RIP, OSPF, BGP

Diego Passos

6 de Junho de 2017

1 Roteamento Intra-AS

O roteamento Intra-AS, também conhecido como IGP (do inglês, *Interior Gateway Protocols*), oferece flexibilidade de escolha aos ASs. Como o protocolo de roteamento Intra-AS é responsável apenas pelas rotas internas, cada AS é livre para escolher o protocolo que julgue mais adequado. Desta forma, não é necessária qualquer padronização em relação a este protocolo.

No entanto, há um subconjunto relativamente pequeno de protocolos que são, de fato, utilizados na prática para o roteamento Intra-AS. Os protocolos mais populares/famosos nesta categoria são o RIP (*Routing Information Protocol*), o OSPF (*Open Shortest Path First*) e o IGRP (*Interior Gateway Routing Protocol*). Ao contrário dos outros dois, o IGPR é um protocolo proprietário da CISCO. Embora o IGRP seja bastante popular, nesta disciplina daremos maior enfoque ao RIP e ao OSPF, por serem alternativas abertas (no sentido de terem suas especificações publicamente conhecidas).

1.1 RIP

O RIP é um protocolo de roteamento antigo, tendo sido especificado originalmente na RFC de 1058 de 1988. Suas primeiras implementações, no entanto, são ainda mais antigas, incluindo seu surgimento no BSD-UNIX em 1982.

O RIP é um protocolo de roteamento baseado em Vetor de Distâncias. Como métrica de roteamento, o RIP utiliza o número de saltos entre origem e destino. Em outras palavras, o custo de cada enlace operacional é sempre a constante 1, independentemente de outras possíveis características de qualidade. Adicionalmente, o RIP limita os caminhos encontrados a um máximo de 15 saltos (*i.e.*, um custo de 16 corresponde a "infinito").

Por padrão, um roteador executando o RIP anuncia seu Vetor de Distâncias a cada 30 segundos. Cada anúncio pode carregar as distâncias do roteador até 25 destinatários diferentes. Cada destinatário corresponde, na verdade, a uma sub-rede alcançável por um dos roteadores da rede, ao invés de dizer respeito a um nó específico da topologia.

O RIP também utiliza os mesmos pacotes de anúncio de vetor de distâncias para detectar enlaces. Em particular, quando um roteador recebe um Vetor de Distâncias por um enlace, ele reconhece aquele enlace como operacional. Por outro lado, se um roteador deixa de receber anúncios em um enlace por pelo menos 180 segundos, o enlace é declarado inoperante. Neste caso, o roteador que detectou este evento deve rever suas decisões de roteamento, invalidando qualquer rota que este havia selecionado utilizando aquele enlace e tentando encontrar caminhos alternativos. Caso este evento resulte em uma mudança no Vetor de Distâncias, o roteador adicionalmente reenvia seu novo Vetor de Distâncias aos seus vizinhos.

Assim como qualquer protocolo baseado em Vetor de Distâncias, o RIP está susceptível ao problema da contagem ao infinito. Para mitigar este problema, o RIP utiliza a técnica de envenenamento reverso.

Uma particularidade de implementação do RIP é que, embora ele seja um protocolo de roteamento — e, portanto, inerente à camada de rede —, o RIP é normalmente implementado na camada de aplicação através de um *daemon* chamado `route-d`. O `route-d` abre um *socket* UDP na porta 520 e o utiliza para enviar e receber anúncios de Vetores de Distância. Todo o processamento dos Vetores de Distância é realizado por esta aplicação que, ao descobrir ou atualizar uma rota, realiza chamadas de sistema ao núcleo do SO requisitando inserções/alterações na tabela de roteamento do sistema. Embora, à rigor, esta seja uma violação do modelo em camadas, há outros exemplos de protocolos de roteamento que operam desta maneira.

1.2 OSPF

O OSPF é um protocolo mais moderno que o RIP, apresentando uma série de funcionalidades avançadas detalhadas a seguir. A letra 'O' na sua sigla significa '*Open*', no sentido de que sua especificação é

publicamente disponível. Em sua versão mais recente, a versão 3, o OSPF é especificado na RFC 5340 de 2008.

Uma diferença fundamental entre o OSPF e o RIP é o fato do OSPF ser baseado em Estado de Enlaces. Isso significa que os anúncios de informações de roteamento utilizados pelo OSPF carregam o estado dos enlaces da vizinhança do roteador, havendo uma entrada para cada vizinho. Além disso, o anúncio de um roteador deve ser propagado por toda a rede (*i.e.*, todo o AS), inundando-a. Também ao contrário do RIP, o OSPF transmite suas mensagens diretamente sobre o IP — ao invés de utilizar um protocolo de camada de transporte.

Um outro protocolo bastante similar ao OSPF é o IS-IS (do inglês, *Intermediate System to Intermediate System*). Ambos são baseados em Estado de Enlaces e disponibilizam conjuntos similares de funcionalidades avançadas. O IS-IS, no entanto, é mais genérico, no sentido de sua especificação e implementação não estar atrelada ao IPv4, como ocorre no caso do OSPF. Enquanto o IS-IS é bastante utilizado para roteamento no *backbone* de ISPs, o OSPF é mais popular para redes institucionais. Dada a semelhança entre ambos os protocolos, nesta disciplina estudaremos em mais detalhes apenas o OSPF.

Uma das funcionalidades avançadas do OSPF é um mecanismo de **autenticação**. As mensagens de controle do OSPF podem passar por um controle de autenticidade, com o objetivo de determinar se aquele conteúdo foi, de fato, gerado por um roteador legítimo do AS. Com isso, torna-se mais difícil — ao menos efetivo — o chamado *ataque de buraco negro*, no qual um atacante poderia inserir um nó na rede se passando por um roteador legítimo anunciando informações falsas de roteamento e, com isso, atraindo o tráfego da rede para si que, posteriormente, é descartado.

O OSPF também suporta o chamado *roteamento multi-path*. A ideia é que, se há múltiplos caminhos de mesmo custo disponíveis entre dois roteadores da rede, ao invés de escolher apenas um, o OSPF permite que todos sejam utilizados. Uma utilidade para isso é a realização de balanceamento de carga — o roteador poderia rotear pacotes diferentes com um mesmo destinatário por cada um dos caminhos disponíveis, reduzindo a carga sobre cada caminho individualmente.

Outra funcionalidade interessante do OSPF é o suporte a diferenciação de tráfego. O protocolo permite o uso de métricas de roteamento diferentes para tipos de tráfego diferentes (de acordo com o campo ToS do cabeçalho IPv4). Assim, o conceito de *melhor rota* pode ser adequado às necessidades de cada tipo de tráfego. Por exemplo, um enlace de satélite — que geralmente tem alta capacidade, porém alta latência — pode ter um peso mais elevado para tráfego com fortes requisitos temporais (como, por exemplo, VoIP) do que para tráfego de melhor esforço.

Outra funcionalidade já presente no OSPF é o seu suporte nativo a roteamento *multicast* — *i.e.*, para comunicações em grupo. Trataremos de roteamento *multicast* em aulas futuras, mas, por hora, basta compreender que as mesmas informações usadas para o roteamento que temos estudado até agora — o *unicast* — são reaproveitadas pelo OSPF para tomar decisões acerca do roteamento *multicast*. Se este não fosse o caso, e, ao contrário, usássemos dois protocolos distintos para *unicast* e *multicast*, é possível que houvesse uma duplicação do *overhead* de roteamento na rede.

Por fim, talvez a funcionalidade avançada mais importante do OSPF é o seu suporte a **roteamento em áreas**. Assim como o roteamento hierárquico da Internet divide o problema de determinar os caminhos em duas porções — Inter- e Intra-AS —, o OSPF é capaz de dividir um grande AS em áreas distintas. Cada área é composta por um conjunto de roteadores interconectados e seus respectivos enlaces. As informações de roteamento são restritas às suas respectivas áreas, tanto em termos de propagação (*i.e.*, os pacotes de anúncio do OSPF relacionados a uma determinada área nunca são propagados para roteadores de outras áreas), quanto em termos de conteúdo (*i.e.*, os anúncios em uma área carregam apenas informações dos enlaces que fazem parte daquela área).

Um roteador pode pertencer, simultaneamente, a mais de uma área OSPF. Neste caso, este roteador passa a ser chamado de *roteador de borda de área*. Os roteadores de borda de área são responsáveis por interconectar áreas diferentes e, portanto, participam do roteamento em todas as áreas às quais pertencem. Estes roteadores descobrem as rotas alcançáveis através de uma área e as anunciam para os outros nós das suas outras áreas. Assim, roteadores de áreas diferentes conseguem descobrir rotas entre si.

Esta funcionalidade de divisão do OSPF em áreas tem como objetivo principal reduzir a escala do problema de roteamento, resultando em menor *overhead* de pacotes de controle na rede. Em geral, define-se uma das áreas como a **área de *backbone***, que interconecta todas as demais áreas.

2 Roteamento Inter-AS: BGP

Ao contrário do roteamento Intra-AS, no roteamento Inter-AS é necessária a padronização na escolha do protocolo de roteamento. Desta necessidade, a Internet convergiu para o estabelecimento do BGP

(*Border Gateway Protocol*) como seu padrão para roteamento Inter-AS.

O BGP pode ser dividido em duas componentes básicas: o eBGP (*External* BGP), responsável pela troca de informações de roteamento entre *gateways* de Borda de ASs vizinhos, o iBGP (*Internal BGP*), responsável pela propagação dos anúncios de rota recebidos de um AS vizinho para os roteadores internos do AS. Através dos anúncios recebidos dos vizinhos, um AS é capaz de aplicar **políticas e critérios** para escolher "boas" rotas para destinatários na Internet. Além disso, um AS utiliza o BGP para anunciar suas sub-redes para o resto da Internet, permitindo, assim, que estas sub-redes sejam inseridas nas tabelas de roteamento dos roteadores do resto da Internet e que datagramas a elas destinados possam ser encaminhados com sucesso.

No cerne do funcionamento do BGP está o conceito de **sessão BGP**. Uma sessão BGP é apenas um jargão para denotar uma conexão TCP estabelecida entre dois roteadores para a troca de informações de roteamento através do BGP. Repare, portanto, que o BGP é implementado na forma de uma aplicação rodando sobre *sockets* TCP. Cada roteador que executa o BGP deve ser configurado para conhecer seus **pares BGP** — *i.e.*, outros roteadores com os quais deve estabelecer uma sessão BGP para trocar informações de roteamento. Um exemplo típico são dois roteadores *gateways* de borda que interconectam dois ASs vizinhos: eles normalmente estabelecerão uma sessão BGP através da qual realizarão anúncios um ao outro de prefixos de sub-rede para os quais cada AS é encaminhar pacotes.

O BGP é dito um protocolo baseado em "Vetor de Caminhos". O conceito é similar ao de um protocolo baseado em Vetor de Distâncias, mas, ao invés de enviar apenas as distâncias para cada destinatário conhecido, uma rota BGP contém a sequência de ASs percorridos. Isso tem algumas vantagens simples, como a possibilidade de evitar o problema de contagem ao infinito (já que um AS facilmente poderia detectar fazer parte do caminho anunciado pelo seu AS vizinho para um dado destinatário). No entanto, a maior utilidade disso para o BGP está em permitir o chamado **roteamento baseado em políticas** (explicado em detalhes mais adiante). Note que, por outro lado, anúncios de rotas no BGP têm o potencial de se tornarem muito grandes, dada a necessidade de listar toda a sequência de ASs percorridos. Uma contramedida para isso é a capacidade do BGP de agregar prefixos próximos em um mesmo anúncio de rota, se valendo da natureza hierárquica do endereçamento na Internet.

Outra diferença do BGP para outros protocolos de roteamento estudados até aqui é o fato de um AS não ser obrigado a anunciar todas as rotas que conhece para seus vizinhos. De fato, como discutiremos a seguir, um AS pode decidir não anunciar rotas para certos destinatários a um ou mais de seus ASs vizinhos. Por outro lado, se um AS anuncia ao seu vizinho uma rota para um determinado prefixo — jargão do BGP para denotar uma sub-rede — **ele está se comprometendo a encaminhar datagramas do seu vizinho para este prefixo**.

Quando um *gateway* de borda recebe o anúncio de uma nova rota a partir de um AS vizinho — o que é feito via eBGP — ele deve propagar esta informação para os demais roteadores do AS. Esta propagação é realizada através do iBGP. A partir do recebimento desta informação pelo iBGP, os demais roteadores do AS avaliam a "qualidade" desta rota e, de acordo com vários critérios, decidem o próximo salto para o destinatário em questão. Além disso, um outro *gateway* de borda do AS, ao receber este anúncio via iBGP *pode* anunciar a nova rota descoberta para outro AS vizinho.

Estes anúncios de rota no BGP carregam uma série de informações. Em particular, uma *rota* no BGP é representada por um prefixo de sub-rede e uma série de atributos. Um exemplo de atributo já discutido é o chamado `AS-PATH`, que é a lista de ASs percorridos pela rota em questão. Outro atributo importante é o `NEXT-HOP`, que armazena o endereço IP da interface de rede através da qual um *gateway* de borda anuncia uma rota para um AS vizinho via eBGP. Quando este anúncio é propagado para os roteadores internos de um AS via iBGP, o atributo mantém seu valor. Para um roteador interno, portanto, o atributo `NEXT-HOP` **armazena o endereço IP do primeiro salto da rota externo ao AS**. Para um roteador interno, portanto, este `NEXT-HOP` **não corresponde ao próximo salto que constará na tabela de roteamento**. O roteador precisará combinar esta informação com as informações do roteamento intra-AS para, de fato, determinar a interface de saída adequada para esta rota.

Repare que como as sessões BGP são transportadas por conexões TCP, não há risco de perda de um anúncio — a menos, é claro, que a conexão TCP seja interrompida, mas neste caso os pares BGP serão avisados da falha e poderão se recuperar de acordo. Com isso, ao contrário de protocolos de roteamento que atual sobre UDP ou diretamente sobre IP, não é necessário que os anúncios de rotas sejam repetidos periodicamente sobre uma sessão BGP. Embora isso possa não parecer uma grande vantagem, é preciso considerar que a quantidade de prefixos anunciados em uma sessão BGP é grande e tem aumentado rapidamente ao longo dos anos. Além disso, lembre-se que uma rota BGP é uma informação composta por vários campos, incluindo uma lista completa de ASs percorridos. Logo, cada prefixo anunciado é, individualmente, grande — ao menos em comparação com as entradas de um Vetor de Distância em um protocolo deste tipo. Logo, não precisar enviar todos os anúncios várias

vezes ajuda a manter o funcionamento do BGP viável na prática.

2.1 Roteamento Baseado em Políticas

Suponha que um roteador de um AS receba múltiplos anúncios de rotas diferentes para alcançar um mesmo prefixo. Como ele deve escolher entre estas várias rotas disponíveis? Ao contrário dos outros protocolos estudados anteriormente, **o BGP baseia fortemente suas decisões de roteamento em políticas**, que nem sempre estão diretamente alinhadas com critérios de desempenho. Estas políticas muitas vezes são determinadas por acordos comerciais — ou pela ausência desses. Por exemplo, um ISP pode decidir que não é interessante para seu AS anunciar certas rotas aos ASs vizinhos, a menos que esses façam um acordo comercial estabelecendo algum tipo de compensação financeira pelo tráfego encaminhado. Um AS pode, também, possuir algum tipo de restrição por encaminhar tráfego através de algum outro AS particular da Internet — por exemplo, desavenças políticas podem levar ASs de um país a querer evitar direcionar seu tráfego por ASs de outro país rival.

Voltando à questão original, quando um *gateway* de borda recebe um anúncio de rota vindo do seu par no AS vizinho, ele utiliza estas políticas pré-configuradas para introduzir — ou alterar o valor de — um atributo numérico chamado `LOCAL_PREFERENCE`. Uma vez configurado o valor deste atributo, o anúncio é disseminado para os demais roteadores do AS. Quando múltiplas rotas para um mesmo prefixo são recebidos por um roteador, o primeiro critério é justamente esse: seleciona-se a rota associada ao maior valor de `LOCAL_PREFERENCE`.

É possível, no entanto, que haja um empate: duas ou mais rotas apresentam o mesmo nível de `LOCAL_PREFERENCE`. Neste caso, o BGP aplica os seguinte critérios, em ordem:

1. `AS-PATH` **mais curto**. Escolhe-se a rota que passa pelo menor número de ASs. Note que, embora este critério tenha uma motivação ligada a desempenho (*i.e.*, minimizar o número de ASs percorridos), **ela não garante que a rota selecionada tenha o menor número de saltos**, já que o `AS-PATH` não especifica a exata sequência de roteadores percorridos em cada AS.

2. **Roteador** `NEXT-HOP` **mais próximo**. Este é o chamado **roteamento batata quente**, brevemente discutido na aula anterior. Cada roteador do AS faz a escolha particular pela rota que tira o datagrama "mais rapidamente" do AS local. Para tomar esta decisão, o roteador utiliza as rotas encontradas pelo roteamento Intra-AS, comparando os custos das melhores rotas encontradas para cada um dos *gateways* de borda utilizados para alcançar o `NEXT-HOP` das rotas BGP.

3. **Critérios adicionais**. Se todos os critérios anteriores falharem, as implementações do BGP ainda são flexíveis o suficiente para permitir uma gama de outros critérios de desempate.

O BGP utiliza também políticas para determinar se deve ou não anunciar certas rotas para ASs vizinhos. Considere, por exemplo, a seguinte situação. Uma grande instituição contrata enlaces redundantes de conexão com a Internet com dois ISPs, A e B. Digamos que os roteadores de borda da instituição executem BGP, estabelecendo sessões eBGP com os *gateways* de borda dos respectivos ISPs. Suponha que através de uma das sessões eBGP o ISP A anuncia para a instituição uma rota para um certo prefixo x na Internet. Assim, a instituição passa a conhecer uma rota para a sub-rede x através do ISP A. Se o BGP anunciasse rotas a ASs vizinhos indiscriminadamente, a instituição anunciaria esta rota recém-aprendida para o ISP B, correndo o risco de que B resolvesse utilizá-la para encaminhar seus pacotes para x.

Neste exemplo, esta situação é obviamente indesejada para a instituição: ela contratou o ISP B para encaminhar seus pacotes, e não o contrário. Neste caso, o administrador da rede da instituição poderia simplesmente configurar seus roteadores de borda para aceitar anúncios de rota vindos dos ISPs, mas enviar rotas para eles — exceto as rotas para as próprias sub-redes da instituição.

Considere agora uma outra situação: três ISPs, A, B e C, se interconectam (*i.e.*, há enlaces entre A e B, B e C, e C e A). Suponha que A possua um AS cliente w. Para que w seja visível para o resto da Internet, o AS A precisa anunciar rotas para os prefixos de w para seus ASs vizinhos — em particular, B e C. Uma vez que B receba este anúncio, ele tem uma escolha: ele deve anunciar para C que sua rota para w? Repare que se B o fizer, ele está se comprometendo a realizar o encaminhamento de pacotes de C para w. Entretanto, qual é a motivação para que B o faça? Estes pacotes não foram gerados pelos clientes de B e não são destinados aos clientes de B. Logo, a menos que haja um acordo comercial estabelecendo algum tipo de compensação por este encaminhamento, pode ser do interesse de B não fazer este encaminhamento, forçando C a encontrar algum caminho alternativo.

3 Roteamento Inter-AS *vs.* Intra-AS

Voltemos uma última vez à discussão sobre as vantagens em se utilizar o roteamento hierárquico na Internet. Já discutimos que a questão da escalabilidade é importante: através do roteamento hierárquico, dividimos o problema de roteamento na Internet em dois sub-problemas menores, resolvidos isoladamente e combinados, resultando na tabela de roteamento de cada roteador. Mas por que precisamos de protocolos de roteamento diferentes? Por que não utilizar um mesmo protocolo para roteamento Inter- e Intra-AS?

É importante notar que os requisitos destes dois problemas são bastante diferentes. Quando discutimos as rotas internas a um AS, normalmente buscamos a maior eficiência possível, tanto em termos das rotas encontradas, quanto em termos do processo utilizado para tanto — *i.e.*, do próprio protocolo de roteamento. Assim, protocolos de roteamento tradicionais, baseados em Vetor de Distâncias ou Estado de Enlaces, que utilizam métricas de roteamento baseadas em algum aspecto de desempenho da rede são as opções imediatas.

Por outro lado, quando falamos de roteamento entre ASs distintos, as prioridades são, muitas vezes, diferentes. Como ASs diferentes são normalmente administrados por entidades diferentes, acordos comerciais e políticas de relacionamento comumente se sobrepõem às métricas puramente baseadas em desempenho. Um protocolo como o RIP, por exemplo, que simplesmente busca minimizar o número de saltos — ou, neste contexto, de ASs — percorridos por uma rota, não seria capaz de capturar estes requisitos. Assim, o BGP com toda a sua flexibilidade e roteamento baseado em políticas se torna bem mais adequado, apesar de sua maior complexidade.