Distribuição Otimizada de Dados IP Multicast sobre ATM

Ana Cristina Afonso Rodrigues – Engenheira da EMBRATEL e mestranda no PESC/COPPE/UFRJ
Rua Arcozelo, 548, VilaValqueire,
Rio de Janeiro, CEP 21321-480
anac@cos.ufrj.br

Paulo Henrique de Aguiar Rodrigues – Analista do NCE/UFRJ e Professor Adjunto do DCC/IM/UFRJ
Av. Brigadeiro Trompowsky, s/n, Bl. C
CCMN – NCE – Cidade Universitária
Rio de Janeiro, CEP 20001-970
aguiar@nce.ufrj.br

Resumo

O artigo propõe uma solução para a distribuição de dados IP multicast sobre ATM, com qualidade de serviço baseada em reserva de recursos usando o protocolo RSVP, utilizando um mecanismo para estabelecer atalhos (*shortcuts*), encontrando assim um caminho mais eficiente através da rede ATM. A solução traz inegáveis benefícios, quando o volume de tráfego ultrapassar determinado limite, ou quando a aplicação utilizar recursos por um período considerável de tempo.

Abstract

This paper describes an RSVP resource reservation based solution to QoS IP multicast data over ATM distribution. The mechanism sets shortcuts to find a more efficient path across the network. The solution is recommended when the traffic volume exceeds specific bounds, or when the application uses network resources over a long time.

1. Introdução

Tradicionalmente, a Internet somente provê uma classe de serviço chamada best effort, onde nenhuma garantia de qualidade de serviço é especificada. Essa classe de serviço tipicamente utiliza escalonamento na ordem de chegada (first-come, first-serve) em cada hop da rede, apresentando boa performance para aplicações como WWW e transferências de arquivos. Para prover um serviço com garantia de qualidade, principalmente para tráfego de tempo real, novos mecanismos e protocolos estão sendo introduzidos atendendo aos novos requisitos dessas aplicações e ao modelo de serviços integrados na Internet [1,7].

O protocolo RSVP, Resource ReSerVation Protocol [2] foi desenvolvido para sinalização na Internet dos requisitos de qualidade de serviço especificados pela aplicação. Ele se baseia na reserva explícita de recursos de rede, memória e recursos de processamento em todos os nós no

caminho ao longo da rota dos pacotes IP, para prover a qualidade de serviço (QoS) especificada pela classe de serviço.

A tecnologia ATM vem rapidamente se firmando como tecnologia de link. O ATM possui a habilidade de requisitar circuitos virtuais (VC) com uma QoS especifica para serviços ponto-a-ponto e ponto-multiponto. No caso de VC ponto-multiponto, nós são adicionados ou removidos dinamicamente do circuito virtual, oferecendo suporte ao protocolo IP multicast.

A distribuição de dados IP multicast implica no envio simultâneo de um pacote para múltiplos destinos usando uma simples operação local. O conjunto de destinos é chamado grupo de multicast. O modelo de multicast é definido como uma abstração de "grupo de endereços", sendo os endereços de grupo de multicast identificados pela classe D de endereçamento, que indiretamente identifica os membros do grupo de multicast.

Nas redes ATM baseadas na UNI3.x [3] e UNI4.0 [4], o serviço de sinalização não provê suporte nativo a grupos de multicast. Quando o protocolo IP está sobre o protocolo ATM, é necessário, então, um mecanismo para prover a resolução de endereço IP multicast. A solução adotada como padrão pelo IETF (Internet Engeneering Task Force) usando LIS, (Logical IP Subnetwork), é dada em termos do modelo MARS, Multicast Address Resolution Server [5], que trabalha como uma extensão ao serviço de ATMARP [6] para resolução de endereço multicast IP para endereço ATM, com acréscimo de pequenas alterações visando uma melhor performance. Em termos de endereçamento de grupos entre subredes é mantida a mesma estrutura do protocolo IP utilizando-se roteadores de multicast, mrouters, de uma maneira similar ao modelo clássico de IP sobre ATM [6], que utiliza roteadores para interconectar subredes para tráfego unicast.

No modelo clássico IP sobre ATM [6], todos os membros de uma mesma LIS têm o mesmo endereço/subrede IP, estão conectados à mesma subrede NBMA, Non Broadcast Multi-Access, e acessam os outros membros da LIS diretamente, sem uso de roteadores. Todavia, um nó de outra LIS tem que ser acessado necessariamente via roteador, o que introduz hops extras no roteamento, se as LIS compartilham da mesma rede física ATM. Esses hops extras impõe um retardo desnecessário devido a filas, segmentação e remontagem de células, uma vez que cada datagrama IP deve ser analisado em cada roteador presente na rota para fins de processamento de decisões de roteamento. O uso de atalho elimina esse retardo, pois dentro do possível os roteadores são evitados, estabelecendo ligações direta entre nós de diferentes LISs operando na mesma rede física.

Para o endereço unicast o mecanismo de atalho é provido pelo protocolo NHRP, NBMA Next Hop Resolution Protocol [8], protocolo apoiado pelo IETF. O protocolo NHRP é baseado em aplicação cliente-servidor, onde um servidor chamado NHS, Next Hop Server, provê a resolução de endereço IP para ATM e endereço ATM para IP, através da função de ARP-Server e resolução de próximo hop para clientes chamados NHC, Next Hop Clients. Para endereçamento multicast algumas propostas já foram apresentadas para estabelecimento de atalhos, nenhuma delas adotada pelo IETF: EARTH, Easy IP Multicast Routing Through ATM Clounds [9] e IMSS: IP Multicast Shortcut Service [10]. A proposta EARTH propõe uma expansão da noção de cluster do MARS para prover resolução de atalhos, e a proposta IMSS descreve um protocolo de mapeamento de endereço IP multicast para endereços de roteadores de multicast ATM para estabelecimento de atalhos entre os roteadores de multicast conectados à rede física ATM.

O artigo concentra-se na distribuição de dados IP multicast sobre ATM sob o ponto de vista do encaminhamento, buscando otimizar a rota dos pacotes de dados, através do uso de atalhos obtidos via protocolo NHRP. O protocolo NHRP provê, originalmente, resolução de próximo hop para endereço unicast, mas é utilizado na proposta do artigo para resolver o próximo hop na

distribuição de dados multicast do ponto de vista dos receptores, para o qual a fonte é um endereço unicast.

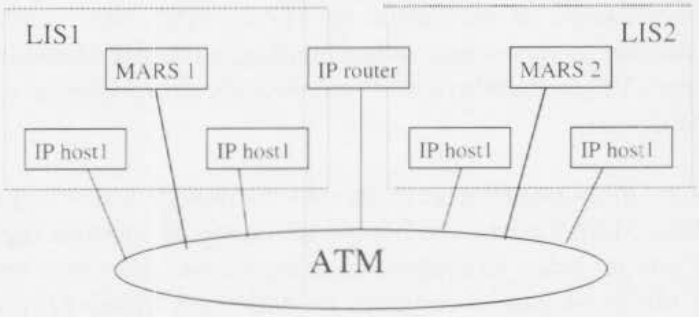
O artigo descreve o modelo MARS, o protocolo NHRP e o protocolo RSVP, respectivamente nas seções 2, 3 e 4. Na seção 5 é apresentada a proposta de distribuição de dados multicast utilizando-se dos protocolos descritos e na seção 6 é dado um exemplo de sua utilização. As conclusões são apresentadas na seção 7.

2. Modelo para resolução de endereços multicast - MARS

O MARS atua como um registrador de membros de grupos multicast, associando grupo de identificadores multicast da camada 3, endereço IP classe D, com as interfaces ATMs de cada membro do grupo em um determinado instante de tempo. O MARS possui mensagens de controle que suportam a distribuição de informação entre o MARS e os nós IP/ATM participantes dos grupos, informando a lista de membros e trocas nos grupos, devido a saídas e entradas de membros.

Cada MARS gerencia um cluster de nós conectados ao ATM, onde cluster é definido como um conjunto de nós que escolhem usar o mesmo MARS, sendo definido em [5], que os administradores de rede devem garantir que cada LIS seja servida por um MARS separado, restringindo assim os limites do cluster aos limites da LIS.

No protocolo IP, um grupo de multicast pode possuir membros de diferentes subredes lógicas (LIS), portanto a distribuição de dados multicast envolve uma combinação de transmissão intra-subrede (dentro da subrede) e inter-subredes (entre subredes). Dentro da subrede da fonte, a facilidade de multicast intra-subrede é usada para propagar os pacotes IP multicast. Já para membros do grupo de outras subredes é requerido roteadores de multicast (*mrouter*) para encaminhamento dos pacotes multicast IP. As operações de multicast intra-subrede e mrouters de inter-subredes são transparentes para a fonte do tráfego IP.



O MARS oferece, através de tabelas, suporte para os dois tipos de emulação de ponto-multiponto: o VC mesh e o MCS (Multicast Server). Para ambos os casos, uma tabela fornece o mapeamento de endereço IP classe D para endereços ATM no formato: {endereço multicast camada de rede, ATM1, ATM2, ATM3, ..., ATMn}. Uma outra tabela, obrigatória no caso de MCS, mapeia em termos de servidores, contendo endereço IP classe D e endereços ATM dos multicast server que suportam esse grupo e usa o formato: {endereço multicast camada de rede, server1, server2, ..., server k}.

Cada nó que deseja receber pacotes de determinado grupo de multicast deve previamente se registrar no MARS de sua LIS através do envio de uma mensagem especifica. Para a transmissão de pacotes para um determinado grupo de multicast esse procedimento não é necessário, bastando consultar o MARS para conhecer os endereços ATM dos possíveis

membros, sem a necessidade de registro. Este procedimento é de certa maneira bastante similar ao procedimento do protocolo IGMP, Internet Group Multicast Protocol [11].

Para propagação de mensagens de alterações nos membros dos grupos de multicast, saídas ou entradas de novos membros, o MARS estabelece um VC ponto-multiponto, o ClusterControlVC, onde cada membro do cluster é uma folha do VC. Adicionalmente, quando o MCS é usado, um VC ponto-multiponto separado é estabelecido para registro dos MCSs, o ServerControlVC, onde cada servidor multicast registrado é membro do ServerControlVC.

Cada nó que possui um pacote para transmissão a um endereço classe D pesquisa em seu cache interno se existem os endereços ATMs dos membros pertencentes ao grupo de multicast, para o qual ele quer enviar pacotes. Se não existirem, ele estabelece um VC ponto-a-ponto para o MARS, e envia uma mensagem solicitando o mapeamento do endereço multicast IP para um conjunto de endereços ATMs dos membros desse grupo. Caso esse mapeamento exista, o MARS envia um mensagem contendo a relação de membros do grupo dentro do cluster ou um conjunto de um ou mais MCSs, e, caso contrário, é enviada uma mensagem indicando que não existem membros para esse grupo. O nó então de posse dos endereços ATM estabelece o VC ponto-multiponto, em caso de VC mesh. Se o solicitante do mapeamento de endereço multicast é um MCS, o MARS envia uma mensagem contendo os membros do grupo dentro do cluster, possibilitando o estabelecimento de um circuito virtual ponto-multiponto do MCS com os membros do grupo.

Quando um membro de um cluster é um membro de um grupo de multicast, ele possui uma entrada na tabela do MARS para esse endereço de grupo de multicast. Caso ele queira sair ou se juntar à um grupo, ele envia ao MARS mensagens indicando saída ou entrada em um ou mais grupos de multicast. Estas mensagens após serem processadas, internamente pelo MARS, são propagadas sobre o ClusterControlVC, para garantir o conhecimento da troca aos membros do grupo. A propagação das mensagens para todos os membros do cluster, em caso de trocas, gera uma pesada carga de sinalização em grupos muito dinâmicos, penalizando a performance da rede. Essa é uma das principais desvantagens do modelo MARS.

No nosso entender, uma alteração no modelo MARS para minimizar o problema acima descrito seria a inclusão de um campo na mensagem de MARS-JOIN, que indicaria se o membro do cluster quer ou não receber as mensagens sobre mudanças nos membros dos grupos, eliminando assim do ClusterControlVC os membros não interessados em receber as alterações, diminuindo assim a carga de sinalização.

Como os roteadores de multicast IP devem receber pacotes de todos os grupos de multicast de maneira transparente, o MARS acrescentou a possibilidade de suportar registro de roteadores de multicast, *mrouters*, em um bloco de endereços de multicast, onde pode ser passado um simples grupo, um pequeno bloco ou todo o conjunto de endereços classe D, provocando com que o roteador de multicast seja membro de todos os grupos, existentes e futuros, no intervalo de endereços especificado no registro.

3. Protocolo NHRP

O protocolo NHRP [8] é usado por um nó IP (ou roteador) conectado a uma rede NBMA que deseje comunicar-se através da interface de rede ATM com outro nó IP, para determinar o endereço ATM do próximo hop NBMA para o nó destino. Se o nó destino está conectado a rede NBMA, então o próximo hop NBMA é o próprio nó destino. Por outro lado se o nó destino não estiver conectado à rede NBMA o próximo hop é então o roteador egresso que esteja mais próximo em termos de roteamento ao nó destino.

A resolução de endereços provida pelo servidor de ATMARP, no modelo clássico IP sobre ATM, atua somente dentro da LIS e resolve o endereço do próximo hop, somente se o nó destino for membro da LIS. Caso contrário, se o nó destino não é membro da LIS, o nó fonte deve encaminhar os pacotes via o roteamento default IP. Os pacotes vão passando de roteador a roteador até chegar ao destino, causando o encaminhamento dos pacotes por vários hops na rede NBMA. O NHRP tenta eliminar esses hops extras eliminando roteadores e estabelecendo um caminho direto entre nó fonte e nó destino, ou nó fonte e roteador egresso mais próximo do destino.

O NHRP possui duas entidades para resolução de atalhos, o NHS, Next Hop Server, e o NHC, Next Hop Client. O NHS desenvolve o serviço de resolução de próximo hop dentro da NBMA de maneira cooperativa e interage com o protocolo de roteamento para troca de informações. Cada NHS serve a um conjunto de nós, os quais ele aprende através de mensagens ou configuração manual. Os NHS na prática atuam nos roteadores egressos da rede NBMA, possibilitando assim troca de informações de roteamento entre a rede NBMA e outras redes conectadas, sendo que opcionalmente ele pode suportar também o serviço clássico de resolução de endereços ARP. O NHC é a entidade que inicia o pedido de resolução de próximo hop ao NHS.

O protocolo funciona da seguinte maneira: um evento ocorre em um nó fonte S disparando um processo de resolução de endereço NBMA para o nó destino D. O nó fonte S primeiro determina o próximo hop para o nó D usando o processo de roteamento normal. Se a informação de resolução de endereço do destino já está presente no cache, então ele usa essa informação para encaminhar os pacotes. Se for determinado que o próximo hop está disponível via rede NBMA, o nó S constrói uma requisição de resolução e a envia para o NHS na rede NBMA que foi configurado como seu servidor. A requisição contém o endereço internet do nó destino e o endereço do nó fonte

A requisição de resolução chega ao NHS que consulta sua base de dados para determinar se ele serve o nó destino. Caso o NHS sirva o nó D, isto é, o nó se registrou previamente no servidor, ou através do roteamento ele saiba que ele é o próprio roteador egresso para o nó D, ou se ele, apesar de não servir à estação D ou a seu roteador egresso, possuir a resolução de próximo hop, armazenada por respostas de requisições anteriores, ele envia uma mensagem com a resolução do próximo hop. Caso contrário, o NHS encaminha a requisição de resolução para outro NHS, determinado através do roteamento ou através de prévia configuração, e assim sucessivamente até o último NHS da rede conectado à rede NBMA . Se é determinado que nenhum NHS possui a resolução de endereço para o nó D ou para o roteador egresso para chegar a ele, é enviada uma mensagem indicando a inexistência de resolução de próximo hop.

A mensagem resposta à requisição de resolução do próximo hop segue a rota contrária à da mensagem de requisição. Cada NHS intermediário da rota extrai a informação de resolução de endereço e a armazena em seu cache de resolução de endereços, fazendo com que ele possa responder a futuras requisições de resolução.

Para determinar roteadores egressos mais próximos de um destino fora da rede NBMA, todos os roteadores egressos da rede NBMA devem trabalhar como NHS, servindo os nós da rede NBMA e os nós destinos além deles. Nesse caso, os roteadores devem trocar informações de roteamento entre a rede NBMA e outras redes a quem eles estão conectados. Se existir algum roteador egresso que não suporte NHRP, não será possível fazer a resolução de próximo hop para os nós servidos por ele fora da rede NBMA, e nesse caso a fonte usará a rota default para encaminhar os pacotes.

4. Protocolo RSVP

O protocolo RSVP [2] é basicamente um protocolo de sinalização de reserva de qualidade de serviço (QoS) designado para suportar serviços integrados sobre a Internet. Ele é usado por hosts e outros elementos da rede para requisitar à rede uma especifica qualidade de serviço (QoS) que atenda aos requisitos de uma aplicação fim-a-fim. O RSVP também é usado por roteadores para propagar requisições de QoS, estabelecer e manter um estado de reserva nos nós ao longo da rota para prover a QoS requerida.

O protocolo RSVP requisita recursos para fluxo de dados simplex, isto é, ele faz a requisição de recursos em apenas uma direção. O RSVP trata a fonte e o receptor logicamente de maneira distinta, embora um mesmo processo de aplicação possa atuar como fonte e receptor simultaneamente.

O RSVP opera no topo do protocolo IP, IPv4 ou IPv6, ocupando parte da camada de transporte, entretanto ele não é uma aplicação de transporte de dados. O protocolo RSVP não é um protocolo de roteamento e ele é designado para operar juntamente com um protocolo de roteamento unicast e multicast do qual ele obtém as rotas.

O RSVP faz reserva de recursos tanto para aplicações unicast como multicast, se adaptando dinamicamente à entrada e saída de membros do grupo, mudanças de rotas e atendendo a requisitos de receptores heterogêneos. É orientado ao receptor, onde é função do receptor iniciar e manter a reserva de recursos a ser usada pelo fluxo.

O mecanismo do protocolo RSVP provê a facilidade para criar e manter estados de reserva distribuídos através de rotas unicast e multicast. Ele transfere e manipula parâmetros de controle de QoS como dados opacos, passando-os para módulos de controle de trafego para interpretação. A estrutura e o conteúdo dos parâmetros de QoS são especificadas pelo Serviços Integrados [1,12].

O RSVP foi desenvolvido para ser bem escalonável em grandes grupos de multicast, onde membros do grupo e a topologia da árvore se alteram com o tempo. Para esse propósito os roteadores usam *soft-state*, onde periodicamente mensagens devem ser enviadas para que se mantenha a reserva de recursos, e, em caso da inexistência de revalidação, a reserva de recursos é automaticamente removida.

A requisição de QoS é feita pela aplicação do nó receptor que é passada ao processo local do RSVP. O protocolo RSVP então carrega a requisição de QoS para todos os nós(hosts e roteadores) no caminho reverso do fluxo de dados até a fonte.

A QoS para um particular fluxo de dados é implementada por mecanismos coletivamente chamados de "controle de trafego". Esses mecanismos incluem um classificador de pacotes, controle de admissão, controle de policiamento e escalonador de pacotes ou algum mecanismo dependente da camada de enlace que determine como os pacotes serão encaminhados. O classificador de pacote determina a QoS para cada pacote. O controle de admissão determina se há recursos disponíveis para suportar a QoS requisitada e o controle de policiamento se o usuário tem permissão administrativa para fazer a reserva.

O RSVP provê diferentes modelos de reserva ou estilos para atender a uma variedade de aplicações, provendo também uma operação transparente através de nós que não o suportem. Não é definido pelo RSVP nenhum tipo de parâmetro para especificação da QoS ou do tráfego, sendo utilizado o modelo IntServ [13]. O IntServ utiliza como métrica para especificar a QoS o retardo do pacote, definindo serviços para suportar áudio, vídeo, aplicações de tempo real e tráfego de dados. Os serviços definidos pelo IntServ são Retardo Garantido e Carga Controlada.

4.1. Modelo de Reserva

O RSVP define sessão como sendo um fluxo de dados com um particular destino e protocolo de transporte e trata cada sessão de maneira independente. Uma sessão RSVP é definida pela tripla: (*DestAddress*; ID do protocolo IP; *Destport*), onde o *DestAddress* é o endereço IP dos pacotes de dados, sendo que para receptores multicast o *DestAddress* é um grupo de endereços e para unicast é um simples endereço. O campo *Destport* é a porta de destino do TCP/UDP ou uma porta equivalente em outro protocolo de transporte, sendo opcional seu uso de maneira geral. Entretanto seu uso é obrigatório quando se quer permitir que mais de uma sessão unicast seja endereçada para um mesmo receptor.

A requisição de reserva elementar do RSVP consiste em um descritor de fluxo, *flow-descriptor*, composto do *flowspec* e *filter-spec*, onde o primeiro especifica a desejada QoS e o segundo quem receberá a QoS definida.

O *flowspec* é composto da especificação da classe de serviço e dois conjuntos de parâmetros numéricos, Rspec, *Reservation Spec*, e Tspec, *Traffic Spec*. O primeiro parâmetro define a QoS desejada e o segundo descreve o fluxo de dados, onde seus formatos e conteúdos são determinados pelo IntServ. O *flowspec* é utilizado para configurar parâmetros no escalonador de pacotes no nó ou outro mecanismo da camada de enlace (assumindo-se que o controle de admissão foi bem sucedido).

O *filterspec*, junto com a especificação da sessão, define o subconjunto de pacotes que terá direito à QoS especificada no *flowspec*, sendo utilizado para configurar parâmetros no classificador de pacotes no nó. Seu formato depende da versão do protocolo IP utilizada, IPv4 ou IPv6, podendo ser definido em termos de fonte (endereço IP e porta), protocolos de alto nível ou em algum campo do header do pacote, mas na versão corrente do RSVP, versão 1, por simplicidade é utilizado a mais restritiva forma de *filterspec*, a definição em termos de fonte, consistindo o *filterspec* de endereço-IP-fonte e opcionalmente o número de porta TCP/UDP.

Do ponto de vista da seleção das fontes há duas maneiras de se fazer a seleção: elas podem ser explicitamente selecionadas através de uma lista, ou implicitamente selecionadas se todas as fontes forem selecionadas, o chamado *wildcard*. Na modelo de reserva com explícita seleção de fontes, cada *filterspec* deve ser corresponder exatamente a uma fonte, enquanto no outro caso nenhum *filterspec* é necessário.

A combinação desses dois fatores: reserva compartilhada ou reserva distinta e seleção de fontes explícita ou *wildcard*, nos leva há quatro estilos de reservas sendo a combinação reserva compartilhada + seleção *wildcard* impossível de definição. Os estilos de reserva são *wildcard-filter*, *fixed-filter* e *shared-explicit*. O estilo de reserva *wildcard-filter* e *shared-filter* são apropriados para as aplicações multicast em que múltiplas fontes de dados transmitem simultaneamente, por exemplo audio-conferência, desde que se limite o número de pessoas falando simultaneamente. O estilo *fixed-filter* que cria reservas distintas para fluxos de diferentes fontes é apropriado para vídeo-conferências.

4.2. Procedimento de Reserva

No RSVP a reserva é feita pelo receptor, mas para isso o receptor tem que conhecer quais as alternativas oferecidas pelas fontes. Em caso de multicast, o receptor deve se juntar previamente ao grupo de multicast especificado pelo campo *DestAddres*. A potencial fonte começa o envio da mensagem RSVP-PATH para todos os receptores do grupo de multicast na árvore de

multicast contendo a especificação do tráfego que será gerado. Os receptores, que são membros do grupo de multicast, originam uma mensagem RSVP-RESV em resposta contendo a requisição de reserva de QoS, que é então enviada na rota inversa da mensagem RSVP-PATH.

Em cada nó intermediário, ao longo do caminho, a requisição de reserva é passada ao controle de admissão e controle de policiamento, que determina se ela pode ser aceita. O controle de admissão determina se o nó possui recursos disponíveis para fornecer a QoS desejada e o controle de policiamento se o usuário tem permissão administrativa para fazer a reserva. Caso a requisição seja rejeitada uma mensagem de erro é gerada para o receptor, mas se ela for bem sucedida, o nó configura o classificador de pacotes para selecionar os pacotes de dados definidos pelo *filterspec*, e interage com a camada de enlace para obter a QoS desejada definida pelo *flowspec*.

A requisição de reserva é encaminhada ao próximo nó na rota reversa até a fonte. A reserva pode não ser necessariamente a que foi propagada por cada receptor, uma vez que os mecanismos de controle de admissão dos nós intermediários podem modificar o *flowspec* e podem fazer junção das requisições de reserva de membros de uma árvore de multicast para uma mesma fonte ou conjunto de fontes.

Os receptores originadores de requisição de reserva podem solicitar uma mensagem de confirmação indicando que a requisição foi provavelmente instalada pela fonte. O modelo básico do RSVP é de um passo, a requisição é enviada a cada nó ao longo do caminho, sendo que cada nó aceita ou rejeita a requisição, o que torna difícil a detecção de aceitação da requisição fim-a-fim. O RSVP suporta também o serviço de um passo conhecido como OPWA, "*One Pass With Advertising*", onde mensagens de avisos sobre a rede são enviadas pelo RSVP para os receptores tornando-os capazes de construir ou ajustar dinamicamente uma requisição de QoS apropriada.

4.3. Mensagens do RSVP

O RSVP possui duas mensagens fundamentais responsáveis pela requisição da reserva, a mensagem RSVP-PATH e a RSVP-RESV e mais quatro mensagens de controle RSVP-PATHTEAR, RSVP-RESVTEAR, RSVP-PATHERR, RSVP-RESVERR e RSVP-RESVCONF. Todas as mensagens possuem um cabeçalho comum e um corpo de tamanho variável contendo objetos.

A mensagem de RSVP-PATH é gerada pela fonte e encaminhada ao próximo nó ao longo de uma rota unicast ou multicast, que é provida pelo protocolo de roteamento, até o nó de destino. A mensagem RSVP-PATH armazena o estado do caminho (*path-state)* ao longo de cada nó por onde ela passa. Esse estado do caminho armazena as informações provenientes da mensagem RSVP-PATH e é utilizado para rotear a mensagem RSVP-RESV hop-a-hop na direção contrária.

. Formato da mensagem de PATH:
<Header Comum> + [<INTEGRITY>] + <SESSION> + <RSVP-HOP> + <TIME-VALUES> + <SENDER-TEMPLATE> + <SENDER-TSPEC> + [<ADSPEC>] + [<POLICY-DATA>]

O objeto SESSION contem o endereço destino IP(*DestAddress*) de uma sessão. O RSVP-HOP contem o endereço do hop anterior da rota dos pacotes de dados. O TIME-VALUES, contem o valor do período de refresh usado pelo originador da mensagem. O SENDER-TEMPLATE, descreve o formato dos pacotes de dados gerados pela fonte, ele possui a forma de um *filter-*

spec e pode ser usado para selecionar pacotes desta fonte de outra dentro da mesma sessão no mesmo link. O SENDER-TSPEC define as características dos fluxo de dado gerado pela fonte.

Opcionalmente a mensagem RSVP-PATH possui o objetos ADSPEC, que carrega informações de avisos do OPWA, o POLICY-DATA, que contem informações sobre política de policiamento a ser adotada pelo processo local e o INTEGRITY que contem informações sobre criptografia de dados para autenticação e verificação das mensagens do RSVP.

Cada fonte envia periodicamente uma mensagem RSVP-PATH contendo a descrição de cada fluxo de dados por ela originado, a qual é encaminhada pela mesma rota usada pelo fluxo de dados.

A mensagem RSVP-RESV é gerada pelo receptor que a encaminha exatamente pela rota reversa dos pacotes de dados hop-a-hop para todas as fontes incluídas na seleção de fontes. Essa mensagem é utilizada para criar e manter o estado de reserva (*reservation-state)* em cada nó ao longo da rota, configurando assim o controle de trafego em cada nó.

. Formato da mensagem de RESV:
<Header Comum> + [<INTEGRITY>] + <SESSION> + <RSVP-HOP> +
<TIME-VALUES> + [<RESV_CONFIRM>] + [<SCOPE>] +
[<POLICY-DATA>] + <STYLE> + <FLOW DESCRIPTOR>

Os objetos SESSION, INTEGRITY, TIME-VALUES e POLICY-DATA têm o mesmo significado do mostrado na mensagem RSVP-PATH. O objeto RSVP_HOP contém o endereço IP da interface para qual a mensagem de RSVP-RESV foi enviada e o LIH da interface lógica para qual a reserva foi requerida. A existência do objeto RESV_CONFIRM sinaliza um pedido para confirmação da reserva e carrega o endereço IP do receptor para o qual a mensagem de RSVP-RESVCONF será enviada. O objeto SCOPE contem a lista explícita de todos os hosts para onde a informação deve ser encaminhada, o objeto STYLE informa o estilo da reserva descrito anteriormente e o FLOW-DESCRIPTOR, também já descrito, contem o *flowspec*, que especifica a QoS desejada, e o *filter-spec*, que determina o subconjunto de pacotes de dados que receberam a QoS desejada.

As mensagem RSVP-PTEAR e RSVP-RTEAR removem os estados do RSVP mantidos em cada nó ao longo da rota dos pacotes de dados. A mensagem RSVP-PTEAR é iniciada explicitamente pelas fontes ou por algum nó cujo o tempo de expiração do estado do caminho estourou. Ela serve para remover o estado do caminho e todos os estados de reservas dependentes em todos os nós do seu ponto de iniciação até os receptores. A mensagem RSVP-RTEAR é iniciada explicitamente pelos receptores ou por algum nó cujo o tempo de expiração do estado de reserva estourou. Ela serve para remover o estado de reserva do ponto de iniciação da mensagem até as fontes. As mensagens podem ser iniciadas por uma aplicação de um nó (fonte ou receptor), ou por um roteador como resultado de um estouro do tempo de expiração de um estado.

As mensagens de erro do RSVP são RSVP-RESVERR e RSVP-PATHERR. A mensagem RSVP-PATHERR reporta erros ocorridos no processamento das mensagens RSVP-PATH. Ela é enviada para a fonte que gerou o erro e não provoca nenhum tipo de alteração do estado do caminho por onde passa, já a mensagem RSVP-RESVERR reporta erros ocorridos no processamento das mensagens RSVP-RESV. Ela é enviada para todos os receptores apropriados. A mensagem RSVP-RESCONF é uma confirmação (probabilística) para uma requisição de reserva feita pela mensagem RSVP-RESV.

5. Proposta para serviço multicast com QoS e atalhos

A proposta de distribuição de dados IP multicast sobre rede ATM, combina o protocolo RSVP a fim de garantir à aplicação a QoS necessária, o protocolo MARS, para resolução de endereço multicast, e o protocolo NHRP para descobrir o próximo hop, a fim de otimizar a rota dos pacotes de dados.

O MARS é utilizado pela fonte, caso a mesma esteja conectada diretamente à rede ATM ou pelo roteador de ingresso da rede ATM, e por cada roteador ao longo da rota para descobrir quais os membros de determinado grupo de multicast IP no âmbito de intra-subrede, LIS, se eles existirem. No âmbito de inter-subrede, quando os pacotes para um determinado grupo de multicast são enviados para fora dos limites da LIS, é utilizado o protocolo IDMR, Inter-Domain Multicast Routing, [14] para encaminhamento dos pacotes IP.

Cada roteador de multicast deve se cadastrar no servidor de MARS em todos os grupos de multicast, conforme sugerido na RFC2022 [5], com o objetivo de monitorar a existência de algum membro conectado em cada LIS da rede NBMA, e para possibilitar o encaminhamento de pacotes para membros que estejam fora da rede NBMA.

O protocolo NHRP é utilizado para resolução do próximo hop dos membro do grupo de multicast para a fonte do grupo de multicast ou para o roteador egresso dessa fonte, caso a fonte não esteja conectada a rede NBMA, sendo necessário que para isso todos os roteadores egressos da rede ATM suportem o protocolo NHRP.

As mensagens de controle RSVP-PATH da fonte do grupo de multicast ou do seu roteador egresso deveram seguir pelo roteamento IP normal, respeitando os limites da LIS, mas a mensagem RSVP-RESV e os dados como proposto deveram seguir pelo atalho identificado via NHRP. Caso não seja possível o estabelecimento de atalho, a mensagem RSVP-RESV e os dados deverão seguir pela rota inversa a da mensagem RSVP-PATH, como especificado no protocolo RSVP[2].

5.1. Anúncio de Tráfego

A fonte que deseja enviar certo tipo de tráfego para um determinado grupo de multicast IP com reserva de recursos gerenciada pelo protocolo RSVP deve enviar a mensagem RSVP-PATH, contendo a especificação do tráfego que ela irá fornecer, para todos os nós presentes na rota dos pacotes de dados. Essa fonte pode estar conectada diretamente ou não à rede ATM.

Se a fonte estiver conectada diretamente à rede ATM, por ela não estar mais numa rede *broadcast,* ela deve primeiro estabelecer um circuito virtual com os membros do grupo de multicast de sua LIS. Para obter a lista desses membros ela deve se conectar ao servidor MARS e enviar uma mensagem MARS-REQUEST contendo o endereço de multicast do grupo alvo. O MARS deverá responder com uma mensagem MARS-MULTI, contendo os membros e roteadores presentes no grupo de multicast. De posse da lista a fonte estabelece então um VC ponto-multiponto best-effort com todos os membros e envia a mensagem RSVP-PATH, anunciando seu tráfego.

Caso a fonte não esteja conectada diretamente à rede ATM, ela envia a mensagem RSVP-PATH, pelo procedimento descrito na RFC1112[11] de envio de tráfego multicast em rede BMA, onde, no caso de Ethernet, é utilizado o endereço especial de multicast do Ethernet. A mensagem vai trafegar pela rota IP até chegar a um roteador ingresso da rede ATM. O roteador verifica se possui recursos para suportar o tráfego, e, se possuir cria o estado do caminho. Ele consulta o MARS, como descrito acima, para verificar os membros do grupo de multicast

presentes na sua LIS, e consulta também as tabelas de roteamento IDMR para obter informação sobre a rota do pacote para fora da LIS, isto é, a árvore de roteamento multicast, que é construída na chegada do primeiro pacote endereçado para o grupo de multicast, no caso a mensagem RSVP-PATH, supondo-se que o roteador trabalhe como roteador de multicast.

5.2. Roteamento da mensagem RSVP-PATH

A mensagem RSVP-PATH é roteada pela rota IP default. A nível de inter-subrede é construída uma árvore de roteamento multicast pelo protocolo IDMR, na chegada da primeira mensagem a um roteador de multicast, que é usada para rotear o pacote que contém a mensagem.

Em cada roteador de multicast por onde a mensagem passa, um procedimento de consulta ao MARS é iniciado, caso o roteador saiba que ele possui membros nesse grupo de multicast em sua rede local. Ele obtém essa informação, através das mensagens de MARS-JOIN e MARS-LEAVE repassadas pelo MARS a todos os membros associados a ele, cada vez que um novo nó se junta ou sai do grupo, e uma vez que todos os roteadores de multicast devem se associar em todos os grupos de multicast, ele pode portanto montar uma tabela com os endereços de grupos de multicast que possuam membros.

Além da consulta ao MARS, os roteadores de multicast consultam suas tabelas internas de roteamento multicast para saber se ele possui algum roteador de multicast vizinho conectado na rede ATM, que é o próximo ramo da árvore de multicast, algum roteador de multicast vizinho e/ou algum membro do grupo de multicast conhecido via protocolo IGMP, conectados fora da rede NBMA, mas alcançados por uma de suas interfaces para rede BMA, possibilitando assim o encaminhamento das mensagem RSVP-PATH para fora da rede NBMA.

Em cada nó que recebe a mensagem RSVP-PATH essa requisição de recursos é passada ao controle de admissão e controle de policiamento, onde, no primeiro é checado a existência de recursos suficientes para atender aos requisitos de tráfego, e, no segundo se quem fez a solicitação da reserva de recursos possui autoridade para faze-lo. Se ambas as checagens são bem sucedidas, é criado o estado do caminho no nó, contendo as informações contidas na mensagem, possibilitando informações sobre o tráfego e roteamento para a mensagem RSVP-RESV, que consulta o estado do caminho para obtenção do próximo hop, não necessitando mais do protocolo de roteamento para conhecer o endereço do próximo hop da sua rota, que pela especificação deve ser exatamente a contrária a rota da mensagem RSVP-PATH.

5.3. Processamento da Reserva

Os nós que receberam a mensagem RSVP-PATH, de acordo com o tipo de nó, devem se preparar para realizar a reserva de recursos necessária para atender às necessidade da aplicação efetivada pelo envio da mensagem RSVP-RESV.

Os roteadores de multicast enviam a mensagem RSVP-PATH para todas as interfaces que possuam membros no grupo de multicast ou algum roteador de multicast vizinho presente na árvore de multicast. Como retorno da mensagem RSVP-PATH, um roteador pode receber uma ou mais mensagens RSVP-RESV para determinada sessão RSVP, que contém a reserva de recursos solicitada. O roteador faz então uma combinação dessas reservas, criando uma reserva única para um dada sessão, um único estado de reserva.

Se um nó membro do grupo de multicast conectado diretamente ao ATM receber a mensagem RSVP-PATH, ele faz a mesma verificação de controle de admissão e controle de policiamento e se ambos forem bem sucedidos é criado o estado de reserva.

Pelo procedimento normal de distribuição de dados multicast IP sobre ATM, os roteadores e nós conectados à rede ATM deveriam consultar o estado do caminho para retirar o endereço do próximo hop para onde encaminhariam a mensagem RSVP-RESV, que seguiria pelo roteamento IP normal. A proposta neste ponto do procedimento propõe que após uma análise do tráfego, se o mesmo ultrapassar certos requisitos, seja tentado o estabelecimento de atalhos interligando os roteadores de multicast e nós participantes do grupo de multicast à fonte do tráfego de multicast ou ao seu roteador egresso na rede ATM.

O nó ou roteador de multicast obtém na mensagem RSVP-PATH, do campo SENDER-TEMPLATE, o endereço da fonte do tráfego multicast, monta uma mensagem de requisição de resolução de endereço do próximo hop e a envia ao NHS que o serve. O NHC verifica se possui em sua tabela a resolução de endereço do próximo hop solicitada; se não possuir ele envia para o próximo NHS configurado. A mensagem vai passando pelos NHSs até que seja encontrada a resolução ou que o número de hops máximo da mensagem seja atingido. Se a resolução for bem sucedida, o NHS envia para o nó que solicitou a resolução de próximo hop a mensagem NHRP-REPLY, contendo o endereço ATM da fonte, caso ela esteja conectada diretamente a rede ATM, ou do roteador egresso, caso contrário.

Se a resolução de próximo hop não foi obtida, pois nenhum NHS possui a informação ou o número de hops máximo da mensagem de requisição estourou, a mensagem RSVP-RESV é encaminhada pela rota contrária da mensagem RSVP-PATH, conforme descrito na RFC2205[2]. Em cada nó, para encaminhamento da mensagem RSVP-RESV, é usado o campo RSVP-HOP presente no estado do caminho, que indica o hop anterior da mensagem RSVP-PATH, portanto o próximo hop da mensagem RSVP-RESV.

O nó ou roteador de multicast de posse da resolução de endereço do próximo hop para a fonte ou roteador egresso do tráfego multicast, estabelece um VC best-effort diretamente com o nó, e envia a mensagem RSVP-RESV, sinalizando a necessidade de confirmação da reserva. Se a reserva for confirmada, é armazenado no estado do caminho a informação relativa ao atalho para ser utilizada nas transmissões futuras da mensagem RSVP-RESV. Se a reserva não for confirmada é seguido o mesmo procedimento descrito acima para inexistência de resolução de próximo hop.

O uso do atalho para a mensagem RSVP-RESV causa assimetria entre a rota da mensagem RSVP-PATH e RSVP-RESV, contrariando a proposta do protocolo RSVP que determina que a rota dos dados e mensagens RSVP-RESV seja a mesma da mensagem RSVP-PATH, sendo em direções opostas. Esta assimetria não causa problemas, podendo ser comparada à solução prevista em [2], que descreve o problema de mensagens do RSVP chegando por roteadores ou interfaces diferentes do previsto.

A fonte ou o roteador egresso da transmissão multicast ao receber a mensagem de RSVP-RESV, cria o estado de reserva, verifica se já existe um VC aberto que atenda as necessidades dessa nova reserva. Se esses dois aspectos forem satisfeitos, o novo receptor é adicionado ao VC. Se o VC não existir ele é estabelecido, e, caso a reserva desse novo receptor não for atendida pelo VC já aberto, é utilizado um dos métodos para resolução de heterogeneidade descritos em [7].

6. Exemplo

Para exemplificação da proposta é dada uma topologia típica de rede IP sobre ATM. Na topologia proposta a fonte S deseja transmitir dados com certa QoS para um determinado grupo de multicast IP. Os roteadores trabalham como roteadores de multicast e em cada LIS é subentendido que existe um servidor de MARS.

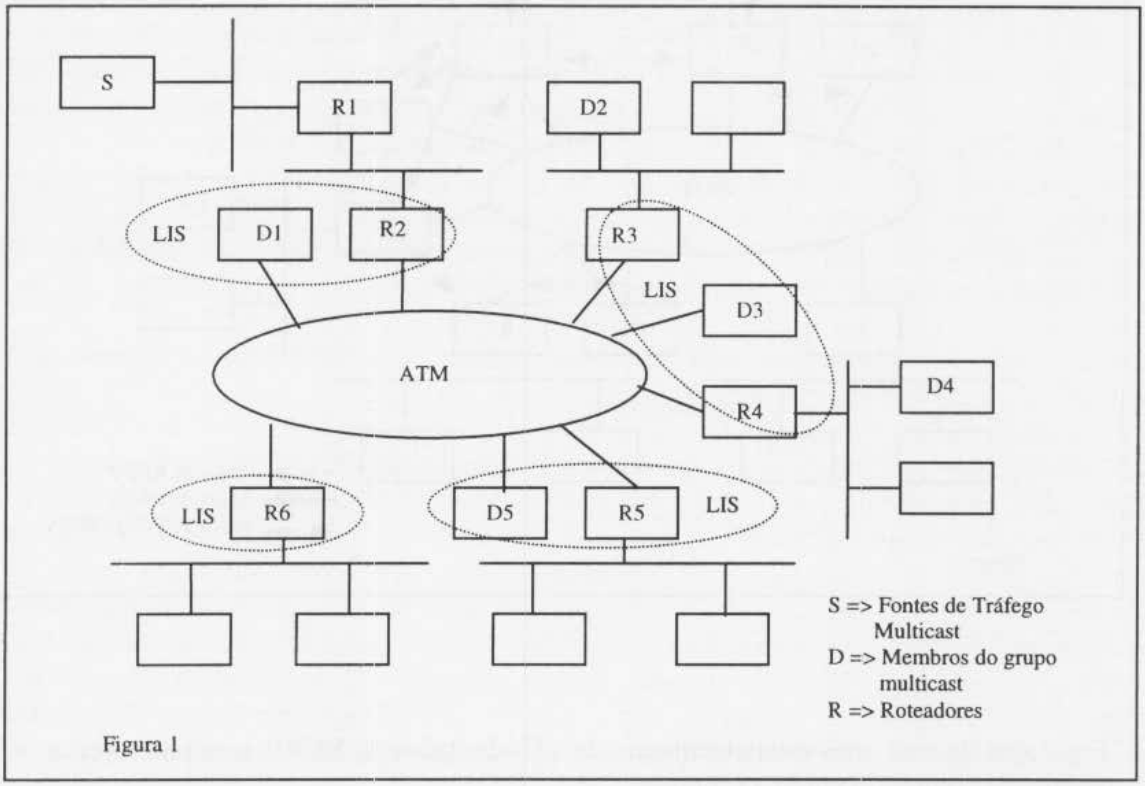


Figura 1

A árvore de multicast IP, obtida segundo um protocolo multicast qualquer:

$R1 \Rightarrow R2 \Rightarrow R3 \Rightarrow R4 \Rightarrow R5$

- Topologia da rede após estabelecimento de VCs de controle do RSVP, sem a utilização da proposta:

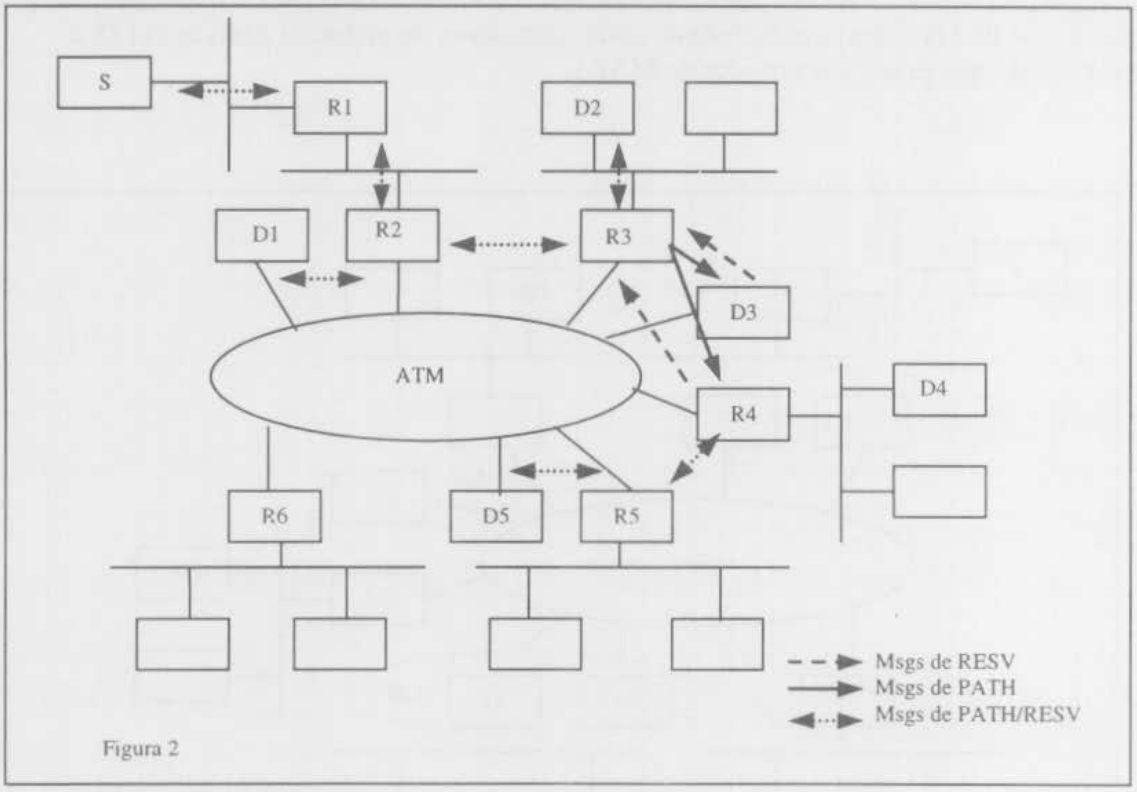


Figura 2

- Topologia da rede após estabelecimento de VCs de dados do RSVP, supondo reserva homogênea, não utilizando a proposta:

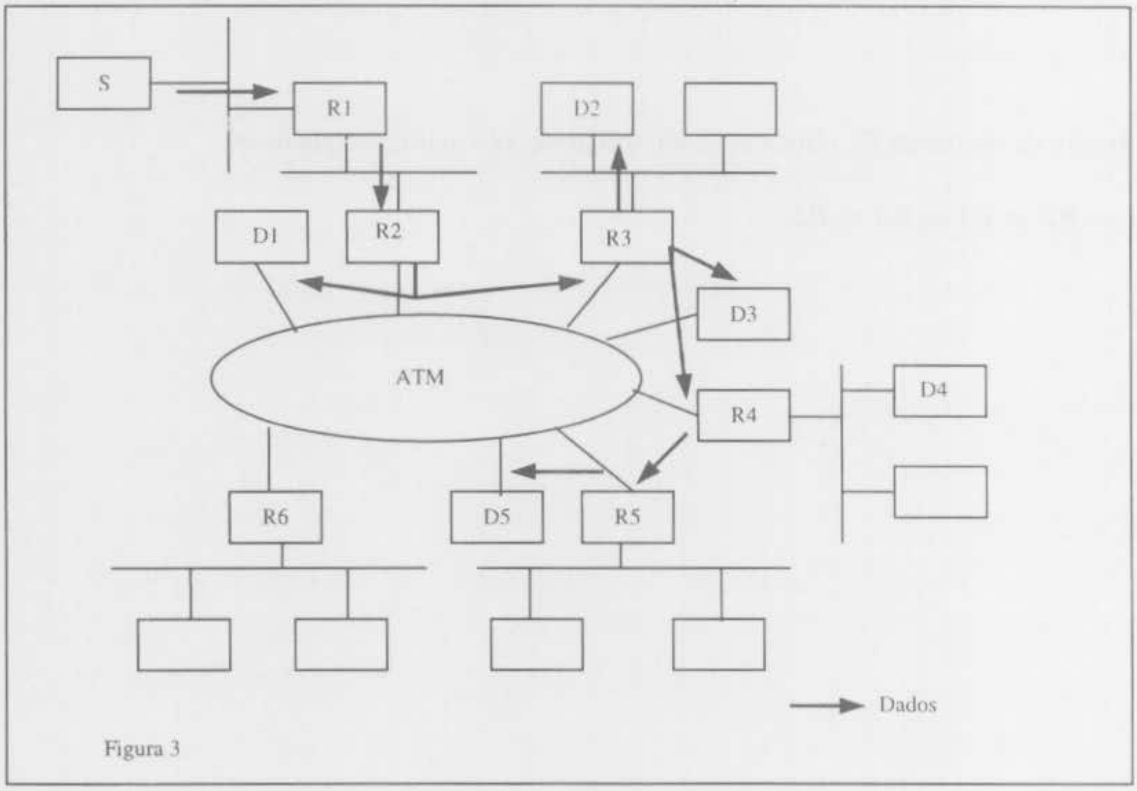


Figura 3

- Topologia da rede após estabelecimento de Vcs de controle do RSVP, utilizando a proposta:

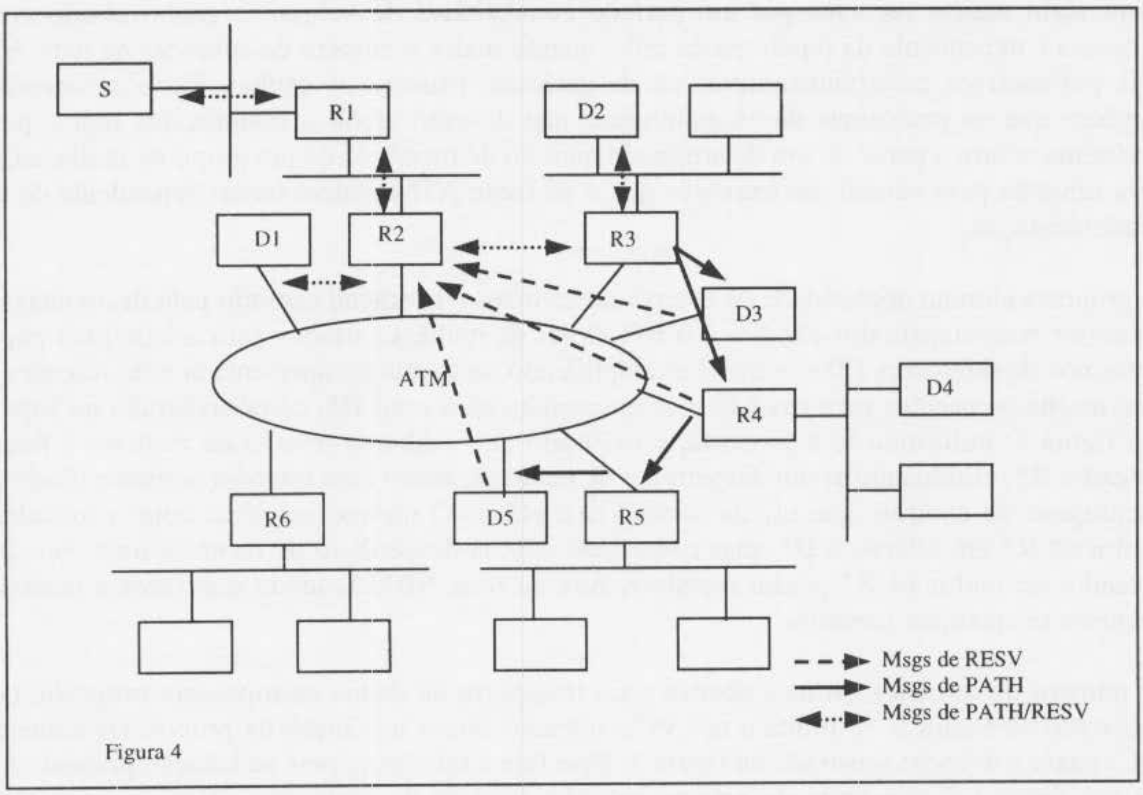


Figura 4

- Topologia da rede após estabelecimento de VCs de dados do RSVP, supondo reserva homogênea, utilizando a proposta:

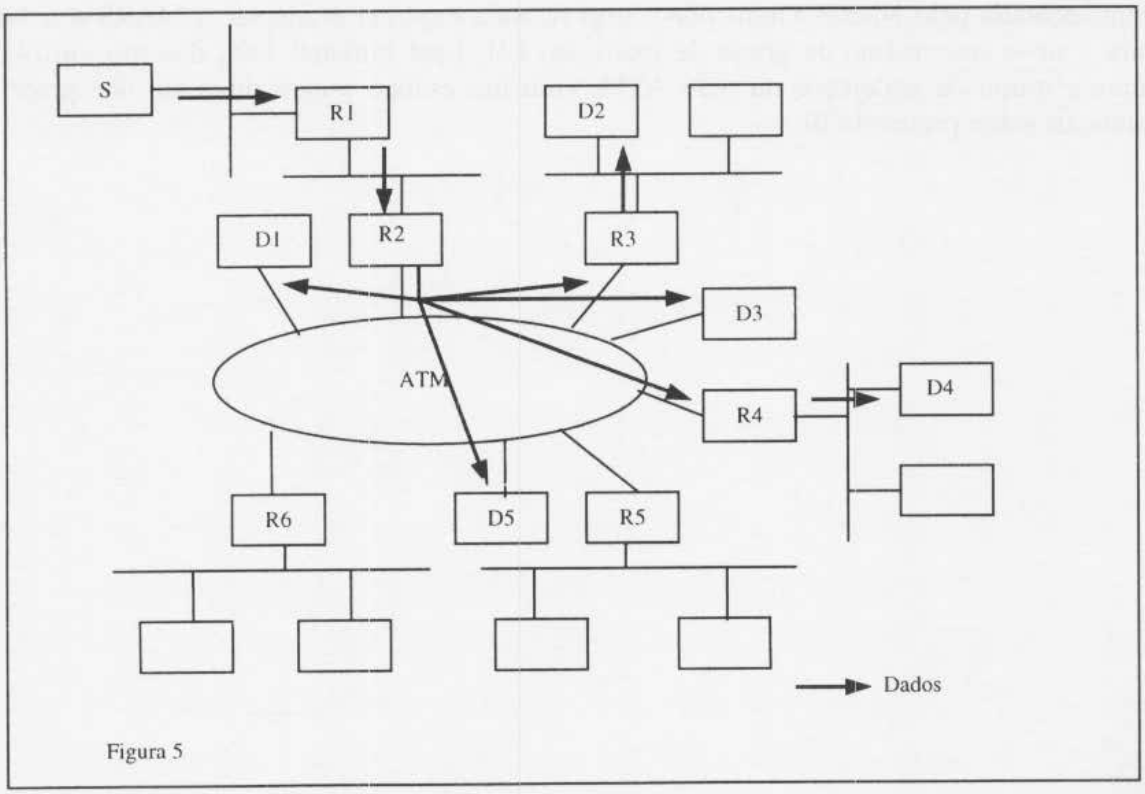


Figura 5

7. Conclusão

A proposta apresenta grandes ganhos para aplicações que gerem um grande volume de dados ou demandem muitos recursos por um período considerável de tempo. O ganho obtido com a proposta é dependente da topologia da rede: quanto maior o número de subredes na rede ATM, seja por motivos de administrativos ou de gerência, maiores os ganhos. Deve se considerar também que os problemas de escalabilidade não deverão afetar a maioria das redes, pois o problema ocorre a partir de um determinado número de membros de um grupo de multicast, que será limitado pelo número de membros que o nó fonte ATM poderá tratar, dependente de cada implementação.

A proposta elimina necessidade de reserva de recursos e overhead causado pela desmontagem e posterior remontagem dos pacotes nos roteadores de multicast usados para encaminhar pacotes entre nós de diferentes LISs, o que é exemplificado na topologia apresentada pelo roteador R5, que recebe os pacotes para sua LIS e os encaminha para o nó D5, como mostrado na figura 3. Na figura 5, utilizando-se a proposta, é mostrado que nenhuma reserva de recursos é feita no roteador R5, eliminando assim desperdício de recursos, sendo esse roteador somente usado para mensagens de controle que utiliza serviço best-effort. O mesmo acontece com o roteador de multicast R3 em relação à D3, mas nesse caso não há desperdício de recursos uma, vez que o roteador de multicast R3 possui membros fora da rede NBMA, tendo que fazer a reserva de recursos de qualquer maneira.

O número de circuitos virtuais abertos para transporte de dados na topologia proposta, como mostrado na figura 5, se limita a um VC, enquanto sem a utilização da proposta o número de VCs passa a 4, como mostrado na figura 3. Esse fato é relevante, pois na solução proposta o uso de um único VC que parte do próprio roteador egresso da fonte gera eficiência máxima no roteamento, pois elimina qualquer tipo de retardo no transporte dos dados causado por roteamento devido aos roteadores de multicast intermediários.

Um trabalho futuro interessante seria unir o modelo MARS ao NHRP, que já suporta ATMARP [6], em um mecanismo único, diminuindo assim o processamento, pois uniria a resolução de endereçamento multicast implementada no MARS com a resolução de próximo hop implementada pelo NHRP. Outro ponto importante a explorar é adaptar o MARS e o NHRP para o novo mecanismo de grupo de multicast, LIJ, Leaf Initiated Join, descrito em [4] que suporta grupo de endereços na rede ATM, com um escopo pouco diferente dos grupos de multicast sobre protocolo IP.

8. Bibliografia

[1] Shenker, S., Wroclawski, J., "Network Element Service Specification Template", RFC2216, Setembro 1997.

[2] Braden, R., Zhang, L., Berson, S., Herzog, S., and Jamin, S., "Resource Reservation Protocol (RSVP) – Version 1 Functional Specification", RFC2205, Setembro 1997.

[3] The ATM Forum, "ATM User-Network Interface (UNI) Specification 3.1", Setembro 1994.

[4] The ATM Forum, "ATM User-Network Interface (UNI) Signalling Specification 4.0", Julho 1996.

[5] Armitage, G., "Support for Multicast over UNI 3.0/3.1 based ATM Networks", RFC2022, Novembro 1996.

[6] Laubach, M., Halpern J., "Classical IP and ARP over ATM", Internet Draft, draft-ion-classic2-03.txt, obsoleta RFC1577 e RFC1626, Outubro 1997.

[7] Berson, S., Berger, L., "IP Integrated Services over RSVP over ATM", Internet Draft, draft-ietf-issll-atm-support-03.ps, Março 1993.

[8] Luciani, J., Katz,D., Pisticello, D., Cole, B., "NBMA Next Hop Resolution Protocol (NHRP)", Internet Draft, draft-ietf-rolc-nhrp-12.txt, Setembro 1997.

[9] Smirnov, M., "EARTH – Easy IP Multicast Routing Through ATM Clounds", Internet Draft, draft-smirnov-ion-earth-02.txt, Março 1997.

[10] Anker, T., Breitgand, D., Dolev, D., Levy, Z., "IMSS: IP Multicast Atalho Service", Internet Draft, draft-anker-congress-00.txt, Julho 1997.

[11] Deering, S., "Host Extensions for IP Multicast", RFC1112, Agosto 1989.

[12] Shenker, S., Wroclawski, J., "General Characterization Parameters for Integrated Service Network Elements", RFC2215, Setembro 1997.

[13] Wroclawski, J., "The use of RSVP with IETF Integrated Services", RFC2210, Setembro 1997.

[14] Maufer, T., Semeria, C., "Introduction to IP Multicast Routing", Internet Draft, draft-ietf-mboned-intro-multicast-02.txt, Março 1997.