

Fundamentos-do-bgp

A proposta deste artigo é dimensionar alguns pontos fundamentais do BGP. Não será mostrado a configuração e muito menos troubleshooting de BGP. Na versão PDF, pode ser acessada clicando neste [link \(https://github.com/iagojonathas/fundamentos-do-bgp\)](https://github.com/iagojonathas/fundamentos-do-bgp).

OBS: Técnica de troubleshooting está sendo elaborado em outro artigo e o conteúdo sobre *BGP attributes* também está em produção.

Índice

O QUE É BGP

CARACTERÍSTICA DO BGP

BGP routing table vs IP routing Table

FUNCIONAMENTO DO BGP

MENSAGEM OPEN

MENSAGEM UPBATE

MENSAGEM NOTIFICATION

MENSAGEM KEEPALIVE

ESTADOS DE VIZINHANÇA DO BGP (FSM)

PRINCIPAIS CAUSAS DE ESTÁGIO ACTIVE

QUANDO USAR E NÃO O BGP

RECURSO DISPONÍVEL

CONCLUSÃO

REFERÊNCIA BIBLIOGRÁFICA



SOBRE AUTOR

O QUE É BGP

BGP é um protocolo de roteamento baseado em política, mas precisamente o BGP se comporta mais como uma aplicação de roteamento do que protocolo de roteamento em si. BGP troca NLRI (Network Layer Reachability Information) entre dois *speakers BGP*. NLRI é uma lista de redes que são compostas por endereço IP e a máscara, ou prefixo em caso do IPv6. NLRI é totalmente conhecido e válido tanto que os prefixos só podem trocados se estiverem presente na IP routing table de um router BGP.

O BGP se situa na borda do ASN, permitindo que o protocolo BGP trace um mapa de conectividade de ASN. Por isso, em algumas literaturas, o BGP é categorizado como um protocolo *ASpath to ASpath* ou *Path Vector*, em suma é uma aplicação interdomain. A classificação do BGP como aplicação respalda-se no fato de ele utilizar a porta de destino 179 para o estabelecimento de uma sessão TCP e os vizinhos não **precisam** estar diretamente conectado para formar uma vizinhança, ao contrário dos protocolos IGP's, tais como: EIGRP, OSPF, ISIS e RIP; devem estar diretamente conectado para forma uma adjacência.

Após, o estabelecimento da sessão TCP, o BGP inicia seu processo de convergência realizando a troca de mensagem entre os peers envolvidos (processo a ser discutido em mais detalhes no decorrer do artigo).

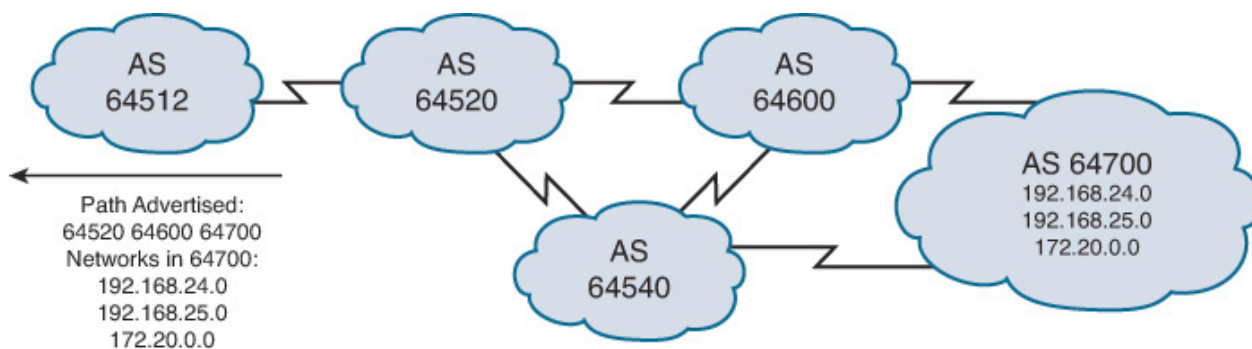
O BGP oferece suporte a várias address family IPv4/IPv6 unicast e multicast, dispondo também de recursos para a utilização de aplicações como *l2vpn, nsn, rtfilter*, VPNv4 e VPNv6. Ou seja, o BGP não foi desenvolvido apenas para IPv4 ou IPv6. Por causa, disso  45.182.104.22  as indústrias de telcos e em algumas empresas corporativas, e podemos afirma que o protocolo é uma

APLICAÇÃO DE ROTEAMENTO.

CARACTERÍSTICA DO BGP

Como dito anteriormente, o BGP é um protocolo Path Vector, que elabora um gráfico de mapeamento por todos ASNs de onde o prefixo foi passado e mitiga looping de roteamento, um exemplo pode ser ilustrado na imagem abaixo:

> Construção de Gráfico de ASN:



FONTE: Implementing Cisco IP Routing (ROUTE) Foundation Learning Guide - Chapter 7

Na perspectiva do ASN 64520 (imagem acima) constrói um gráfico de ASN por onde o prefixo foi deslocado, para ele alcançar os prefixos contidos no ASN 64700, precisa passar por ASNs:

- 64520 64600 64700
- 64520 64540 64700

Então, o speaker BGP do ASN 64520 armazena esses dados na sua BGP routing table para selecionar o melhor caminho, dependendo do vendor sendo usado pode ter entre 10 a 11 critérios para determinar o melhor caminho (discutido no artigo *BGP Attributes*). Após, o BGP escolher o melhor caminho, esse caminho é ofertado na IP routing table e em seguida o router pode encaminhar pacotes aprendido via BGP. O router BGP consegue fazer esse gráfico graças ao atributo AS-PATH presente na mensagem UPDATE. O formato da mensagem UPDATE será discutido ao longo desse material.

Lembra do primeiro seção onde um speaker BGP só envia NLRI se estiver na IP routing table? Embora, que a BGP routing table tenham informações analítica sobre NLRI recebido do seu vizinho, o router não enviará essa tabela para o seu vizinho adjacente e envia informações presentes na IP routing table. Ou seja, o 64512 terá um caminho possível na sua BGP routing table:

- 64512 64520 64600 64700

Caso o ASN 64512 queira uma BGP routing table mais robusta e/ou sofisticada, eliminando o *single point of failure* (ponto único de falha) pode contratar outra telco ou ISP, formando uma topologia *single dualhomed* (demostrado na imagem ao lado) Com isso, a BGP routing table fica:

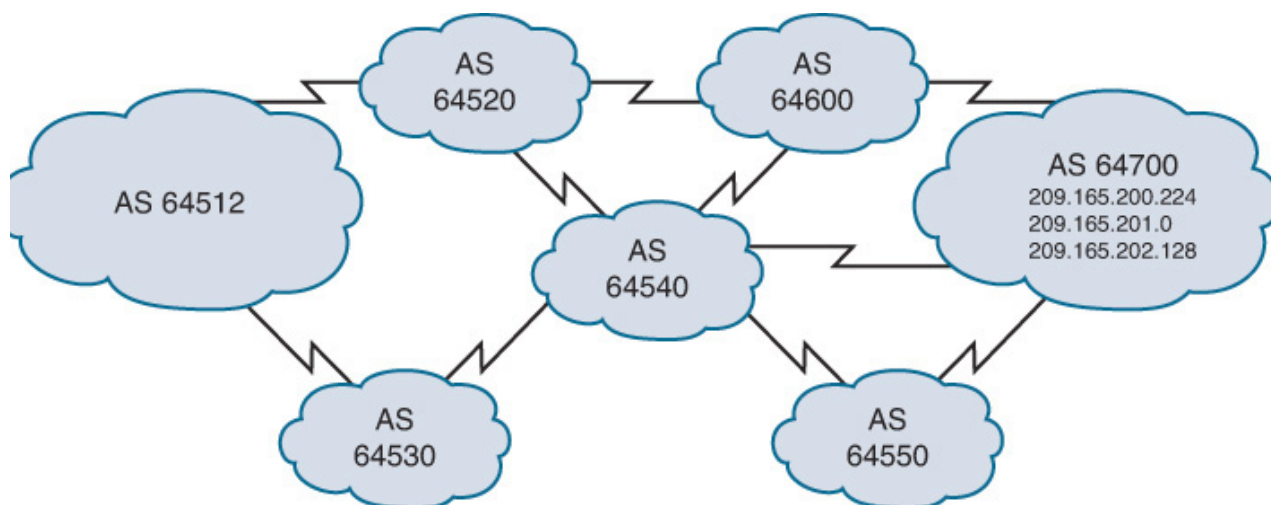
- 64512 64520 64600 64700
- 64512 64530 64540 64700

Vale enfatizar que o BGP é uma aplicação de roteamento, permitindo-a realizar roteamento assimétrico, ou seja tanto *upload* quanto *download* pode vir por rotas distintas, por exemplo, o ASN 64512 pode ter um upload via ASN 64520 e o download via ASN 64530 dependendo da sua política de roteamento, para mais informações de como influenciar o upload e download no seu vizinho, pode consultar este link (https://wiki.brasilpeeringforum.org/w/O_Minimo_que_Voce_precisa_saber_sobre_o_BGP).

OBS: BGP routing table e IP routing table serão discutido ao longo deste artigo!

> Mitigação de looping:

O nome DV (*distance vector*) é derivado do fato de que as rotas são anunciadas como vetores (distance, vector), onde a distance é



ONTE: Implementing Cisco IP Routing (ROUTE) Foundation Learning Guide - Chapter 7

definida em termos de uma métrica e a vector é definida em termos do roteador next-hop. Por exemplo, "O destino A está a uma **distance** de 5 saltos, no **vector** do roteador X next-hop". Como essa declaração implica, cada roteador aprende rotas a partir das perspectivas dos roteadores vizinhos e, em seguida, anuncia as rotas a partir de sua própria perspectiva. Como cada roteador depende de seus vizinhos para obter informações, o que os vizinhos podem ter aprendido com seus vizinhos, e assim por diante, o roteamento do vetor de distância às vezes são chamados de "roteamento por rumor", segundo ciscopress.com (<https://www.ciscopress.com/articles/article.asp?p=24090&seqNum=3>). Se você observar minuciosamente o speaker BGP não passa a informação topológica sobre da rede BGP (BGP routing table) e um speaker confia nas informações recebida do seu neighbor (IP routing table). Constituindo o conceito **Distance Path** (DP). Vejamos a diferença entre eles:

- **Distance Vector**: Distance = quantidade de salto; Vector = roteador;
- **Path vector**: Path = Atributos e Vector = ASN.

Então, podemos afirma que o BGP é semelhante ao RIP, mas com algumas particularidades e melhorias!

Ambos protocolos respeitam a regra **split horizon** evitando looping ou flapping de roteamento, por exemplo, na última imagem tem ASN 64700 e ele envia um update com NLRI 209.165.200.224 aos neighbors simultaneamente, como o BGP é um DP, o ASN 64550 recebe um update constando NLRI 209.165.200.224 e repassa isso para todos neighbors adjacente, exceto o neighbor que enviou. O ASN 64550 envia NLRI 209.165.200.224 para o ASN 64540, no entanto, o ASN 64540 **aceitará** esse update, pois o atributo as-path não tem o seu próprio ASN e manterá essa informação na BGP routing table. Todavia, o ASN 64540 enviará esse NLRI para o neighbor 64700 e o mesmo vai **descarta** porque no atributo as-path já consta seu próprio ASN; mitigando a possibilidade de looping ou flapping em toda topologia BGP.

BGP routing table vs IP routing Table

BGP routing table é uma estrutura analítica do BGP, onde um roteador fazem uma análise complexa de 10 a 11 critérios, dependendo do fornecedor sendo usado. IP routing table é a tabela de roteamento que um roteador possa encaminhar tráfego para redes de destino. Assumindo que a IP routing table é uma tabela confiável elaborado por protocolo de roteamento, o BGP usa essa tabela para publicar os NLRIs aos seus vizinhos.

A tabela BGP pode ter seguintes nomes:

- BGP table;
- BGP topology table;
- BGP topology database;
- BGP routing table;
- BGP forwarding database.

OBS: BGP routing table e IP routing table têm outras relevâncias quando a ISP trabalha com atributos *atomic aggregate* e *aggregator*, porém está além do escopo deste artigo.

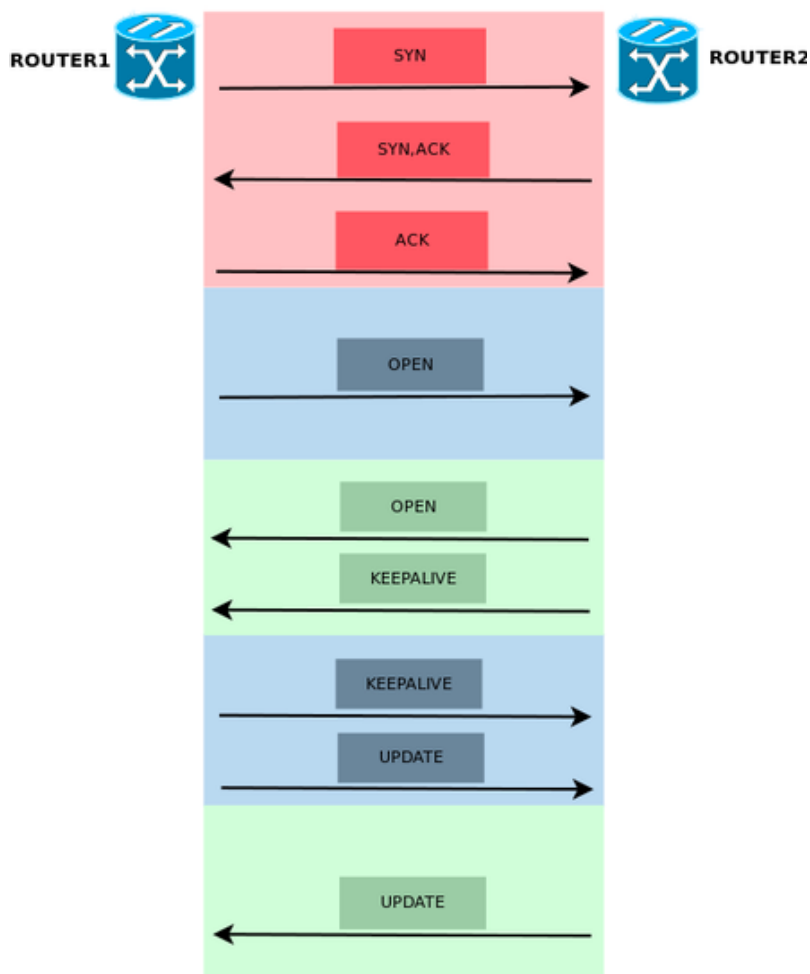
FUNIONAMENTO DO BGP

Assim como qualquer outro protocolo de redes, o BGP utiliza mensagens para realizar a sua convergência. O BGP usa quatro mensagens, elas são open, keepalive, update e notification. E na camada de transporte do modelo OSI, usufrui da confiabilidade e rapidez, a característica mais significativa do TCP, em termos de troca de rotas entre dois peerings. As trocas de mensagens são trocadas após de estabelecer uma sessão TCP na porta 179.

Como TCP deixa o processo de convergência do BGP mais rápido? Protocolo de roteamento do tipo IGP, como OSPF, não é aplicação de roteamento, atuando diretamente na camada de rede do modelo OSI e possuindo o seu próprio protocolo na camada de transporte. No entanto, tem *window one-for-one*, ou seja, se um roteador OSPF precisa enviar mais de uma LSU para o seu vizinho, ele precisa enviar apenas um update e esperar um LSAck. Se não receber um LSAck, ele enviará o primeiro update até receber um LSAck. Também, único update do OSPF suporta no máximo 10.000 rotas. Portanto, se a sua rede corporativa for de 100.000 rotas OSPF, o roteador precisa enviá 10 updates. Isso é muito ineficiente e demorado se você tiver uma grande rede, como a IP routing table global têm 840.791 rotas, segundo o site *cidr report* na data de 01/07/2020. OSPF precisaria enviar aproximadamente 84 updates. Deixando-a a convergência menos eficiente em qual tange 840.791 rotas.

Já o BGP precisa lidar com a 840.791 rotas, com a função TCP implementada na camada de transporte do modelo OSI; quebrando o paradigma *window one-for-one* e usa *sliding window*, permitindo que único update possa enviar milhares de rota, deixando-a muito mais eficiente e eficaz.

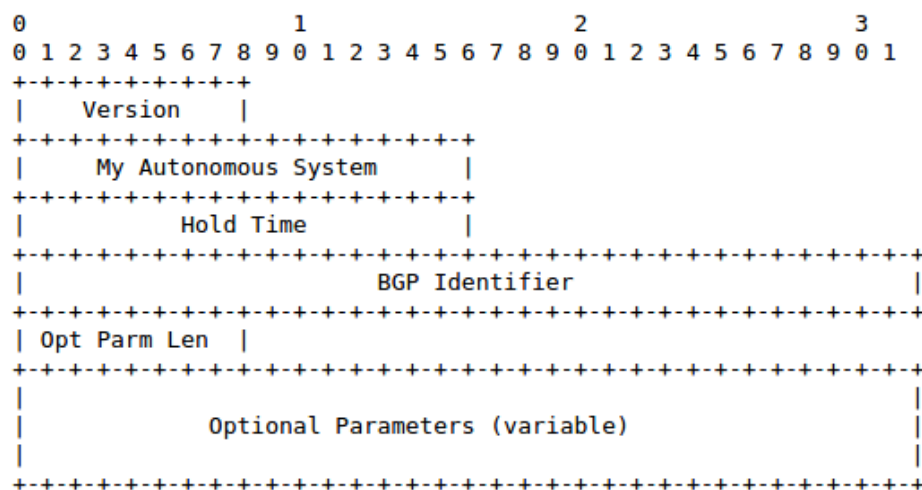
OBS: as mensagens do BGP são trocados APÓS do estabelecimento da sessão BGP, como mostrado na imagem a seguir.



Trocas de mensagens BGP, extraído do wireshark

MENSAGEM OPEN

Mensagem open, é a primeira mensagem a ser trocada assim de estabelecer uma sessão TCP. Se uma mensagem open é aceitável, uma mensagem keepalive é enviado como resposta. No cabeçalho da mensagem open possuem seguintes campos:



FONTE: RFC 4271 - Página 12

- **Version:** campo de 8-bit indica o número da versão mensagem BGP. Atualmente o número da

versão é o 4 (BGP-4);

- **My Autonomous System:** campo de 16-bit indica o ASN do remetente;
- **Hold Time:** campo de 16-bit indica o maior número de segundos que pode decorrer entre a mensagem keepalive ou update sucessivamente pelo remetente. Com um aceite de uma mensagem open, o roteador BGP deve calcular o valor do hold time a ser usado com o seu neighbor, por padrão, o hold time são de 180 segundos. Se o hold time local de um roteador é menor do que holdtime mínimo, a relação de vizinhança não pode ser formada;
- **Version: BGP Identifier:** um campo composto por 32-bit, identificando o router-id do speaker BGP, um endereço IPv4 é constituído por 32-bit e por isso é usado. A designação de 32-bit tem a seguinte ordem: router-id explícito dentro da configuração; o maior endereço IPv4 de uma interface loopback ativa; maior endereço IPv4 de qualquer interface física ativa;
- **Version: Optional Parameters:** identifica o tamanho total do comprimento do optional parameters. Esse parâmetros são Type, Length e Value (TLV). Um exemplo de um optional parameters é o fator de autenticação.

MENSAGEM UPDATE

A mensagem update tem a finalidade de trocar informação de roteamento entre os peerings BGP e da mesma forma construir um gráfico em que os prefixos já foi passado.

Uma mensagem update é usado para propagar uma rota ou prefixo presente na tabela de roteamento e com os atributos desse prefixo. A mensagem update sempre terá um tamanho fixo no cabeçalho BGP e também campos de informações, como demonstrado abaixo.

Withdrawn Routes Length (2 octets)
Withdrawn Routes (variable)
Total Path Attribute Length (2 octets)
Path Attributes (variable)
Network Layer Reachability Information (variable)

FONTE: RFC 4271 - Página 14

Atenção: alguns desses campos podem estar ausente em todas mensagens updates. Por isso que enfatizamos os principais campos, para mais informações pode consultar a RFC 4271! A justificativa da ausência será discutida no próximo artigo sobre BGP attributes.

- **Withdrawn Routes:** é um valor variável que consta uma lista de prefixo de endereço IP para as rotas são withdran "retirada" do serviço . O valor 0 indica que não há nenhuma rota sendo retirada do serviço e o campo Withdrawn Routes não está sendo constatado nessa mensagem update;
- **Path Attributes:** uma sequência de atributos do BGP sendo transportado nesse campo, tais como: AS-path, origin, local preference, next hop e assim por diante. Cada atributo pode ter atributo type, atributo length e atributo value (TLV). O atributo type há flags, seguido por código do atributo, de acordo com [iana.com \(https://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml\)](https://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml) :

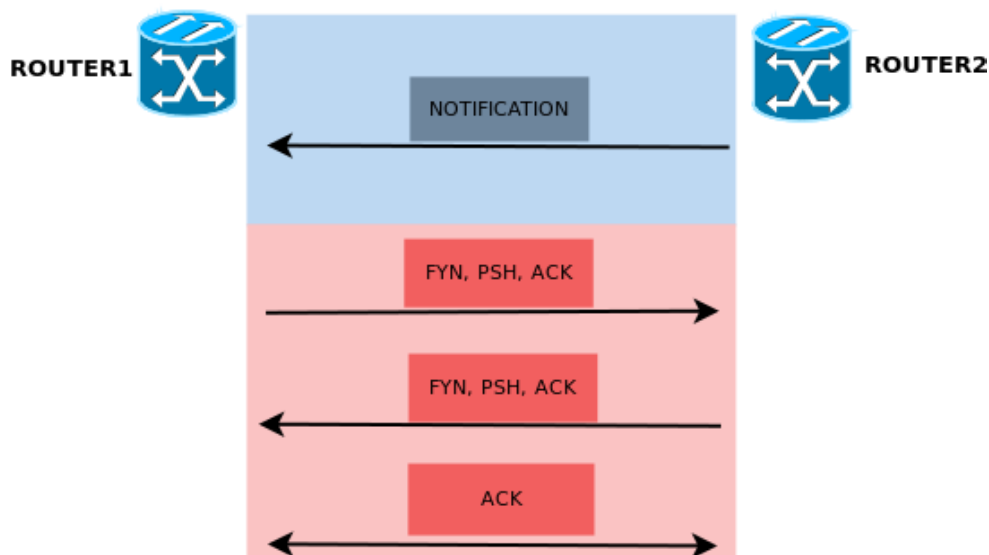
1. ORIGIN (Type code 1);
2. AS-PATH (Type code 2);
3. Next Hop (Type code 3);
4. MED (Type code 4);
5. Local Prefer (Type code 5);
6. Atomic Aggregate (Type code 6);
7. Aggregator (Type code 7);
8. Community (Type code 8);

9. Originator_ID (Type code 9);

- **Network Layer Reachability Information:** Uma lista de redes que pode ser alcançada por esse update.

MENSAGEM NOTIFICATION

É enviado quando há uma condição de erro ou um router BGP restabeleceu a sua sessão TCP e a sessão é encerrada **imediatamente**. Um exemplo do processo notification:



Processo de Mensagem Notification

MENSAGEM KEEPALIVE

Keepalive têm duas funções no BGP. Uma é manter a sessão TCP estabelecida por padrão, é enviada a cada 60 segundos. E atua como resposta da mensagem open.

ESTADOS DE VIZINHANÇA DO BGP (FSM)

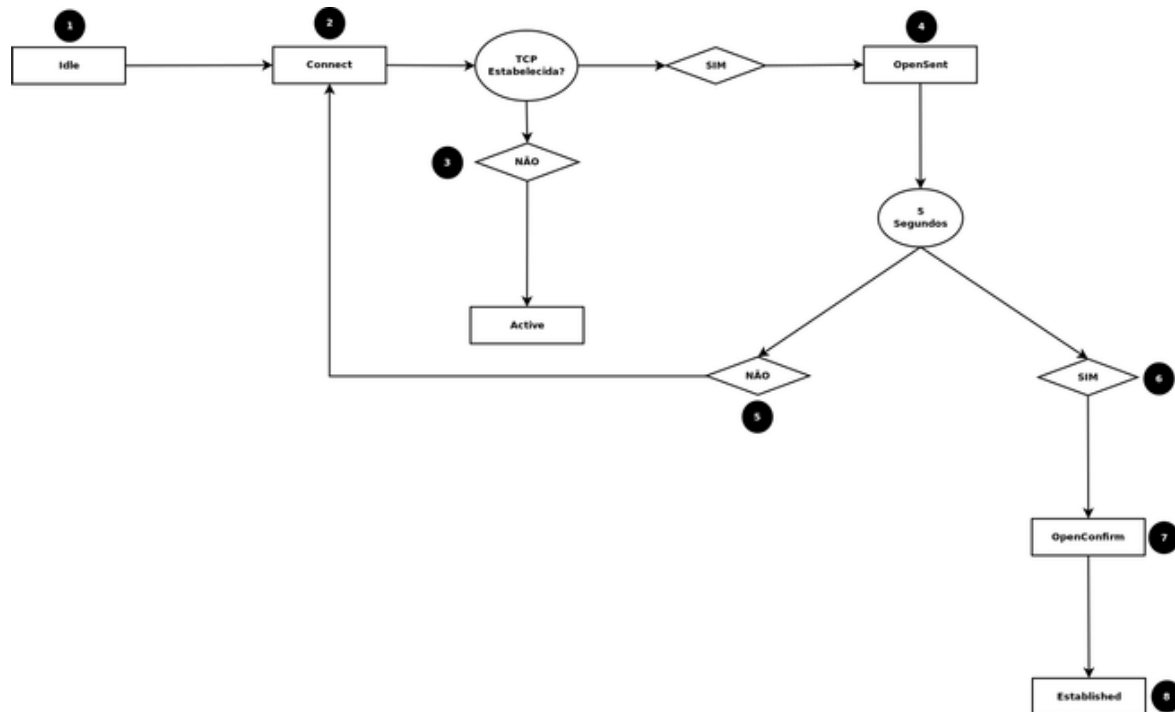
BGP é protocolo *Finite State Machine*, ou seja, precisa passar por vários estágio com o seu neighbor. *Finite State Machine* em curta palavra é: "o estágio atual em um FSM é determinado pelos estágios anteriores e pelas operações executadas para fazer a transição entre os estágios. Uma transição de um estágio para outro depende do sucesso ou fracasso de uma operação", segundo a [cisco.com](https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/ts/guide/UCSTroubleshooting/UCSTroubleshooting_chapter_010.pdf) (https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/ts/guide/UCSTroubleshooting/UCSTroubleshooting_chapter_010.pdf). Então, o BGP desloca por vários estágio ou estados até alcançar a sua convergência completa/established.

Os estágios de vizinhança são:

- **Idle:** o roteador procura uma rota para o endereço IP do neighbor configurado, em sua tabela de roteamento e envia um SYN;
- **Connect:** o roteador encontra uma rota para o neighbor e finaliza o three-way handshake TCP;
- **Open Sent:** uma mensagem Open foi enviada, com os parâmetros da sessão;
- **Open Confirm:** o roteador recebeu um aceite nos parâmetros para estabelecer uma sessão;
- **Established:** os speakers BGP estão trocando informação de roteamento - envio e recebimento de mensagem Update;
- **Active:** um problema persiste no peering BGP e está impactando do roteador de avançar para os próximos estágios.

Como podemos observar um flowchart / fluxograma dos estágios state machine do protocolo BGP, na imagem subsequente. E uma

numeração para descrever cada estágio.



FSM do BGP

1. Quando o administrador insere o endereço IP do neighbor dentro do processo ID BGP. o FSM do BGP fica em idle e nesse estágio o roteador começa procurar uma rota em sua tabela de roteamento, do endereço IP configurado. Se tiver uma rota o roteador envia um SYN e passa para estágio Connect em poucos segundos ou milésimos de segundos, se não tiver rota em sua tabela de roteamento o estágio permanecerá em Idle;
2. Após de encontrar uma rota e enviar um SYN, entrará no estado Connect. Nesse estágio espera que o three-way handshake TCP na porta 179 seja completada;
3. Se o roteador não receber um SYN, ACK do seu neighbor, o estágio vá para o Active;
4. Após do um speaker BGP estabelecer uma sessão TCP, o roteador local precisa enviar uma mensagem Open (já discutido anteriormente neste artigo) e quando o speaker BGP já enviou uma Open, entra no estágio Open Sent;
5. Se ele não receber uma mensagem Open como resposta do primeiro Open dentro de 5 segundos, o estágio é movido para o Idle;
6. No entanto, se o roteador local receber um Open dentro de 5 segundos, o estágio FSM do BGP é movido para o Open Confirm;
7. Open Confirm começa a escanear a tabela de roteamento dos caminhos para enviar para o neighbor e nesse estágio o BGP espera receber uma mensagem keepalive ou notification;
8. Um vez, alcançando o estágio Established, os peering BGP começa trocar rotas (prefixos ou NLRI) entre si. Ou seja, há trocas de updates.

PRINCIPAIS CAUSAS DE ESTÁGIO ACTIVE

Se um roteador BGP alcançou esse estágio quer dizer que o mesmo tem uma rota para o endereço IP do seu vizinho, todavia, é que o vizinho não tenha uma rota para o endereço IP para a origem, certifica-se que o roteador tenha uma rota presente na tabela de roteamento do speaker BGP.

Um outro problema muito corriqueiro é quando o roteador consegue estabelecer uma sessão TCP na porta 179 (estágio Connect) e o roteador envia uma mensagem Open, porém o speaker BGP de destino não tem configuração para a origem ou tem algum parâmetro errado do BGP, como por exemplo, endereço IP do peering, ASN, autenticação, holdtimer da mensagem configurado erroneamente.

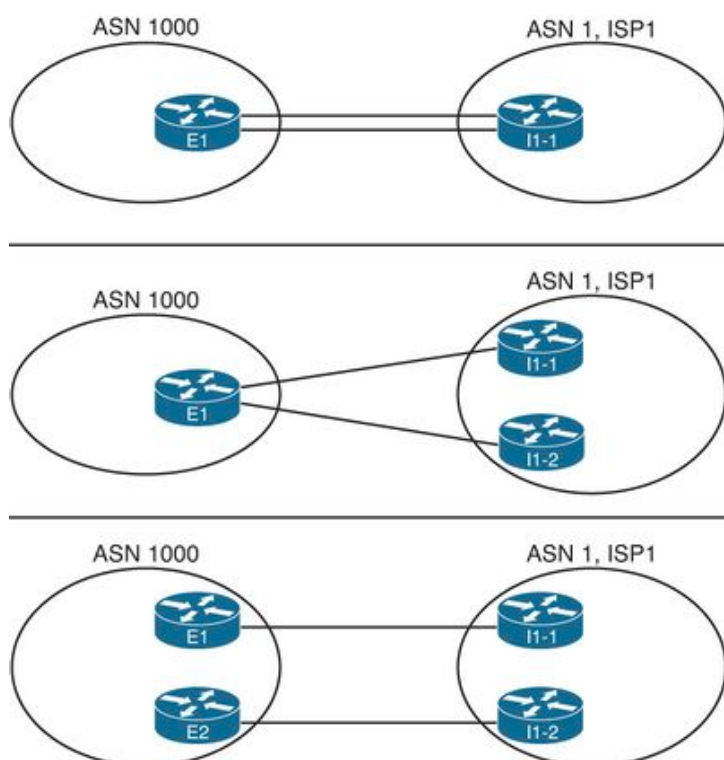
QUANDO USAR E NÃO O BGP

Adoção ou não adoção do BGP, depende de muitos aspectos e do que a empresa precisa, mas o fator crucial é quantidade link ou tipos de conexões:

- Single Homed;
- Dual-Homed;
- Single-Multihomed;
- Dual-Multihomed.

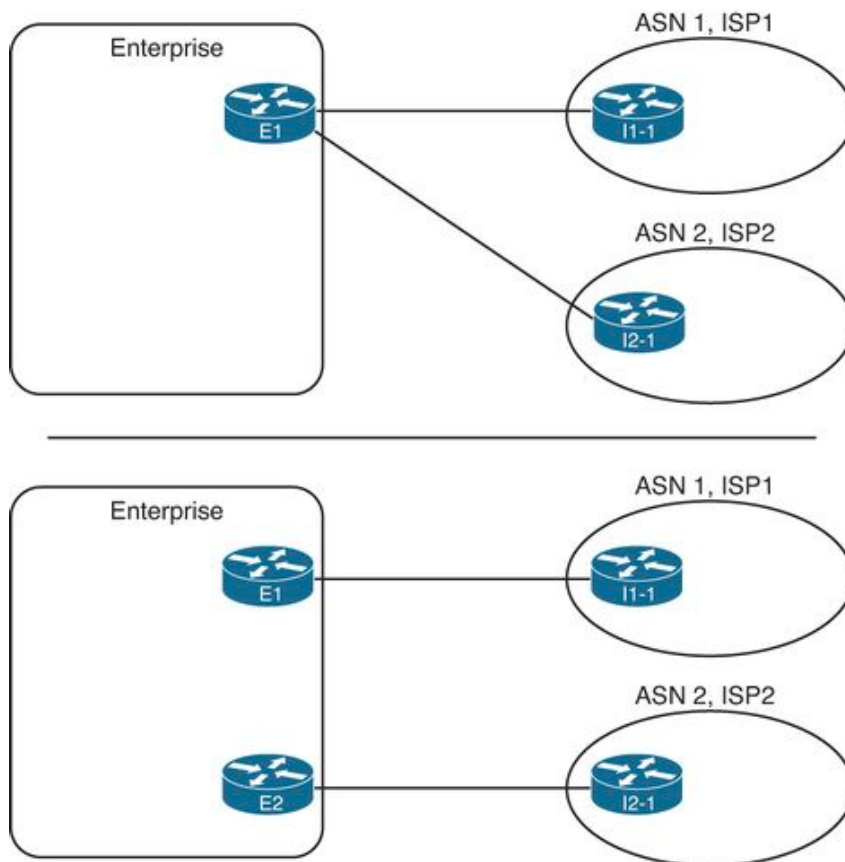
Single-Homed: é um tipo de conexão que só tem um caminho de saída para internet, não precisa executar o BGP. Uma rota default é mais do que suficiente. Single homed é conhecido como rede *stub*;

Dual-homed: a empresa ou a telco tem caminho redundante por mais de um link ou por mais de dois roteadores, porém para o mesmo ISP, se a ISP cair toda o acesso a internet é paralisada. Esse tipo de rede, pode não precisar do BGP em si, mas se torna uma tarefa desafiante para o administrador de rede, podemos ver uma topologia na imagem a seguir:



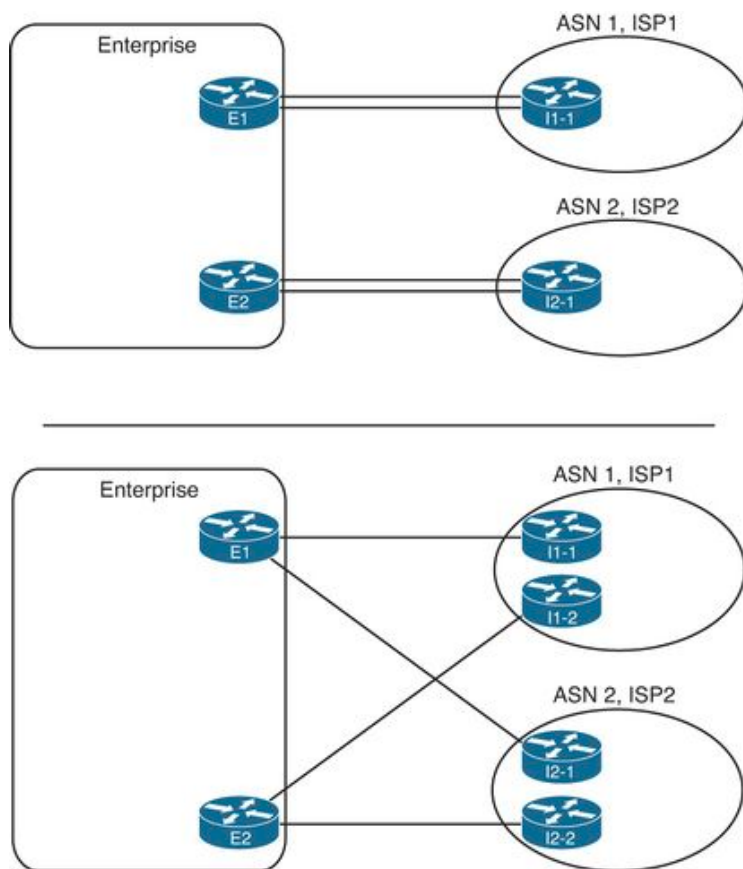
CCNP Routing and Switching ROUTE 300-101 Official Cert Guide- Chapter 13

Single-MultiHomed: um link para vários ISPs (pelo menos dois), como demonstrado na imagem abaixo. Aqui fica impossível ter caminhos redundantes e o acesso a internet ficar ininterrupto usando apenas rota default, nesta topologia é necessário a implementação do BGP;



CCNP Routing and Switching ROUTE 300-101 Official Cert Guide-
Chaper 13

Dual-Multihomed: uma topologia de rede que a empresa possui vários link de conexão para vários telco, se tornando imaginável usar rota default, o uso do BGP se torna essencial nessa rede para fazer a uma rápido *reroute*.



RECURSO DISPONÍVEL

Mapa menta e versão em PDF: <https://github.com/iagojonathas/fundamentos-do-bgp>

CONCLUSÃO

O BGP não é apenas um protocolo de roteamento simples, pelo ao contrário ele se estende para diversos protocolos, cujo é conhecido como MP-BGP (Multiprotocol BGP). Também, foi desenvolvida diversas features (next-hop-self, update-source, eBGP multihop, eBGP multipath e dentre outros) para afirmar que seu comportamento é de uma aplicação. O BGP é um protocolo **manualizado** que é super bom, pois o analista de rede consegue ter um total controle desse magnífico protocolo e eles (analistas) podem usar esses filtros para manipular o que ser transmitido ou não, e influencia o fluxo de tráfego de outro ASN.

Esse artigo foi escrito o mais breve possível e deixando o conceito pontuais, cirúrgico e transparente. Então, alguns aspectos não foram mencionadas por se tratar de uma linguagem mais técnico, necessitando-a de um estudo mais robusto. Entretanto, é o básico para quem administra um ASN.

Espero ter colaborado para a wiki e seja útil - para dúvida, esclarecimento e elogio só acionar no [linkedin](https://www.linkedin.com/in/iagojonathas/) (<https://www.linkedin.com/in/iagojonathas/>) ou [telegram](https://t.me/iagojonathas) (<https://t.me/iagojonathas>) e ABS!

“E lembre-se: você é seu próprio general. Então, tome agora a iniciativa, planeje e marche decido para a vitória”- Sun Tzu - A Arte da Guerra

REFERÊNCIA BIBLIOGRÁFICA

FUNCIONAMENTO DO BGP:

rfc:4271;

O QUE É BGP, FUNCIONAMENTO DO BGP e ESTADOS DE VIZINHANÇA DO BGP (FSM):

Implementing Cisco IP Routing (ROUTE) Foundation Learning Guide: (CCNP ROUTE 300-101);

QUANDO USAR E NÃO USAR O BGP:

CCNP Routing and Switching ROUTE 300-101 Official Cert Guide.

SOBRE AUTOR

Autor: [Iago Jonathas \(https://www.linkedin.com/in/iagojonatas/\)](https://www.linkedin.com/in/iagojonatas/)

Contatos: [Telegram \(https://t.me/iagojonathas\)](https://t.me/iagojonathas) ou email: iagojonathas.ij@gmail.com

Disponível em "<https://wiki.brasilpeeringforum.org/index.php?title=Fundamentos-do-bgp&oldid=2625>"

Esta página foi modificada pela última vez em 20 de julho de 2020, às 20h16min