoritmos eficientes de CA sem, no entanto, introduzir perda de desempenho ao nível do encaminhamento.

Não obstante a possibilidade de distribuir de forma hierárquica algumas das funções atribuídas ao BB, as tarefas de negociação com os domínios adjacentes são implementadas apenas por um único BB em cada domínio. Assim, sob pena de introduzir uma latência considerável, nomeadamente, no processo de admissão de uma aplicação, a máquina que suporta este elemento centralizador deve ser dimensionada com capacidade suficiente para fazer face ao volume de solicitações de que é alvo.

Por outro lado, estratégia adoptada pelo BB na gestão dos recursos do domínio influencia tanto o nível de eficiência na utilização dos mesmos recursos como o grau de compromisso que a rede pode estabelecer relativamente à qualidade do serviço solicitado. Desta forma, ao domínio é possibilitado o estabelecimento de firmes garantias relativamente à qualidade do serviço fornecido se os recursos forem atribuídos dinamicamente através de um processo de reserva. Se, pelo contrário, os recursos do domínio são estaticamente distribuídos pelos serviços disponibilizados e partilhados pelo tráfego suportado o domínio não tem capacidade para estabelecer tais garantias mas, em contrapartida, é obtida uma maior eficiência na utilização dos seus recursos (ver Tabela 1).

Conceptualmente opostas, as propostas referidas em [12], [13] e [14] distribuem o processo de CA pelos encaminhadores de ingresso do domínio. Enquadradas no contexto do modelo *DiffServ*, estas estratégias têm em comum o facto de se basearem

no valor indicativo do nível de congestão obtido por tráfego de prova para suportar a decisão do CA.

Devido à imprecisão intrínseca ao processo de medição por tráfego de prova ao paradigma *connectionless* característico das redes IP e ao comportamento, de uma forma geral, não constante das aplicações (designadamente em termos de débito), esta abordagem não permite que sejam fornecidas garantias firmes no que respeita à performance do serviço fornecido pela rede. Desta forma, estas estratégias são adoptadas apenas no âmbito de serviços onde não sejam estabelecidos limites rigorosos para as características de QoS consideradas – *soft real-time services* [15].

Por outro lado, esta estratégia requer que seja atribuída ao tráfego de prova determinada percentagem da largura de banda disponível o que na prática se traduz numa diminuição dos recursos utilizáveis pelo tráfego de dados das aplicações.

Seguindo a mesma linha de abordagem, foi ainda apresentada uma estratégia de CA que utiliza como valor indicativo do nível de congestão a informação associada pelo mecanismo de encaminhamento QoSR no cálculo do caminho mais adequado às características do tráfego.

Esta abordagem contorna a necessidade da existência de qualquer preliminar ao processo de estabelecimento de uma aplicação, já que a informação que serve de base ao critério do CA é acessível aos encaminhadores de ingresso logo que seja conhecido o caminho seleccionado pelo mecanismo QoSR.

Ao basear-se num valor de congestão obtido por medição, esta estratégia proporciona uma utilização eficiente dos recursos da rede não permitindo, no entanto, estabelecer garantias rigorosas em relação a qualquer característica de QoS.

Esta é também a estratégia adoptada para o mecanismo de CA desenvolvido no âmbito do modelo de serviço do Laboratório de Comunicações e Telemática da Universidade de Coimbra [19].

## **VI. Agradecimentos**

Trabalho parcialmente financiado pelo Programa de Investigação POSI do Ministério da Ciência e Tecnologia (Projecto QoS II - Qualidade de Serviço em Sistemas de Comunicação por Computador).

## **VII. Conclusão**

Devido à especificidade dos modelos de QoS propostos pelo IETF, não é possível generalizar a definição de um procedimento de CA.

Uma estratégia de CA definida no contexto *IntServ* apresenta-se ineficiente e inadequada no âmbito do modelo de serviço *DiffServ*. Neste sentido, é de referir que no modelo *IntServ*, ao contrário do que acontece no *DiffServ*, a garantia da qualidade do serviço fornecido é mais importante do que a eficiência de utilização dos recursos da rede. Desta forma, no contexto do modelo *IntServ* o mecanismo de CA deverá dotar o sistema de comunicação da capacidade de estabelecer garantias rigorosas relativamente à qualidade do serviço fornecido.

Por seu turno, o modelo *DiffServ* dota as redes da capacidade de diferenciação de tráfego, não pretendendo contudo estabelecer garantias firmes em relação a qualquer característica de QoS. Com o intuito de aumentar o volume de tráfego suportado, o mecanismo de CA de um domínio *DiffServ* pode adoptar uma estratégia menos restritiva mesmo que para isso as aplicações fiquem expostas a flutuações moderadas da QoS de que são alvo.

Desta forma, pode-se afirmar que a especificação de uma estratégia adequada de CA depende do modelo de serviço implementado pelo sistema de comunicação alvo, já que, quer o nível de qualidade dos serviços fornecidos, quer o grau de compromisso assumido pela rede no fornecimento dos mesmos serviços condicionam a sua definição.

Por outro lado, factores como o posicionamento do mecanismo, a utilização (ou não) de protocolos de sinalização, o carácter associado ao pedido de admissão entre outros influenciam a definição do mecanismo de CA a adoptar em determinado sistema de comunicação.

### VIII. Referências

- [1] Shenker, S. *et al.*, “Specification of Guaranteed Quality of Service”, Internet Engineering Task Force, Request for Comments 2212, September 1997.
- [2] Braden, R., *et al.*, “Integrated Services in the Internet Architecture: an Overview”, Internet Engineering Task Force, Request for Comments 1633, June 1994.
- [3] Blake, S., *et al.*, “An Architecture for Differentiated Services”, Internet Engineering Task Force, Request for Comments 2475, December 1998.
- [4] Braden, R., *et al.*, “Resource ReSerVation Protocol (RSVP) – version 1 Functional specification”, Internet Engineering Task Force, Request for Comments 2205, September 1997.
- [5] Feher, G., *et al.*, “Boomerang – A Simple Protocol for Resource Reservation in IP Networks”, IEEE Workshop on QoS Support for Real-Time Internet Applications, Vancouver, Canada, June 1999.
- [6] Jamin, S., *et al.*, “A Measurement-based Admission Control Algorithm for Integrated Services Packet Networks (Extended Version)”, ACM/IEEE Transactions on Networking, 1996.
- [7] Huston, G., “Next Steps for the IP QoS Architecture”, Internet Engineering Task Force, Request for Comments 2990, November 2000.
- [8] Nichols, K., *et al.*, “A Two-bit Differentiated Services Architecture for the Internet” Internet Engineering Task Force, Request for Comments 2638, July 1999.
- [9] Terzis, A., *et al.*, “A Prototype Implementation of the Two-Tier Architecture for Differentiated Services”, RTAS99 Vancouver, Canada, 1999.
- [10] Zhang, Z., *et al.*, “Decoupling QoS Control from Core Routers: A novel Bandwidth Broker Architecture for Scalable Support of Guaranteed Services” In proceedings of IEEE INFOCOM 2000, March 2000.
- [11] Salsano, S., “COPS Usage for Outsourcing DiffServ Resource Allocation”, Internet Engineering Task Force, Internet Draft, February 2000. Work in progress.

- [12] Elek, V. *et al.*, “Admission Control Based on End-to-End Measurements” In proceedings of IEEE INFOCOM 2000, March 2000.
- [13] Bianchi, G. *et al.*, “Throughput analysis of end-to-end measurements-based admission control in IP” In proceedings of IEEE INFOCOM 2000, Março 2000.
- [14] Cetinkaya, C. *et al.*, “Egress Admission Control” In proceedings of IEEE INFOCOM 2000, March 2000.
- [15] Breslau, L., *et al.*, “End Admission Control: Architectural Issues and Performance”, ACM SIGCOMM 2000, August 2000.
- [16] Floyd, S. and Jacobson, V., “Random Early Detection gateways for Congestion Avoidance”, IEEE/ACM Translation on Networking, V.1 N.4, August 1993.
- [17] Floyd, S. and Ramakrishnan, K. “A Proposal to Add Explicit Congestion Notification (ECN) to IP”, Internet Engineering Task Force, Request for Comments 2481, January 1999.
- [18] Floyd, S. and Ramakrishnan, K. “The Addition of Explicit Congestion Notification (ECN) to IP”, Internet Engineering Task Force, Internet Draft, June 2001. Work in progress.
- [19] Lourenço, D. *et al.*, “Definição do Mecanismo de Controlo de Admissão para o Modelo de Serviço do LCT-UC”, In proceedings of CRC’2000, Viseu, November, 2000.