



BELLA

Building the Europe Link to Latin America

Roteamento para Internet: Uma abordagem de dentro da rede para fora

Setembro/2019

Instrutores

Bruno Ramos <brunoramos@pop-ba.rnp.br>

Ibirisol Fontes <ibirisol@pop-ba.rnp.br>

Thiago Bomfim <thiagobomfim@pop-ba.rnp.br>

O Curso

- ✓ Preparar o aluno para projetar e implantar um esquema de roteamento para redes de diversos tamanhos, interligando redes a outras redes, sob administração própria ou de terceiros
- ✓ O programa do curso abrange o protocolo OSPF para roteamento interno e o protocolo BGP para roteamento externo
- ✓ Atividades práticas utilizando emulador de redes, open-source, CORE

Sumário

Dia 1, manhã - fundamentos

- Introdução
- Resumo TCP/IP e OSI
- Endereçamento IPv4 e IPv6
 - Modelo de Classes
 - CIDR
 - VLSM
- Noções Roteamento
 - IGP x EGP
 - RIB e FIB
 - Roteamento IP na Internet

Dia 1, tarde - OSPF

- Protocolo OSPF
- Conceito de Estado do Enlace (Link State)
- Algoritmo Shortest Path First (SPF)- Dijkstra
- Funcionamento e configuração do protocolo OSPF
- Roteadores de borda e de área
- Área única e Multi-área

Dia 2 - BGP

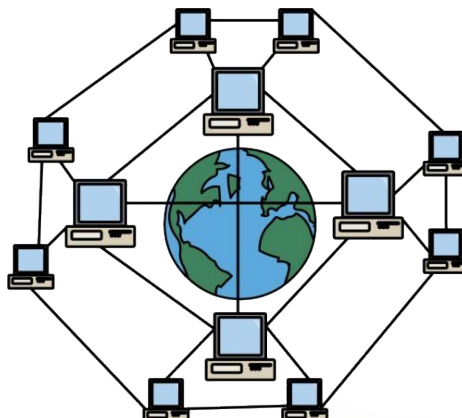
- Como a Internet funciona
- Sistemas Autônomo
- BGP na Internet
- Protocolo BGP
- Roteamento IP com BGP
- Configuração do BGP
- Um pouco mais de BGP
 - Multihoming
 - Políticas de roteamento
 - Mapas de rotas
 - Route servers
- Pontos de troca de tráfego

Como a Internet funciona

“Conjunto de computadores autônomos interconectados por uma única tecnologia” Tanenbaum, 2003

“É uma rede de computadores mundial, isto é, uma rede que interconecta milhões de equipamentos de computação em todo o mundo.” Kurose, 2006

“Redes de pessoas conectadas por computadores.” Conrado Adolpho - Google Marketing



Como a Internet funciona



Como a Internet funciona



Como a Internet funciona

✓ Autoridades da Internet

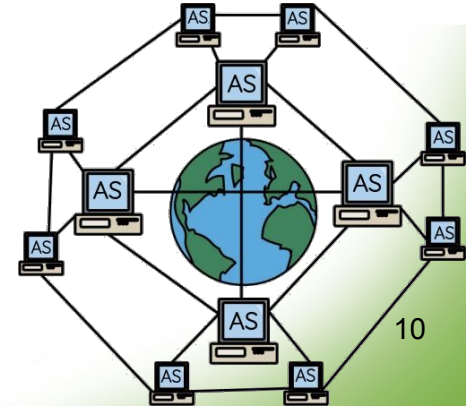
- ✓ IANA (Internet Assigned Numbers Authority)
 - ✓ ARIN (América do Norte)
 - ✓ Estados Unidos, Canadá, algumas ilhas do Caribe
 - ✓ AFRINIC (África)
- ✓ RIPE NCC (Europa/Parte da Ásia/Oriente Médio)
 - ✓ Europa, Parte da Ásia
- ✓ LACNIC (México, América Central e Sul)
 - ✓ NIC México
 - ✓ **NIC.br (BRASIL)**
- ✓ AAPNIC (Ásia/Pacífico)
 - ✓ APJII - Indonésia
 - ✓ CNNIC - China
 - ✓ JPNIC - Japão
 - ✓ KRNIC - Coreia
 - ✓ VNNIC - Vietnã
 - ✓ TWNIC - Taiwan

Sistemas autônomos

- ✓ Sistema Autônomo (AS) é grupo de redes IP sob uma única gerência técnica e que possui uma mesma política de roteamento
 - ✓ Sistema:
 - ✓ Estrutura mínima possível de rede, atualmente medida pela necessidade de endereços IPv4 ou IPv6.
 - ✓ Autônomo:
 - ✓ Entidade com autonomia ou poder de decisão sobre como será realizado o encaminhamento de seu tráfego para caminhos externos (outros ASes).

Sistemas autônomos

- ✓ Todo AS possui um número identificador (ASN)
 - ✓ Identificação única e global ampliada de 16 para 32 bits (de 0 a 4.294.967.295)
 - ✓ ASes reservados:
 - ✓ Uso privado: ASN 64512 até 65535 e 4.200.000.000 a 4.294.967.294
 - ✓ Documentação: ASN 64496 a 64511 e 65536 até 65551
- ✓ Atualmente cerca de 65 mil ASNs distintos
 - ✓ <http://www.cidr-report.org/as2.0/>
- ✓ No Brasil são cerca de 6541 ASNs registrados
 - ✓ <https://ix.br/localidades/brasmap>



Sistemas autônomos

- ✓ Espaço de endereçamento IP próprio

- ✓ Recebe-se uma faixa de IPv4 (/22) e IPv6 (/32) do sistema de registro de IPs nacional (Registro.br): <https://registro.br/tecnologia/numeracao/regras-e-politicas/>

- ✓ Permite alocação de endereços IP válidos diretamente a máquinas ou clientes
 - ✓ Melhora a utilização de algumas aplicações e clientes (segurança), etc.
 - ✓ Facilita o processo de rastreabilidade e endereçamento IP Portável (PI)

- ✓ Independência de provedor
 - ✓ Troca do provedor de acesso a Internet é mais simples
 - ✓ Não envolve mudanças de configuração interna
 - ✓ Maiores possibilidade de negociação de preço de trânsito IP

Sistemas autônomos

- ✓ Redundância
 - ✓ Possibilita a implementação de redundância do acesso Internet pela conexão com 2 ou mais provedores
 - ✓ Aumento da disponibilidade dos serviços prestados

- ✓ Acordos de troca de tráfego
 - ✓ Possibilita a conexão da entidade com pontos de troca de tráfego (e.g. IX.br)
 - ✓ Estabelecimento de acordos multilaterais e bilaterais
 - ✓ Economia de recursos com contratação de banda para internet
 - ✓ Melhor qualidade de interconexão

Sistemas autônomos

- ✓ Quando se tornar um AS?
 - Comprovar necessidade de IP (IPv4 esgotados!)
 - Duas ou mais saídas de acesso a internet
 - Equipe técnica capacitada para operar o roteamento da instituição via BGP
 - Recursos de hardware e software suficientes para utilizar o protocolo BGP
 - Condições financeiras para implantação e operação da nova estrutura

Sistemas autônomos

- ✓ Como se tornar um AS?
 - Preencher formulário abaixo e enviar para numeracao-pedido@registro.br
<https://registro.br/tecnologia/pedido-form.txt>
 - Solicitar ASN, IPv4 e IPv6 todos no mesmo formulário
 - O prazo médio para análise é de 2 semanas
 - Maiores informações gerais sobre custo, benefícios, como tirar, etc:
<https://registro.br/tecnologia/provedor-acesso.html>

BGP na Internet

BGP na Internet

- ✓ Principal protocolo de roteamento na Internet
- ✓ Provê a cada AS os meios para:
 - Obter informações de alcance de subrede dos ASes vizinhos
 - Propagar informações de alcance para todos os roteadores internos ao AS
 - Determinar “boas” rotas para as subredes baseado em informações de alcance e política
 - Permite que uma subrede comunique sua existência para o resto da internet (“Estou aqui!”)

BGP na Internet

- ✓ O fluxo do tráfego que entra e sai do AS deve ser controlado (engenharia de tráfego)
 - ✓ As informações sobre as rotas da internet encontram-se na tabela BGP:
 - Tabela Global IPv4: ~511.000 entradas (CIDR REPORT Aug/14)
 - Tabela Global IPv6: ~18.000 entradas (CIDR REPORT Aug/14)
- ↓
- Tabela Global IPv4: ~794.000 entradas (CIDR REPORT Set/19)
 - Tabela Global IPv6: ~74.000 entradas (CIDR REPORT Set/19)

BGP na Internet

- ✓ Quando não usar BGP?
 - ✓ Se você não tem / é um sistema autônomo
 - ✓ Tiver uma única conexão com a Internet ou outro AS
 - ✓ Não há preocupação / necessidade com políticas e seleção de rotas
 - ✓ Mesma política de roteamento utilizada pelo seu ISP
 - ✓ Bloco de rede pertencente ao ISP, ao invés de bloco próprio (sem AS)
 - ✓ Equipamento incompatível, hardware e/ou software, com o protocolo
 - ✓ Não suporta BGP (software) ou
 - ✓ Reduzida capacidade de memória e processamento (full-routing)

Protocolo BGP

- ✓ O BGP é um protocolo de roteamento dinâmico baseado em distância (vetor distância), variação denominada Vetor de Caminhos (Path Vector)
 - ✓ depende que os vizinhos anteriores passem suas rotas das tabelas de roteamento para roteadores a frente
- ✓ O BGP faz seus cálculos de rota baseado nas rotas anunciadas recebidas
 - ✓ Com base no resultado dos cálculos, passa informação para restante dos vizinhos
- ✓ Apesar de ser baseado no vetor distância, não quantifica a distância como apenas um número
 - ✓ Usa lista de ASNs que indicam o caminho que o pacote deve passar para alcançar seu destino (Atributo chamado de AS_PATH)
 - Quanto menor a lista de ASNs, maior a preferência a rota BGP terá

Protocolo BGP

- ✓ AS_PATH é o principal, mas é somente um dos atributos associados às rotas BGP
 - ✓ Pela flexibilidade, BGP é conhecido como protocolo baseado em políticas

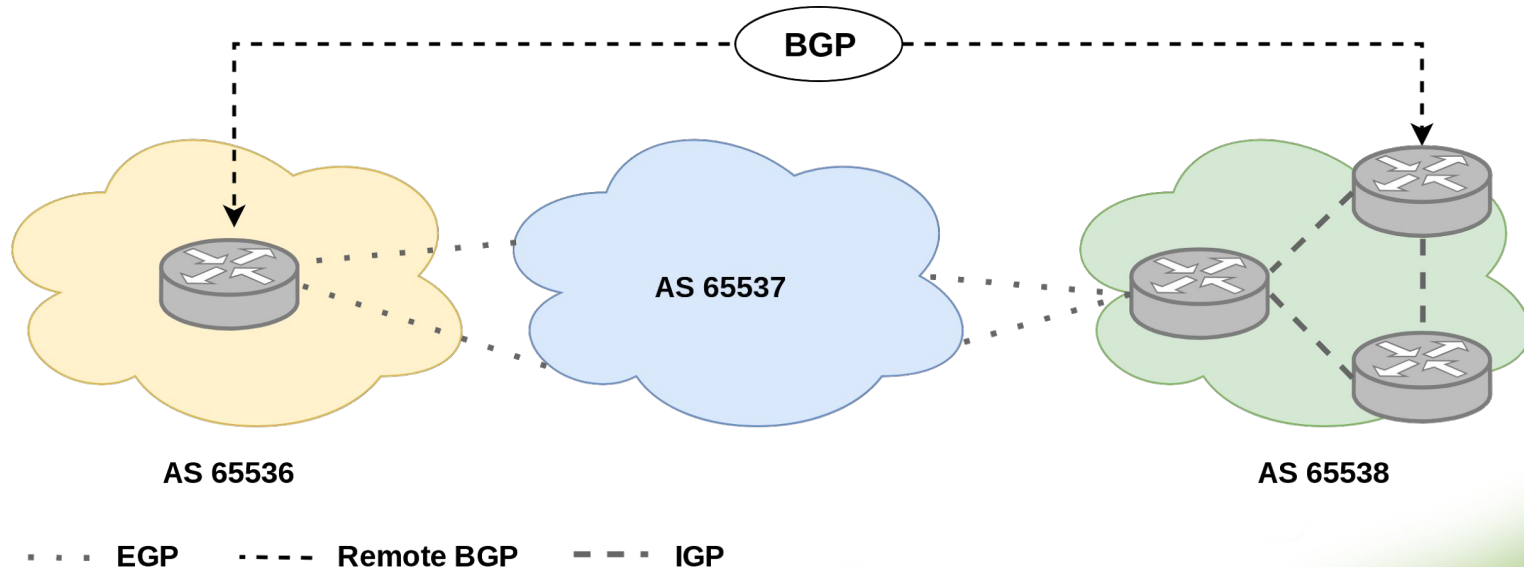
- ✓ A comunicação entre os vizinhos BGP se dá através da porta 179/TCP
 - ✓ Uma conexão individual única de um roteador para cada um de seus vizinhos
 - ✓ Ao se utilizar o TCP:
 - Maior garantia na comunicação entre vizinhos
 - Diminuição da complexidade de implementação do protocolo
 - Ack/retransmissão/sequenciamento de pacotes BGP é feito pelo TCP

Protocolo BGP

- ✓ Vizinhos/Pares BGP:
 - ✓ Neste curso vamos considerar Vizinhos e Pares (Peer) como a mesma coisa!
 - ✓ BGP não possui descoberta de vizinhos
 - Vizinhança/Peering BGP deve ser configurada manualmente
 - Após subir vizinhança:
 - Por padrão, redes anunciadas são divulgadas automaticamente para todos os vizinhos

Protocolo BGP

- ✓ Sessão BGP
 - ✓ Quando dois roteadores estabelecem vizinhança BGP entre si com sucesso
 - ✓ As sessões BGP não necessariamente correspondem aos enlaces (links físicos)

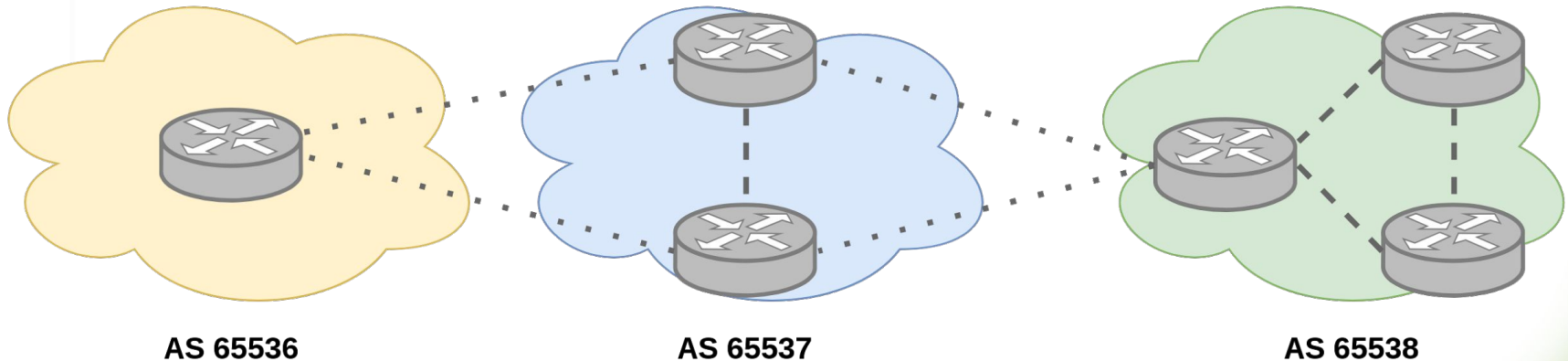


Protocolo BGP

- ✓ Sessões BGP trocam Network Layer Reachability Information (NLRI)
 - ✓ Informações de Alcançabilidade da Camada de Rede
 - ✓ Se referem a um ou mais Prefixo + Máscara de bits
 - Exemplo de NLRI: prefixo 204.149.16.128 e máscara /25

Protocolo BGP

- ✓ Anúncio de prefixos IP através de sessão BGP:
 - ✓ Quando o AS 65537 anuncia um prefixo do AS 65536, ele está prometendo que encaminhará todos os datagramas destinados a esse prefixo em direção ao roteador que originou o prefixo dentro do AS 65536:



..... EGP

--- IGP

Protocolo BGP

Características	OSPF	BGP
Funcionamento intra ou inter-AS?	Intra-AS	iBGP (intra-AS); eBGP (inter-AS)
Algoritmo	Menor caminho (SPF)	Menor caminho (menor AS-PATH)
Método	Estado de link	Vetor distância Avançado (Path Vector)
Possui detalhes da topologia de cada Área/AS?	Sim, convergência depende de topologia	Não
Métrica para cálculo de melhor rota	Custo da interface (velocidade)	Diversas, a principal o AS-PATH
Tabela de rotas	Somente rotas instaladas pelo SPF	Todas as rotas recebidas e anunciadas
Modelo de hierarquia	Em Áreas, sendo 0 a de backbone	Sem hierarquia
Método de atualizações de rotas	Pacotes IP Multicast	Pacotes IP Unicast via TCP porta 179
Possibilita vizinhos não pertencentes ao mesmo segmento?	Não	Sim*
Vizinhança criada automaticamente?	Sim	Não
Ao estabelecer vizinhança, redes divulgadas automaticamente?	Sim	Sim
Modo de recálculo de rotas em caso de atualização:	Recálculo completo (todas as rotas)	Cálculo somente do que mudou
Prevenção contra loops	Ponto-a-ponto lógico por todas as áreas se conectarem diretamente à Área 0	Comportamento de Split Horizon (iBGP) Atributo AS_PATH (eBGP)

* Solução com iBGP (Route Reflector) e eBGP-Multihop.

Protocolo BGP

- ✓ Processo de Peering:
 - ✓ Inicialmente, um roteador BGP deve reconhecer e se possível autenticar o seu peer
 - ✓ Os dois peers estabelecem uma conexão TCP
 - ✓ Cada peer envia uma informação de alcançabilidade positiva ou negativa
 - ✓ Ocorre uma divulgação completa das rotas ativas e inativas de cada um
 - ✓ A troca de mensagens continua para confirmação do recebimento das rotas e também da conexão entre os roteadores
 - ✓ Atualizações de anúncios são feitos individualmente, à medida que mudanças ocorrem na rede

Protocolo BGP

- ✓ BGP Finite State Machine (FSM)
 - ✓ A máquina de estados do BGP controla a conectividade com seu peers
 - ✓ É instanciada uma máquina para cada conexão TCP iniciada
 - Há a possibilidade de ter mais de uma máquina de estados para um mesmo peer (múltiplas sessões configuradas)
 - ✓ Ao receber diversos sinais as variáveis de controle de estado são mudadas
 - Administrativos, Temporizadores (Timers), TCP connection (three-way handshake), Mensagens BGP (Recebendo ou enviando)

Protocolo BGP

✓ Estados das sessões BGP:

✓ **Idle**

- Estado inicial, não há conectividade com o vizinho
- São feitas as inicializações dos recursos para manter uma sessão
- Tentativas de conexões BGP para os pares configurados não são aceitas

✓ **Connect**

- A máquina de estados do BGP tenta estabelecer, ao passo que espera, uma conexão TCP para firmar a sessão com o peer

Protocolo BGP

✓ Estados das sessões BGP:

✓ **Active**

- Nesse estado, de forma passiva, se espera conexão TCP do peer. Comandos de inicialização são descartados

✓ **OpenSent**

- A conexão TCP existe e este roteador local está tentando criar uma conexão com o dispositivo vizinho
- Nesse estado a FSM espera por uma mensagem OPEN oriunda de seu peer
- Comandos de inicialização são descartados

Protocolo BGP

✓ Estados das sessões BGP:

✓ **OpenConfirm**

- A conexão TCP existe e o roteador local recebeu uma mensagem de volta do roteador vizinho
- Nesse estado o BGP espera uma mensagem de KEEPALIVE ou NOTIFICATION
- Comando de inicialização são ignorados

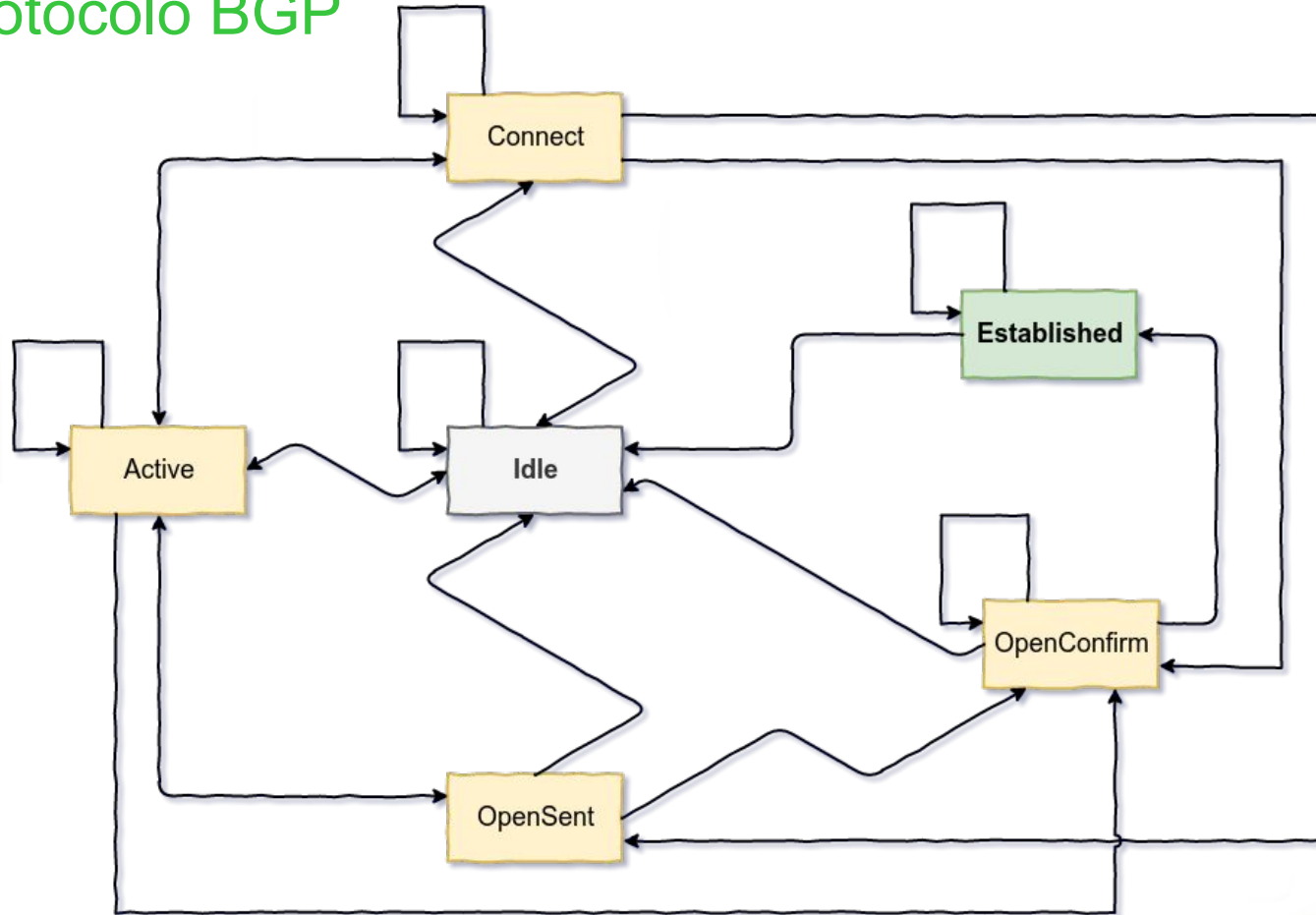
Protocolo BGP

✓ Estados das sessões BGP:

✓ **Established**

- Ambos os roteadores concordaram em formar uma relação e eles conseguem trocar mensagens BGP entre si
- A FSM do BGP pode trocar mensagem de UPDATE, NOTIFICATION e KEEPALIVE com seu peer
- Comandos de inicialização são ignorados

Protocolo BGP



Protocolo BGP

- ✓ Formato das mensagens BGP
 - ✓ Comprimento variável
 - ✓ Compostas no mínimo de um cabeçalho, sendo opcional a parte de dados
 - ✓ Mensagens do BGP
 - OPEN, NOTIFICATION, KEEPALIVE, UPDATE (RFC4271)
 - ROUTE-REFRESH (RFC2918)
 - A extensão MP-BGP (RFC4760, RFC7606)

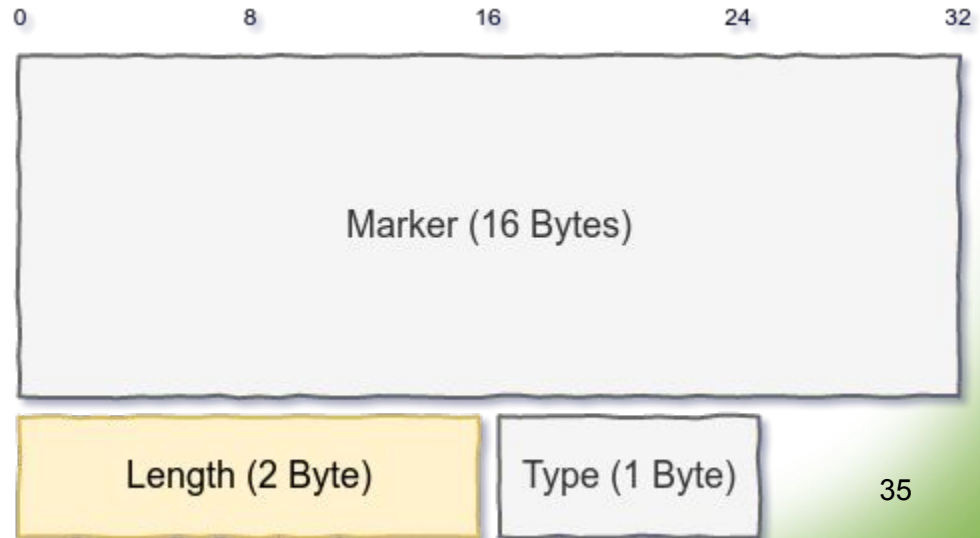
Protocolo BGP

- ✓ Campo marcador (Maker)
 - ✓ Tem objetivo de verificar se há algo errado com a mensagem recebida e se ocorreu perda de sincronização com o vizinhos BGP
 - ✓ Geralmente preenchida com “1”



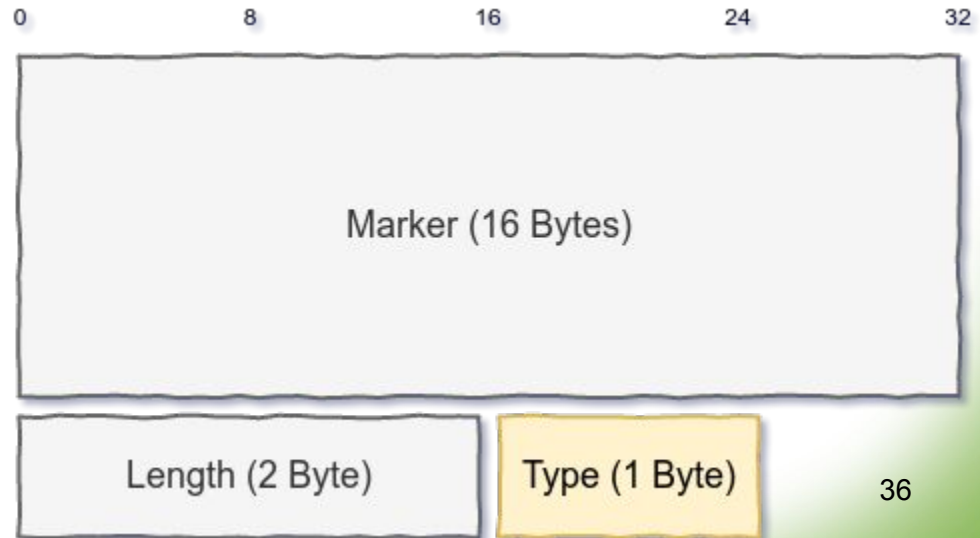
Protocolo BGP

- ✓ Campo Tamanho (Length)
 - ✓ Indica o tamanho total da mensagem
 - Mínimo de 19 bytes
 - Máximo de 4096 bytes



Protocolo BGP

- ✓ Campo Tipo (Type)
 - ✓ Contém um número representando o código de um tipo de mensagem
 - OPEN (1)
 - NOTIFICATION (2)
 - KEEPALIVE (3)
 - UPDATE (4)
 - ROUTE-REFRESH (5)



Protocolo BGP

✓ OPEN

- ✓ Enviada para iniciar a abertura da sessão BGP entre neighbors ou peers BGP
 - **Versão:** Número de versão do BGP sendo usado na origem. O valor mais alto suportado pelos roteadores da sessão será utilizado (geralmente 4)
 - **Número do AS:** Contém o número do AS ao qual o roteador pertence, informação usada pelo peer para determinar se trata-se de uma sessão eBGP ou iBGP. A sessão pode ser encerrada se não for o número do AS esperado
 - **Tempo de espera:** Valor, em segundos, do maior tempo de espera permitido entre mensagens do tipo UPDATE ou KEEPALIVE. O tempo de espera padrão é de 180 segundos

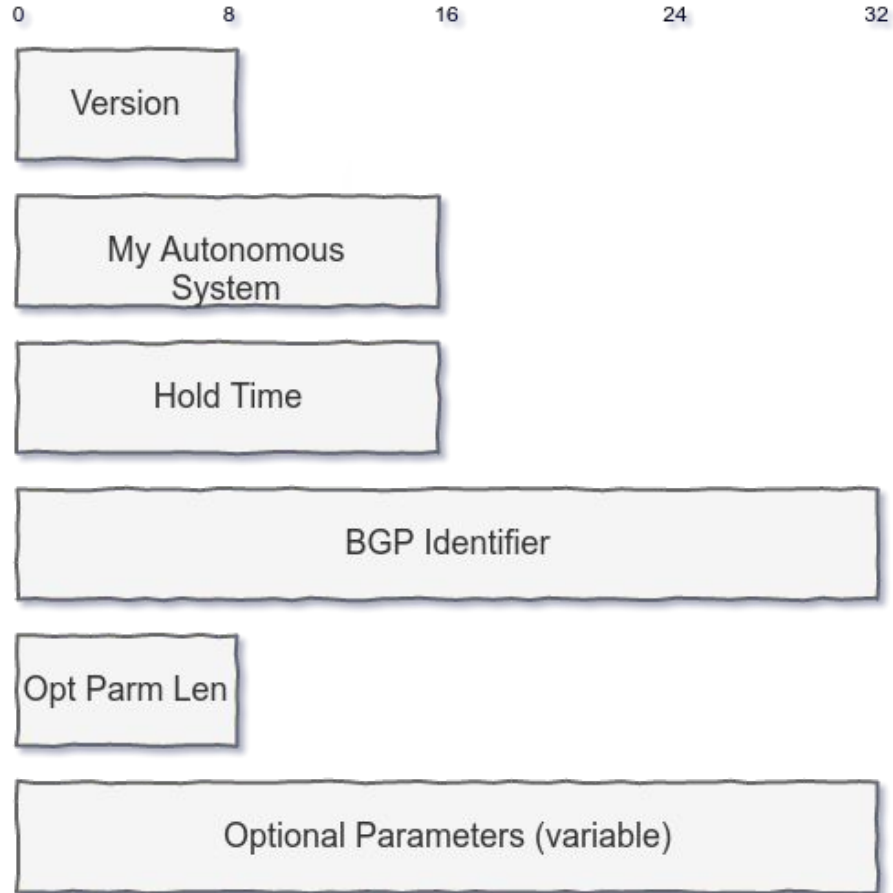
Protocolo BGP

✓ OPEN

- ✓ Enviada para iniciar a abertura da sessão BGP entre neighbors ou peers BGP
 - **Identificador BGP:** Convencionalmente o IP do roteador, atribuído no início do processo de configuração do BGP, pode ser automático ou estático.
 - **Comprimento dos parâmetros opcionais:** Comprimento total do campo de parâmetros opcionais
 - **Parâmetros opcionais:** Possui comprimento variável e trás uma lista de parâmetros. Cada parâmetro é descrito por uma tupla (Tipo, Tamanho, Valor)
 - Multiprotocol Extension, Route Refresh, Multiple Routes, Graceful restart, 4-byte AS number, entre outros

Protocolo BGP

✓ OPEN



Protocolo BGP

✓ NOTIFICATION

- ✓ Mensagem enviada quando há detecção de erros durante ou após a sessão BGP
 - Campo erro (Error): contem o tipo de notificação
 - Campo subcódigo de erro: contém um valor que fornece mais informações sobre o erro
 - Campo de dados: contem dados referentes ao erro (cabeçalho mal formado, ASN inválido, etc).



Protocolo BGP

✓ NOTIFICATION

ERROR CODE	ERROR SUBCODE
1	Erro no cabeçalho da mensagem 1. Conexão não sincronizada 2. Comprimento da mensagem inválido 3. tipo de mensagem inválido
2	Erro na mensagem OPEN 1. Número de versão não suportado 2. Número de AS vizinho inválido 3. Identificador BGP inválido 4. Parâmetro opcional não suportado 5. Falha na autenticação 6. Tempo de espera inaceitável
3	Erro na mensagem UPDATE 1. Lista de atributos mal formada 2. Atributo Well-Know desconhecido 3. Atributo Well-Know faltando 4. Erro nas flags de atributos 5. Erro no comprimento do atributo 6. Atributo de origem inválido

ERROR CODE	ERROR SUBCODE
3	Erro na mensagem UPDATE 7. Loop de roteamento em AS 8. Atributo NEXT_HOP inválido 9. Erro no atributo Opcional 10. Campo de rede inválido 11. AS_path mal formado
4	Hold Time Expired No Subcodes
5	Finite State Error No Subcodes
6	Cease No Subcodes

“Error code” fonte ESR (adaptado)



Protocolo BGP

✓ KEEPALIVE

- ✓ Mensagens trocadas para verificar se o peer está operacional (vivo)
- ✓ Não é enviado ao peer se uma mensagem de UPDATE já tiver sido trocada no intervalo válido
- ✓ Tempo $\frac{1}{3}$ do contador Holdtime
- ✓ Se o holdtime for zero os routers não enviam Keepalive

Protocolo BGP

✓ UPDATE

- ✓ Mensagens trocadas entre peers contendo informações para atualização da tabela de rotas mantida pelo BGP
 - Contém informações dos prefixo anunciado no BGP (eBGP e iBGP)
- ✓ Inserção/remoção de anúncios à medida que se modificam, de um roteador para cada vizinho
 - lista de prefixos de endereços para rotas que devem ser atualizadas na tabela de rotas do BGP
- ✓ Possui campos obrigatórios e “discricionários” (opcionais)

Protocolo BGP

✓ UPDATE

0 8 16 24 32

Withdrawn Routes Length

Withdrawn Routes (variable)

Total Path Attribute Length

Path Attributes (variable)

Network Layer Reachability Information (variable)

0 8 16 24 32

Length

Prefix (variable)

Attr. Flags

Attr. Type Code

Length

Prefix (variable)

Roteamento IP com BGP

- ✓ **ConnectRetryTimer:** Define o intervalo de tempo antes de tentar novamente uma solicitação de conexão (padrão 120s) condições.
- ✓ **HoldTimer:** Indica o tempo máximo em que um peer será considerado alcançável mesmo sem o recebimento de UPDATE ou KEEPALIVE (padrão 90s, mínimo de 3s, valor 0 remove o limite)
- ✓ **KeepAliveTimer:** Define a frequência das mensagens de KEEPALIVE (padrão é 30s, geralmente 1/3 do HoldTime)
- ✓ **MinRouteAdvertisementIntervalTimer (MRAI Timer):** Tempo mínimo de espera antes que um peer possa anunciar e/ou retirar rotas, possibilita a definição por prefixo IP, mas normalmente é mantido com base no peer (padrão 5s para iBGP, 30s para eBGP). Por possibilitar a remoção de prefixos serem acionadas esse contador, geralmente, não é usado na prática

Roteamento IP com BGP

- ✓ **MinASOriginationIntervalTimer:** Tempo mínimo de espera antes que alterações no AS possam ser enviadas ao peer através de mensagens UPDATE. (Padrão 15 segundos). Geralmente esse contador não é implementado
- ✓ **DelayOpenTimer (opcional):** Tempo de espera antes que a mensagem OPEN seja enviada.
- ✓ **IdleHoldTimer (opcional):** Tempo no estado IDLE antes de reiniciar uma sessão para um determinado peer, geralmente usado para inibir oscilações.

Roteamento IP com BGP

✓ Atributos

- ✓ Informações carregas nas mensagens de UPDATE do BGP
- ✓ Well-known
 - Mandatórios (mandatory): AS_PATH, NEXT_HOP, ORIGIN
 - Discricionário (discretionary): LOCAL_PREFERENCE, ATOMIC_AGGREGATE
- ✓ Opcionais
 - Transitivos: AGGREGATOR, COMMUNITY
 - Não transitivos: MED (Multi Exit Discriminator)

Roteamento IP com BGP

✓ Atributos Obrigatórios na mensagem de UPDATE: **Origem**

- ✓ Indica a origem da informação do caminho;
- ✓ Pode assumir os seguintes valores:
 - ‘i’ de IGP - Originado no interior do próprio AS
 - ‘e’ de EGP - Aprendido por meio de um Exterior Gateway Protocol (no nosso caso, o BGP)
 - ‘?’ de Incomplete - Aprendida por outros meios (ex: redistribuição)

Obrigatórios no
BGP UPDATE:

Origem

Roteamento IP com BGP

✓ Atributos Obrigatórios: **Origem**

BGP table version is 1660291, local router ID is 192.168.20.30

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal

Origin codes: i - IGP, e - EGP, ? - incomplete

Network	Next Hop	Metric	LocPrf	Weight	Path
*>i12.0.48.0/20	192.168.252.254	100	0		11537 10578 1742 i
*>i12.6.208.0/20	192.168.252.254	100	0		11537 10578 1742 i
*>i12.6.252.0/24	192.168.252.254	100	0		11537 10578 14325 ?
*>i12.16.126.192/26	192.168.252.254	100	0		11537 10578 14325 ?
*>i12.144.59.0/24	192.168.252.254	100	0		11537 10466 13778 i

Exemplo de saída do comando: # **show ip bgp**

Roteamento IP com BGP

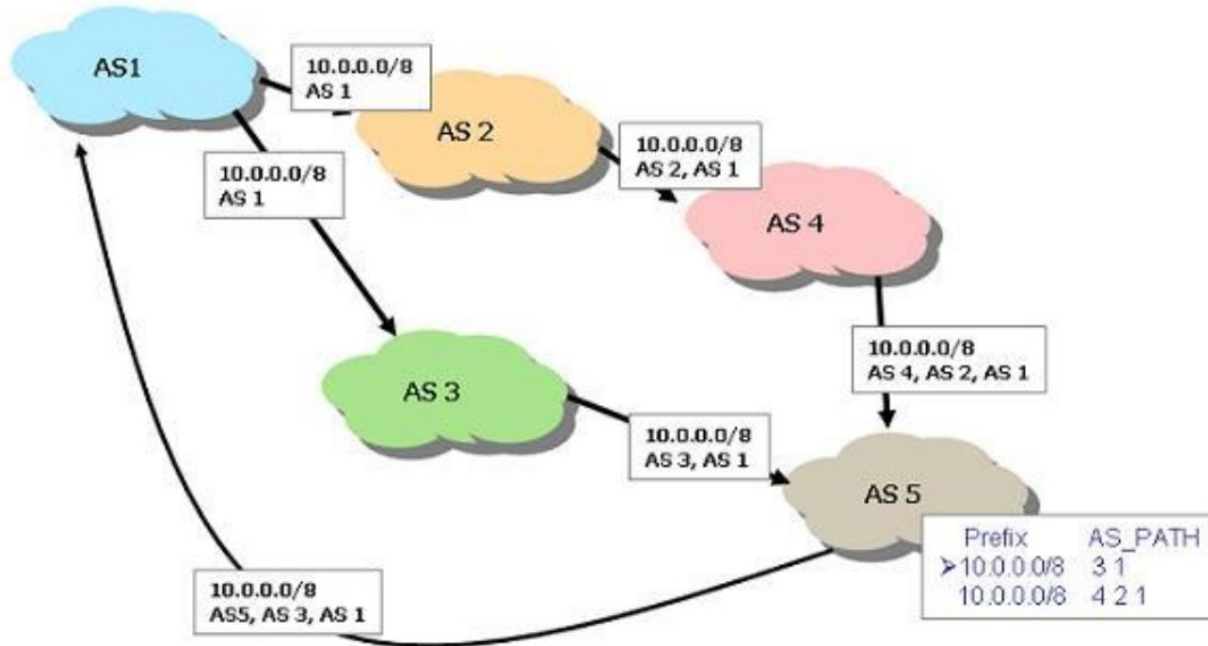
✓ Atributos Obrigatórios: **AS-Path**

64996 65000 i

- ✓ Atributo composto de uma sequência de ASNs, que representam o caminho a ser percorrido até a origem de um determinado prefixo.
- ✓ Sempre que um anúncio de prefixo é realizado entre AS's, o número local do AS anunciante é anexado do lado esquerdo na atualização a ser enviada.
- ✓ Caso seja realizado o anúncio de um prefixo a um vizinho interno (dentro do próprio AS) este atributo não é modificado

Roteamento IP com BGP

✓ Atributos Obrigatórios: **AS-Path**



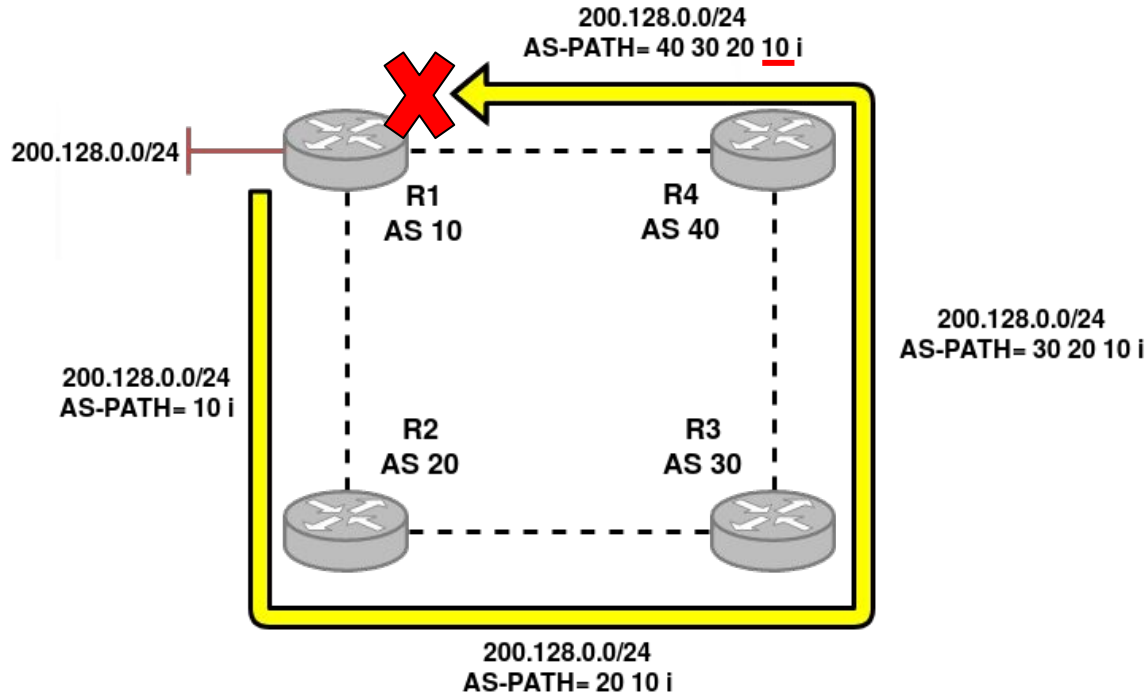
Obrigatórios no BGP UPDATE:

Origem

AS-PATH

Roteamento IP com BGP

✓ Atributos Obrigatórios: **AS-Path**

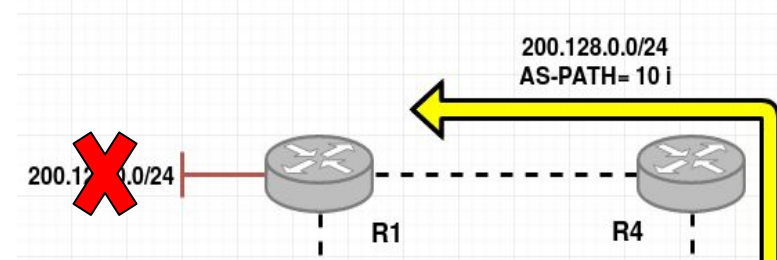
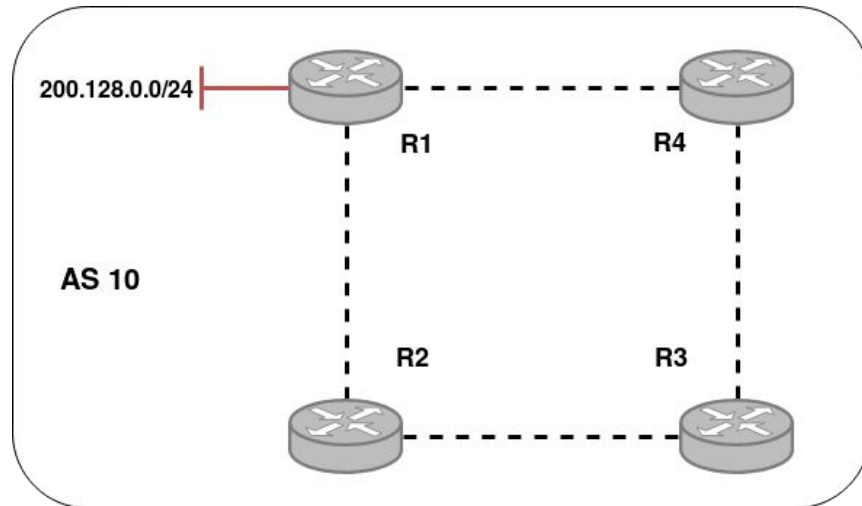


Controle de loop
do eBGP é feito
com o atributo
AS-PATH!

Roteamento IP com BGP

- ✓ Mas, como limitar o Loop em iBGP?
 - ✓ Como? se no iBGP o atributo AS-PATH não muda?

iBGP não possui um atributo para resolver o problema.

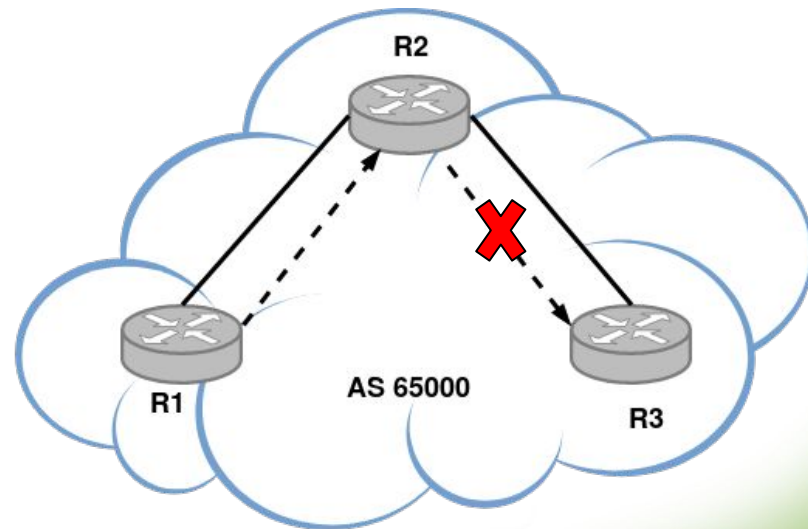


Ainda, caso R1 deixe de anunciar sua rota iBGP, existe a chance de R1 aprender seu antigo anúncio através de seus vizinhos.

Roteamento IP com BGP

- ✓ Como limitar o Loop em iBGP?
 - ✓ Solução dada não por atributo, mas por regra de anúncios iBGP
 - iBGP por padrão não permite que o anúncio de um roteador seja repassado para outro
 - Comportamento conhecido como **'Split Horizon'**

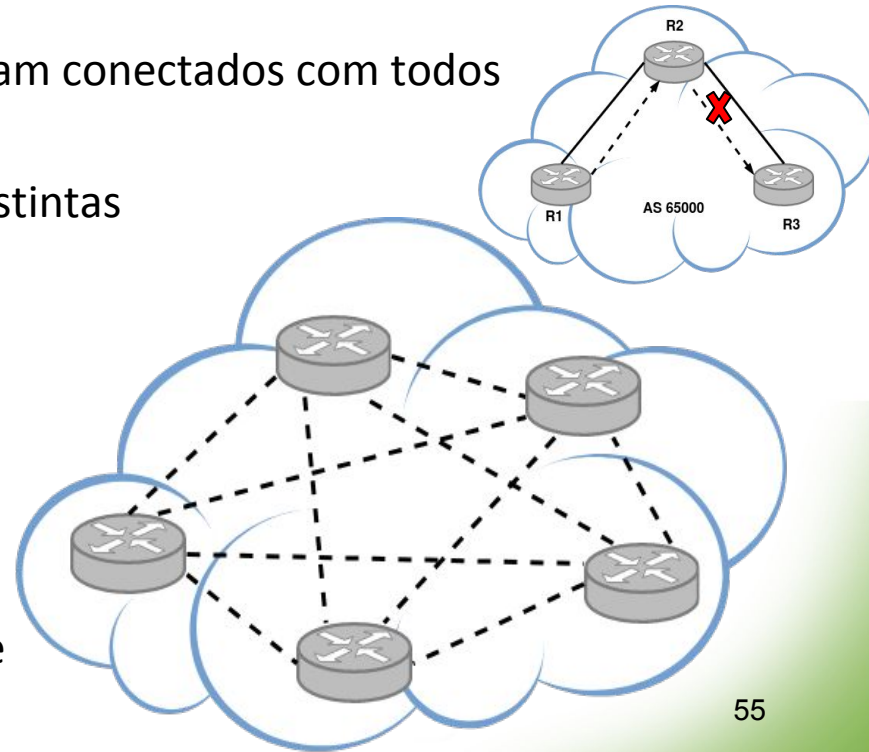
Split Horizon é um comportamento padrão do **iBGP** que garante topologia de rede sem loop. Mas, paga-se o preço de poder causar falha de roteamento intra-AS.



Roteamento IP com BGP

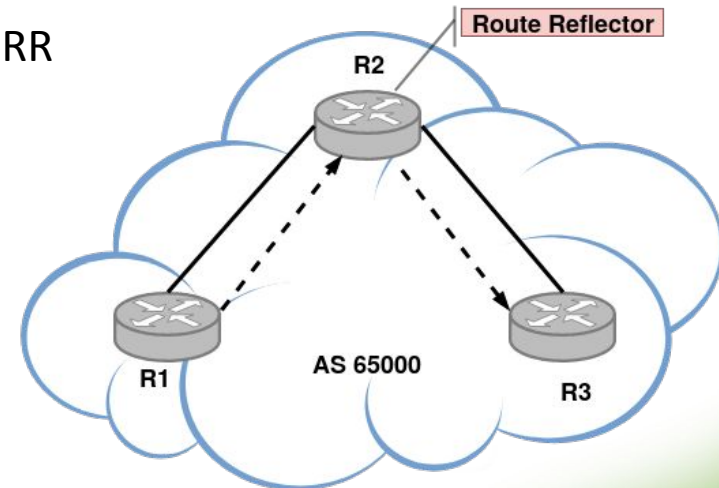
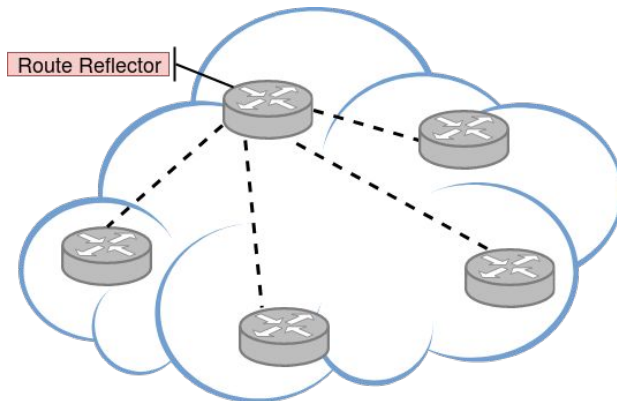
✓ Split Horizon

- ✓ Força com que todos os roteadores estejam conectados com todos
 - Topologia Full mesh
 - 5 roteadores = 10 sessões iBGP distintas
 - 6 roteadores = 15
 - 10 roteadores = 45
 - Claramente isto não escala...
 - Desvantagem:
 - Aumenta complexidade e quantidade de mensagens na rede



Roteamento IP com BGP

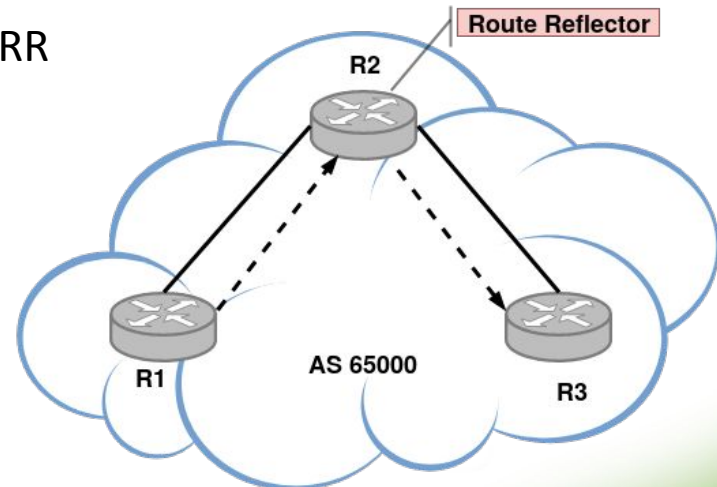
- ✓ Surge solução com **Refletores de Rotas** (RR ou Route Reflectors)
 - ✓ Um RR é um roteador iBGP que é permitido quebrar a regra de encaminhamento limitado de rotas iBGP ao 'refletir' os prefixos entre clientes iBGP
 - ✓ Permite iBGP operar com menor número de sessões (interconexões físicas entre roteadores)
 - ✓ Configurações são aplicadas apenas no RR



Roteamento IP com BGP

- ✓ Surge solução com **Refletores de Rotas** (RR ou Route Reflectors)
 - ✓ Um RR é um roteador iBGP que é permitido quebrar a regra de encaminhamento limitado de rotas iBGP ao 'refletir' os prefixos entre clientes iBGP
 - ✓ Permite iBGP operar com menor número de sessões (interconexões físicas entre roteadores)
 - ✓ Configurações são aplicadas apenas no RR

```
Configurar roteador R2 como route-reflector de R3:  
# configure terminal  
# router bgp 65000  
# neighbor a.b.c.d route-reflector-client
```



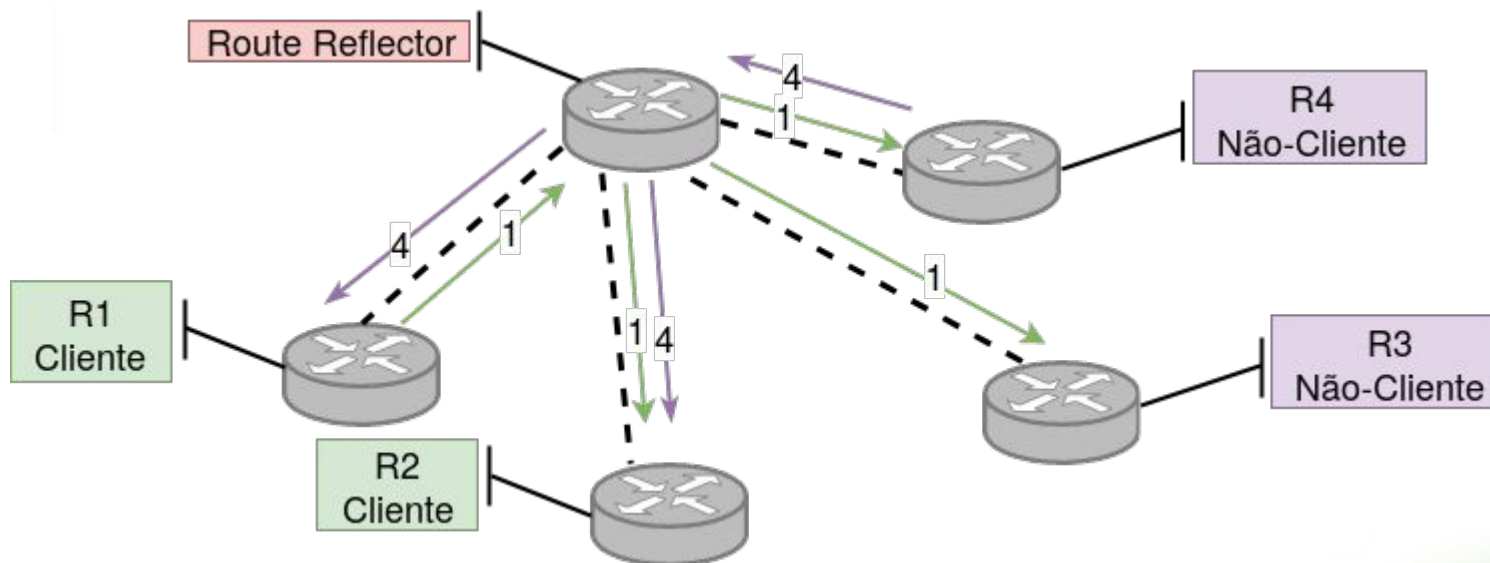
Roteamento IP com BGP

- ✓ Refletores de Rotas (RR ou Route Reflectors)
 - ✓ Possui esquema cliente e não-cliente
 - ‘Cliente’ é o roteador que terá suas rotas quebrando a regra de Split Horizon
 - ✓ Na visão do RR, ao receber rotas:
 - 1) Uma rota de um par BGP Não-cliente:
 - Reflete para todos os clientes.
 - 2) Rota de um par BGP Cliente:
 - Reflete para todos os pares Não-clientes e também para os pares Clientes
 - Por isso que os pares Clientes não necessitam ser totalmente interconectados (full mesh)

Consumo de memória e processamento é maior no RR do que nos outros roteadores.

Roteamento IP com BGP

- ✓ Refletores de Rotas (RR ou Route Reflectors)
 - ✓ Como seria a distribuição de rotas de um roteador Cliente (R1) e um roteador não-cliente (R4)? Exemplo:



Roteamento IP com BGP

- ✓ Ainda é possível ocorrer Loop com RRs por má configuração
- ✓ Definiu-se métodos Anti-Loop para Route reflectors
 - ✓ 1. Route reflector insere o atributo ORIGINATOR_ID em seus anúncios para identificar quando uma rota faz loop por outros roteadores e retorna para ele (não instalando a rota 'loopada')
 - ✓ Atributo ORIGINATOR_ID == Router ID
 - ✓ 2. PORÉM: Somente isso não evita que ocorra loop quando um RR divulga a rota para outro RR que conecte outros roteadores clientes diferentes. Portanto, existe também o conceito de CLUSTER_LIST com CLUSTER_IDs.
 - ✓ Cluster LIST e CLUSTER_ID é conceitualmente similar ao AS_PATH e ASN
 - ✓ Um Cluster é formado por um conjunto de um ou mais RR e um ou mais clientes
 - ✓ Cluster com 1 RR, todos os roteadores do RR receberão o cluster_ID == Router_ID do RR
 - ✓ Mas em caso de 2 RRs no cluster, pode-se manualmente configurar o Cluster_ID e até fazer regras de descarte específicas de um RR para o outro.
 - Comando: 'reflector cluster-id'
- ✓ Maiores informações:
<https://www.youtube.com/watch?v=RKJD5d-MI9Y&list=PLsYGHuNuBZcZaCOYysss4xgWb9oMdnVOWw&index=6>

Roteamento IP com BGP

✓ Atributos Obrigatórios: **Next-hop**

- ✓ Endereço IP do próximo salto para onde os pacotes devem ser encaminhados visando alcançar o prefixo recebido em uma mensagem UPDATE.

```
R1#sh ip bgp

  Network          Next Hop          Metric LocPrf Weight Path
*> 20.20.20.0/24   10.0.0.5          0           0 200 i
*> 30.30.30.0/24   0.0.0.0           0           32768 i

R5#sh ip bgp

  Network          Next Hop          Metric LocPrf Weight Path
*> 20.20.20.0/24   0.0.0.0           0           32768 i
*> 30.30.30.0/24   10.0.0.1          0           0 100 i
```

Obrigatórios no BGP UPDATE:

Origem

AS-PATH

Next-Hop

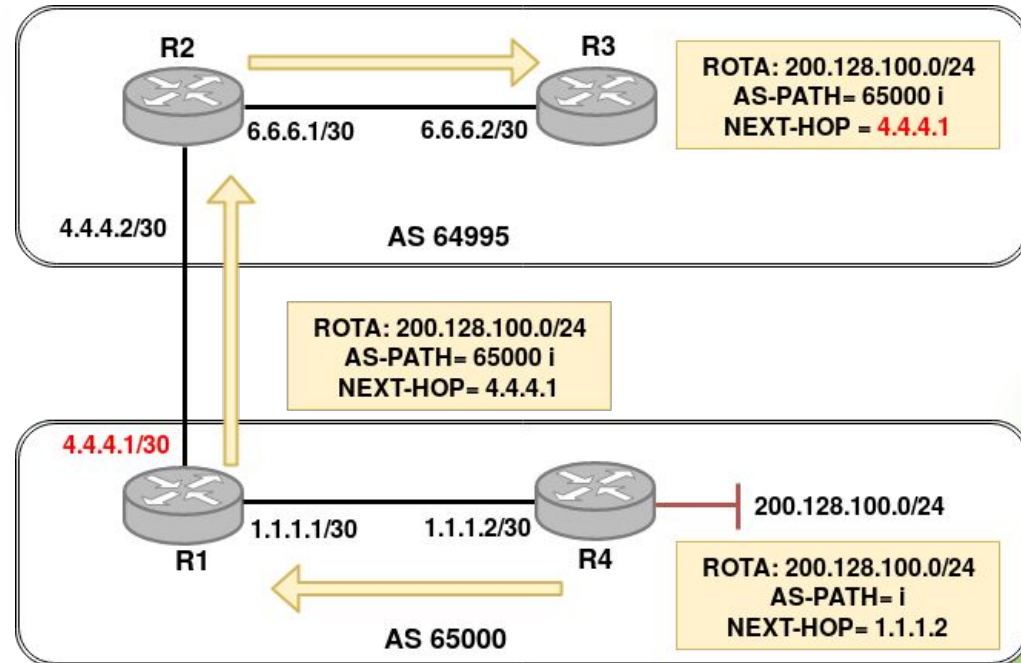
Next-hop de valor '0.0.0.0' significa que a rota é anunciada pelo roteador local. Outra indicação para isto é o valor 'i' na coluna Path.

Roteamento IP com BGP

✓ Comportamento **padrão** do atributo **Next-hop**:

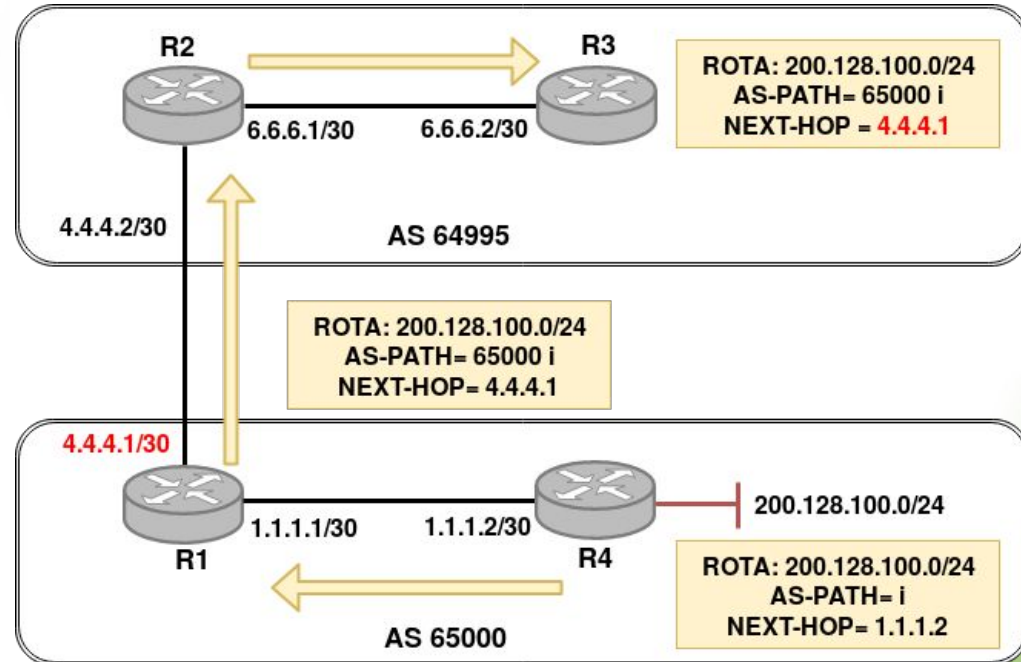
- ✓ No eBGP, é atualizado por salto
- ✓ No iBGP, é mantido o mesmo

Cenário comum:
Um roteador iBGP recebe anúncio de prefixo eBGP, mas não tem alcance ao IP de Next-hop do anúncio.



Roteamento IP com BGP

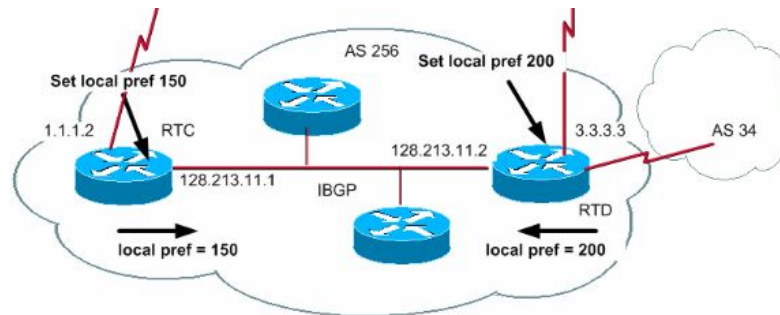
- ✓ Solução: R3 ter rota para 4.4.4.0/30
- ✓ Possíveis formas:
 - ✓ Inserindo rota estática ou default em R3
 - ✓ Via OSPF, R2 se divulgar como gateway default para vizinhos
 - ✓ Via BGP, R2 se anunciar como Next-hop de R3: 'next-hop-self'
 - ✓ via BGP, R2 anunciar o prefixo 4.4.4.0/30



Roteamento IP com BGP

✓ Atributos Discricionários: **Local Preference**

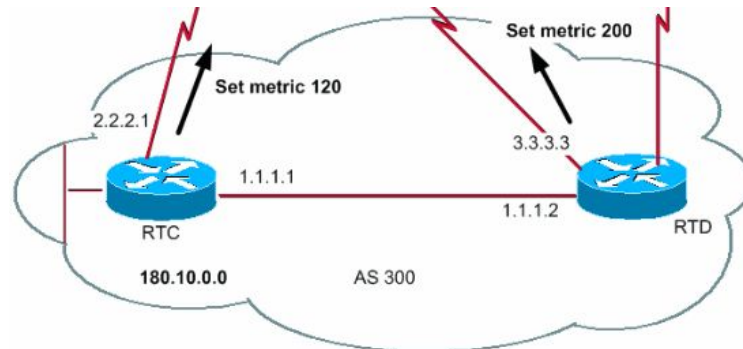
- ✓ Um roteador utiliza este atributo para informar aos seus vizinhos internos o grau de preferência de uma rota.
- ✓ Utilizado para escolher a rota de saída de um AS.
- ✓ Quanto maior o valor, maior a preferência da rota.



Roteamento IP com BGP

✓ Atributos Discricionários: **MED**

- ✓ É utilizado como sugestão para um AS vizinho quando há mais de uma entrada possível para o AS que está originando o prefixo.
- ✓ Quanto mais baixo o valor, maior a preferência da rota.



Roteamento IP com BGP

- ✓ Atributos Discricionários: **Communities**
 - ✓ Uma community é um grupo de destinos, ou prefixos IP, que compartilham uma propriedade em comum.
 - ✓ Cada AS pode definir quais communities um determinado prefixo irá pertencer.
 - ✓ Por padrão, todos os prefixos pertencem à community genérica Internet.
 - No export: Todos os prefixos marcados com esta community NÃO DEVEM ser anunciados a vizinhos eBGP.
 - No advertise: Todos os prefixos marcados com esta community NÃO DEVEM ser anunciados a qualquer vizinho BGP.
 - Local-as: Previne o envio de pacotes para fora do AS confederado.

Off-Topic: Preferência de valores menores ou maiores?

Atributo	Preferência sobre valor (menor ou maior)
Distância administrativa	Menor = maior preferencia
Prioridade de interface OSPF (DR & BDR)	Maior = maior preferência
RouterID (DR & BDR)	Maior = maior preferência
Custo OSPF	Menor = maior preferência
BGP AS-PATH	Menor = maior preferência
BGP Weight	Maior = maior preferência
BGP Local Preference	Maior = maior preferência
BGP Multi Exit Descriptor (MED)	Menor = menor preferência

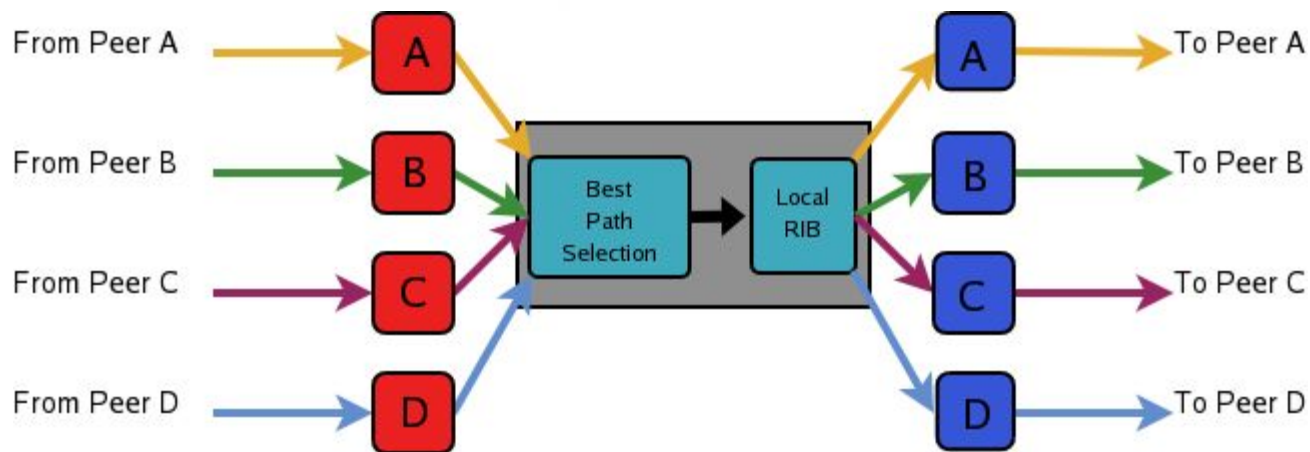
Roteamento IP com BGP

✓ Critério de desempate na Escolha de Rotas (tie-breakers):

1. Rota com **MAIOR** valor de **WEIGHT** (Cisco)
2. Rota com **MAIOR** valor de **LOCAL_PREF**
3. Rotas originadas localmente (comando network)
4. Rota com o **MENOR AS_PATH**
5. Rota com **MENOR** tipo de origem:
 - a. IGP (i) < EGP (e) < INCOMPLETE (?)
6. Rota com **MENOR** métrica multi-exit discriminator (**MED**).
7. Escolhe a rota **recebida** por (**eBGP**) em relação a uma (iBGP).
8. Rota com a **MENOR** métrica **IGP** para o next-hop BGP
9. Checagem multi-pathing (se habilitado)
10. Rota **EXTERNA** mais **ANTIGA** (oscilação)
11. Rota recebida com **MENOR Router ID**
12. Rota com **MENOR Cluster List length** (Route-Reflector)
13. Rota com o **MENOR neighbor address**

Roteamento IP com BGP

- ✓ Plano de controle do BGP (Adj-RIB-In, Loc-RIB, Adj-RIB-Out)



Processamento dos anuncios dentro de um normal BGP speaker (Adaptado a partir Quagga Documentation Figure 12.1)



Roteamento IP com BGP

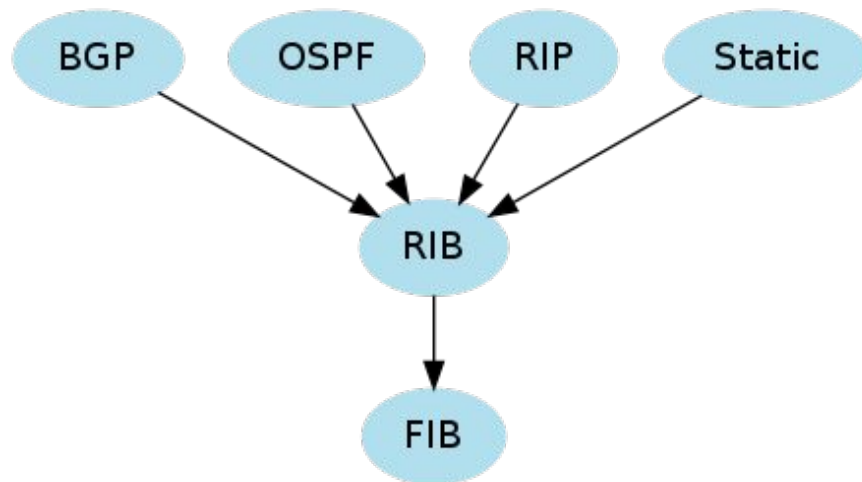
✓ Plano de controle do BGP (RIB)

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 100.1.20.0/24	200.223.79.7			0	7738 60200 ?
*> 100.1.102.0/24	200.223.79.7			0	7738 60200 ?
*> 100.1.128.0/20	200.223.79.7			0	7738 60200 ?
*> 100.1.201.0/24	200.223.79.7			0	7738 60200 ?
*> 177.126.112.0/24	0.0.0.0	0		32768	?
*> 177.126.113.0/24	0.0.0.0	0		32768	?
*> 187.84.144.0/20	200.223.79.7			0	7738 60200 53071 i
*> 187.84.150.0/23	200.223.79.7			0	7738 60200 53071 i
*> 189.130.1.0/24	200.223.79.7	0		0	7738 ?

```
187.84.0.0/16 is variably subnetted, 2 subnets, 2 masks
B   187.84.150.0/23 [20/0] via 200.223.79.7, 00:18:34
B   187.84.144.0/20 [20/0] via 200.223.79.7, 00:18:34
100.0.0.0/8 is variably subnetted, 4 subnets, 2 masks
B   100.1.102.0/24 [20/0] via 200.223.79.7, 00:18:34
B   100.1.20.0/24 [20/0] via 200.223.79.7, 00:18:34
B   100.1.201.0/24 [20/0] via 200.223.79.7, 00:18:34
B   100.1.128.0/20 [20/0] via 200.223.79.7, 00:18:34
189.130.0.0/24 is subnetted, 34 subnets
B   189.130.50.0 [20/0] via 200.223.79.7, 00:19:03
B   189.130.49.0 [20/0] via 200.223.79.7, 00:19:03
```

Roteamento IP com BGP

- ✓ Plano de controle do BGP (Protocolo, RIB e FIB)



Roteamento IP com BGP

- ✓ Configuração básica do BGP
 - ✓ Habilitar o BGP Routing (obrigatório)
 - ✓ Configurar os vizinhos BGP (obrigatório)
 - ✓ Usar BGP Soft Reconfiguration (recomendado)
 - ✓ Permitir interações entre BGP e IGP
 - ✓ Realizar filtragem de prefixos para os vizinhos BGP
 - ✓ Realizar a seleção de caminhos a ser transmitido para cada vizinho

Configuração do BGP

✓ COMANDOS DE CONFIGURAÇÃO DO PROTOCOLO BGP

Sintaxe	Descrição
router bgp <AS>	Habilita o BGP
bgp router-id <router-id>	Modifica manualmente o router ID
bgp default local-preference <number>	Modifica o valor padrão do local preference
neighbor <ip-address peer-group-name> remote-as <AS>	Configura um vizinho BGP
neighbor <ip-address peer-group-name> ebgp-multihop [maximum-hop-count]	Configura um vizinho que não está diretamente conectado.
neighbor <ip-address peer-group-name> update-source <interface>	Especifica uma interface de origem para a conexão TCP
neighbor {ip-address} soft-reconfiguration inbound	Habilita o “soft reset” na sessão BGP, evitar a instabilidade causada pelo reset das conexões BGP/TCP
neighbor <ip-address peer-group-name> route-map <name> <in out>	Aplica um filtro do tipo route-map às rotas de entrada / saída

Configuração do BGP

✓ COMANDOS DE CONFIGURAÇÃO DO PROTOCOLO BGP

Sintaxe	Descrição
neighbor <ip-address peer-group-name> prefix-list <name> <in out>	Aplica um filtro do tipo prefix-list às rotas de entrada / saída
neighbor <ip-address peer-group-name> filter-list <name> <in out>	Aplica um filtro do tipo as-path list às rotas de entrada / saída
neighbor <ip-address peer-group-name> next-hop-self	Configura o roteador como o próximo salto (next hop) dos prefixos
neighbor <ip-address peer-group-name> password <password>	Habilita a autenticação MD5
network <network [mask <mask>]> [route-map <name>]	Especifica as redes a serem anunciadas pelo protocolo
no synchronization	Impedir que rotas aprendidas sejam instaladas na RIB após ser encaminhada via iBGP
redistribute <connected kernel ospf static [metric <valor>] [route-map <nome>]	Redistribui as rotas aprendidas por um outro protocolo de origem

Configuração do BGP

✓ COMANDOS DE CONFIGURAÇÃO DO PROTOCOLO BGP

Sintaxe	Descrição
neighbor <ip-address peer-group-name> prefix-list <nome> <in out>	Aplica uma prefix-list na entrada ou saída de anúncios de um vizinho
neighbor <ip-address peer-group-name> distribute-list <nome> {in out weight}	Aplica uma distribute-list na entrada ou saída de anúncios de um vizinho
neighbor <ip-address peer-group-name> filter-list <nome> {in out weight}	Aplica uma filter-list na entrada ou saída de anúncios de um vizinho
neighbor <ip-address peer-group-name> route-map <nome> {in out weight}	Aplica uma route-map na entrada ou saída de anúncios de um vizinho
ip as-path access-list <numero> {permit deny} expressão-regular-as	Aplica uma access-list na entrada ou saída de anúncios de um vizinho

Configuração do BGP

✓ COMANDOS DE VERIFICAÇÃO DO PROTOCOLO BGP

Sintaxe	Descrição
<code>show ip protocol</code>	Mostra filtros (route-map) aplicados aos protocolos
<code>show ip bgp</code>	Mostra informações do protocolo BGP
<code>show ip bgp prefix-list <prefix-list></code>	Exibe rotas casando com uma prefix-list
<code>show ip bgp regexp <regexp></code>	Exibe rotas casando a regex por as-path
<code>show ip bgp route-map <route-map></code>	Exibe rotas casando com route-maps
<code>show ip bgp <community community-info community-list></code>	Exibe informações sobre as communities
<code>show ip bgp neighbors [ip] [subcomando]</code>	Mostra os vizinhos BGP
<code>show ip bgp summary</code>	Mostra um resumo das informações sessões BGP
<code>show ip bgp flap-statistics</code>	Exibe informações de flap em rotas

Configuração do BGP

✓ COMANDOS DE INTERVENÇÃO DO PROTOCOLO BGP

Sintaxe	Descrição
<code>clear ip bgp <ip> [soft-reconfiguration {in out}]</code>	Reinicializa a conexão BGP entre pares, usado quando se deseja que modificações em filtros, pesos, distâncias, versões, ou temporizadores tenham efeito
<code>clear ip bgp * [soft-reconfiguration {in out}]</code>	Reinicializa, assim como comando acima, mas para todos os pares
<code>show ip bgp prefix-list <prefix-list></code>	Exibe rotas casando com uma prefix-list
<code>show ip bgp regexp <regexp></code>	Exibe rotas casando a regexp por as-path
<code>show ip bgp route-map <route-map></code>	Exibe rotas casando com route-maps
<code>show ip bgp <community community-info community-list></code>	Exibe informações sobre as communities
<code>show ip bgp neighbors [ip] [subcomando]</code>	Mostra os vizinhos BGP
<code>show ip bgp summary</code>	Mostra um resumo das informações sessões BGP

Configuração do BGP

✓ EXEMPLO DE CONFIGURAÇÃO DO PROTOCOLO BGP

```
router bgp 52720
  bgp log-neighbor-changes
  neighbor 177.126.112.2 remote-as 60200
  neighbor 177.126.112.2 description EMPRESA-XYZ
  neighbor 200.223.79.7 remote-as 7738
  neighbor 200.223.79.7 description ISP-A
  !
  address-family ipv4
    redistribute connected
    neighbor 177.126.112.2 activate
    neighbor 177.126.112.2 send-community
    neighbor 177.126.112.2 soft-reconfiguration inbound
    neighbor 177.126.112.2 route-map BGP-IN-SET_CLIENTES in
    neighbor 200.223.79.7 activate
    neighbor 200.223.79.7 send-community
    neighbor 200.223.79.7 soft-reconfiguration inbound
    no auto-summary
    no synchronization
  exit-address-family
  !
```

Configuração do BGP

✓ EXEMPLO DE CONFIGURAÇÃO DO PROTOCOLO BGP

```
R6# show ip bgp
BGP table version is 46, local router ID is 200.223.79.6
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop          Metric LocPrf Weight Path
*> 100.1.20.0/24    200.223.79.7          0  7738 60200 ?
```

Configuração do BGP

✓ EXEMPLO DE CONFIGURAÇÃO DO PROTOCOLO BGP

```
RG#show ip bgp summary
BGP router identifier 200.223.79.6, local AS number 52720
BGP table version is 46, main routing table version 46
45 network entries using 5400 bytes of memory
46 path entries using 2392 bytes of memory
5/4 BGP path/bestpath attribute entries using 620 bytes of memory
3 BGP AS-PATH entries using 72 bytes of memory
2 BGP community entries using 48 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
Bitfield cache entries: current 1 (at peak 1) using 32 bytes of memory
BGP using 8564 total bytes of memory
BGP activity 48/3 prefixes, 49/3 paths, scan interval 60 secs

Neighbor      V   AS MsgRcvd MsgSent   TblVer  InQ  OutQ Up/Down  State/PfxRcd
177.126.112.2 4 60200     0      0       0    0    0 never    Active
200.223.79.7  4  7738    21     21      46    0    0 00:14:58    43
```


Configuração do BGP

✓ EXEMPLO DE CONFIGURAÇÃO DO PROTOCOLO BGP

```
BGP neighbor is 200.223.79.7, remote AS 7738, external link
Description: ISP-A
BGP version 4, remote router ID 189.130.50.1
BGP state = Established, up for 00:27:16
Last read 00:00:16, last write 00:00:16, hold time is 180, keepalive interval is 60 seconds
Neighbor capabilities:
  Route refresh: advertised and received(old & new)
  Address family IPv4 Unicast: advertised and received
Message statistics:
  InQ depth is 0
  OutQ depth is 0

                Sent      Rcvd
Opens:           1         1
Notifications:  0         0
Updates:         4         4
Keepalives:     29        29
Route Refresh:  0         0
Total:          34        34
Default minimum time between advertisement runs is 30 seconds
```

Configuração do BGP

✓ EXEMPLO DE CONFIGURAÇÃO DO PROTOCOLO BGP

```
For address family: IPv4 Unicast  
BGP table version 46, neighbor version 0/0  
Output queue size: 0  
Index 1, Offset 0, Mask 0x2  
1 update-group member  
Inbound soft reconfiguration allowed  
Community attribute sent to this neighbor  
Inbound path policy configured
```

Configuração do BGP

✓ EXEMPLO DE CONFIGURAÇÃO DO PROTOCOLO BGP

```
Prefix activity:          Sent      Rcvd
-----
Prefixes Current:       45        43 (Consumes 2236 bytes)
Prefixes Total:         45        43
Implicit Withdraw:       0          0
Explicit Withdraw:      0          0
Used as bestpath:       n/a        42
Used as multipath:      n/a         0

Local Policy Denied Prefixes:
-----
AS_PATH loop:           n/a         2
Total:                   0          2
Number of NLRI in the update sent: max 36, min 2

Connections established 1; dropped 0
Last reset never
Connection state is ESTAB, I/O status: 1, unread input bytes: 0
Connection is ECN Disabled, Minimum incoming TTL 0, Outgoing TTL 1
Local host: 200.223.79.6, Local port: 26202
Foreign host: 200.223.79.7, Foreign port: 179
Connection tableid (VRF): 0
```

Configuração do BGP

✓ EXEMPLO DE CONFIGURAÇÃO DO PROTOCOLO BGP

```
R6#show ip bgp neighbors 200.223.79.7 received-routes
BGP table version is 46, local router ID is 200.223.79.6
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop          Metric LocPrf Weight Path
*> 100.1.20.0/24    200.223.79.7          0  7738 60200 ?
*> 100.1.102.0/24  200.223.79.7          0  7738 60200 ?
*> 100.1.128.0/20  200.223.79.7          0  7738 60200 ?
*> 100.1.201.0/24  200.223.79.7          0  7738 60200 ?
*> 187.84.144.0/20 200.223.79.7          0  7738 60200 53071 i
*> 187.84.150.0/23 200.223.79.7          0  7738 60200 53071 i
*> 189.130.1.0/24  200.223.79.7          0              0  7738 ?
```

Configuração do BGP

✓ EXEMPLO DE CONFIGURAÇÃO DO PROTOCOLO BGP

```
R6#show ip bgp neighbors 200.223.79.7 advertised-routes
BGP table version is 46, local router ID is 200.223.79.6
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network                Next Hop                Metric LocPrf Weight Path
*> 100.1.20.0/24          200.223.79.7              0  7738 60200 ?
*> 100.1.102.0/24         200.223.79.7              0  7738 60200 ?
*> 100.1.128.0/20         200.223.79.7              0  7738 60200 ?
*> 100.1.201.0/24         200.223.79.7              0  7738 60200 ?
*> 177.126.112.0/24      0.0.0.0                   0                32768 ?
*> 177.126.113.0/24      0.0.0.0                   0                32768 ?
*> 187.84.144.0/20       200.223.79.7              0  7738 60200 51071 i
```

Duvidas?

¿...?

Hora da prática

Vamos ao roteiro de laboratório 3

Um pouco mais de BGP

✓ Multihoming:

- ✓ É o termo usado para descrever quando um AS está conectado a mais de um ISP.
 - Aumentar a confiabilidade da conexão com a Internet, porque se uma conexão falhar a outra ainda estará disponível
 - Aumentar o desempenho e, assim, caminhos melhores poderão ser usados para determinados destinos

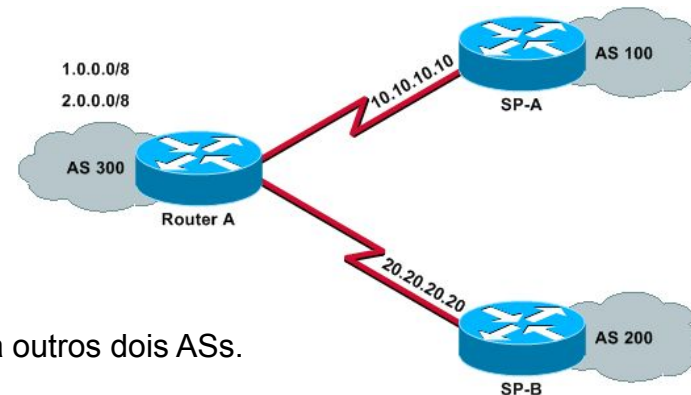


Imagem: Roteador A do AS 300 em Multihoming, conectado a outros dois ASs.

Fonte: <https://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/23675-27.html>

Um pouco mais de BGP

- ✓ Política de roteamento:
 - ✓ Uma política de roteamento bem definida começa no plano de endereçamento IP (v4 e v6)
 - Escalabilidade
 - Flexibilidade
 - Consistência

Um pouco mais de BGP

- ✓ Plano de endereçamento
 - ✓ Distribuição dos serviços e servidores entre partes distintas do bloco IP
 - Facilita a influência do tráfego de entrada e saída de seu AS
 - Exemplo: Seu AS possui um prefixo /24 distribuído em quatro /26, onde cada /26 é para um serviço: servidores web; nuvem de arquivos; pool de NAT dos clientes internos; e IPs públicos de roteadores de seu AS.
 - Não adianta concentrar todo o tráfego principal atrás do mesmo prefixo /24 ou /48 (no caso IPv6) anunciado na Internet
 - ✓ A má distribuição restringe a influência do tráfego de entrada ou de saída

Um pouco mais de BGP

- ✓ Dado um bom planejamento, pode-se gerar políticas de entrada e saída
- ✓ Política de roteamento de entrada:
 - ✓ Dependendo de como você anuncia o seu prefixo, afeta a resposta (download) dos outros ASes para você
 - ✓ Exemplo: Balanceamento de tráfego de download ao anunciar seu prefixo /24 quebrado em dois /25. Ou seja, um /25 por um AS e o /25 por outro AS
 - Como habilitar redundância automática e prover maior resiliência?
 - Apenas também divulgar o prefixo /24 para ambos os ASes vizinhos

Um pouco mais de BGP

- ✓ Política de roteamento de saída:
 - ✓ Depende dos anúncios recebidos da Internet
 - Interfere no tráfego de saída (upload)
 - ✓ Pode-se também influenciar o tráfego de saída alterando o valor do LOCAL_PREFERENCE
 - O atributo WEIGHT também pode ser usado, mas não se propaga no iBGP
 - ✓ Exemplo: Seu AS precisa influenciar seu tráfego de saída, de modo que o tráfego com destino ao primeiro /21 do AS 64513 saia preferencialmente pelo link com o AS 64511, e o tráfego com destino ao segundo /21 do AS 64513 saia preferencialmente pelo link com o AS 64512

Um pouco mais de BGP

- ✓ Mapa de rotas
 - ✓ Listas de acesso que permitem BGP controlar e modificar as condições de propagação de rotas entre roteadores
 - Condição testada através do comando **match**
 - Caso condição seja satisfeita, uma ação pode ser tomada pelo comando **set**
 - Vantagens:
 - Flexibilidade na inserção/remoção, pois cada entrada é numerada
 - Criar uma política de roteamento complexa

Comando: **# route-map <nome-mapa> [permit|deny] seq <numero>**

Um pouco mais de BGP

- ✓ Mapa de rotas
 - ✓ Listas de acesso que permitem BGP controlar e modificar as condições de propagação de rotas entre roteadores

Exemplo prático: **# route-map <nome-mapa> [permit|deny] seq <numero>**

```
# route-map BGP-OUT-AS7738 permit 10  
# match ip address 200.100.0.0/24  
  
# route-map BGP-OUT-AS7738 permit 15  
# match ip address 200.200.0.0/24           // BLACKHOLE  
# set community 7738:666 additive  
  
# route-map BGP-OUT-AS7738 deny 100  
# match as-path <AS-PATH-ACL>
```

Um pouco mais de BGP

- ✓ Route server (RS)
 - ✓ **Equivalente** em funcionalidade a um **Route Reflector** para **eBGP**
 - ✓ Infraestruturas, tradicionalmente servidores, que permitem que sistemas autônomos troquem tráfego, através do anúncio de rotas
 - ✓ Estrutura simplificada que atua como facilitador das interconexões
 - ✓ É possível que cada participante estabeleçam bilaterais entre si, mas tende a ser mais complexo e difícil de escalar

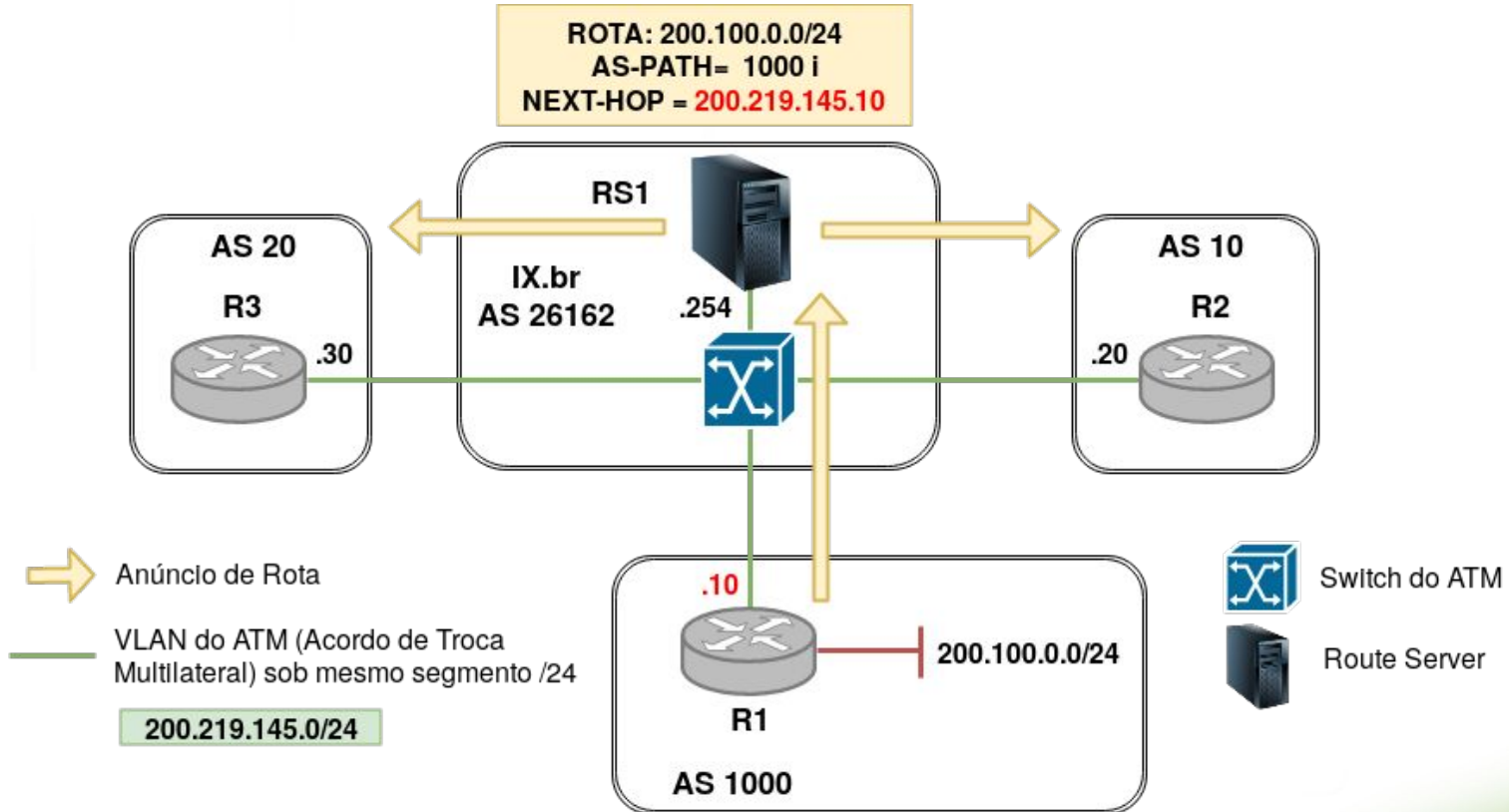
Um pouco mais de BGP

- ✓ Route server (RS)
 - ✓ Tradicionalmente presente nos pontos de trocas de tráfego (IX), os ASes participantes são interconectados em camada 2 e cada um desses fecham sessões BGP com os RSeS para anunciar suas rotas e receber as rotas de todos demais.
 - Prefixos recebidos de cada um serão repassados aos demais, com exceção do próprio participante que originou o anúncio, mas sem acrescentar o ASN do servidor de rotas, e sem modificar o atributo `next_hop`.

Um pouco mais de BGP

- ✓ Route server (RS)
 - ✓ Se um participante não deseja trocar prefixos com outro(s), filtros de rotas podem ser implantadas nos RSes
 - Feita manualmente pelos operadores dos RSes, após solicitação do sistema autônomo;
 - Ou através de comunidades BGP (communities) predefinidas

Um pouco mais de BGP



Um pouco mais de BGP

- ✓ Route server (RS)
 - ✓ Servidores de rotas são totalmente neutros, não interferindo no encaminhamento do tráfego;
 - Não acrescentam seu próprio AS ao **AS-PATH**, nem alteram o **NEXT-HOP**
 - Cada participante recebe rotas dos demais de forma direta diretamente
 - ✓ Alta disponibilidade e resiliência
 - Possibilidade de uso de servidores redundantes, multi-vendor e distribuídos

Um pouco mais de BGP

- ✓ Route server (RS)
 - ✓ Exemplo: IX.br - Salvador e AS 53164 (PoP-BA | UFBA)

```
rs1.ba.ptt.br> show ip bgp neighbors 200.219.145.9 received-routes
BGP table version is 0, local router ID is 200.219.145.253
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, R Removed
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop          Metric LocPrf Weight Path
*> 192.188.11.0     200.219.145.9      0         0 53164 i
*> 200.128.0.0/17   200.219.145.9      1         0 53164 i
*> 200.128.0.0/18   200.219.145.9      1         0 53164 i
*> 200.128.64.0/18 200.219.145.9      1         0 53164 i

Total number of prefixes 4
```

Um pouco mais de BGP

- ✓ Route server (RS)
 - ✓ Exemplo: IX.br - Salvador e AS 53164 (PoP-BA | UFBA)

```
rs1.ba.ptt.br> show ip bgp neighbors 200.219.145.9 advertised-routes
BGP table version is 0, local router ID is 200.219.145.253
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, R Removed
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop          Metric LocPrf Weight Path
* > 2.22.84.0/22    200.219.145.45          0 14840 20940 20940 i
* > 2.23.98.0/23    200.219.145.45          0 14840 20940 20940 i
* > 23.36.112.0/20  200.219.145.46          0 52871 i
      .
      .
      .
* > 201.182.246.0/23 200.219.145.7          0 28186 267550 267550 i
* > 201.182.252.0/22 200.219.145.59          0 267613 i
* > 201.182.252.0    200.219.145.59          0 267613 i
* > 201.222.20.0/22  200.219.145.43          0 61902 i
* > 216.75.220.0     200.219.145.45          0 14840 26363 i
* > 216.98.216.0/22  200.219.145.67          0 53087 265290 266226 i
* > 216.98.218.0/23  200.219.145.67          0 53087 265290 266226 i

Total number of prefixes 5012
```

Um pouco mais de BGP

✓ Route server (RS)

✓ Looking Glass

- Serviço que permite consultas sobre roteamento (protocolo BGP), testes de alcance (ping) e do caminho (traceroute)

```
lg.ba.ptt.br$ show ip bgp 8.8.8.0/24
BGP routing table entry for 8.8.8.0/24
Paths: (2 available, best #1, table Default-IP-Routing-Table)
Not advertised to any peer
53242 15169
 200.219.145.23 from 200.219.145.23 (177.21.0.1)
  Origin IGP, localpref 100, valid, external, best
  Community: 53242:300
  Last update: Thu Jun 24 10:18:54 1971

14026 7738 15169
 200.219.145.2 from 200.219.145.4 (200.219.145.4)
  Origin IGP, localpref 100, valid, external
  Last update: Tue Jun 8 21:16:25 1971
```



```
lg.ba.ptt.br$ show ip bgp summary
BGP router identifier 200.219.145.252, local AS number 20121
RIB entries 1433218, using 131 MiB of memory
Peers 120, using 534 KiB of memory

Neighbor      V  AS MsgRcvd MsgSent  TblVer  InQ OutQ Up/Down State/PfxRcd
200.219.145.1 4 1916 263571 44821    0 0 0 18:20:48 14047
200.219.145.3 4 266345 20790 18215    0 0 0 2d22h14m Active
200.219.145.4 4 14026 205687 22412    0 0 0 02w1d13h 45843
200.219.145.5 4 18881 0 0        0 0 0 never Connect
200.219.145.6 4 25832 0 0        0 0 0 never Connect
200.219.145.7 4 28186 45823 44822    0 0 0 02w1d13h 0
200.219.145.8 4 25832 0 0        0 0 0 never Connect
200.219.145.9 4 53164 25312 22412    0 0 0 02w1d13h 4
200.219.145.10 4 265028 25620 22407    0 0 0 22:39:22 2
200.219.145.13 4 28186 0 0        0 0 0 never Connect
200.219.145.16 4 23106 25802 22413    0 0 0 02w1d13h 0
200.219.145.17 4 264437 39278 37123    0 0 0 2d16h00m Active
200.219.145.18 4 262843 25454 22291    0 0 0 2d04h30m 10
200.219.145.19 4 53239 124341 132216    0 0 0 6d08h45m 0
200.219.145.20 4 268539 0 8595    0 0 0 never Active
200.219.145.21 4 20144 0 0        0 0 0 never Connect
200.219.145.22 4 265019 25989 22412    0 0 0 02w1d13h 6
200.219.145.23 4 53242 11708340 43807    0 0 0 00:27:43 778051
200.219.145.24 4 28573 22422 22412    0 0 0 02w1d13h 111
```

```
lg.ba.ptt.br$ show ip bgp 8.8.8.0/24
BGP routing table entry for 8.8.8.0/24
Paths: (2 available, best #1, table Default-IP-Routing-Table)
Not advertised to any peer
53242 15169
 200.219.145.23 from 200.219.145.23 (177.21.0.1)
  Origin IGP, localpref 100, valid, external, best
  Community: 53242:300
  Last update: Thu Jun 24 10:18:54 1971

14026 7738 15169
 200.219.145.2 from 200.219.145.4 (200.219.145.4)
  Origin IGP, localpref 100, valid, external
  Last update: Tue Jun 8 21:16:25 1971
```


Um pouco mais de BGP

✓ Route server (RS)

✓ Looking Glass

- Serviço que permite consultas sobre roteamento (protocolo BGP), testes de alcance (ping) e do caminho (traceroute)

```
lg.ba.ptt.br$ show ip bgp 8.8.8.0/24
BGP routing table entry for 8.8.8.0/24
Paths: (2 available, best #1, table Default-IP-Routing-Table)
Not advertised to any peer
53242 15169
 200.219.145.23 from 200.219.145.23 (177.21.0.1)
  Origin IGP, localpref 100, valid, external, best
  Community: 53242:300
  Last update: Thu Jun 24 10:18:54 1971

14026 7738 15169
 200.219.145.2 from 200.219.145.4 (200.219.145.4)
  Origin IGP, localpref 100, valid, external
  Last update: Tue Jun 8 21:16:25 1971
```



```
lg.ba.ptt.br$ show ip bgp summary
BGP router identifier 200.219.145.252, local AS number 20121
RIB entries 1433218, using 131 MiB of memory
Peers 120, using 534 KiB of memory

Neighbor      V  AS MsgRcvd MsgSent  TblVer  InQ OutQ Up/Down  State/PfxRcd
200.219.145.1 4 1916 263571 44821    0 0 0 18:20:48 14047
200.219.145.3 4 266345 20790 18215    0 0 0 2d22h14m Active
200.219.145.4 4 14026 205687 22412    0 0 0 02w1d13h 45843
200.219.145.5 4 18881 0 0        0 0 0 never Connect
200.219.145.6 4 25832 0 0        0 0 0 never Connect
200.219.145.7 4 28186 45823 44822    0 0 0 02w1d13h 0
200.219.145.8 4 25832 0 0        0 0 0 never Connect
200.219.145.9 4 53164 25312 22412    0 0 0 02w1d13h 4
200.219.145.10 4 265028 25620 22407    0 0 0 22:39:22 2
200.219.145.13 4 28186 0 0        0 0 0 never Connect
200.219.145.16 4 23106 25802 22413    0 0 0 02w1d13h 0
200.219.145.17 4 264437 39278 37123    0 0 0 2d16h00m Active
200.219.145.18 4 262843 25454 22291    0 0 0 2d04h30m 10
200.219.145.19 4 53239 124341 132216    0 0 0 6d08h45m 0
200.219.145.20 4 268539 0 8595    0 0 0 never Active
200.219.145.21 4 20144 0 0        0 0 0 never Connect
200.219.145.22 4 265019 25989 22412    0 0 0 02w1d13h 6
200.219.145.23 4 53242 11708340 43807    0 0 0 00:27:43 778051
200.219.145.24 4 28573 22422 22412    0 0 0 02w1d13h 111
```

```
lg.ba.ptt.br$ show ip bgp 8.8.8.0/24
BGP routing table entry for 8.8.8.0/24
Paths: (2 available, best #1, table Default-IP-Routing-Table)
Not advertised to any peer
53242 15169
 200.219.145.23 from 200.219.145.23 (177.21.0.1)
  Origin IGP, localpref 100, valid, external, best
  Community: 53242:300
  Last update: Thu Jun 24 10:18:54 1971

14026 7738 15169
 200.219.145.2 from 200.219.145.4 (200.219.145.4)
  Origin IGP, localpref 100, valid, external
  Last update: Tue Jun 8 21:16:25 1971
```

Pontos de troca de tráfego

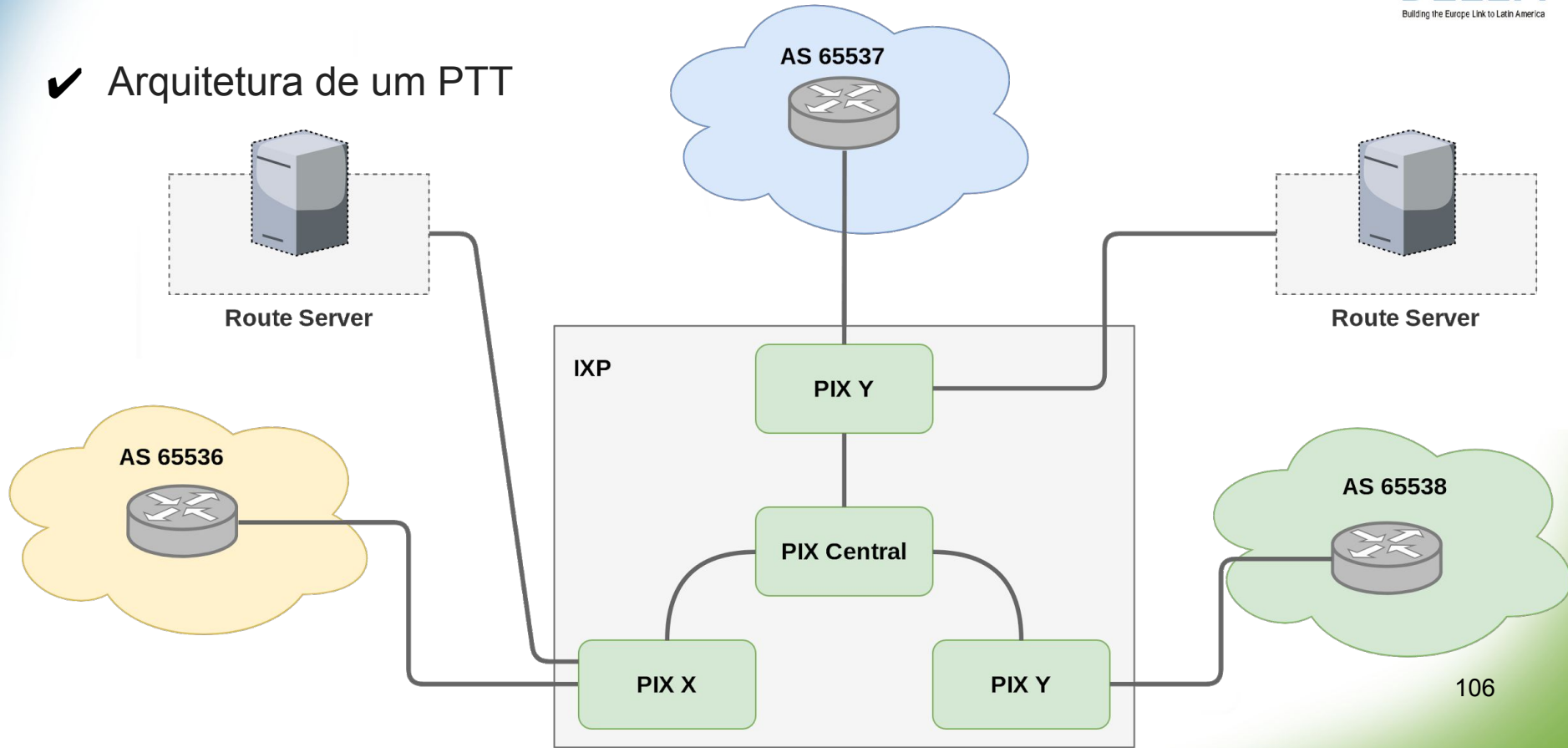
- ✓ Ponto de troca de Tráfego (PTT) / Internet Exchange Point (IXP)
 - ✓ Solução de rede com o objetivo de viabilizar a conexão direta entre as entidades que compõem a internet, os Sistemas Autônomos (AS): provedores de Internet e de conteúdo;
 - ✓ O IX otimiza a interconexão entre AS, pois possibilita:
 - Melhor qualidade (menor latência), evitando intermediários externos
 - Menor custo
 - Maior organização da estrutura de rede regional (pontos concentradores)

Pontos de troca de tráfego

- ✓ Características
 - ✓ Neutralidade - independência de provedores comerciais
 - ✓ Qualidade - troca de tráfego eficiente
 - ✓ Economia - baixo custo das alternativas, com alta disponibilidade
 - ✓ Matriz de troca de tráfego regional única
 - ✓ Troca de tráfego local e direta, sem a necessidade de passar por redes de terceiros
 - ✓ Alternativa para escoamento de tráfego do AS

Pontos de troca de tráfego

✓ Arquitetura de um PTT



Pontos de troca de tráfego

✓ Acordos de troca de tráfego

✓ **ATM**

- Domínio de broadcast compartilhado (VLAN por família de protocolo)
- Troca de tráfego entre todos os participantes
- Sessão BGP estabelecida entre os participantes e servidores de rotas

✓ **ATB**

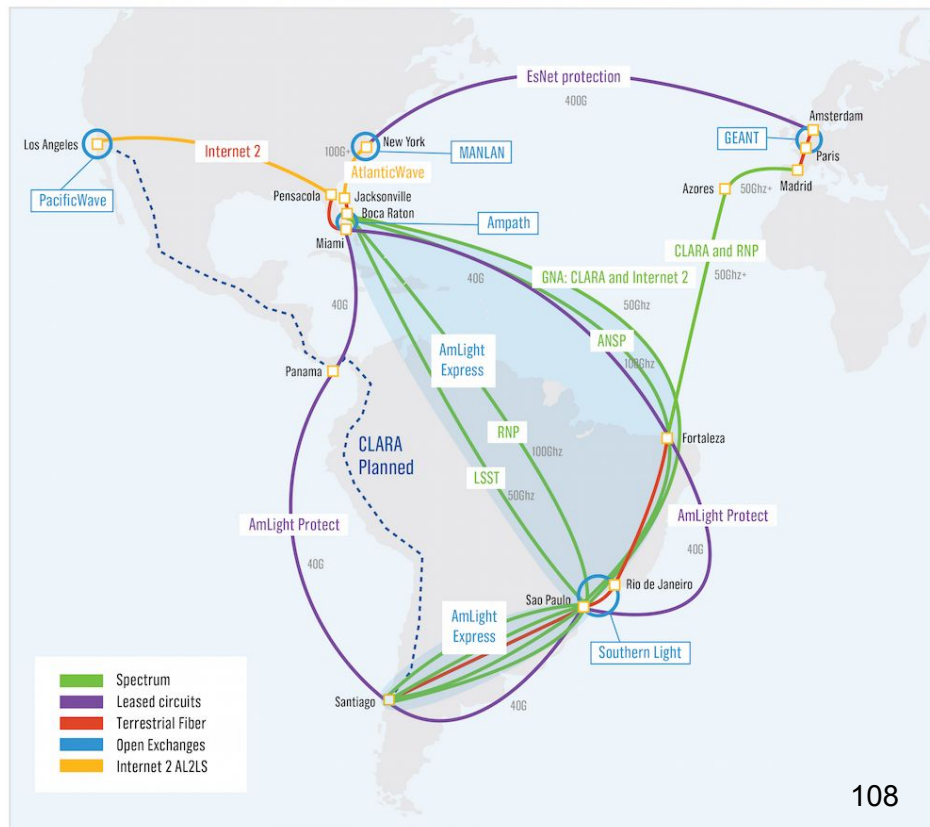
- Troca de tráfego apenas entre dois participantes (**B**ilateral)
- VLAN compartilhada (mesmo domínio de broadcast do ATM)
- VLAN dedicada

Pontos de troca de tráfego

✓ Pelo mundo

✓ AMPATH

- Interconecta as redes de pesquisa e educação para troca de tráfego
- Localizado em Miami/FL

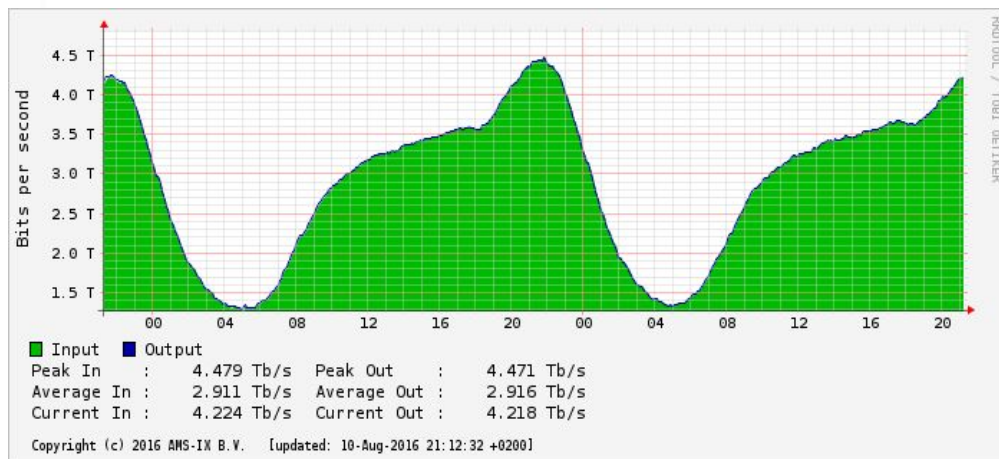


Pontos de troca de tráfego

✓ Pelo mundo

✓ Amsterdam Internet Exchange (AMSIX)

- Interconecta cerca de 800 redes em quase 1450 interfaces
- Capacidade de até 20 Tbps



Pontos de troca de tráfego

- ✓ Pelo mundo
- ✓ Diversas localidades pelo mundo
- ✓ Melhora o grau de conectividade do AS

Fonte: <http://www.internetexchangemap.com/>



Fonte: <http://www.ixp toolkit.org/ixps>

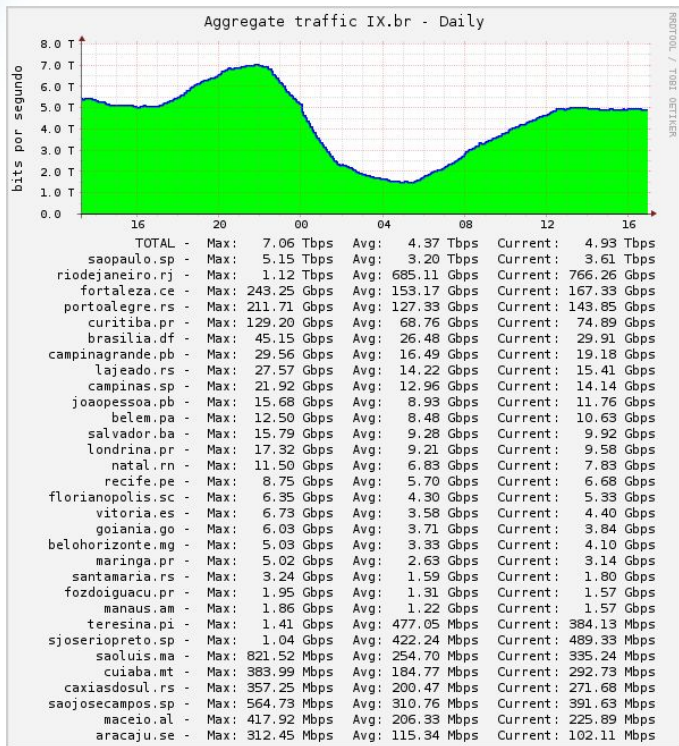
Pontos de troca de tráfego

- ✓ IX.br
 - ✓ Projeto do Comitê Gestor da Internet no Brasil (CGI.br) que promove e cria a infraestrutura necessária para a interconexão direta entre as redes (ASs) que compõem a Internet Brasileira;
 - ✓ Principais vantagens deste modelo, é a racionalização dos custos, uma vez que os balanços de tráfego são resolvidos direta e localmente e não através de redes de terceiros, muitas vezes fisicamente distantes.

Pontos de troca de tráfego

- ✓ IX.br
 - ✓ Maior controle que uma rede pode ter com relação a entrega de seu tráfego o mais próximo possível do seu destino, o que em geral resulta em melhor desempenho e qualidade para seus clientes e operação mais eficiente da Internet como um todo.
 - ✓ Um PTTMetro é, assim, uma interligação em área metropolitana de pontos de interconexão de redes (PIXes), comerciais e acadêmicos, sob uma gerência centralizada.

Pontos de troca de tráfego

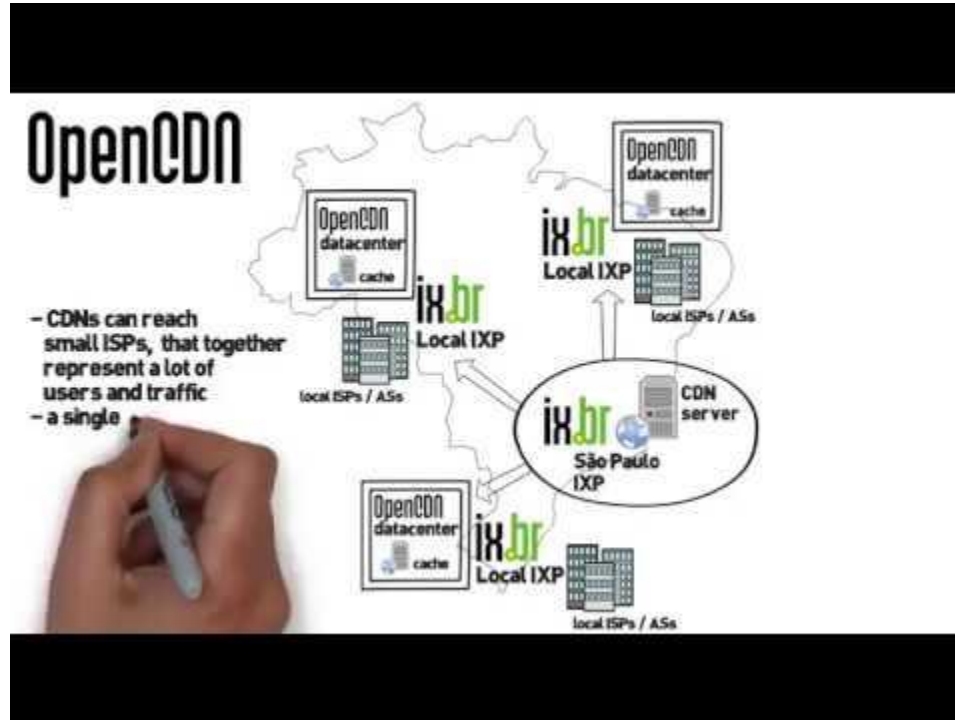


IX.br - Salvador																
ASN	NOME	ATM		LG		TRÁNSITO		IPV6								
		V4	V6	V4	V6	V4	V6									
1916	RNP	✓	✓	✓	✓			✓	262801	Viva	✓	✓		✓	✓	✓
7738	Oi						✓		262843	Lognet	✓	✓	✓			✓
10906	b.dns.br anycast - Nic.br	✓							263903	INFORBARRA	✓	✓	✓	✓	✓	✓
10954*	SERPRO	✓	✓	✓	✓				263912	PixelTelecom	✓	✓	✓	✓	✓	✓
14026	Simet	✓	✓	✓	✓				264304	SenalCimatec	✓	✓	✓	✓	✓	✓
14840	Commcorp	✓	✓	✓	✓		✓	✓	264404	NETGOOL	✓	✓	✓	✓	✓	✓
18881	GVT	✓	✓	✓	✓		✓	✓	264437	TELEMAC	✓	✓	✓	✓	✓	✓
20144	L-Root - ICANN	✓							264479	TURBOZONE	✓	✓		✓	✓	✓
23106	CEMIGTelecom	✓	✓	✓	✓	✓	✓	✓	264592	EnlaceTelecom	✓	✓	✓	✓	✓	✓
25832	Prodeb	✓	✓	✓	✓			✓	264906	PlanetInternet	✓	✓	✓	✓	✓	✓
25933	Vogel Telecom	✓	✓	✓	✓	✓	✓	✓	265019	NexTelecom	✓	✓	✓	✓	✓	✓
28186	ITS	✓	✓	✓	✓	✓	✓	✓	265028	LINKVIDA	✓	✓	✓	✓	✓	✓
28269	COMPUS	✓	✓	✓	✓			✓	265499	GS TELECOM	✓	✓	✓	✓	✓	✓
28573	NET	✓	✓	✓	✓			✓	265947	INTERADIONET	✓	✓	✓	✓	✓	✓
52696	JSX TELECOM	✓	✓	✓	✓	✓	✓	✓	265998	CLICK NET	✓	✓	✓	✓	✓	✓
52720	Webfoco	✓	✓	✓	✓	✓	✓	✓	266045	ARTCOMPUS	✓	✓	✓	✓	✓	✓
52871	TASCOM	✓	✓	✓	✓			✓	266065	VANGUARDA Telecom	✓	✓	✓	✓	✓	✓
52872	Voanet	✓	✓	✓	✓	✓	✓	✓	266327	CentralSolicoes	✓	✓	✓	✓	✓	✓
52873	SOFTCOMP	✓	✓	✓	✓	✓	✓	✓	266345	IBT	✓	✓	✓	✓	✓	✓
52999	STARTNET	✓	✓			✓	✓	✓	266390	SingularCDN	✓	✓	✓	✓	✓	✓
53071*	Tecnoasp	✓				✓	✓	✓	266413	Viva Tecnologia	✓	✓	✓	✓	✓	✓
53075	Holistica	✓	✓	✓	✓	✓	✓	✓	266487	ISM TELECOM	✓	✓	✓	✓	✓	✓
53087	TELY	✓	✓	✓	✓	✓	✓	✓	266524	ONBAHIA.NET	✓	✓	✓	✓	✓	✓
53164	UFBA/PoPBA	✓	✓	✓	✓			✓	266551	CONEXAObA	✓	✓		✓	✓	✓
53236	ACESSO.NET	✓	✓	✓	✓			✓	266611	NOVAINFONET	✓	✓	✓	✓	✓	✓
53239	COGEL	✓	✓	✓	✓			✓	267050	RedelInfor	✓	✓	✓	✓	✓	✓
53242	Teletalk	✓	✓	✓	✓	✓	✓	✓	267209	PIRAJANET	✓	✓	✓	✓	✓	✓
61580	OpenCDN							✓	267217	SWINTERNET	✓	✓	✓	✓	✓	✓
61597	GalizaNET	✓	✓	✓	✓	✓	✓	✓	267320	AJNET	✓	✓	✓	✓	✓	✓
61727	Speedzone	✓	✓	✓	✓	✓	✓	✓	267369	MAFRELINE TELECOM	✓	✓	✓	✓	✓	✓
61798	PontoNet							✓	267613	Eletronet	✓	✓	✓	✓	✓	✓
61902	Bahialink Technology	✓	✓	✓	✓	✓	✓	✓	267964	RCSERVICOS	✓	✓	✓	✓	✓	✓
262668	News Tellecom	✓	✓	✓	✓	✓	✓	✓	268234	VALERI@Net	✓	✓	✓	✓	✓	✓
262687	PDN	✓	✓	✓	✓			✓	268393	LINKTECH	✓	✓	✓	✓	✓	✓
262769	GEEKNET	✓	✓					✓	268477	PROVEDORJUNIORWEB	✓	✓	✓	✓	✓	✓
262784	DSTECH	✓	✓					✓	268539	Intra Telecom	✓	✓	✓	✓	✓	✓

* ASN em ativação

Pontos de troca de tráfego

- ✓ Content Delivery Network (CDN)



Duvidas?

¿...?

Vamos ao roteiro de laboratório 4

Referências

- ✓ VAN BEIJNUM, Iljitsch. **BGP: Building reliable networks with the Border Gateway Protocol**. " O'Reilly Media, Inc.", 2002.
- ✓ BEZERRA, Jerônimo; GALIZA, Humberto. **TREINAMENTO EM REDES IP/MPLS MÓDULO 2 – ROTEAMENTO IP COM BGP**, 2013.
- ✓ BERTHOLDO, Leandro; TAROUCO, Liane. **BGP Avançado**. Rio de Janeiro: RNP/ESR, 2011.
- ✓ LOBATO, Luiz. **Protocolos de Roteamento IP**. Rio de Janeiro: RNP/ESR, 2013.
- ✓ MEDHI, Deep; RAMASAMY, Karthik. **Network routing: algorithms, protocols, and architectures**. Morgan Kaufmann, Second Edition, 2018.
- ✓ Quagga. Quagga Documentation (Standalone HTML page). URL <https://www.nongnu.org/quagga/docs/docs-multi/index.html>, 2019.
- ✓ BARRETO, Luiz; BOMFIM, Thiago. **Administração de Redes e Conectividade ao PoP-BA**, 2012.
- ✓ BAKER, F. **Requirements for IP version 4 routers**; RFC-1812. Internet Request For Comments, v. 1812, 1995.
- ✓ CISCO. **BGP Case Studies**. URL <http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/26634-bgp-toc.html>, 2012.
- ✓ CISCO. **BGP Best Path Selection Algorithm**. URL <http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13753-25.html>, 2012.
- ✓ HUSTON, Geoff. **Autonomous System (AS) Number Reservation for Documentation Use**; RFC-5398. Internet Request For Comments, v. 5398, 2008.
- ✓ Maila Networks. **Entendendo o BGP**. URL <https://suporte.maila.net.br/kb/faq.php?id=4>, 2019.
- ✓ Bates et al. **BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)**; RFC 4456. Internet Request For Comments, v.4456, 2006.
- ✓ BATES, T. et al. **RFC 4760: Multiprotocol Extensions for BGP-4**. IETF, January, 2007.
- ✓ CHEN, E. Scudder, J., Ed., Mohapatra, P., and K. Patel," **Revised Error Handling for BGP UPDATE Messages**. RFC 7606, DOI 10.17487/RFC7606, August 2015,< <http://www.rfc-editor.org/info/rfc7606>.