

Roteando redes IPv6 e Multicast com MP-BGP

Andrey Vedana Andreoli, Leandro Márcio Bertholdo, Liane Tarouco

Ponto de Presença da RNP no Rio Grande do Sul – POP-RS
RSiX – Ponto de Troca de Tráfego do Rio Grande do Sul
Centro de Processamento de Dados da Universidade Federal do Rio Grande do Sul
Rua Ramiro Barcelos, 2574 – Porto Alegre – RS – Brasil
{andrey,bertholdo,liane}@penta.ufrgs.br

Resumo. Este artigo apresenta as extensões feitas no protocolo BGP que formam o Multiprotocol BGP, tendo o intuito de fornecer suporte para o roteamento global da Internet para novas tecnologias como Multicast e IPv6. São apresentados detalhes sobre as mudanças no funcionamento e operação do BGP para o novo MP-BGP. Ao final, são dados exemplos de utilização e uma visão geral de sua utilização no Brasil.

Palavras chave: Roteamento, BGP, Multiprotocol BGP, MP-BGP, IPv6, Multicast.

1. Informações Gerais

Com o grande crescimento da Internet desde seus primórdios, o modelo conhecido como hierárquico, constituído por uma rede centralizada que conectava redes secundárias, foi substituído por um modelo distribuído. Para tanto, adotou-se o conceito de sistemas autônomos, também chamados de ASes. Isso torna a Internet um conjunto de “células” conectadas entre si, chamadas de ASS. Essa formação garante que todos os ASs sejam alcançáveis entre si, visto que estarão conectados direta ou indiretamente. Outro fato importante é que o que ocorre internamente a um AS não é conhecido pelos demais ASes, diminuindo a complexidade da Internet global. O protocolo que iniciou a utilização desse conceito e permanece até os dias de hoje é o BGP (Border Gateway Protocol) [RFC 1771].

O funcionamento básico do BGP é baseado no estabelecimento de sessões entre dois roteadores, utilizando para isso o protocolo TCP, sob a porta 179. De forma muito freqüente o termo “peers” é atribuído a roteadores que estabelecem entre si uma sessão BGP e o termo “peering” é dado à troca de tráfego entre dois sistemas autônomos.

O protocolo BGP tem sido utilizado basicamente no roteamento IPv4, mas com o advento do IPv6, Multicast, MPLS e outras tecnologias, surgiu a demanda para que fossem feitas modificações afim de suportar essa nova gama de protocolos. Sabendo que o funcionamento do BGP tem se consolidado e que o funcionamento da Internet depende dele, surgiu uma extensão do protocolo BGP chamada de Multiprotocol BGP, tendo como principal objetivo fornecer suporte às novas tecnologias, mantendo a compatibilidade com o BGP atual.

O Multiprotocol BGP, erroneamente citado como Multicast BGP [HALABI 2000], tornou-se padrão em Fevereiro de 1998 [RFC 2858], atualmente já sendo implementado por fabricantes como Cisco, Juniper, entre outros. Como o padrão do protocolo não determina nenhuma abreviação específica para identifica-lo, encontram-se na literatura os termos MBGP e/ou MP-BGP para identificar o Multiprotocol BGP.

2. Descrição do MP-BGP

O protocolo MP-BGP apresenta-se como um complemento do protocolo BGP. Todo o seu funcionamento, políticas e atributos são mantidos. O que de fato muda é que na mesma sessão BGP podem ser trocados anúncios de múltiplos protocolos. A figura 1 apresenta o formato de uma mensagem tipo UPDATE convencional do BGP.

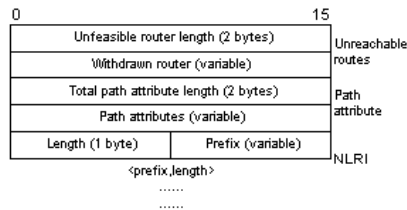


Figura 1 – Mensagem UPDATE do BGP

Conforme a figura acima, a mensagem de UPDATE é formada basicamente por 3 partes: rotas que devem ser removidas, atributos de rotas alcançáveis que estão sendo divulgadas e os prefixos que estão sendo anunciados. A mudança em si para o MP-BGP é feita nos atributos de rotas (Path attribute). Normalmente, os atributos utilizados são: AS_PATH, Origin, Next_Hop, entre outros. São apresentados então os novos atributos para suporte ao MP-BGP, listados abaixo:

- MP_REACH_NLRI: Utilizado para anúncios de redes alcançáveis a serem incluídas na tabela BGP;
- MP_UNREACH_NLRI: Utilizado para redes que devem ser retirados da tabela BGP;

Os campos pertencentes a cada um dos novos atributos são apresentados na figuras 2:

Address Family Identifier - AFI (2)
Subsequent Address Family Identifier – Sub-AFI (1)
Length of the Next-Hop Address (16 or 32)
Network Address of Netx-Hop (global and/or Link Local)
Number of SNPA's and Lenght (may be zero)
NLRI (Length and Prefix)

Address Family Identifier - AFI (2)
Subsequent Address Family Identifier – Sub-AFI (1)
WithDrawn Routes (Length and Prefix)

Figura 2 - Campos dos atributos MP_REACH_NLRI e MP_UNREACH_NLRI, respectivamente.

Como pode-se conferir na figura 2, surge o conceito de “Address Family Information” (AFI) e Subsequent Address Family Identifier (Sub-AFI). Estes campos são responsáveis por identificar os diversos protocolos em que o MP-BGP suporta.

Como exemplo mais prático, a figura 3 apresenta uma mensagem de UPDATE com o atributo MP_REACH_NLRI, onde são anunciados 7 prefixos de redes IPv6 Unicast:

```

❑ UPDATE Message
  Marker: 16 bytes
  Length: 139 bytes
  Type: UPDATE Message (2)
  Unfeasible routes length: 0 bytes
  Total path attribute length: 116 bytes
❑ Path attributes
  ❑ ORIGIN: IGP (4 bytes)
  ❑ AS_PATH: 65502 65504 (9 bytes)
  ❑ MP_REACH_NLRI (103 bytes)
    ❑ Flags: 0x80 (Optional, Non-transitive, Complete)
    Type code: MP_REACH_NLRI (14)
    Length: 100 bytes
    Address family: IPv6 (2)
    Subsequent address family identifier: unicast (1)
  ❑ Next hop network address (32 bytes)
    Subnetwork points of attachment: 0
  ❑ Network layer reachability information (63 bytes)
    ❑ 3ffe:2620:14:1::/64
      MP Reach NLRI prefix length: 64
      MP Reach NLRI prefix: 3ffe:2620:14:1::
    ❑ 3ffe:2620:14:2::/64
    ❑ 3ffe:2620:14:3::/64
    ❑ 3ffe:2620:14:4::/64
    ❑ 3ffe:2620:14:5::/64
    ❑ 3ffe:2620:14:100::/64
    ❑ 3ffe:2620:14:666::/64

```

Figura 3 - Mensagem de UPDATE com atributo MP_REACH_NLRI

Uma vez que a sessão MP-BGP entre dois peers seja estabelecida, são mantidas diferentes tabelas de roteamento devido aos diferentes protocolos que podem ser tratados [NANOG 2003]. As duas tabelas são: Unicast Routing Information Base (U-RIB) e Multicast Routing Information Base (M-RIB);

Algumas das tecnologias suportadas pelo MP-BGP, que são as mais utilizadas atualmente são descritas de forma mais detalhada abaixo:

a) Multicast

Conhecida também como comunicação de grupo, é uma tecnologia muito utilizada principalmente em transmissões de um único sentido destinadas a um grupo de hosts distribuídos em diferentes redes. Seu objetivo principal é otimizar o transporte de pacotes, evitando que diversas cópias de um mesmo fluxo de transmissão sejam transmitidas, economizando banda e recursos, além de garantir que o fluxo de dados seja entregue para todos que desejarem recebe-lo. É formado por um conjunto de protocolos de nível 2 e 3. Utilizado amplamente no Brasil pela Rede Nacional de Pesquisa na transmissão de eventos e cursos à distância. O MP-BGP interage com o Multicast e permite que anúncios multicast sejam feitos para seus peers [CISCO 2003a].

b) MPLS VPN

Sabendo que o funcionamento básico do protocolo MPLS é baseado em “labels” - onde cada roteador de uma rede MPLS provê um tratamento diferenciado de acordo com o label - torna-se possível estabelecer uma VPN (Virtual Private Network) [RFC 2547]. Alguns dos recursos que o MPLS fornece que podem ser úteis para uma VPN são: os níveis de garantia de QoS, latência e jitter, além de permitir o re-roteamento inteligente se necessário, que trata-se de um recurso que permite que o MPLS faça automaticamente o remanejo de um caminho, caso este não esteja mais atendendo aos recursos mínimos estabelecidos para seu funcionamento. Adicionalmente, o MPLS provê a devida segurança, visto que todos os pacotes estarão enquadrados sob a mesma classe de equivalência, tendo o mesmo tratamento e fazendo o mesmo caminho.

Como suporte, o MP-BGP permite que os anúncios de rotas sejam distribuídos entre peers na forma de <VPN-prefixo>.

c) IPv6

Com a crescente utilização e crescimento da Internet, o endereçamento IPv4 tem se tornando escasso. O que reforça essa tendência é a intenção de empregar endereçamento IP em aparelhos móveis, como celulares e computadores portáteis, aumentando ainda mais a necessidade de uma faixa de endereços maior. Diante desse problema, foi concebido o IPv6, passando dos 32 bits existentes no IPv4 para 128 bits, propondo também um conjunto de melhoras para limitações encontradas no IPv4. Com isso, o MP-BGP permite que prefixos IPv6 sejam anunciados, da mesma forma que prefixos IPv4 já são anunciados com o BGP.

d) 6PE

Também conhecida como IPv6 Provider Edge Router, o 6PE trata-se de uma funcionalidade presente em roteadores Cisco que tem por objetivo fornecer suporte a IPv6 sem a necessidade de que um backbone baseado em MPLS e IPv4 seja modificado. Para tanto, ele baseia-se na propagação de prefixos IPv6 e seus labels MPLS utilizando o BGP sobre IPv4, tendo seu atributo Next-hop identificado por um endereço IPv4.

Para a operação de uma rede com 6PE, os roteadores de distribuição que conectam sites/clientes IPv6 devem implementar o 6PE, sendo então aptos a anunciarem prefixos da família v6, juntamente com seu Next-hop IPv4, que é mapeado como “IPv4-mapped IPv6 address”, já que o Next-hop deve ser da mesma família que o prefixo anunciado nas mensagens da UPDATE do MP-BGP. O backbone em si não sofre nenhuma alteração, já que irá rotear os pacotes 6PE baseados apenas em seu label, até o próximo roteador 6PE.

Esta abordagem é muito semelhante a técnicas de tunelamento como é o caso do Generic Routing Encapsulation (GRE). Mesmo assim, a técnica de 6PE é muito utilizada em backbones europeus [6NET 2003], apresentando como principal benefício o menor overhead, além de não exigir que o backbone MPLS sofra modificações.

3. Negociação das chamadas capabilities

Outra característica do MP-BGP é a negociação de capabilities entre peers. As chamadas capabilities são features que podem ser muito úteis na operação de peerings BGP. Algumas delas já são conhecidas e utilizadas anteriormente ao MP-BGP. Tais features são negociadas ainda no início da sessão BGP, através das mensagens de OPEN. Para a negociação ser bem sucedida, ambos peers devem reconhecer e suportar a feature em questão. Abaixo são apresentadas 3 tipos que podem ser negociados:

- Multiprotocol extension (AFI e SAFI)

No início da sessão BGP, são negociados quais os protocolos que serão suportados por cada peer, determinando assim qual a AFI e SAFI que ambos estarão trocando prefixos.

- Route refresh

A motivação para esta capacidade é que o BGP não possui um recurso dinâmico para requisitar um novo envio dos prefixos contidos na tabela de um peer. De acordo com [RFC 2918], isso é desejável quando a política no recebimento de prefixos é alterada. O máximo que é feito até então era fazer uso do recurso de “soft-reconfiguration” que mantinha em memória todos os anúncios recebidos de determinado peer, afim de saber o que foi recebido e re-aplicar nos anúncios a nova política, sem ter que derrubar a sessão BGP. O recurso de

route refresh permite que os prefixos sejam enviados novamente por um peer, sendo disparado por uma requisição de route refresh, especificando qual a AFI e Sub-AFI desejada. Um dos benefícios desse método é a economia de memória e CPU utilizada com o recurso de soft-reconfiguration.

- Outbound route filtering

O processo de geração e processamento de prefixos entre peers BGP é uma operação que consome recursos valiosos. Em alguns casos, enquanto um dos peers - geralmente o upstream - gera os prefixos, o outro peer que recebe aplica sua prefix-list na entrada, afim de receber apenas os prefixos desejados. Em situações como esta o recurso de outbound route filtering permite que o peer que irá fazer os anúncios aplique como prefix-list de saída a mesma prefix-list que é utilizado na entrada do peer que irá receber os anúncios [CISCO 2003b]. Isso pode reduzir de forma significativa a troca de mensagens de UPDATE, além de poupar recursos valiosos de ambos.

4. Exemplos de sessão MP-BGP

Para demonstrar a configuração básica deste protocolo, foi criada uma topologia simples, onde dois roteadores estabelecem uma sessão MP-BGP, ilustrada pela figura 4:

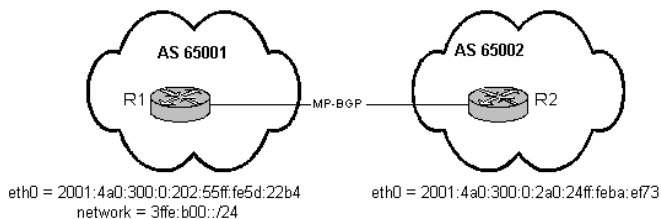


Figura 4 - Exemplo de sessão MP-BGP

Abaixo, na figura 5, é apresentada a configuração aplicada no roteador R1.

```
router bgp 60001
bgp router-id 10.0.0.1
no bgp default ipv4-unicast
neighbor 2001:4a0:300:0:2a0:24ff:feba:ef73 remote-as 60002
neighbor 2001:4a0:300:0:2a0:24ff:feba:ef73 description connection to R2
address-family ipv6 unicast
network 3ffe:b000::/24
neighbor 2001:4a0:300:0:2a0:24ff:feba:ef73 activate
neighbor 2001:4a0:300:0:2a0:24ff:feba:ef73 soft-reconfiguration inbound
exit-address-family
```

Figura 5 - Configuração MP-BGP em R1

Depois de ambos roteadores estarem configurados, pode-se verificar o status da sessão MP-BGP pela ilustração na figura 6. Verifica-se então que a sessão está estabelecida (UP) e que 1 prefixo foi recebido de R1.

```

R2> sh bgp ipv6 sum
BGP router identifier 10.0.0.2, local AS number 60002
Neighbor          V    AS  MsgRcvd MsgSent  TblVer  InQ OutQ Up/Down  State/PfxRcd
2001:4a0:300:0:202:55ff:fe5d:22b4 4 60001   24317   24696    0    0    0 14:42:25    1
Total number of neighbors 1

```

Figura 6 - Exemplo de verificação de sessão MP-BGP em R2

5. Visão geral da utilização do roteamento IPv6 e Multicast

Por serem tecnologias novas, backbones comerciais, em sua maioria, ainda não tem identificado o nicho de mercado para justificar tais inovações sobre o IPv4. Essa mudança deverá ser gradual, conforme aumentar a demanda por tais tecnologias. Já no meio acadêmico, tais tecnologias já têm sido amplamente empregadas, justificando-se pelo próprio caráter de pesquisa e inovação que tais instituições apresentam. No Brasil, assim como na Internet2, o MP-BGP tem sido utilizado pela Rede Nacional de Ensino e Pesquisa em caráter de produção, mais precisamente na conexão internacional afim de fazer peerings IPv6 e Multicast. Internamente ao sistema autônomo da RNP, é utilizado o protocolo PIM-SM para distribuição do tráfego Multicast e RIP6 para o roteamento entre os Pontos de Presença Estaduais que participam do projeto piloto de IPv6 [RNP 2003]. A escolha pelo RIP6 foi feita baseada na situação da época onde outros protocolos IGP, como era o caso do OSPF, não estavam inteiramente estáveis no suporte a essa tecnologia.

6. Referências bibliográficas

- [HALABI 2000] Sam Halabi, Danny McPerson. Internet Routing Architectures, Second Edition. Indianapolis – USA : Cisco Press, 2000
- [RFC 1771] A Border Gateway Protocol 4 (BGP-4) – RFC 1771 – The Internet Engineering Task Force – IETF
- [RFC 2858] Multiprotocol Extensions for BGP-4 – RFC 2858 - The Internet Engineering Task Force – IETF
- [RFC 2918] Route Refresh Capability for BGP-4 – RFC 2918 – The Internet Engineering Task Force – IETF
- [RFC 2547] BGP/MPLS VPNs – RFC 2547 – The Internet Engineering Task Force - IETF
- [NANOG 2003] Meeting of NANOG – on line – 2003 - www.nanog.org/mtg-0306/pdf/mcbride.ppt
- [CISCO 2003a] Cisco Enterprise : Multicast – on line – 2003 - <http://www.cisco.com/warp/public/779/largeent/learn/technologies/multicast.html>
- [6NET 2003] Initial Ipv4 do Ipv6 migration Cookbook for organizational/ISP (NREN) and backbone networks. on line – 2003 - www.6net.org/publications/deliverables/D2.2.2.pdf
- [CISCO 2003b] BGP Prefix-Based Outbound Route Filtering – on line – 2003 - http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120limit/120s/120s22/s_bpborf.htm#xtocid0
- [RNP 2003] Site da Rede Nacional de Pesquisa – on line – 2003 – <http://www.rnp.br>