3
Difusão Seletiva em Redes IP sobre WDM

3.1. Introdução

Nos últimos anos, em função das diversas questões mal resolvidas do IP Multicast e da maior demanda por aplicações de grupo com grande consumo de largura de banda, um número crescente de pesquisadores têm se dedicado a discutir e desenvolver propostas de soluções para as questões relacionadas à difusão seletiva em redes de comutação de pacotes (Perlman et al., 1999; Levine et al., 2000; Diot, et al., 2000; Holbrook & Cain, 2003).

Com a evolução da tecnologia WDM e a tendência de sua adoção nos principais backbones IP, novos desafios estão surgindo no contexto da difusão seletiva. Alguns desses novos desafios são decorrentes da perspectiva de se passar a realizar a difusão seletiva diretamente na camada WDM, como, por exemplo, a construção de árvores de luz multiponto em redes onde nem todos os nós são aptos a realizar a divisão dos feixes de luz, como também não são capazes de efetuar a conversão de lambdas.

Outros desafios surgem da interação entre IP e WDM, pois algumas soluções adotadas no domínio IP não são adequadas no contexto óptico. Um bom exemplo são as árvores multiponto compartilhadas por todos os emissores de um grupo, construídas por alguns protocolos de roteamento do IP Multicast (e.g. PIM-SM). Tais árvores, da forma como são construídas no IP Multicast, não são adequadas ao domínio óptico, principalmente se a comutação baseada em lambdas for utilizada, pois lambdas não são aglutináveis.

Neste capítulo, são discutidas as principais questões para a adequação da difusão seletiva às futuras redes ópticas. Inicialmente é feita uma avaliação do IP Multicast. Em seguida, são discutidos pontos polêmicos da interação da difusão seletiva com a comutação baseada em rótulos. Por fim, são apresentados os principais desafios para a adequação da difusão seletiva, em especial do IP Multicast, às redes baseadas em comutação óptica, com uma discussão das questões relacionadas ao roteamento e à construção de árvores multiponto.

3.2. Avaliação do IP Multicast

Como já discutido em (Diot, et al., 2000; Barros & Stanton, 2001) e mencionado no Capítulo 1, após quase 15 anos desde sua proposição, o modelo do IP Multicast continua sujeito a muitos questionamentos. A sua lenta adoção pelos principais operadores e provedores de serviços vem estimulando o surgimento de propostas que alteram completamente o paradigma da difusão seletiva na Internet (Pendarakis et al., 2001; Francis, 2000). Contudo, as melhores soluções ainda vêm sendo aquelas que incrementam e melhoram o IP Multicast, sem alterar a filosofia de implementar a difusão seletiva na camada de rede.

O modelo de serviço do IP Multicast baseado em uma abordagem aberta é inerentemente complexo (Diot, et al., 2000). Nenhum tipo de controle de acesso ou mecanismo restringe as estações ou usuários na criação dos grupos, na recepção de dados de um grupo, ou mesmo no envio de dados para um determinado grupo. Para receber dados de um grupo multiponto, as estações devem solicitar adesão ao grupo, contactando os roteadores aos quais estão diretamente conectadas através do protocolo IGMP ("Internet Group Membership Protocol") (Deering, 1989). Uma vez que uma estação tenha efetuado a adesão ao grupo, ela passa a receber todos os dados enviados para aquele endereço de grupo. Para enviar dados para um grupo, as estações não precisam fazer parte do grupo. Emissores não podem reservar endereços, ou mesmo evitar que outros emissores escolham os mesmos endereços. O número de estações que efetuaram adesão a um grupo é dinâmico e desconhecido. O status das entidades (i.e., emissor, receptor ou ambos) também é desconhecido.

As árvores de distribuição multiponto são geradas e mantidas por um protocolo de roteamento multiponto. Muitos protocolos têm sido propostos ao longo desses quase quinze anos; entre os principais pode-se citar: DVMRP ("Distance Vector Multicast Routing Protocol") (Waitzman et al., 1988) MOSPF ("Multicast OSPF") (Moy, 1994), PIM ("Protocol Independent Multicast") modo denso (PIM-DM) (Deering et al., 1996) e modo esparso (PIM-SM) (Estrin et al., 1998), CBT ("Core Base Tree") (Ballardie, 1997) e BGMP ("Border Gateway Multicast Protocol") (Thaler et al., 2000b). A diferença entre esses protocolos reside essencialmente no tipo de árvore multiponto que eles constróem. Os protocolos DVMRP, MOSPF e PIM-DM, por exemplo, constróem árvores

multiponto originadas no emissor. Já os protocolos PIM-SM[1] e CBT constróem árvores multiponto compartilhadas por todos os emissores baseadas em um ponto central conhecido, chamado no PIM-SM de "rendezvous-point" (RP). O PIM-SM utiliza uma árvore compartilhada unidirecional onde os pacotes são enviados primeiro para o RP, o qual, por sua vez, envia os pacotes via difusão seletiva através da árvore multiponto para todos os membros do grupo. Os protocolos CBT e BGMP utilizam árvores compartilhadas bidirecionais, nas quais os pacotes podem ser enviados via difusão seletiva a partir de qualquer ponto da árvore.

Baseada nos padrões recentes da IETF ("Internet Engineering Task Force"), pode-se dizer que a arquitetura IP Multicast "de fato" engloba os protocolos IGMP (versão 2) para o gerenciamento de grupo, o PIM-SM para roteamento no interior dos domínios e o MSDP ("Multicast Source Discovery Protocol") (Meyer & Fenner, 2001), juntamente com o MBGP ("Multicast Border Gateway Protocol") (Bates et al., 1998) para o roteamento entre domínios. O MSDP tem a função de anunciar o grupo e os emissores para todos os outros RPs localizados em outros domínios. Tais informações são divulgadas por rotas fornecidas pelo MBGP (também chamado de BGP4+). A partir dessas informações, os RPs remotos que possuírem estações pertencentes ao grupo no seu domínio formam um túnel até o roteador do emissor. Como cada RP em cada domínio precisa saber sobre todos os emissores do grupo, essa solução não é escalável, tendo sido proposta apenas como alternativa de curto prazo.

O modelo proposto por Deering, além da acentuada sobrecarga na construção de árvores multiponto, requer que os roteadores mantenham informações de estado por grupo. Tal procedimento viola o princípio original da Internet de não manter informações de estado nos elementos centrais de rede, introduzindo maior complexidade e restrições de escalabilidade. Outras duas grandes críticas ao IP Multicast são em relação à falta de um gerenciamento de grupo efetivo, com maior controle sobre as adesões e transmissões a um grupo, e a ausência de controle na alocação de endereços de grupos. Por fim, o IP Multicast oferece um serviço de melhor esforço ("best effort") e deixa para as camadas superiores funções como, confiabilidade, controle de congestionamento, controle de fluxo e segurança.

[1]- O PIM-SM pode chavear entre árvores compartilhadas e árvores originadas no emissor.

Diversas propostas para solucionar alguns desses problemas vêm sendo apresentadas ao longo desses anos. As mais interessantes são analisadas e discutidas em "surveys" apresentados por (Diot et al., 2000; Barros & Stanton, 2001; Pragyansmita & Raghavan, 2002). A Tabela 3.1 apresenta um resumo das principais propostas apresentadas ao IP Multicast. É interessante mencionar que, apesar da tabela estar dividida por tópicos, algumas propostas apresentaram contribuições em diferentes tópicos. Contudo, apenas as contribuições mais relevantes estão sendo destacadas na Tabela 3.1.

Como alternativas no gerenciamento da comunicação entre estações, merecem destaque o AIM ("Addressable Internet Multicast ") (Levine & Garcia, 1997) e o PGM ("Pretty Good Multicast") (Speakman et al., 1998). Na questão da confiabilidade, o PGM também se destaca juntamente com o SRM ("Scalable Reliable Multicast") (Floyd et al., 1995), o RLM ("Receiver-driven Layer Multicast") (McCanne et al., 1996), entre outros (Abelém, 2000). No que diz respeito à alocação de endereços, além do MAAA ("Multicast Address Allocation Architecture") (Thaler et al. 2000a), merecem destaque, o GLOP (Meyer & Lothberg, 2001), a alocação por fonte (ou canal) proposta no EXPRESS (Holbrook et al., 1999) e de maneira similar no SM ("Simple Multicast") (Perlman et al., 1998) e o endereçamento do IPv6 (Deering et al., 1998). Aperfeiçoamentos no esquema de gerenciamento de grupo também estão sendo proposto no IGMP versão 3 (Cain, 2001), permitindo que um membro pertencente ao grupo especifique explicitamente quais os emissores de quem ele deseja receber dados, ou de quem ele não quer receber mensagens.

Quanto à redução de complexidade e melhorias na escalabilidade do modelo, que são questões mais diretamente ligadas à proposta MIRROR, as alternativas mais interessantes são: o EXPRESS[2] e seu sucessor SSM ("Single Source Multicast") (Holbrook & Cain, 2003), o REUNITE ("REcursive UNIcast TrEes") (Stoica et al., 2000) e seu aprimoramento HBH ("Hop By Hop multicast") (Costa et al., 2001a), assim como o XCAST ("eXplicit multiCAST") (Boivie et al., 2002). Estas propostas serão melhor discutidas no Capítulo 5, onde são analisados os trabalhos relacionados.

[2] - Perlman apresentou propostas semelhantes, denominadas SM ("Simple Multicast") (Perlman et al., 1998) e RAMA ("Root Administered Multicast Addressing") (Perlman & Raman, 1999).

Propostas	Características
Alocação de endereços Multiponto	
MAAA	Esquema hierárquico de gerenciamento do espaço de endereços multiponto ao custo de um projeto complexo.
GLOP	Usa os números de AS ("Autonoms Systems") como base para restringir os endereços disponíveis para cada domínio.
EXPRESS, SSM	Propõe que os grupos sejam composto por apenas um emissor (modelo de canal) e a identificação destes seja através de uma dupla (S, G), onde S é o endereço IP do emissor e o G é um endereço IP classe D.
IPv6	Aumenta drasticamente o espaço de endereços IP, ao custo de alterar completamente a estrutura do pacote IP.
Gerenciamento de comunicação, confiabilidade e controle de congestionamento	
AIM	Habilita os emissores a restringir a entrega de pacotes para um subgrupo de receptores, como também permite que os receptores selecionem subgrupo de emissores. Além disso, fornece roteamento para descoberta de recursos.
PGM	Baseado na abordagem "iniciada pelo receptor". Envia NACKs através de ligações ponto a ponto. Os roteadores PGM intermediários repassam estes para o próximo elemento e confirmam o seu recebimento via difusão seletiva.
SRM	Baseado na estratégia "iniciada pelo receptor". Garante entrega confiável de pacotes sem controle de ordenação sobre eles. Um pacote perdido pode ser retransmitido pelo receptor mais próximo apto para tal.
RLM	Também baseado na estratégia "iniciada pelo receptor". Supõe que os fluxos de dados podem ser divididos em camadas de diferentes qualidades.
Redução da complexidade do modelo e melhoria na escalabilidade	
EXPRESS, SSM	Restringe o modelo de distribuição para apenas 1 emissor por grupo.
REUNITE, HBH	Não usa endereços IP classe D. Mantém as informações de estado dos grupos apenas nos roteadores de ramificação das árvores multiponto, os quais são os responsáveis por criar cópias dos pacotes de dados e alterar os endereços de destino dos mesmos.
XCAST	Implementa a distribuição multiponto baseada exclusivamente na infra-estrutura ponto a ponto. os endereços IP dos destinatários são encapsulados nos cabeçalhos de cada pacote.
Difusão seletiva implementada na camada de aplicação	
ALMI	Oferece serviço de difusão seletiva através de uma árvore multiponto virtual, baseada em conexões um para um entre as estações, formando uma árvore geradora mínima.
YOID	Implementação mista. Usa IP Multicast, sempre quando este está disponível. Transforma o IP Multicast em pequenas e disjuntas ilhas e fornece uma arquitetura rudimentar para a difusão seletiva global.
RMX/Scattercast	Divide uma sessão multiponto confiável em grupos menores com receptores homogêneos. Organiza esses grupos em uma árvore geradora, usando conexões TCP.
End System Multicast/ Narada	Todas as funcionalidades da difusão seletiva, incluindo gerenciamento de grupo e replicação dos pacotes, ficam a cargo das estações finais. Além disso, propõe a organização das estações através de um protocolo distribuído que procura otimizar os custos.

Tabela 3.1 – Sumário das principais propostas para alterar o modelo tradicional do IP Multicast.

Apesar de sugerirem algumas mudanças, as propostas anteriores, de um modo geral, não implicam na reformulação do modelo do IP Multicast de uma forma mais ampla. Recentemente, alguns pesquisadores resolveram questionar os argumentos usados por Deering em 1989, quando este alegou que a inserção do IP Multicast na camada IP era a melhor solução, pois traria benefícios significativos em termos de desempenho, os quais suplantariam e justificariam qualquer custo adicional de complexidade nesta camada. Esses pesquisadores propuseram que a difusão seletiva fosse implementada na camada de aplicação, ou seja, nas estações dos usuários finais, e argumentaram que, apesar de não conseguir resultados tão bons quanto o IP Multicast em termos de desempenho, o novo modelo ofereceria menos entraves tecnológicos e, em função disso, teria maior apelo comercial.

De acordo com essa nova visão, merecem destaque as seguintes propostas: a ALMI ("Application Level Multicast Infrastructure") (Pendarakis, et al. 2001), a Yoid (Francis, et al.,2000), a "End System Multicast"/Narada (Chu et al., 2000) e a Scattercast (Chawathe et al., 2000). Todas elas apresentam arquiteturas alternativas ao modelo tradicional do IP Multicast e, de um modo geral, são baseadas na premissa de oferecer um serviço de difusão seletiva através de uma árvore multiponto virtual, a qual consiste de conexões ponto a ponto entre as estações, formando uma árvore geradora mínima.

Apesar de serem consideradas por seus autores como solução para todos os problemas do IP Multicast, a grande maioria dessas novas arquiteturas tem se mostrado mais adequada para aplicações de grupos relativamente pequenos (algumas dezenas de participantes), esparsos e com uma semântica muitos para muitos, em virtude, entre outras coisas, da alta sobrecarga para se formar a malha virtual de interconexão entre os membros.

3.3. IP Multicast no contexto da Comutação Baseada em Rótulos

O MPLS, apesar de estar sendo estendido para trabalhar também com comutação por divisão de tempo, por comprimento de onda e por divisão de espaço, ainda possui uma série de questões em aberto no que diz respeito à difusão seletiva. Na verdade, a grande maioria das definições e das padronizações sobre MPLS foi feita apenas para a comunicação ponto a ponto ("unicast") (Rosen et al, 2001), deixando o contexto da difusão seletiva para trabalhos futuros. Algumas dessas questões começaram a ser discutidas apenas recentemente por

Acharya (1999) e por Omms et al.(2002). Esta seção pretende apresentar e discutir alguns dos principais aspectos necessários para a adequação do MPLS para comunicação via difusão seletiva.

3.3.1. Formas de Disparar o Estabelecimento de LSPs

Na difusão seletiva o estabelecimento de um LSP pode ser disparado de três formas (Omms et al., 2002): orientado por requisição, orientado por topologia ou orientado pelo tráfego. No primeiro caso, a formação de um LSP é disparada a partir da interceptação de mensagens de controle, as quais, no IP Multicast, podem ser de dois tipos: mensagens de roteamento multiponto e mensagens de reserva de recursos (e.g. RSVP). No caso de mensagens de roteamento multiponto, apenas os protocolos que usam sinalização explícita (e.g. PIM-SM ou CBT) podem ser usados. Os protocolos de modo denso, como o PIM-DM ou o DVMRP, precisam ser adaptados, como proposto por Farinacci et al. (2000). A desvantagem desse esquema é que os cálculos realizados pelos protocolos de roteamento no nível 3, para determinar a tabela de roteamento multiponto, são repetidos pelo módulo MPLS[3].

No estabelecimento orientado pela topologia, a criação de um LSP para fluxos multiponto é feita através do mapeamento da árvore de roteamento do nível 3, a qual está disponível nas tabelas de roteamento multiponto dos roteadores, para uma árvore no nível MPLS. O mapeamento neste caso é feito mesmo se não houver tráfego. A desvantagem desse método é que rótulos são consumidos mesmo quando nenhum tráfego existe.

No estabelecimento orientado pelo tráfego, a formação dos LSPs é disparada com a chegada de dados. Este esquema consome menos rótulos que o orientado por topologia. Entretanto, problemas podem ocorrer caso a rede não suporte nenhuma combinação de encaminhamento entre os níveis MPLS e de rede, ou algum esquema alternativo de distribuição de rótulo, pois para estabelecer um novo LSP em um ramo da árvore multiponto é preciso que dados cheguem por aquele ramo. Porém, os dados só chegarão se existir um LSP naquele ramo. Logo,

[3] - Apesar do MPLS ser um "multiprotocolo" tanto nível 3 como nível 2, consideraremos neste trabalho apenas o IP como um protocolo nível 3, situando o MPLS no topo das tecnologias nível 2. Postura semelhante foi adotada em (Omms et al., 2002) e (Papadimitriou et al., 2001).

cria-se um impasse se as alternativas de encaminhamento citadas acima não existirem. A Seção 3.3.3 apresenta um bom exemplo para esta questão.

Baseados nas alternativas acima, alguns autores vêm propondo variações e esquemas alternativos para adequar o estabelecimento de LSPs ao IP Multicast. Uma dessas alternativas bastante debatida foi proposta por Farinacci et al. (2000) e consistia em agrupar as mensagens de atribuição de rótulos com os pedidos de adesão aos grupos. Contudo, foi constatado que este esquema possuía uma série de desvantagens e de limitações, como a necessidade de adaptações nos protocolos de roteamento existentes e a limitação na forma de atribuição de rótulos, que seria necessariamente a partir do nó de saída do enlace ("downstream"), na variação não solicitada.

3.3.2. Controle Independente versus Controle Ordenado

Um aspecto diretamente relacionado com as formas de disparos para o estabelecimento de LSPs é a questão da distribuição de rótulos. No contexto da comunicação ponto a ponto, duas alternativas são oferecidas: o controle independente e controle ordenado. Entretanto é interessante mencionar novamente que tais alternativas são definidas apenas no contexto da comunicação ponto a ponto ("unicast"). Para o caso da difusão seletiva existem apenas propostas em discussão sobre a questão (Omms, et al., 2002).

Em uma rede baseada em MPLS, os pacotes que chegam são classificados de acordo com determinadas características (e.g. endereço IP de destino, ou em algum critério qualitativo) em classes de equivalência ("Forwarding Equivalence Class - FEC"). Os pacotes pertencentes à mesma FEC recebem o mesmo rótulo e são encaminhados da mesma maneira. A associação das classes de equivalência com rótulos para construir caminhos comutados por rótulos (LSPs) pode ser feita de duas formas, através de controle independente ou de controle ordenado. Na primeira, cada roteador comutado por rótulo ("Label Switching Router – LSR") toma uma decisão independente de associar um rótulo com àquela classe de equivalência e distribuir essa associação para seus pares. Neste caso, um LSP é definido como no roteamento salto a salto, de forma distribuída. Cada LSR toma uma decisão independente de como tratar cada pacote e baseia-se no algoritmo de roteamento existente para convergir rapidamente e garantir a entrega correta dos pacotes. Na segunda, um LSR somente faz a associação de uma classe de

equivalência a um rótulo, se ele for o nó egresso (de saída) da rede ou se ele tiver recebido a associação para aquela FEC do nó egresso (Rosen, et al., 2001). Como destacado na Seção 2.3, o GMPLS estende esta funcionalidade, permitindo que os LSRs de entrada dos enlaces ("LSRs upstream") enviem uma sugestão de rótulo para seus pares na saída dos enlaces ("LSRs downstream")[4]. De qualquer forma, caso deseje-se garantir que um tráfego de uma FEC específica siga por um caminho com determinadas propriedades, o controle ordenado deve ser utilizado.

A princípio, as três formas de disparar o estabelecimento de LSPs descritas na seção anterior combinam com controle independente (Omms, et al., 2002). No entanto, adaptações podem ser feitas para adequá-las ao controle ordenado. O estabelecimento de LSPs orientado pelo tráfego talvez seja o que menos se adapta ao controle ordenado, principalmente se os nós envolvidos não suportarem nenhuma combinação de encaminhamento nos níveis MPLS e 3. A Figura 3.1 ilustra melhor esta questão. Suponha que se deseje estabelecer um novo LSP para o ramo de R2 e que o estabelecimento orientado pelo tráfego e o controle ordenado estejam sendo utilizados. Quando o LSR X receber o tráfego, baseado na abordagem orientada pelo tráfego, ele irá solicitar um rótulo para Y. Este por sua vez, em função do controle ordenado, só poderá estabelecer um mapeamento de rótulo para o enlace X-Y se já tiver sido atribuído o rótulo para o enlace Y-Z. Todavia, o rótulo para o enlace Y-Z só poderá ser atribuído quando Y receber o tráfego no enlace X-Y. Cria-se então um impasse caso os nós envolvidos não suportem nenhuma combinação de encaminhamento nos níveis MPLS e de rede, como mencionado anteriormente.

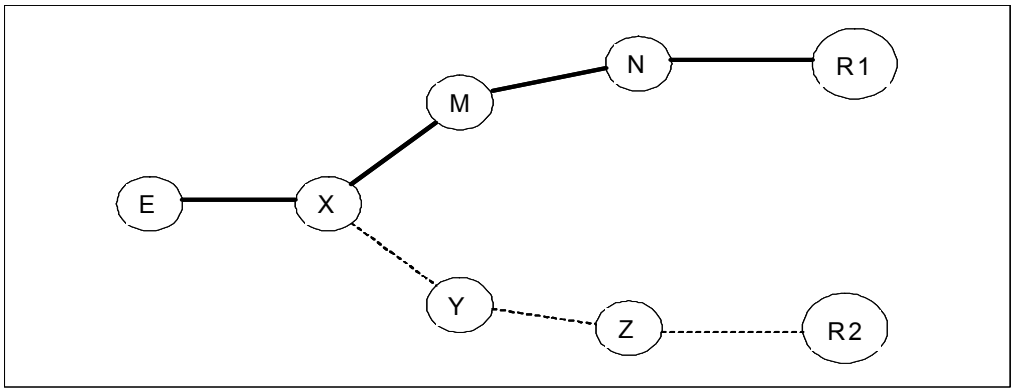

Figura 3.1 –Estabelecimento de LSPs orientado pelo tráfego com controle ordenado.

[4] - Os termos LSR de entrada e LSR de saída dos enlaces são relativos e empregados considerando o sentido do fluxo de datagramas no LSP.

No que diz respeito aos outros dois métodos, tanto o orientado por requisição como o orientado por topologia se mostram facilmente adaptáveis ao contexto de controle ordenado. No primeiro caso, até mesmo em situações onde o controle independente esteja sendo adotado, a distribuição de rótulos poderá ser feita como no controle ordenado, ou seja, com as mensagens de controle fluindo a partir dos nós egressos, de saída dos enlaces, principalmente se mensagens RSVP forem usadas para disparar o estabelecimento de LSPs.

Quanto à abordagem orientada por topologia, pode-se dizer que a maior diferença ocorrerá quando mecanismos de engenharia de tráfego estiverem sendo usados. Neste caso, as informações de rotas que são mapeadas pelos módulos MPLS para o nível 2 serão obtidas, provavelmente, através de algum critério de engenharia de tráfego, ao invés de serem oriundas das tabelas de roteamento multiponto mantidas pelos protocolos de roteamento multiponto convencionais.

3.3.3. Atribuição de Rótulos

Para a comunicação ponto a ponto, a arquitetura MPLS define que a decisão de atribuir um rótulo específico para uma determinada classe de equivalência seja feita apenas pelo LSR de saída do enlace (“downstream”) para o qual a associação está ocorrendo. Desta forma, o LSR de saída do enlace, após atribuir o rótulo para a classe de equivalência em questão (“dowstream assigned”), informa o seu par na entrada do enlace (“upstream”) sobre a associação. No que diz respeito à requisição das atribuições, a arquitetura MPLS permite duas variações, a sob demanda (“on-demand”) e a não solicitada (“unsolicited”). Na primeira, o LSR de entrada do enlace requisita explicitamente ao seu par na saída do enlace um rótulo para uma determinada classe de equivalência. Enquanto, na segunda, o LSR de saída do enlace realiza a atribuição sem solicitação prévia e informa seu par em seguida (Rosen, et al., 2001).

Na difusão seletiva, a princípio, a atribuição de rótulos pode ocorrer, tanto a partir do LSR de saída do enlace, nas variações sob demanda ou não solicitada, como a partir do LSR de entrada do enlace, nas formas sob demanda, não solicitada ou implícita. As duas formas de atribuição de rótulos têm diferentes vantagens e deficiências (Omms, et al., 2002). Por exemplo, a atribuição a partir do LSR de saída do enlace seria a mesma utilizada na comunicação ponto a ponto, eliminando assim a necessidade de se desenvolver novos procedimentos de

distribuição de rótulos. A atribuição de rótulos a partir do LSR de entrada do enlace, por sua vez, oferece um estabelecimento mais rápido dos LSPs quando estes são disparados pelo tráfego.

3.3.4. Outras Questões

Além das questões discutidas nas subseções anteriores, outras de menor relevância, ou menos controversas, também vêm sendo discutidas na área (Omms, et al., 2002). Entre elas pode-se destacar o tratamento dado ao campo TTL ("time-to-live"), o roteamento explícito e o modo de retenção dos rótulos.

Como é de conhecimento geral na área, o campo TTL do cabeçalho IP, no contexto da comunicação ponto a ponto, é tipicamente usado para descartar pacotes em caminhos cíclicos ("loops"). No contexto da difusão seletiva, ele costuma ser usado também para limitar o escopo da comunicação: local, regional ou de longa distância (Deering, 1991). Quando se utiliza comutação baseada em rótulos, junto com a difusão seletiva, alguns LSRs internos (e.g. LSRs ATM, ou mesmos os LSRs ópticos) não suportam a função de decremento do campo TTL. A questão é importante, pois tal fato pode afetar as funções de restrições de escopo adotadas na difusão seletiva. Contudo, propostas para solucionar tal questão já existem. As primeiras propunham o decremento do campo TTL apenas nos LSRs de saída das redes MPLS, o que provocaria desperdício de banda caso os pacotes tivessem que ser descartados. As propostas mais recentes, dentro dos estudos de generalização da arquitetura MPLS, estão sugerindo uma operação de decremento prévio nos LSRs de entrada da rede (Mannie, 2003).

No contexto da difusão seletiva, o roteamento explícito pode ser interpretado de duas maneiras. Na primeira, o roteamento explícito multiponto substitui a árvore gerada pelos protocolos de roteamento multiponto por uma árvore de LSPs, gerada, por exemplo, por ferramentas autônomas ("off-line"). Na segunda maneira, o roteamento explícito multiponto baseia-se nos protocolos de roteamento multiponto, os quais geram a árvore a partir de rotas ponto a ponto explícitas, ao invés das rotas de menor custo geradas pelo protocolos IGP.

Quanto ao modo de retenção dos rótulos, a comunicação ponto a ponto trabalha com duas formas, a conservativa e a liberal. Na primeira, são alocados apenas os rótulos anunciados pelos pares (próximo salto) de um determinado LSR e que serão usados por ele para encaminhar os dados. Na segunda, são alocados e

mantidos os rótulos anunciados não só pelos pares, mas também por "possíveis" pares. Na difusão seletiva, o modo liberal não faz muito sentido por duas razões. Primeiro, nem todos os LSRs vizinhos sabem encaminhar os pacotes de uma determinada FEC, enquanto na comunicação ponto a ponto isso não ocorre. Segundo, na difusão seletiva um LSR sempre sabe para qual vizinho mandar a requisição de rótulos ou a mensagem de mapeamento de rótulos, o que nem sempre é verdade para a comunicação ponto a ponto, onde um LSR é obrigado a divulgar os rótulos para todos os seus vizinhos.

3.4. IP Multicast e a Difusão Seletiva em Redes IP sobre WDM

Como ressaltado no início deste capítulo, o conceito de difusão seletiva para redes de comutação de pacotes tem sido largamente estudado nos últimos anos, devido, entre outros motivos, à demanda crescente por aplicações de grupo com grande consumo de largura de banda. Atualmente, as redes WDM oferecem um excelente campo de desenvolvimento para tais aplicações. Estendendo o conceito de difusão seletiva para redes ópticas, podem ser viabilizadas aplicações como Internet TV e vídeo (quase) sob demanda, entre outras.

Para oferecer serviços de difusão seletiva em redes IP sobre WDM existem três alternativas possíveis para sua implementação: via difusão seletiva somente na camada IP, via múltiplos caminhos de luz na camada WDM, ou através de difusão seletiva na camada WDM (ver Figura 3.2). Dessas três formas, a difusão seletiva diretamente na camada WDM é a mais vantajosa, pois permite, entre outras virtudes, a construção de árvores de distribuição multiponto mais eficientes e um maior grau de transparência, em termos de taxa de bit e de formato de codificação (Qiao et al.,1999b).

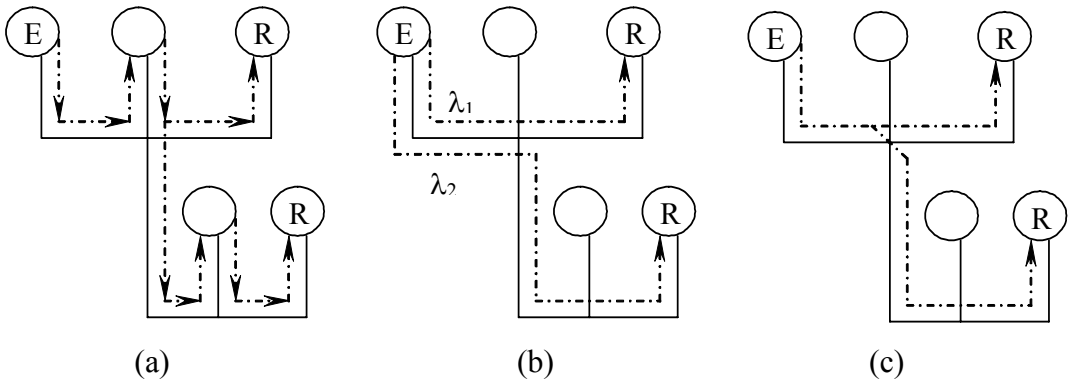

Figura 3.2 – Alternativas de implementação de serviços de difusão seletiva em redes IP sobre WDM: (a) difusão seletiva na camada IP; (b) via múltiplos caminhos de luz na camada WDM; (c) difusão seletiva direto na camada WDM.

No entanto, a difusão seletiva na camada WDM possui uma série de desafios. Um dos principais vem do fato que os comutadores ópticos com capacidade de divisão de feixes de luz são muito caros e a operação de divisão também causa a divisão da potência do sinal, limitando o alcance deste. Uma rede WDM onde apenas parte dos comutadores ópticos possuem capacidade de difusão seletiva é dita de divisão esparsa ("sparse splitting"), enquanto uma rede em que todos os nós são aptos a realizar a divisão, só que através de um número limitado de saídas, é dita de divisão limitada ("limited splitting") (Murthy & Gurusamy, 2002).

Em uma rede com capacidade de divisão limitada a construção da árvore multiponto é similar àquela realizada em uma rede baseada em comutação de pacotes. Apenas que, na rede óptica, os limitadores serão o número de transceptores ("transceivers") e de lambdas disponíveis. Em outras palavras, ao gerar a árvore multiponto os algoritmos de construção devem levar em consideração a disponibilidades desses recursos.

Em uma rede com capacidade de divisão esparsa nem sempre pode ser possível incluir todos os receptores de uma sessão multiponto em apenas uma única árvore multiponto. Dependendo da topologia da rede, pode ser necessário um conjunto de árvores multiponto para atender a todos os receptores. Isto significa que o emissor pode ter que transmitir os dados via difusão seletiva por mais de um canal, talvez até mesmo em diferentes fibras, ou com diferentes comprimentos de onda. Ao conjunto de árvores construídas para uma sessão multiponto, convencionou-se chamar de floresta multiponto (Qiao et al., 1999; Murthy & Gurusamy, 2002).

Além dos problema típicos de redes WDM, existem também aqueles surgidos da interação entre IP e WDM. Nas inter-redes IP, por exemplo, os protocolos de roteamento do IP Multicast podem criar tanto árvores de distribuição a partir de cada emissor de um grupo como árvores compartilhadas por todos os emissores de um grupo. O primeiro tipo de árvore tem a virtude de ser mais eficiente, enquanto o segundo é mais escalável. No contexto óptico, contudo, alguns novos fatores precisam ser levados em consideração. Esta seção irá analisar tais questões, enfatizando a questão do problema de roteamento multiponto e das árvores de distribuição multiponto.

3.4.1. Roteamento Multiponto em Redes WDM

Projetar redes WDM tradicionais envolve, entre outras tarefas, tratar de dois problemas relevantes no contexto deste trabalho, o primeiro denominado projeto da topologia dos caminhos ópticos ("LTD – Lightpath Topology Design") e o segundo conhecido como roteamento e atribuição de comprimento de onda ("RWA – Routing and Wavelength Assignment") (Ramaswani & Sivarajan, 2002). No primeiro, o objetivo é definir a topologia de caminhos ópticos que será oferecida pela camada óptica para a camada superior, ou seja, definir os caminhos ópticos que irão interconectar os dispositivos da camada superior (e.g. roteadores IP ou comutadores SONET/SDH). Esta topologia de caminhos ópticos é também chamada na literatura de topologia lógica ou virtual. O problema de LTD está diretamente ligado ao problema de roteamento nas camadas superiores e vem sendo largamente investigado pela comunidade de redes na última década, com diversas soluções propostas (Kershenbaum, 1993; Cahn, 1998).

O problema de RWA, por sua vez, envolve equacionar a topologia de caminhos ópticos dentro da camada óptica e costuma ser dividido em dois sub-problemas (Ramaswani & Sivarajan, 2002). O primeiro é o problema de roteamento propriamente dito, onde deve-se determinar o caminho pelo qual pode-se estabelecer uma conexão. O segundo está relacionado à questão da atribuição de um comprimento de onda em cada enlace ao longo do caminho selecionado e consiste em determinar as rotas requisitadas usando o menor número possível de trocas de comprimento de onda.

No contexto da difusão seletiva, o problema de RWA está diretamente relacionado com à capacidade dos nós da rede em dividir os feixes de luz e em converter lambdas. Nas redes com capacidade de divisão limitada, como mencionado anteriormente, a construção da árvore multiponto se limita a questões similares às existentes nas redes de pacotes, com o agravante de ter como limitadores o número de transceptores e de lambdas disponíveis. Neste contexto, Sahasrabuddhe & Mukherjee (1999) e Sahin & Azizoglu (2000), entre outros, apresentaram propostas para construção de árvores multiponto minimizando o número de tranceptores e de conversões de lambdas utilizados.

Já para as redes WDM com capacidade de divisão esparsa, onde apenas alguns nós possuem capacidade de divisão dos feixes de luz, as tarefas dos

algoritmos geradores da árvore multiponto são um pouco mais complexas e envolvem minimizar um ou mais dos seguintes parâmetros: número máximo de lambdas utilizadas na árvore ou floresta, o número de lambdas usadas por enlace, o número de saltos entre a fonte e os receptores e o tempo de ajuste necessário para construir a árvore ou floresta. O esquema comumente adotado para tratar destes casos de redes com capacidade de divisão esparsa é o baseado na abordagem originada no emissor. Neste, a árvore multiponto é construída com o emissor na raiz da árvore. Dependendo do objetivo, emprega-se dois diferentes métodos para construir a árvore; um deles é através da geração de árvores originadas no emissor, que procura minimizar os custos dos caminhos individuais do emissor para cada receptor, o outro é baseado na construção de árvores geradoras mínimas, o qual objetiva minimizar o custo total da árvore. Neste último caso, sabe-se que, o problema da árvore geradora mínima é NP-completo (Winter, 1987), de modo que heurísticas vêm sendo empregadas para aproximá-la (Malli, et al., 1998, He et al., 2001).

Recentemente, Sreenath et al. (2001) propuseram um esquema alternativo baseado em uma abordagem onde a raiz da árvore é uma fonte virtual ("VS – Virtual Source rooted approach"). Essa fonte virtual (VS) nada mais é que um nó com capacidade de divisão de feixes de luz e de conversão de lambdas. A proposta consiste em identificar na rede todos os nós com essa capacidade e interligá-los, de modo que exista um caminho de luz para cada par de nós VS. A rede então é particionada em regiões em função da vizinhança destes nós especiais e a árvore é construída baseadas nestas conexões entre nós VS e tendo o nó VS mais próximo do emissor como raiz da árvore. Apesar de apresentar melhorias em alguns aspectos em relação à abordagem tradicional originada no emissor, a proposta baseada em fontes virtuais tem contra si o aumento considerável dos custos ("overhead") com o estabelecimento de caminhos entre os nós VS.

É interessante mencionar que o problema de RWA pode ser atenuado com a utilização de esquemas alternativos à comutação de lambdas, como a comutação de rajadas ópticas (OBS), destacada no Capítulo 2. Isto porque na comutação de lambdas, os caminhos roteados por comprimento de ondas devem ser estabelecidos antes que os dados possam ser transmitidos e cada lambda deve ficar dedicado a um ramo da árvore de distribuição até que os caminhos sejam liberados. No paradigma OBS, não há necessidade de lambdas ficarem dedicados

a cada ramo de uma árvore multiponto. Os lambdas são alocados sob demanda, na medida que pacotes de controle de rajadas vão sendo processados, e são liberados tão logo a rajada de dados passe através do enlace, ou automaticamente ou através de indicação explícita. Isto implica que rajadas de diferentes fontes para diversos destinatários podem efetivamente utilizar a largura de banda de um mesmo lambda, através do compartilhamento de tempo. Essas características tornam a tecnologia OBS mais eficiente, em termos de utilização de largura de banda, bem como menos suscetível ao problema de atribuição de comprimentos de onda (RWA). Algumas propostas nesse sentido foram recentemente investigadas por Zhang et al. (1999) e por Qiao et al. (1999), com resultados bastante satisfatórios.

3.4.2. IP Multicast em Redes WDM

No IP Multicast, como mencionado no início desta seção, os protocolos de roteamento podem construir, tanto árvores de distribuição originadas a partir de cada emissor de um grupo, como árvores compartilhadas por todos os emissores de um grupo. As árvores originadas a partir do emissor têm a virtude de serem mais eficientes, enquanto as árvores compartilhadas são mais escaláveis.

No contexto óptico, entretanto, as árvores compartilhadas são pouco recomendadas da forma como são construídas no IP Multicast, pois caso se utilize comutação baseada em rótulos, será necessário construir árvores de luz multiponto a multiponto, em função da possibilidade de diversos emissores no grupo, com rótulos que não são "aglutináveis", como é o caso de lambdas (Papadimitriou et al., 2001). A Figura 3.3 ilustra tal situação, com dois emissores, E1 e E2[5], enviando tráfego para o grupo através de lambdas diferentes. Supõe-se, neste caso, que um nó interno seja capaz de realizar as funções necessárias a um RP, como por exemplo, ter capacidade de conversão de comprimentos de onda (lambdas), entre outras.

[5] - No contexto deste trabalho, os termos membros do grupo, emissor e receptores, referem-se , a priori, a roteadores.

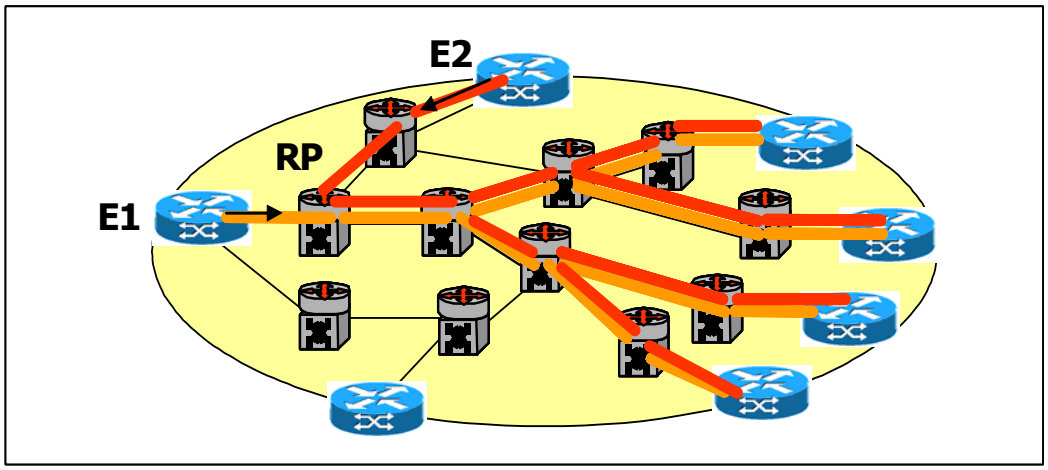

Figura 3.3 – Ilustração do problema de utilização de árvores compartilhadas no contexto da comutação de lambdas.

Outro fator que desabona a utilização de árvores compartilhadas no contexto da comutação de rótulos, da forma como são construídas no IP Multicast, é que o protocolo de roteamento utilizado no IP Multicast (PIM-SM) permite que alguns receptores optem por receber o tráfego multiponto de um determinado emissor a partir de árvores originadas na fonte, enquanto os outros continuam recebendo o tráfego pela árvore compartilhada. Suponha um situação como a ilustrada na Figura 3.4, onde os receptores estão recebendo tráfego de dois emissores através de uma árvore compartilhada comutada por rótulos (no caso lambdas). Suponha também que RP seja capaz de realizar a conversão de comprimento de onda. Se um receptor optar por receber o tráfego do emissor 'E2' através de uma árvore originada no emissor, enquanto continua recebendo o tráfego do emissor 'E1' através da árvore compartilhada. Então cria-se um impasse pois, não há como a árvore compartilhada comutada por rótulos podar aquele ramo da árvore, já que o receptor ainda está recebendo o tráfego enviado por 'E1' por ela.

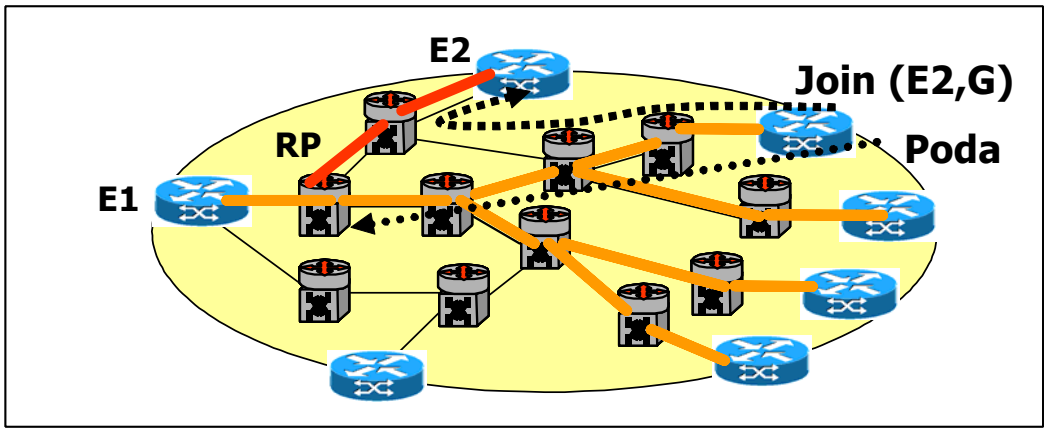

Figura 3.4 – Ilustração do problema de utilização de árvores compartilhadas da forma com são construídas no protocolo PIM-SM.

Além disso, no contexto óptico, pode ser interessante considerar abordagens de construção de árvores multiponto diferentes daquelas adotadas no IP Multicast tradicional (i.e. PIM-SM), as quais são baseadas no caminho reverso mais curto (ou seja, no caminho mais curto do receptor para o emissor). Entre outros motivos, pode-se ressaltar, além da assimetria dos enlaces das redes na Internet (Costa, et al., 2001b), a capacidade de divisão esparsa de feixes de luz das redes ópticas, o problema de RWA e, também, o fato de que os dispositivos de comutação internos à rede, por falta de conhecimento global sobre a rede, podem não estar aptos a encaminhar os pedidos de adesão aos grupos (e.g. PIM join), caso os emissores estejam localizados fora do domínio óptico. Neste sentido, a utilização de mecanismos baseados em engenharia de tráfego para a geração de árvores de distribuição multiponto começa a ser bastante incentivada e investigada (Chung et al., 2002). O próprio GMPLS é essencialmente baseado em extensões de engenharia de tráfego ao MPLS, incluindo, entre outros tópicos, novos recursos aos dois principais protocolos de sinalização definidos para o MPLS-TE, o RSVP-TE (Berger, 2003b) e o CR-LDP (Ashwood-Smith et al., 2003).

Outra questão importante diz respeito à manutenção de informações de estado relativas aos grupos em cada roteador pertencente às respectivas árvores de distribuição multiponto. Este esquema foi adotado quando os pacotes eram roteados apenas através do caminho mais curto. Atualmente, com o emprego de paradigmas como Diffserv, MPLS e engenharia de tráfego, onde a escolha das rotas pode levar em consideração outras questões como escassez de largura de banda, menor retardo, entre outras, as informações pertinentes a qual o melhor caminho a ser tomado e qual a preferência que um determinado tráfego deve ter em relação a outros são mantidas essencialmente nos roteadores de borda das redes. Enquanto aos roteadores internos cabe a tarefa de comutar o tráfego eficientemente. Em função disso, o armazenamento nos roteadores internos dos backbones de informações de estado, tais como as informações sobre as árvores de distribuição multiponto, tornou-se de certa forma desnecessário e improdutivo, pois, esses nós internos, em geral, não terão informações suficientes para tomar decisões de encaminhamento mais elaboradas, uma vez que essas informações estão sendo mantidas apenas nos dispositivos de borda da rede. Estratégias alternativas vêm sendo investigadas e incentivadas (Xiao, et al., 1999; 2000; Bernet et al., 2000).

No contexto óptico, o armazenamento de informações de estado sobre a árvore multiponto apenas nos dispositivos de borda da rede se torna ainda mais desejável, pois facilitaria a construção otimizada dessas árvores, já que nem todos os nós internos das redes ópticas são aptos a ramificar feixes de luz. Obviamente os nós de bordas teriam que utilizar um protocolo de roteamento baseado no conhecimento da topologia da rede, da capacidade e da disponibilidade dos recursos, com as devidas extensões às redes ópticas.