
CAPÍTULO 6

Aplicação ao Problema de Encaminhamento em Redes Integradas de Comunicações

1. Introdução

As tradicionais arquitecturas de redes de comunicações, foram estruturadas para suportar utilizadores com requisitos de qualidade de serviço (“Quality of Service” — QoS) simples e homogéneos [16]. No entanto, com o aumento da procura dos diversos serviços, as tecnologias de rede que floresceram num ambiente de economia de escala (isto é, quanto maior a dimensão, menor o custo unitário) são, gradualmente, substituídas por novas tecnologias que oferecem economia de alcance (isto é, abrangendo muitas áreas). As modernas redes de comunicações devem poder adaptar-se a utilizadores com diversos requisitos de QoS, com elevado grau de granularidade e subjectivos. Esta tarefa é realizada através de um conjunto de mecanismos de controlo da rede, que actuam em várias escalas de tempo.

O desempenho de uma rede depende de vários factores, tais como a configuração, o número de dados que pode ser transferido e os métodos de gestão da rede. Numa rede de telecomunicações, as funções da rede são responsáveis pela transferência de informação entre dois utilizadores finais, seleccionando e fixando os recursos de rede ao longo de um caminho. Uma **chamada** é a associação lógica entre a transmissão e os utilizadores finais. Uma **ligação** é a “cadeia” de recursos de rede que suporta uma chamada [16].

Considerem-se as operações sobre uma rede em tempo real. Sempre que uma chamada chega a um ponto da rede, é necessário determinar se existe um caminho através do qual a chamada pode ser encaminhada para o seu destino. Se for determinado um caminho, é então

necessário decidir se é ou não utilizado; se não existe nenhum caminho disponível, tem que se decidir o que fazer com a chamada. Cada passo destes está relacionado com um aspecto diferente do encaminhamento.

O **encaminhamento** é um elemento muito importante na gestão de redes, que consiste num conjunto de regras de decisão (geralmente denominada por técnica de encaminhamento) para ligar as chamadas, quando chegam a um ponto da rede. Desta forma, o encaminhamento está relacionado com operações sobre a rede, em tempo real, e com um mecanismo de controlo ao nível das chamadas, através do qual é escolhido um dos caminhos disponíveis (se existir mais do que um) para estabelecer a comunicação entre dois pontos (origem e destino) da rede.

Alguns dos tradicionais algoritmos de encaminhamento consistem em resolver problemas de caminho mais curto com um só objectivo, em que a função objectivo se baseia numa métrica simples, tal como custo, atraso, número de arcos, probabilidade de perda, etc., ou numa única função de diferentes métricas. No entanto, o problema torna-se mais complexo quando se introduzem restrições para satisfazer os vários requisitos de QoS.

Geralmente, estas restrições pertencem a duas categorias, consoante a entidade onde são impostas : nos arcos e nos caminhos. As **restrições nos arcos** são limitações na utilização dos arcos do caminho escolhido (por exemplo, a capacidade disponível num arco deve ser maior ou igual ao requerido pela chamada). As **restrições nos caminhos** são limites predefinidos de uma métrica de desempenho ao longo do caminho escolhido (por exemplo, o atraso ponto a ponto na rede não pode exceder o que a chamada pode tolerar) [16].

Devido à subjectividade dos actuais requisitos de QoS e aos complexos compromissos entre eles, tornou-se mais difícil definir uma adequada métrica de encaminhamento. Além disso, como as características dos vários tipos de tráfego são distintas, a mesma métrica não é aplicável a todos os casos.

Se, por um lado, não existe grande vantagem em ter um caminho em que a QoS associada seja melhor do que a especificada pelo utilizador, por outro, é muito difícil determinar um caminho com a QoS especificada pelo utilizador. Por esta razão, é que para redes integradas de comunicações, é necessário um novo paradigma de encaminhamento, que realce a procura de um caminho aceitável e que satisfaça os vários requisitos de QoS.

Devido à crescente procura de novos e mais sofisticados serviços de telecomunicações, entra em jogo a heterogeneidade dos requisitos de QoS, nas redes integradas de comunicações. Consequentemente, os operadores de redes não podem continuar a confiar nas

técnicas de encaminhamento que se baseiam numa métrica simples, sendo necessário considerar explicitamente métricas distintas nos algoritmos [16].

Geralmente, o problema da escolha do caminho é formulado como um problema de caminho mais curto com uma única função objectivo (uma única métrica ou uma função envolvendo diferentes métricas), cujos requisitos de QoS podem ser incorporados nestes modelos matemáticos, através da utilização de restrições adicionais. No entanto, e uma vez que os modelos matemáticos têm inerentemente uma estrutura de rede (podendo, desta forma, utilizar eficazmente algoritmos específicos e eficientes), a introdução de restrições adicionais destrói algumas propriedades interessantes, o que implica um elevado esforço computacional.

2. Tecnologias de encaminhamento

O encaminhamento alternativo é muito utilizado em redes telefónicas, para fornecer serviços fiáveis. Existem duas classes de algoritmos de encaminhamento alternativo : determinístico e aleatório.

“Dynamic Non-Hierarchical Routing” (DNHR) é um algoritmo determinístico, cuja atribuição de caminhos alternativos é estática e sujeita a alterações com as horas do dia [16].

No algoritmo denominado por “Dynamic Traffic Management” (DTM), também determinístico, as decisões de encaminhamento são produzidas em grupos de chamadas, em vez de chamada a chamada, permitindo-lhes testar um caminho directo e um caminho alternativo recomendado, com o maior número de circuitos livres [16].

Nos algoritmos de encaminhamento alternativo aleatórios, cada caminho é escolhido com uma dada probabilidade por chamada. A probabilidade atribuída a cada caminho é constantemente actualizada, de forma a reflectir a possibilidade do caminho ser testado com sucesso. Quanto maior for a possibilidade de um caminho alternativo ser testado com sucesso, maior é a probabilidade de ser escolhido. Na versão conhecida por “Dynamic Alternate Routing” (DAR) os encaminhamentos escolhidos são restabelecidos sempre que uma chamada falha.

Os protocolos de encaminhamento entre domínios, tal como o “Open Shortest Path First” (OSPF), suportam métricas de encaminhamento de diferentes tipos de serviço, baseadas em diferentes combinações de atraso, débito e segurança.

Nas actuais redes integradas, o encaminhamento sujeito a várias restrições nos caminhos (por exemplo, restrições no custo e no atraso), apesar de ser intratável, tem uma característica apetecível. É possível formular um problema de encaminhamento sujeito a várias restrições nos caminhos, como um problema de caminho mais curto multicritério, onde cada métrica de caminho com restrições é tomada como um objectivo do encaminhamento. Uma abordagem simples de encaminhamento multicritério, é supor que o caminho “óptimo” é não dominado, tal que qualquer outro caminho tem pelo menos para uma das funções objectivo, um valor pior do que o correspondente valor daquele caminho. Com esta suposição, pode-se gerar o conjunto de caminhos não dominados, e através de uma abordagem baseada em função utilidade determinar o caminho desejado.

3. Restrições de qualidade de serviço

A QoS pode ser considerada como um grau de adaptação para os critérios de serviço especificados pelo utilizador [16]. Muitas das estruturas de QoS existentes, tanto concentram os critérios de desempenho na gestão do tráfego, como fornecem requisitos de qualidade orientados pelo utilizador com elevado grau de granularidade. Um elemento importante na arquitectura de redes integradas é a capacidade de oferecer uma diversidade de requisitos de QoS às diferentes aplicações que utilizam a rede, e assim, tornar eficiente o uso dos recursos da rede.

Em [16] é proposto uma estrutura de QoS ao nível da chamada, na qual a QoS é especificada em termos de três classes de restrições, que podem depender do tipo de serviço da chamada : *restrições de desempenho* (por exemplo, atraso), *restrições de recursos* (por exemplo, transferência média, canal de segurança) e *restrições de prioridade* (por exemplo, estabelecimento de prioridades, retenção das prioridades).

Uma **restrição de desempenho** é uma restrição sobre uma medida de qualidade de informação completamente perceptível, transferida durante uma ligação [16]. Uma restrição de desempenho pode, ou não, ser negociada entre a rede e os utilizadores finais. Uma restrição negociável é especificada em termos de um conjunto de valores limitados por um valor requerido e um aceitável, em que o valor requerido é o nível de desempenho mais desejado, que o utilizador gostava de alcançar se os recursos estivessem logo disponíveis, e o valor aceitável é o nível de desempenho menos desejado, mas que o utilizador pode tolerar. Uma restrição não negociável é especificada em termos de apenas um valor aceitável. Apesar

de cada restrição de desempenho ser basicamente uma restrição de QoS dependente do caminho, ela pode ser implementada como um atributo de arco (por exemplo, débito) ou como um atributo de caminho (por exemplo, atraso). Um *atributo de arco* é um parâmetro de arco, considerado individualmente, servindo para determinar se um dado arco é aceitável para transportar uma dada ligação. Um *atributo de caminho* é o somatório de um parâmetro de arco ao longo de um dado caminho, servindo para verificar se o caminho é aceitável para transportar uma dada ligação [16].

Uma **restrição de recursos** é uma limitação na utilização de um dado tipo de recurso da rede, com um conjunto de características específicas [16]. As restrições de recurso, que estão sujeitas à definição do utilizador, podem estar directa ou indirectamente relacionadas com a QoS (por exemplo, segurança e escolha do transportador, respectivamente). Os recursos podem referir-se a elementos básicos da rede (por exemplo, arcos), ou aglomerados de elementos da rede (por exemplo, domínios administradores). A eficiência da topologia utilizada na escolha do caminho é determinada pela disponibilidade e aceitabilidade dos recursos da rede com vários atributos de recursos. Cada *atributo de recursos* (por exemplo, transmissão média) está associado a um conjunto de possíveis valores de atributos discretos (por exemplo, satélite, microondas). Uma restrição de recurso pode ser especificada em termos de um subconjunto deste conjunto. Um recurso da rede não é aceitável para uma chamada, a não ser que cada um dos valores dos atributos de recurso pertença ao correspondente conjunto de restrições de recurso. As restrições de recurso mais simples, que são predeterminados e não dependem do estado da rede, determinam, ou não, se um dado recurso é aceitável para o encaminhamento de uma chamada.

Uma **restrição de prioridade** é uma condição imposta na distribuição dos recursos da rede, para fornecer diferentes probabilidades de bloqueio ao tráfego das diferentes classes de prioridade. As restrições de prioridade podem ser implementadas em uma ou duas abordagens numa arquitectura de encaminhamento : encaminhamento que provoca a preempção e encaminhamento que não provoca a preempção. No *encaminhamento que provoca a preempção*, os recursos da rede que já foram atribuídos a chamadas existentes, podem ser recuperados e usados para acomodar novas chamadas de grande importância [16]. Assim, a eficácia do bloqueio das chamadas de alta prioridade, é melhorada à custa da interrupção das chamadas de baixa prioridade. No *encaminhamento que não provoca a preempção*, as chamadas de alta prioridade têm acesso preferencial e garantido aos recursos da rede, sem interferir com as

chamadas existentes fora dela. Desta forma, a eficácia do bloqueio das chamadas de alta prioridade é melhorada à custa da eficácia do bloqueio das chamadas de baixa prioridade.

4. Arquitectura de uma chamada

A arquitectura de uma chamada, que pode ser utilizada para suportar a adaptação da QoS, é composta por **processamento** e **encaminhamento da chamada**. Estes processos são duas componentes chave do controlo da rede, localizados em cada nó e representando um sistema de comutação numa rede. Estes processos são utilizados em simultâneo, para fornecer a cada chamada uma ligação que satisfaça todas as restrições de QoS e para manter um nível aceitável de QoS durante a chamada. A Fig. 1 apresenta o fluxo de controlo de sinais entre um par de utilizadores finais durante a instalação de uma chamada.

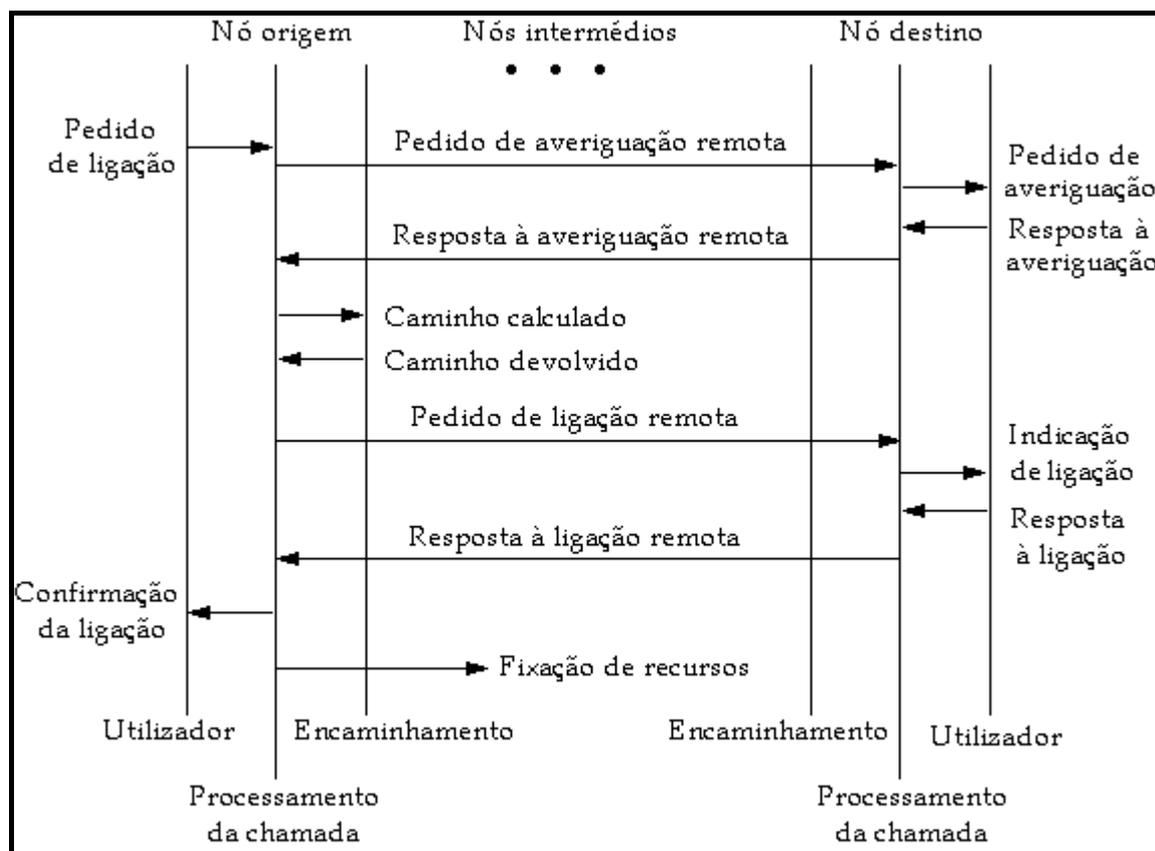


Fig. 1 – Arquitectura de uma chamada [16].

Quando o utilizador na origem faz um pedido de ligação, o *processamento da chamada* da origem transfere o controlo para o *processamento da chamada* do destino, o qual averigua a QoS ali especificada. Depois de averiguar sucessivamente o utilizador, o *processamento da chamada*

do destino devolve a informação obtida ao *processamento da chamada* da origem, para que as restrições de QoS obtidas a partir dos requisitos de QoS dos utilizadores finais sejam consolidadas num conjunto consistente de restrições, as quais são atribuídas à ligação a ser estabelecida. Note-se que um conjunto de restrições de QoS é aceitável para a instalação de uma chamada, se for aceitável para ambos os utilizadores. Consequentemente, o *processamento da chamada* da origem obtém, a partir do encaminhamento, um caminho aceitável que satisfaça as restrições de QoS consolidadas. A QoS correspondente ao caminho seleccionado é referido como a QoS disponível [16].

A negociação de QoS é um processo que envolve a rede e os utilizadores finais, e que ocorre durante a instalação da chamada, para determinar o nível de QoS que foi estipulado pelos utilizadores finais e que a rede pode suportar. Logo que seja concluída a negociação de QoS, um *pedido de ligação remota* é enviado através da rede e indicado ao utilizador no destino. A resposta é transmitida ao *processamento da chamada* do nó origem, que notifica o utilizador na origem da conclusão da negociação de QoS. Por fim, o *processamento da chamada* percorre o caminho, fixando os recursos desde a origem até ao destino.

5. Gestão da ligação

A **gestão da ligação** é um mecanismo de controlo da rede ao nível da ligação, responsável pelo estabelecimento, continuação e libertação das ligações. A Fig. 2 mostra as transições entre quatro estados distintos de uma ligação : *estabelecimento*, *restabelecimento*, *transferência de informação* e *libertação*.

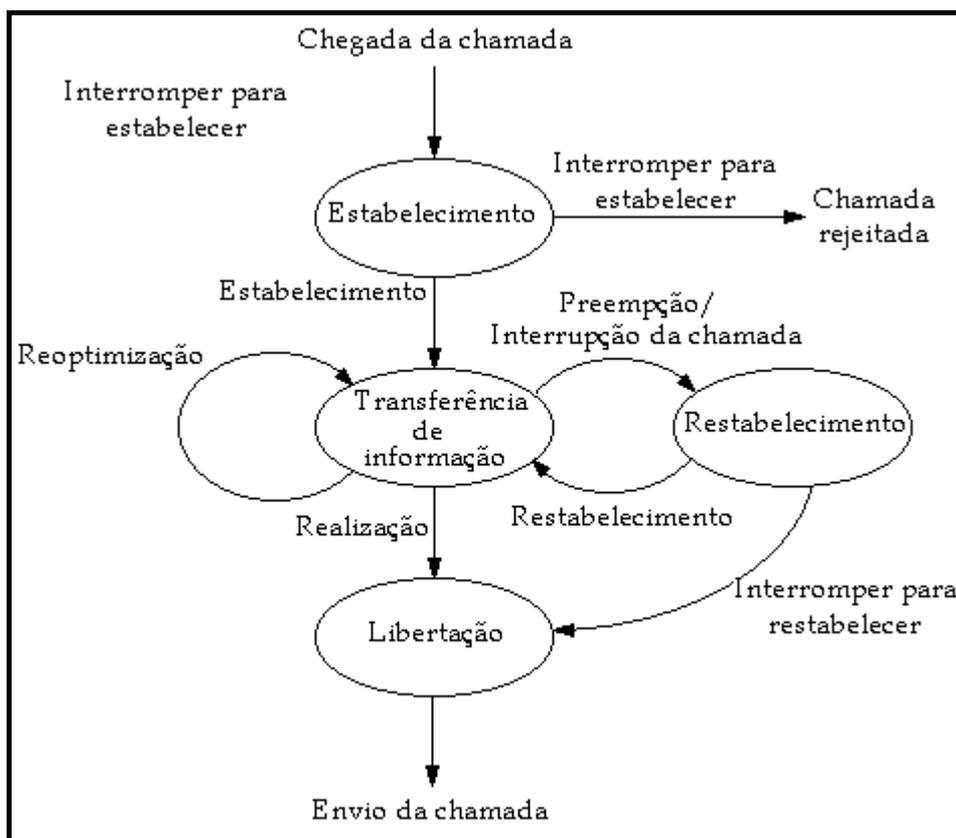


Fig. 2 – Estados de uma ligação [16].

O **estabelecimento** está associado à instalação da ligação.

O **restabelecimento** é necessário após a ligação existente ter sido interrompida, devido a uma falha da rede ou por provocar a preempção. A *reoptimização* é executada pela rede para conservar os recursos utilizados para estabelecer as ligações.

Quando a chamada chega a um nó da rede, regista o estado de *estabelecimento*. Se não existem recursos suficientes para suportar a chamada, esta é rejeitada; caso contrário, ao estabelecer com êxito a ligação (isto é, utilizar um caminho em que são satisfeitas todas as restrições de QoS), a chamada regista o estado de **transferência de informação**.

Quando a chamada no estado de *transferência de informação* é concluída, regista o estado de **libertação** e a ligação é imediatamente libertada. Quando uma chamada está no estado de *transferência de informação*, a sua ligação pode ser interrompida ou ficar sujeita a preempção. Se isto acontecer, a chamada regista o estado de *restabelecimento* até que a rede possa tentar, automaticamente, restabelecer a ligação, determinando um novo caminho aceitável. Durante o restabelecimento, os requisitos de QoS associados à chamada afectada, são ajustados de acordo com as regras de actualização, que têm em consideração o valor previamente estabelecido. Caso o *restabelecimento* tenha êxito, a chamada volta ao estado de *transferência de*

informação. O intervalo de tempo no qual a tentativa de *restabelecimento* de uma ligação pode ser repetido, é limitado pelo atraso do restabelecimento da chamada. Além deste atraso, o processo de *restabelecimento* é abortado e a chamada entra no estado de *libertação*.

A *reoptimização* é concluída pela rede sem o envolvimento directo do utilizador, determinando um caminho mais económico ou satisfazendo as restrições de QoS mais rigorosas. O processo da *reoptimização* é semelhante ao do *restabelecimento*.

Durante a preempção da chamada, a chamada que provoca a preempção deve estar no estado de *estabelecimento* ou de *restabelecimento* da ligação. Durante o cálculo do caminho, a chamada apenas pode provocar a preempção nas chamadas de mais baixa prioridade que estejam no estado de *transferência de informação*. No entanto, para provocar a preempção devido a conflitos nos recursos, quando estes são atribuídos à nova ligação, a chamada sujeita a preempção pode estar em qualquer estado.

6. Encaminhamento sujeito a restrições de QoS

Nesta secção são examinados os problemas de encaminhamento sujeitos a restrições de desempenho, de recursos e de prioridade.

No **encaminhamento sujeito a restrições de desempenho**, uma restrição no caminho é obtida a partir do valor aceitável de cada parâmetro de desempenho. Cada métrica associada a um caminho com restrições, pode ser ela própria um critério de minimização para aumentar a possibilidade de encontrar um caminho que satisfaça essa restrição específica no caminho. No entanto, segundo [16] “um problema de caminho mais curto multicritério não é um problema de optimização bem definido, a não ser que todos os critérios estejam envolvidos numa única função utilidade, utilizada como a única função objectivo do problema”.

No **encaminhamento sujeito a restrições de recursos**, é prática corrente atribuir pesos aos arcos e incorporá-los na função objectivo do encaminhamento, de forma a adaptar a selecção preferencial dos recursos da rede. Neste caso, contudo, pode não ser possível satisfazer as preferências utilizador, atendendo aos coeficientes de ponderação que atribui aos recursos, dado que a minimização de uma função objectivo soma ponderada ao longo do caminho, não garante a minimização do uso dos recursos num dado arco. Além disso, o desempenho do encaminhamento é muito sensível aos pesos atribuídos aos arcos [16].

No **encaminhamento sujeito a restrições de prioridade**, os algoritmos de encaminhamento que provoca a preempção, que contam com o transbordo para determinar os

caminhos desejáveis, não são muito eficientes na utilização dos recursos da rede [16]. Os algoritmos do encaminhamento que não provoca a preempção, que reservam os recursos da rede nas chamadas de alta prioridade, não são muito eficientes, uma vez que os recursos reservados não reclamados não podem ser utilizados pelas chamadas de baixa prioridade.

Na abordagem que provoca a preempção, pode ser evitado o transbordo, se toda a informação relevante para o encaminhamento ficar disponível em cada nó, através de um protocolo eficiente de distribuição da topologia. As *restrições de prioridade* podem ser especificadas em relação aos estados das ligações, se forem implementadas como restrições nos arcos. Especificamente, a cada chamada é atribuído um nível de prioridade para os estados de estabelecimento da ligação (*prioridade de estabelecimento*), de restabelecimento da ligação (*prioridade de restabelecimento*) e de transferência de informação (*prioridade de retenção*) [16]. Apenas é permitida a preempção, quando a prioridade da chamada que provoca a preempção é mais elevada do que a da chamada que fica sujeita a preempção. O nível de prioridade adequado utilizado na comparação, é aquele que está associado ao estado de *ligação* de uma dada chamada. Assim, a admissibilidade da chamada num dado arco, depende do nível adequado de prioridade da chamada. Os valores da prioridade são também utilizados para resolver conflitos dos recursos, quando várias chamadas tentam, simultaneamente, utilizar os mesmos recursos da rede. Neste caso, a chamada com a mais elevada prioridade é processada em primeiro.

Ocorre um efeito cíclico indesejado, quando uma chamada com elevada prioridade de restabelecimento, provoca a preempção a uma chamada existente com prioridade de retenção muito baixa, e posteriormente, fica sujeita a preempção por esta última, porque a prioridade de retenção da primeira é menor do que a prioridade de restabelecimento da última. Consequentemente, as duas chamadas podem, alternativamente, comutar entre os estados de restabelecimento da ligação e o de transferência de informação. O efeito cíclico pode ser evitado, se a prioridade de retenção requerida por cada chamada for, pelo menos, tão elevada como a sua prioridade de restabelecimento [16].

7. Estruturas de encaminhamento

Nesta secção, faz-se uma breve referência a algumas estruturas de encaminhamento, comparando-se as suas capacidades em redes integradas de comunicações. Para tal, consideram-se cinco aspectos específicos : economia de escala, economia de alcance, elevada

granularidade, optimalidade do caminho e capacidade adaptativa. As comparações estão sumariadas na seguinte tabela [16] :

Estrutura do encaminhamento	economia de alcance	economia de escala	elevada granularidade	optimalidade do caminho	capacidade adaptativa
chamada a chamada	×		×	×	×
controlado por tabela		×			
na origem	×		×	×	
passo a passo		×			×
com "fallback"	×		×	×	×
alternativo		×			×

O **encaminhamento chamada a chamada** tem flexibilidade para determinar um caminho com as características e os requisitos de QoS de cada chamada [16]. É permitido o controlo da rede com elevada granularidade, já que o cálculo do caminho é feito à chegada de cada chamada. Como o cálculo total é independente do número de tipos de chamada, então existe economia de alcance. Desde que as redes integradas actuais estejam bem definidas (de forma a suportar as relações requeridas de QoS), o encaminhamento chamada a chamada apresenta grande potencial para o encaminhamento sensível à QoS.

No **encaminhamento controlado por tabela predefinida**, todos os caminhos são predeterminados e guardados em tabelas de encaminhamento [16]. Existe economia de escala, porque muitas chamadas são encaminhadas pelo mesmo caminho. Os caminhos escolhidos não podem ser óptimos para todos os tipos de tráfego, a não ser que tabelas diferentes sejam utilizadas em cada tipo de tráfego. As tabelas de encaminhamento têm que se adaptar às alterações topológicas.

No **encaminhamento na origem**, as decisões para encaminhar uma chamada são totalmente produzidas na origem, baseadas na configuração geral e no estado da rede, as quais são actualizadas através de um protocolo de distribuição da topologia, o que faz com que as funções de encaminhamento nos nós intermédios sejam relativamente simples [16]. Com suficiente informação da topologia disponível na origem, é possível ter um controlo da rede com elevada granularidade e otimizar os caminhos desejados. Existe economia de alcance, já que a mesma informação da topologia pode ser utilizada por muitos tipos de tráfego. Por isso, é que nas redes integradas modernas o encaminhamento na origem é muito

prometedor para o caminho escolhido onde existe um aumento crescente de muitos tipos de tráfego, devido às aplicações “multimedia”.

No **encaminhamento passo a passo**, as decisões de encaminhamento são distribuídas e o cálculo do encaminhamento nos nós intermédios não é trivial [16]. Com a utilização de tabelas de encaminhamento, existe economia de escala, apesar destas tabelas exigirem armazenamento. Neste tipo de encaminhamento, são necessários passos para prevenir ciclos. Embora o processo possa responder rapidamente a interrupções, a recuperação é sub-ótima, porque retém o subcaminho da origem ao ponto de interrupção. É difícil utilizar o encaminhamento passo a passo para suportar a adaptação da QoS ao nível da chamada (por exemplo, chamada que provoca a preempção), porque não existe informação suficiente, relativa às chamadas específicas, de onde as decisões de encaminhamento são produzidas.

O **encaminhamento com “fallback”** (isto é, com possibilidade de voltar ao início do processo, calculando outro caminho) determina, sequencialmente, os caminhos baseados numa sequência de instâncias de encaminhamento, até uma delas estar disponível, ou a chamada ficar bloqueada ao realizar uma sequência predeterminada de “fallbacks” (Fig. 3). O encaminhamento com “fallback” adapta-se às alterações do estado da rede, permitindo a determinação de caminhos alternativos, para adaptar as restrições de recursos preferidas, às chamadas que provocam a preempção, e de outras marcas de encaminhamento que exijam o cálculo de vários caminhos por cada chamada instalada. Como o encaminhamento com “fallback” oferece grande economia de alcance, adapta utilizadores heterogéneos através do cálculo de caminhos alternativos quando necessário. A sequência de instâncias de encaminhamento com “fallback” é predeterminada ou seleccionada em tempo real, de acordo com as regras estabelecidas. As instâncias de encaminhamento podem ser especificadas para suportar a elevada granularidade, desde que a informação pertinente de QoS esteja disponível.

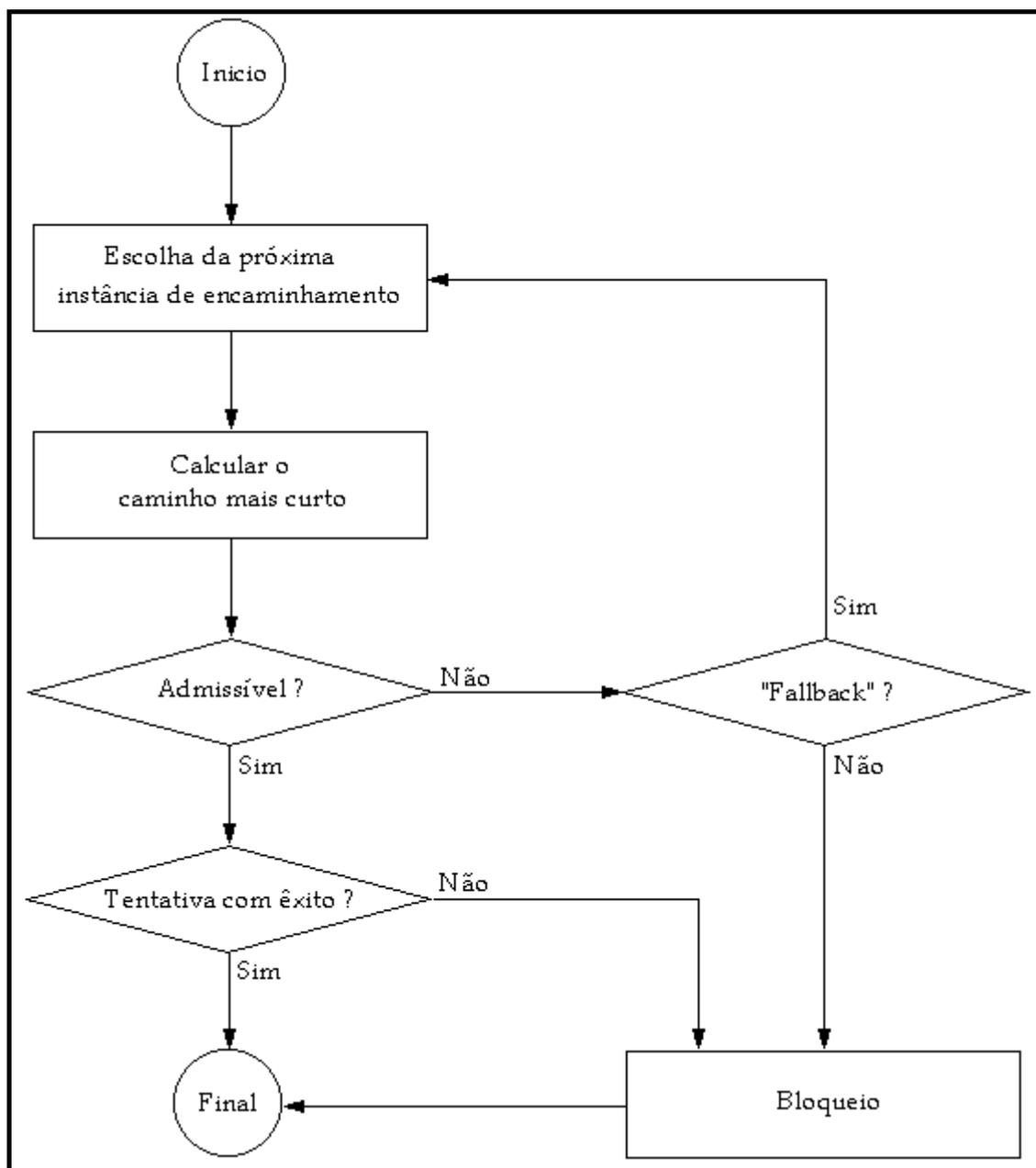


Fig. 3 – Encaminhamento com “fallback” [16].

No **encaminhamento alternativo**, um conjunto de caminhos predeterminados e armazenados em tabelas de encaminhamento é testado sequencialmente, durante a instalação de cada chamada para fixar os recursos, até uma tentativa ter sucesso ou a chamada ficar bloqueada na realização de uma seqüência de tentativas [16]. Os caminhos alternativos podem depender das classes de tráfegos e da hora do dia. A utilização de tabelas de encaminhamento apresenta economia de escala, mas é necessário uma significativa armazenagem. As tabelas de encaminhamento podem ser actualizadas periodicamente, de modo a se adaptarem às condições existentes da rede. O modo como os caminhos são testados

depende das variantes de encaminhamento alternativo existentes. Enquanto sob cargas leves o encaminhamento alternativo minimiza a probabilidade de bloqueio, sob cargas intensas a probabilidade de bloqueio pode aumentar drasticamente, quando os caminhos alternativos utilizados tendem a consumir mais recursos da rede.

8. Encaminhamento chamada a chamada na origem baseado em regras

A estratégia proposta em [16] de encaminhamento chamada a chamada na origem com “fallbacks” baseado em regras, em redes de comunicações com tráfegos integrados sujeitos a diversos requisitos de QoS, consiste em determinar, eficientemente, um caminho aceitável para cada chamada, de acordo com o estado corrente da rede. Em oposição ao paradigma do encaminhamento tradicional, onde o passo inicial consiste em minimizar o valor de uma função objectivo, o paradigma de encaminhamento que se descreve a seguir realça a conjugação de várias restrições. Apesar disso, o algoritmo de caminho mais curto é utilizado habitualmente como uma forma de identificar caminhos aceitáveis.

8.1. Encaminhamento com “fallback” baseado em regras

A estratégia de encaminhamento baseado em regras utiliza toda a informação disponível para modificar, dinamicamente, a sequência de caminhos com “fallback”, determinados de acordo com os estados da rede e da ligação. A estratégia começa por calcular o caminho com uma instância de encaminhamento inicial, o qual é determinado através do estado da ligação e dos requisitos de QoS utilizados. Se não for determinado qualquer caminho admissível, então são utilizadas regras de paragem para decidir se o “fallback” calculado está em ordem. No “fallback” escolhe-se uma nova instância de encaminhamento, com as restrições relaxadas de acordo com as regras de “fallback”. Se for encontrado um caminho admissível, então a rede tentará fixar os seus recursos para a chamada, ao longo do caminho. Por vezes esta tentativa falha, porque há contenção dos recursos devido à latência na actualização da topologia, o que provoca a interrupção da chamada.

A alternativa mais sofisticada é permitir “crankback” (isto é, seleccionar mediante certas condições), até que a origem escolha uma nova instância de encaminhamento por “fallback”, utilizando a última informação obtida a partir do teste sem sucesso. O encaminhamento com “crankback” é semelhante ao encaminhamento alternativo, excepto

que os caminhos alternativos não são predeterminados, mas calculados um de cada vez, após cada tentativa sem êxito para estabelecer uma ligação.

A estratégia proposta para o encaminhamento baseado em regras segue um modelo genérico que consiste em três módulos : **base de dados**, **base de conhecimento** e **mecanismo de inferência** [16]. A *base de dados* contém a informação da topologia, que é actualizada através de um protocolo de distribuição da topologia. A *base de conhecimento* contém regras que são utilizadas para gerar instâncias de encaminhamento, baseadas numa política de encaminhamento predeterminada e da QoS pedida pelos utilizadores. O *mecanismo de inferência* é basicamente um cálculo sequencial de caminhos mais curtos, a qual verifica também a admissibilidade do caminho.

“Fallback” com restrições de desempenho — Em vez de calcular um caminho óptimo para o problema de caminho mais curto com restrições, a estratégia de encaminhamento baseado em regras utiliza a seguinte heurística para as “fallbacks” [16] :

Inicialmente, resolve-se o problema original de encaminhamento sem restrições nos caminhos, verificando-se, a seguir, a admissibilidade através da verificação do caminho escolhido e comparando as restrições nos caminhos. Se estas não forem satisfeitas para o caminho escolhido, a “fallback” permite que a chamada tenha uma ou mais oportunidades adicionais para procurar um outro caminho admissível.

Em vez de utilizar uma função utilidade, a estratégia do encaminhamento baseada em regras pressupõe que os critérios de encaminhamento são ordenados pelos utilizadores da rede, de acordo com as importâncias relativas. Estes são então utilizados para determinar as funções objectivo de caminho mais curto para o cálculo dos caminhos inicial e com “fallback”.

“Fallback” com restrições de recursos — A estratégia de encaminhamento baseado em regras pode introduzir preferências nas restrições de recursos. As preferências utilizadas nos recursos podem-se traduzir por privilégios a atribuir a conjuntos de restrições de recursos. Cada um destes conjuntos pode ser escolhido para definir uma instância de encaminhamento, de acordo com algumas regras predeterminadas. No encaminhamento com “fallback”, primeiro são calculados os caminhos que apenas utilizam os recursos que satisfazem as restrições (de recursos) requeridas, com um “fallback” para as restrições de recursos aceitáveis menos restritivas [16].

“Fallback” com restrições de prioridade — Nas redes com classes de prioridade, o encaminhamento que provoca a preempção pode ser interrompido, se o débito total não for aumentado, devido às chamadas que provocam a preempção. Apesar disso, e uma vez que

nenhum recurso está reservado para as chamadas de alta prioridade, não é provável que as chamadas de baixa prioridade sejam bloqueadas quando a rede está ligeiramente sobrecarregada. Muitas vezes, é possível relaxar as restrições de recurso de uma chamada, até que as suas restrições de desempenho possam ser encontradas, sem interrupções desnecessárias nas chamadas de mais baixa prioridade. Uma vez que a preempção está associada a interrupções, o encaminhamento que não provoca a preempção deverá ser testado antes da preempção ser considerada. Esta abordagem é referida como preempção de “look-around-first”. A abordagem com “fallback” para o encaminhamento, permite a preempção de “look-around-first” para evitar a preempção desnecessária, na primeira tentativa para determinar um caminho admissível sem provocar preempção em qualquer chamada existente [16]. Se isto falhar, então na procura de um caminho admissível, a instância de encaminhamento com “fallback” apenas dá conta dos recursos consumidos pela mesma chamada ou pelas chamadas de mais alta prioridade.

8.2. Regras de “fallback” dependente do estado

Para facilitar a adaptação da QoS ao nível da chamada, as estratégias existentes de encaminhamento alternativo utilizam regras simples para seleccionar os caminhos alternativos predeterminados. Alguns protocolos de encaminhamento entre domínios, tal como OSPF, confiam nas instâncias predeterminadas de encaminhamento que dependem do tipo de serviço. A estratégia proposta de encaminhamento baseado em regras utiliza uma base de conhecimento, que contém um conjunto de regras para seleccionar, em tempo real, as instâncias de encaminhamento, quando necessárias para os cálculos dos caminhos inicial e com “fallback”. Nas suas principais características incluem-se regras para seleccionar restrições nas instâncias de encaminhamento, regras de “fallback” de uma instância para outra e regras para determinar quando fazer “fallback” e quando parar.

9. Análise como problema multiobjectivo

Como se pode concluir das secções anteriores, é vantajoso considerar o problema de encaminhamento sujeito a restrições de qualidade de serviço como um problema multiobjectivo. Para tal, restrições nos caminhos, tais como custo, atraso (ou probabilidade de bloqueio) e largura de banda, são introduzidas explicitamente nos modelos matemáticos como funções objectivo de encaminhamento a optimizar.

Desta forma, os modelos de encaminhamento com múltiplos objectivos permitem compreender os compromissos entre os vários requisitos de QoS, os quais podem, eventualmente, estar dependentes das aplicações, racionalizando a comparação entre os diferentes encaminhamentos alternativos, de modo a confrontar os critérios conflituosos envolvidos no cálculo para seleccionar os caminhos aceitáveis. Assim, os modelos tornam-se mais realistas, assumindo a natureza multiobjectivo do problema, permitindo compreender os conflitos inerentes e os compromissos entre os vários objectivos, ao seleccionar o melhor plano de compromisso a partir do conjunto de soluções não dominadas [4].

10. Uma abordagem multiobjectivo ao problema do encaminhamento

Em redes integradas de comunicações, o problema de encaminhamento é modelado como um problema de caminho mais curto com múltiplos objectivos, no qual os vários aspectos do cálculo são explicitamente incorporados na formulação do problema, dando ênfase à procura de caminhos de compromisso satisfatórios, relacionados com os vários requisitos de QoS.

Para cada métrica, os requisitos de QoS são expressos como restrições adicionais “leves” (isto é, não incorporadas explicitamente na formulação matemática do problema) nos valores da função objectivo, em termos de limites requerido e aceitável, os quais definem regiões de preferência no espaço das funções objectivo. Estas regiões são então analisadas para determinar os caminhos (se existirem) que satisfaçam aqueles requisitos.

10.1. As métricas para os requisitos de QoS

Geralmente, as métricas de encaminhamento consideradas são o atraso, o custo, o número de arcos, a probabilidade de perda, a taxa de erro e a largura de banda. A função de agregação para calcular o valor de um caminho depende da métrica utilizada. Como as métricas são representações de uma rede no encaminhamento, então as maiores dificuldades encontram-se não só na complexidade do cálculo do caminho, mas também na cadeia de requisitos de QoS que se pode suportar. As características desejáveis de cada métrica são as seguintes [26]:

- i) Para qualquer métrica escolhida, devem existir algoritmos eficientes para calcular os caminhos, para que o protocolo de encaminhamento seja capaz de abranger as grandes redes. A complexidade destes algoritmos deve ser, de preferência, comparável com a dos

algoritmos de encaminhamento actuais. Qualquer algoritmo deve também ser capaz de funcionar, tanto em ambientes centralizados, como em ambientes distribuídos.

- ii) As métricas devem reflectir as características básicas da rede. A informação que elas contém deve permitir que suportem os requisitos básicos de QoS. Note-se que qualquer requisito de QoS tem que ser planeado de acordo com restrições num caminho expresso em termos de métricas; portanto, as métricas determinam, até certo ponto, os tipos de QoS que a rede pode suportar.
- iii) As métricas devem ser ortogonais entre si, de modo a não existir informação redundante entre elas. A informação redundante pode introduzir interdependência entre as métricas, tornando impossível o cálculo de cada métrica, separadamente. A avaliação recursiva entre métricas pode complicar substancialmente o cálculo dos caminhos.

Os algoritmos com apenas uma métrica para determinar caminhos, tais como atraso e número de arcos, são bem conhecidos e têm sido muito utilizados nas redes actuais. Desta forma, será que uma única métrica poderá suportar os requisitos de QoS utilizados ?

Uma possível abordagem consiste em definir uma função e gerar uma única métrica com vários parâmetros. A ideia é fixar vários fragmentos de informação numa única medida e utilizá-la como base para as decisões de encaminhamento.

No entanto, uma única métrica fixa pode apenas ser utilizada, na melhor das hipóteses, como um indicador, visto que ela não contém informação suficiente para estabelecer se os requisitos de QoS utilizados podem ser encontrados ou não. Outro problema é utilizar parâmetros fixos de diferentes regras de composição de métricas individuais.

Apesar de várias métricas poderem modelar uma rede com mais precisão, a dificuldade está em arranjar algoritmos eficazes para resolver problemas de caminho mais curto com várias restrições adicionais, uma vez que a introdução destas restrições destrói algumas propriedades fundamentais do modelo.

O problema no encaminhamento com QoS é muito mais complicado, uma vez que os requisitos dos recursos especificados são muito diversos e dependentes da aplicação. Inicialmente, a complexidade do cálculo é determinado pelas regras de composição das métricas. Basicamente, existem três tipos de métricas : aditiva, multiplicativa e côncava [26]. Seja c_{ij} a métrica para o arco (i, j) e p um caminho numa rede; então,

- a) a métrica é aditiva se

$$V_p = \sum_{(i,j) \in p} c_{ij}$$

b) a métrica é multiplicativa se

$$V_p = \prod_{(i,j) \in p} c_{ij}$$

c) a métrica é côncava se

$$V_p = \min_{(i,j) \in p} \{ c_{ij} \}$$

Enquanto o atraso, o número de arcos e o custo seguem a função de agregação aditiva, a largura de banda segue a regra de agregação côncava.

As métricas de probabilidade de perda (e de taxa de erro) seguem a função de agregação

$$V_p = 1 - \prod_{(i,j) \in p} (1 - c_{ij})$$

No entanto, esta métrica pode ser transformada, primeiro numa métrica multiplicativa, e depois numa aditiva (podendo assim ser usada numa abordagem de caminho mais curto), da seguinte forma :

$$V_p = 1 - \prod_{(i,j) \in p} (1 - c_{ij})$$

$$1 - V_p = \prod_{(i,j) \in p} (1 - c_{ij})$$

Esta expressão dá a probabilidade de não bloqueio, que segue a função de agregação multiplicativa.

No entanto, aplicando logaritmos, obtém-se :

$$\log (1 - V_p) = \log \prod_{(i,j) \in p} (1 - c_{ij})$$

$$\log (1 - V_p) = \sum_{(i,j) \in p} \log (1 - c_{ij})$$

A maximização da probabilidade da transmissão ter êxito é então equivalente a

$$\max \log (1 - V_p) = \max \sum_{(i,j) \in p} \log (1 - c_{ij})$$

que é equivalente a

$$\min \sum_{(i,j) \in p} -\log (1 - c_{ij})$$

Uma vez que c_{ij} é uma probabilidade, o coeficiente associado ao arco (i, j) , que é positivo, é

$$-\log (1 - c_{ij})$$

Qualquer métrica aditiva pode ser usada numa abordagem de caminho mais curto. No entanto, a abordagem proposta para o caso multiobjectivo pode ser utilizada sempre que

forem consideradas métricas aditivas, multiplicativas e probabilidade de perda, em que as duas últimas são transformadas no caso aditivo.

10.2. O algoritmo proposto

A metodologia geral para resolver problemas de encaminhamento quando modelados como problemas de caminho mais curto com h objectivos ($h \geq 2$), é a seguinte :

1. Determinar as soluções não dominadas que optimizam cada função objectivo individualmente, resolvendo h problemas de caminho mais curto. Isto produz informação relativa ao conjunto de valores de cada função objectivo, no conjunto de soluções não dominadas.
2. Indicar os requisitos de QoS para cada uma das h métricas (associadas às h funções objectivo), os quais são especificados através dos seguintes limites :

valor requerido (nível de aspiração) : M_{req}

valor aceitável (nível de reserva) : M_{ac} ($M_{req} < M_{ac}$).

O facto de acrescentar estes tipos de restrições “leves” define regiões de prioridade, nas quais as soluções não dominadas são pesquisadas de acordo com os requisitos de QoS. De forma a facilitar a ilustração gráfica deste método, a Fig. 4 mostra a forma de utilizar os requisitos de QoS para definir as regiões de prioridade de um problema com duas funções objectivo.

Esta abordagem tanto pode ser implementada de uma maneira interactiva (pedindo ao AD informação para conduzir a procura de soluções não dominadas), como de uma maneira automática (seguindo regras de procura predefinidas). O algoritmo consiste em optimizar funções objectivo de somas pesadas, para calcular as soluções não dominadas que pertencem ao contorno convexo, e em utilizar um algoritmo eficiente dos k caminhos mais curtos, para procurar as soluções não dominadas pertencentes às zonas do interior do respectivo invólucro convexo.

Neste algoritmo, os pesos associados à função escalar soma ponderada são calculados a partir das componentes, normalizadas, do gradiente do hiperplano que passa pelos pontos formados com os valores requeridos e com os que optimizam as funções objectivo separadamente. Com estes pesos, a região de maior interesse (a de primeira prioridade) é pesquisada de uma forma mais uniforme, uma vez que aquele gradiente traduz a relação existente entre os valores requeridos.

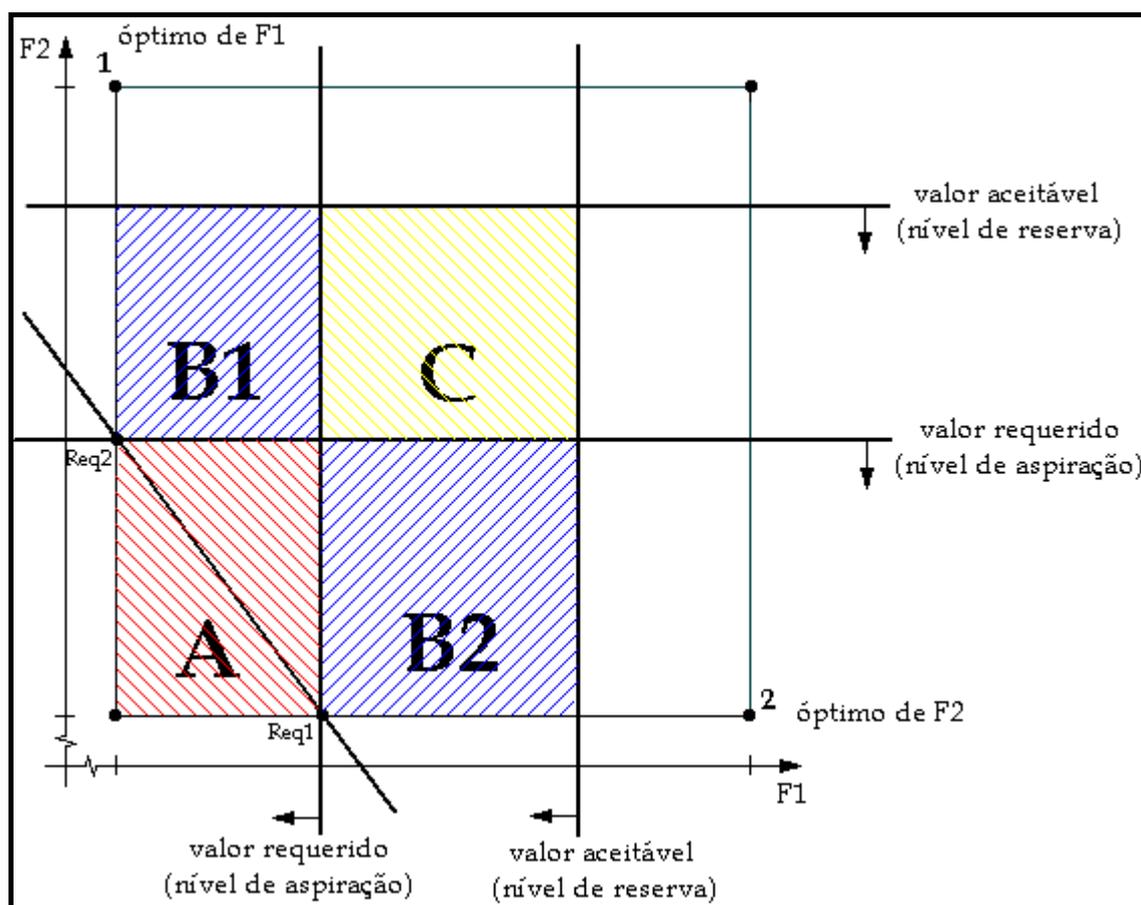


Fig. 4 – Os requisitos de QoS utilizados para definir as zonas de prioridade.

O passo principal deste algoritmo, cuja descrição pormenorizada se encontra em Algoritmo 1, consiste no seguinte :

1. determinar uma solução (não dominada) utilizando o algoritmo dos k caminhos mais curtos (MPS) aplicado à função escalar soma ponderada construída da forma descrita,
2. analisar a que região de prioridade pertence a solução,
3. se esta solução é melhor do que a melhor encontrada até ao momento (se o nível de prioridade desta nova solução for menor do que o da melhor solução encontrada até ao momento), então actualizar a melhor solução (que passa a ser esta última),
4. se todas as r regiões de maior prioridade ficaram totalmente analisadas, verificar se a melhor solução, encontrada até ao momento, pertence a uma dessas regiões ou à de prioridade $r+1$: se pertence, então o processo termina (sendo esta a solução escolhida), caso contrário continua a pesquisa da solução desejada.

Na descrição de Algoritmo 1, considere-se a seguinte notação :

$h \leftarrow$ número de funções objectivo (métricas) do problema;

NZonas \leftarrow número de regiões de prioridade;

Req(i), Ac(i) \leftarrow valores requerido e aceitável para o objectivo **i** ($i = 1, \dots, h$);

Solução(i) \leftarrow solução que optimiza a função objectivo **i** ($i = 1, \dots, h$)

Opt(i) \leftarrow valor da função objectivo **i** da Solução(i) ($i = 1, \dots, h$)

Se não existe qualquer solução **Ou** existe apenas uma **Então**

STOP

Indicar os requisitos de QoS associados a cada função objectivo : Req(i) e Ac(i)

Zona(j) \leftarrow região de prioridade **j** ($j = 1, \dots, NZonas$) — utiliza Req(i) e Ac(i)

MelhorSolução $\leftarrow \emptyset$ (a melhor solução encontrada até ao momento)

MelhorNível $\leftarrow \infty$ (nível de prioridade da região a que pertence a MelhorSolução)

FObjPesada \leftarrow função escalar soma ponderada, construída à custa de Req e de Opt

Solução \leftarrow caminho mais curto relativo à função objectivo FObjPesada

REPETIR

Prioridade \leftarrow nível de prioridade da região a que pertence a Solução

Se Prioridade = 1 **Então**

MelhorSolução \leftarrow Solução

STOP { encontrou a solução desejada }

Se Prioridade < MelhorNível **Então**

MelhorSolução \leftarrow Solução

MelhorNível \leftarrow Prioridade

Para j desde 1 até NZonas Fazer

Se Zona(j) foi pesquisada na totalidade e MelhorNível = j + 1 **Então**

STOP { encontrou a solução desejada }

Solução \leftarrow próximo caminho mais curto relativo à função objectivo FObjPesada

ATÉ encontrar a solução desejada

Algoritmo 1. Problema de encaminhamento : determinação da melhor solução.

10.3. A largura de banda como métrica

A largura de banda (ou débito) de um caminho é definida como a largura de banda residual mínima entre todos os arcos do caminho, ou a largura de banda segundo a regra do “funil” [26].

Se for considerada a largura de banda requerida, então, e uma vez que é uma métrica côncava (isto é, $b_{st} = \min \{ b_{ij} : (i, j) \in p \}$ em que p é um caminho), os requisitos de QoS para esta métrica, em termos dos mesmos limites

valor requerido (nível de aspiração) : b_{req}

valor aceitável (nível de reserva) : b_{ac} ($b_{ac} < b_{req}$)

pode ser tratada do seguinte modo :

- constrói-se uma nova rede, removendo os arcos (i, j) para os quais $b_{ij} < b_{req}$;
- se a rede resultante é conexa, então existem caminhos entre os nós origem e destino que satisfazem a largura de banda requerida (nível de aspiração);
- se a rede se torna não conexa, então removem-se apenas os arcos (i, j) da rede original, tais que $b_{ij} < b_{ac}$;
- se esta nova rede é conexa, então existe um caminho entre os nós origem e destino que satisfaz a largura de banda aceitável (nível de reserva);
- se a rede não é conexa, então não existe caminho para a largura de banda especificada pelos parâmetros de QoS.

10.4. Problema com duas funções objectivo

Para o problema com duas funções objectivo, podem definir-se quatro regiões associadas a três níveis de prioridade : A (nível 1), B1, B2 (nível 2) e C (nível 3) — Fig. 5. A cada função objectivo (métrica) estão associados dois requisitos de QoS : valor requerido e valor aceitável.

Na região A, de primeira prioridade, são satisfeitos os valores requeridos das duas métricas. Nas regiões B1 e B2, de segunda prioridade, é satisfeito o valor requerido de apenas uma das métricas e o valor aceitável da outra. Na região C, de terceira prioridade, o valor requerido das duas métricas não são satisfeitos, apenas são os seus valores aceitáveis.

Pode-se considerar uma outra região, formada por todos os restantes locais onde possam existir soluções não dominadas do problema com a menor prioridade (*última chance*). Nesta região, o valor aceitável (e o requerido) de pelo menos uma métrica não é satisfeito.

Por outro lado, se for indicada alguma preferência de um objectivo sobre o outro, isto pode ser traduzido fazendo uma distinção entre regiões com a mesma prioridade — neste caso apenas entre B1 e B2. Desta forma, se $F1 \succ F2$ ($F1$ tem preferência em relação a $F2$) então $B1 \succ B2$ ($B1$ é prioritária em relação a $B2$); se $F2 \succ F1$ então $B2 \succ B1$ (Fig. 5).

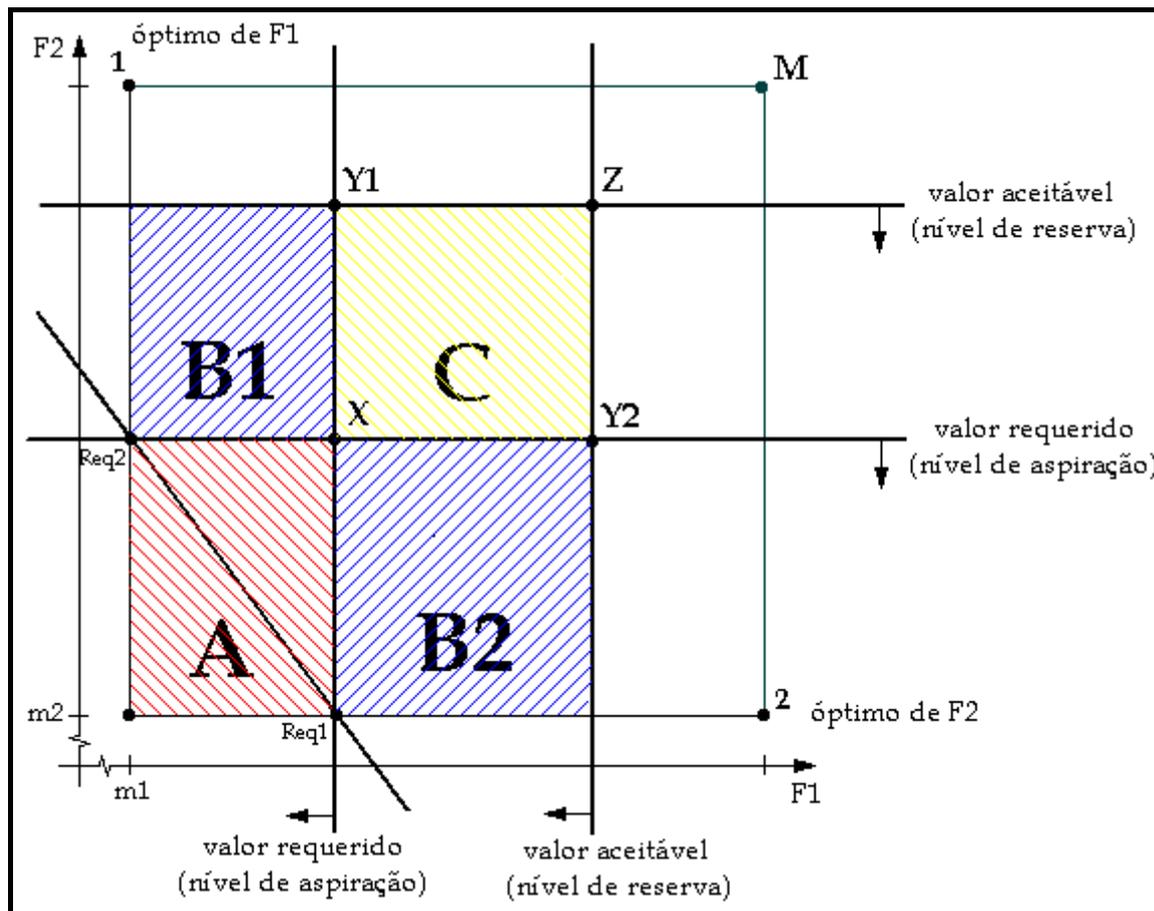


Fig. 5 – Encaminhamento (bi-objectivo) : construção das zonas de prioridade.

Note-se que qualquer solução de A não é dominada por soluções de B e C, e domina qualquer solução de C. Por outro lado, não existe solução de C que domine soluções de B. Da mesma maneira, se conclui que qualquer solução de *última chance* não domina as soluções das regiões A, B e C. Desta forma, fica justificada a não dominância da solução final.

Relativamente aos pesos envolvidos na construção da função escalar soma ponderada, estes são as componentes normalizadas do gradiente do hiperplano que passa pelos pontos Req1 e Req2, os quais são definidos através da conjugação dos valores requeridos e dos ótimos de F1 e F2, respectivamente m1 e m2, da seguinte forma :

$$\text{Req1} = (M_{\text{req}}(\text{F1}), m2),$$

$$\text{Req2} = (m1, M_{\text{req}}(\text{F2}))$$

como se pode ver na Fig. 5. Com estes pesos, a região A (a de primeira prioridade) é pesquisada de uma forma mais uniforme, uma vez que aquele gradiente traduz a relação existente entre os valores $M_{\text{req}}(\text{F1})$ e $M_{\text{req}}(\text{F2})$.

Na descrição apresentada em Algoritmo 1, considera-se que B1 tem preferência em relação a B2, o que não acarreta perda de generalidade. Desta forma, no desenvolvimento deste algoritmo, a primeira solução que se encontrar dentro da região A é imediatamente seleccionada. No entanto, se a região A foi totalmente analisada sem determinar qualquer solução não dominada (o que acontece quando a recta de nível correspondente à função escalar soma ponderada passar o ponto X — Fig. 5), então verifica-se se já foi encontrada alguma solução na região B1. Da mesma maneira, se a região B1 foi totalmente analisada (recta de nível passou o ponto Y1), então verifica-se se já foi encontrada alguma solução em B2. O mesmo se passa em relação às regiões B2 (ponto Y2) e C (ponto Z).

No entanto, só se pode verificar se B2 foi totalmente analisada, quando se tiver a certeza que B1 já o foi, porque se isto não for garantido, pode acontecer que B2 seja totalmente analisada antes de B1 (o que acontece quando a recta de nível passar por Y2 sem ainda ter passado por Y1). Consequentemente, pode-se escolher, como solução final, uma que pertença a C, apesar de B1 ainda não ter sido totalmente analisada. Relativamente às regiões A e C, não existe qualquer problema, uma vez que a região A é a primeira a ser analisada na sua totalidade e a região C é a última (o ponto X é o primeiro a ser passado e o Z o último, qualquer que sejam os valores requeridos e aceitáveis indicados).

Desta forma, o processo termina apenas quando uma das seguintes situações ocorrer (analisar a Fig. 5) :

- (1) foi encontrada uma solução pertencente à região A,
- (2) foi analisada totalmente a região A e já foi encontrada uma solução em B1,
- (3) foi analisada totalmente a região B1 e já foi encontrada uma solução em B2,
- (4) foram analisadas totalmente as regiões B1 e B2 e já foi encontrada uma solução em C,
- (5) foi analisada totalmente a região C, logo a solução escolhida é uma de *última chance*.

Por outro lado, é garantido que o processo termina, já que pelo menos uma solução é determinada : a solução 1 ou a 2 (Fig. 5). Se o problema não tiver qualquer solução, ou apenas uma, o algoritmo ao verificar esse facto termina imediatamente (Algoritmo 1).

10.5. Problema com três funções objectivo

Para o problema com três funções objectivo, podem ser definidas oito regiões correspondentes a quatro níveis de prioridade : A (nível 1), B1, B2 e B3 (nível 2), C1, C2 e C3 (nível 3) e D (nível 4), como mostram as figuras 6, 7 e 8.

A cada uma das três funções objectivo (métricas) estão associados dois requisitos de QoS : valor requerido e valor aceitável.

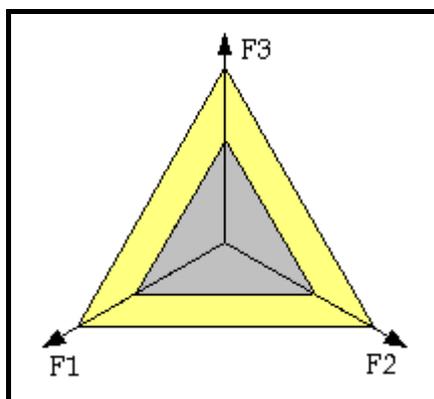


Fig. 6 – Encaminhamento (tri-objectivo) : zonas de primeira e quarta prioridade.

Na Fig. 6 encontram-se representadas a região A, de primeira prioridade (triângulo interior) e a região D, de quarta prioridade (coroa triangular exterior). Na região A são satisfeitos o valor requerido de todas as funções objectivo. Na região D, não é satisfeito o valor requerido de qualquer função objectivo, apenas é o valor aceitável de todas elas.

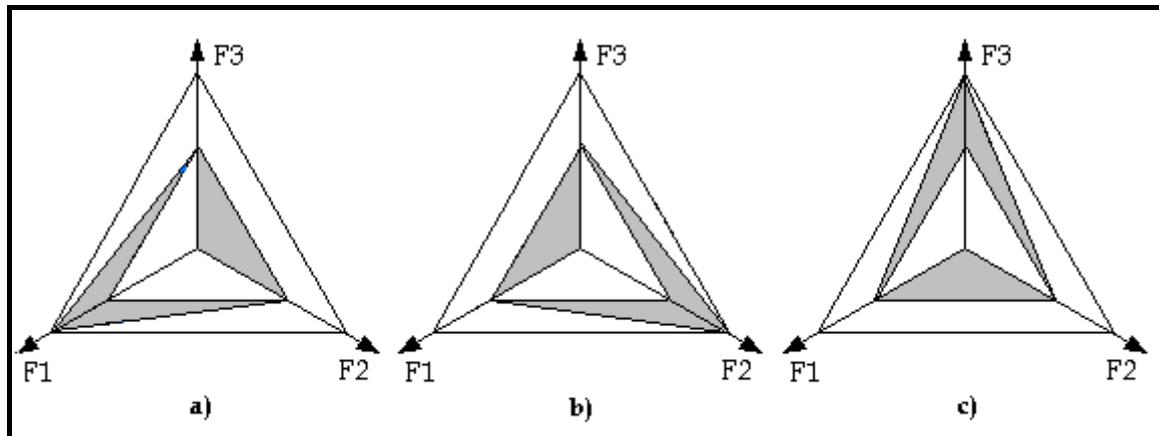


Fig. 7 – Encaminhamento (tri-objectivo) : zonas de segunda prioridade.

Nas regiões de segunda prioridade, B1, B2 e B3, são satisfeitos os valores requeridos de duas funções objectivo e apenas o aceitável da outra; ou seja, apenas o valor requerido de uma função objectivo não é satisfeito. Portanto, cada uma das três regiões formadas está relacionada com a função objectivo cujo valor requerido é satisfeito : B1 (Fig. 7.a)), B2 (Fig. 7.b)) e B3 (Fig. 7.c)), relacionado com a primeira, segunda e terceira função objectivo, respectivamente.

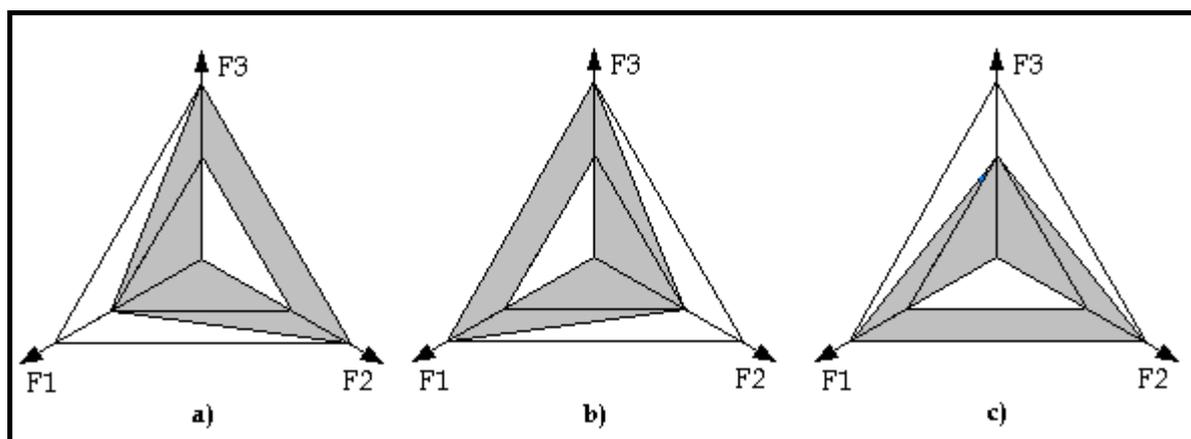


Fig. 8 – Encaminhamento (tri-objectivo) : zonas de terceira prioridade.

Nas regiões de terceira prioridade, C1, C2 e C3, são satisfeitos apenas o valor requerido de uma função objectivo e os aceitáveis das outras duas; ou seja, os valores requeridos de duas funções objectivo não são satisfeitos. Portanto, cada uma das três regiões formadas está relacionada com a função objectivo cujo valor requerido não é satisfeito : C1 (Fig. 8.a)), C2 (Fig. 8.b)) e C3 (Fig. 8.c)), relacionadas com a primeira, segunda e terceira função objectivo, respectivamente.

Pode-se considerar uma outra região, com a menor prioridade, associada às soluções não dominadas que não pertençam a qualquer das regiões descritas (*última chance*). Esta região está associada às soluções que não satisfazem os valores aceitáveis de pelo menos uma função objectivo. Isto é, o valor de pelo menos uma função objectivo é superior ao valor aceitável correspondente.

Também neste caso, a indicação de preferências entre os objectivos envolvidos, implica fazer-se distinção entre as regiões com a mesma prioridade. Por exemplo, se $F1 \succ F2 \succ F3$ então $B1 \succ B2 \succ B3$ e $C1 \succ C2 \succ C3$.

10.6. Exemplo ilustrativo

Para ilustrar o algoritmo proposto, atenda-se à análise de seis casos, que traduzem todas as situações genéricas possíveis de acontecer num problema com dois objectivos, de acordo com a Fig. 9.

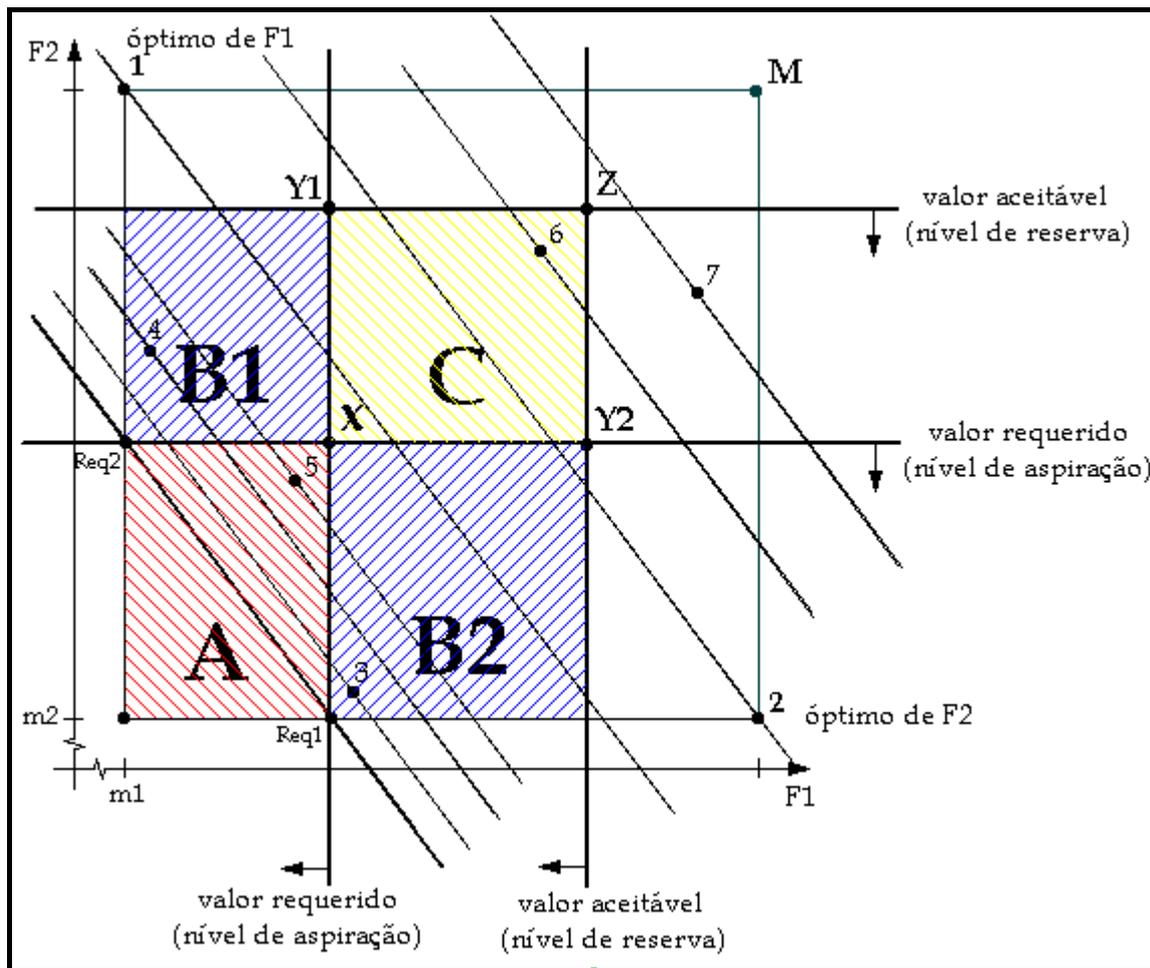


Fig. 9 – Determinação de soluções não dominadas nas várias zonas de prioridade.

Em primeiro lugar, são determinadas as soluções 1 e 2, que otimizam separadamente cada função objectivo. Depois, a partir dos requisitos associados a cada função objectivo indicados são construídas as regiões de prioridade. Finalmente, a primeira fase termina com a construção da função escalar soma ponderada, a partir dos requisitos de QoS e dos valores associados às soluções já determinadas, cujas rectas de nível são mostradas na Fig. 9.

1º Caso.

Usando aquela função escalar suponhamos que são alcançadas sucessivamente as soluções 3 e 4 (ambas fora da zona de primeira prioridade). A pesquisa prossegue e é obtida a solução 5. Desta forma, o processo termina com a escolha da solução 5, já que esta pertence à região de primeira prioridade e as soluções já encontradas (3 e 4) pertencem a uma zona de menor prioridade (B2 e B1, respectivamente).

2º Caso. Suponha-se que a solução 5 não existe.

Usando aquela função escalar são alcançadas sucessivamente as soluções 3 e 4 (ambas fora da zona de primeira prioridade). Prosseguindo com a pesquisa, obtém-se a solução 1. Desta forma, como a região de primeira prioridade foi totalmente pesquisada (a recta de nível associada à solução 1 passou X) e já foram determinadas soluções nas duas regiões de segunda prioridade, então o processo termina. Desta forma, se $F1$ tiver preferência relativamente a $F2$ ($F1 \succ F2$), então a solução escolhida é a 4; se for $F2$ a preferencial, é a solução 3 a escolhida; se não existir qualquer preferência entre os objectivos, a solução escolhida é a 3, pois foi a determinada em primeiro lugar.

3º Caso. Suponha-se que as soluções 4 e 5 não existem e $F1 \succ F2$ ($B1 \succ B2$).

Usando aquela função escalar é alcançada a solução 3. Depois a pesquisa prossegue e é obtida a solução 1. Nesta altura, a região A ficou totalmente pesquisada (a recta de nível que passa pela solução 1 está para além de X), mas como não foi encontrada qualquer solução na região de segunda prioridade de maior preferência ($B1$) e esta ainda não está totalmente pesquisada, a solução 3 não pode ser escolhida como solução final. Continuando a pesquisa obtém-se a solução 2. Desta forma, como a região $B1$ ficou totalmente pesquisada (a recta de nível que passa pela solução 2 está para além de $Y1$) o processo termina, uma vez que se pode escolher a solução 3 que pertence à região $B2$.

4º Caso. Suponha-se que as soluções 3 e 5 não existem e $F2 \succ F1$ ($B2 \succ B1$).

Usando aquela função escalar são alcançadas as soluções 4 e 1, por esta ordem. No entanto, a solução 4 não pode ser escolhida, pois apesar da região A estar totalmente pesquisada, ainda não foi encontrada qualquer solução na região $B2$ e esta ainda não está totalmente pesquisada. Prosseguindo a pesquisa, são obtidas sucessivamente as soluções 2 e 6. Agora, como a região $B2$ ficou totalmente pesquisada (a recta de nível que passa pela solução 6 está para além de $Y2$), o processo termina escolhendo-se a solução 4 como solução final, uma vez que pertence à região $B1$.

5º Caso. Suponha-se que as soluções 3, 4 e 5 não existem.

Usando aquela função escalar, são determinadas as soluções 1 e 2, sucessivamente. Como estas duas soluções pertencem às regiões de menor prioridade (*última chance*) e ainda existem regiões com maior prioridade por pesquisar, o processo não pode terminar. A pesquisa prossegue e é obtida a solução 6. Desta forma, como as regiões de primeira e segunda prioridades estão totalmente pesquisadas (a recta de nível que passa pela solução 6 está para além de $Y2$) e a solução 6 pertence à região de terceira prioridade (é a única), então o processo termina com a escolha da solução 6 como solução final.

6º Caso. Suponha-se que as soluções 3, 4, 5 e 6 não existem.

Usando aquela função escalar são alcançadas sucessivamente as soluções 1 e 2. Ao prosseguir com a pesquisa, obtém-se a solução 7. Como as regiões de primeira, segunda e terceira prioridades estão totalmente pesquisadas (a recta de nível que passa pela solução 7 está para além de Z) e a solução 7 é de última chance, então o processo termina escolhendo a solução 1 (foi a primeira do seu tipo) como solução final.

10.7. Aplicação prática

Nesta secção são resolvidos dois problemas de encaminhamento : com duas e com três funções objectivo.

10.7.1. Caso bi-objectivo

Considere-se uma rede não orientada, gerada aleatoriamente, com 21 nós e 39 arcos, em que cada arco tem associados dois valores. Pretende-se determinar um caminho entre os nós 1 e 18 nesta rede. A Fig. 10 apresenta o resultado da resolução do problema especificado, para o qual se introduziram os seguintes dados :

- o objectivo 2 tem preferência em relação ao objectivo 1 ($F2 \succ F1$);
- requisitos de QoS para a função objectivo 1 : 500 (requerido) e 600 (aceitável);
- requisitos de QoS para a função objectivo 2 : 300 (requerido) e 380 (aceitável).

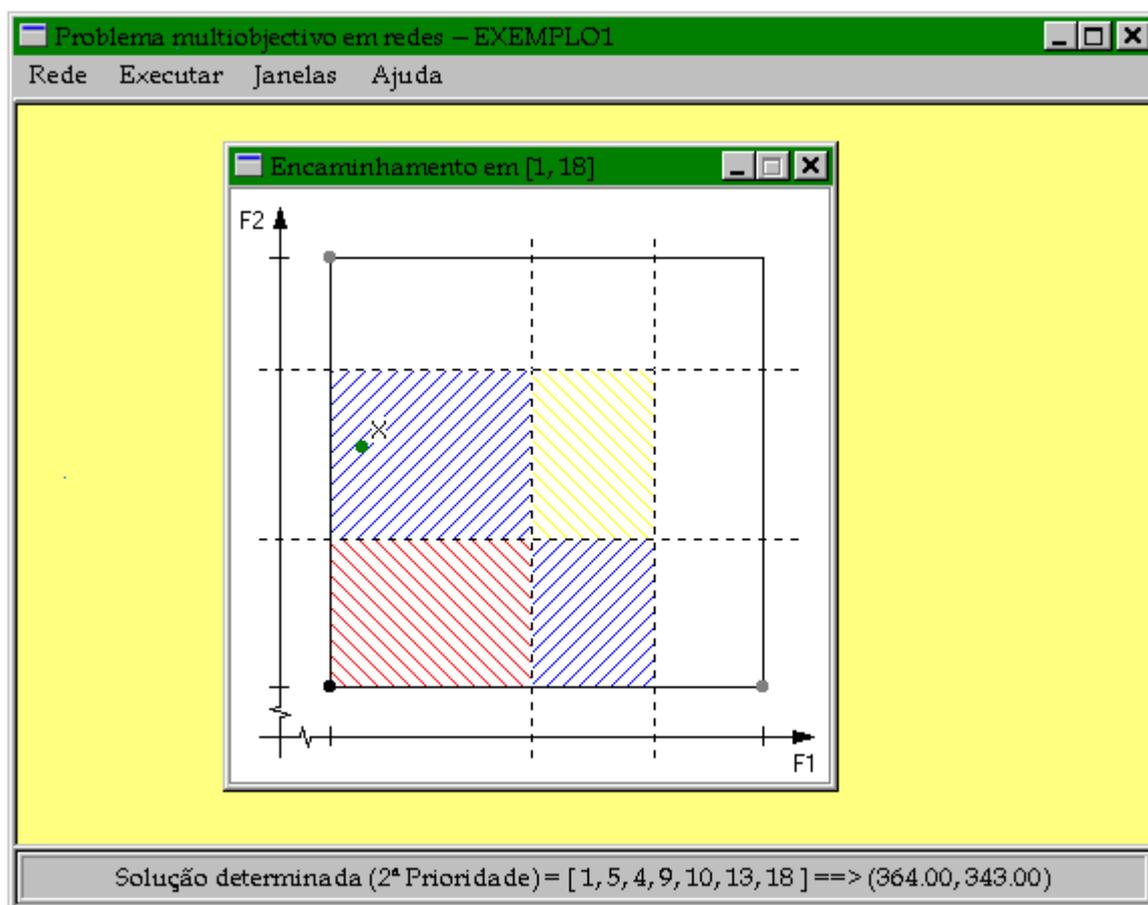


Fig. 10 – Problema de encaminhamento (bi-objectivo) : exemplo.

A solução determinada (ver Fig. 10) foi a seguinte :

$$p = [1, 5, 4, 9, 10, 13, 18] \text{ com um custo de } c(p) = (364, 343),$$

que pertence à região de segunda prioridade de menor preferência (B1).

10.7.2. Caso tri-objectivo

Considere-se uma rede orientada, gerada aleatoriamente, com 50 nós e 250 arcos, em que cada arco tem associados três valores. Pretende-se determinar um caminho entre os nós 15 e 45 nesta rede. A Fig. 11 apresenta o resultado da resolução do problema especificado, para o qual se introduziram os seguintes dados :

- o objectivo 1 tem preferência em relação ao 2 e este em relação ao 3 ($F1 \succ F2 \succ F3$);
- requisitos de QoS para a função objectivo 1 : 1500 (requerido) e 2000 (aceitável);
- requisitos de QoS para a função objectivo 2 : 600 (requerido) e 1000 (aceitável);
- requisitos de QoS para a função objectivo 3 : 1400 (requerido) e 2000 (aceitável).

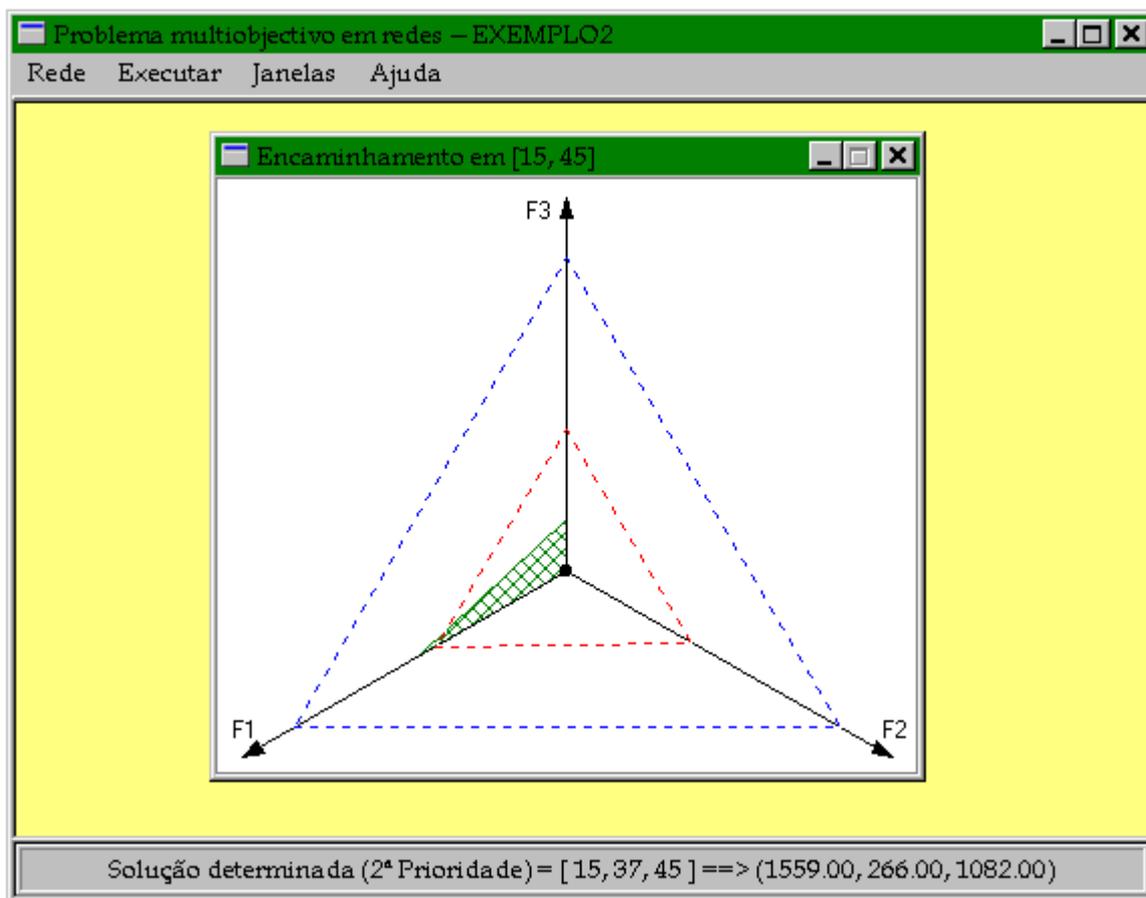


Fig. 11 – Problema de encaminhamento (tri-objectivo) : exemplo.

A solução determinada (ver Fig. 11) foi a seguinte :

$$p = [15, 37, 45] \text{ com um custo de } c(p) = (1559, 266, 1082),$$

que pertence à região de segunda prioridade (apenas não foi satisfeito o valor requerido da função objectivo 1, e portanto foram satisfeitos os outros dois) de menor preferência.