# ARTIGO Nº 10

# Fundamentos de Segment Routing

**1 – OBJETIVO**

O objetivo deste artigo é apresentar, em linhas gerais, a nova tecnologia de rede comutada modo pacote denominada *Segment Routing* (SR). Trata-se de um novo paradigma da tecnologia de comutação modo pacote baseado em *source routing*, que objetiva otimizar, simplificar e adicionar valor a outras tecnologias de comutação modo pacote, cujos exemplos já explicitados são as diferentes opções de MPLS e o IPv6.

A padronização de *Segment Routing*, a cargo do IETF (*IETF SPRING Working Group*), encontra-se ainda em fase incipiente, tendo sido emitida pelo IETF praticamente apenas a RFC 8402, em julho de 2018, estando todos os demais padrões ainda sob a forma de *RFC drafts.*

As vantagens imediatas de *Segment Routing* para certas aplicações são tão evidentes, que, conforme informações de fornecedores, não foi possível aguardar o final da padronização da tecnologia pelo IETF, obrigando os fornecedores a antecipar a sua implementação.

Duas aplicações se destacam nesse ponto. Na interconexão de *Data Centers,* a forma simples e escalável com que SR define diferentes caminhos assumiu grande significado. Os maiores *Data Centers* mundiais, como o próprio Google, utilizam ou planejam utilizar *Source Routing*.

Uma outra condição observada, é a enorme importância por SR na implementação da tecnologia 5G, com a grande expectativa da crescente implantação de SR pelos Provedores de Serviço (SPs) em âmbito mundial.

A nossa intenção é escrever, em breve futuro, um tutorial mais detalhado sobre *Segment Routing*, dando tempo para que a sua padronização evolua e se consolide.

Pretendemos também antes concluir a série de tutoriais dedicados a aplicações do GMPLS, complementando a série já escrita sobre conceituação, padronização e operação no/do GMPLS que se encontra integralmente publicada no portal *WirelessBrasil.*

Pretendemos, ao mesmo tempo, elaborar uns poucos artigos introdutórios sobre algumas novas tecnologias, a exemplo de EVPN (*Ethernet Virtual Private Network*) e de VXLAN (*Virtual Extensible LAN*), com o propósito de orientar os leitores, a despeito de nossas flagrantes limitações.

Neste tutorial*,* usamos figuras retiradas dos seguintes documentos: *Introduction to Segment Routing* (da Cisco), *Introduction to Segment Routing* (Youssef El Fathi), *Segment*

*Routing (SR) and Traffic Engineering (TE),* da Juniper e *Fundamentals of Egress Peering Engineering* (da Juniper).

## 2 - INTRODUÇÂO

A comutação IGP das redes IP (IPv4 e IPv6) processa-se no paradigma de comutação modo pacote descentralizada, onde todos os roteadores da rede operam um algoritmo de definição de rotas que conduzem ao destino dos quadros e hospedam as resultantes tabelas de roteamento. O algoritmo normalmente é o SPF, que conduz à definição do melhor (mais curto) para alcançar o destino do quadro, ressalvada a possibilidade de aplicação de ECMP.

Como vimos em nosso tutorial Nº 1, o MPLS apresenta-se sob três diferentes opções, que são o MPLS Básico (protocolos OSPF/IS-IS e LDP), o MPLS-TE (protocolos OSPF-TE/ISIS-TE e RSVP-TE) e o MPLS-TP (protocolos GMPLS OSPF-TE/GMPLS ISIS-TE e GMPLS RSVP-TE).

No caso do MPLS Básico, o roteamento IGP, com base no OSPF ou no IS-IS, ocorre da mesma forma descentralizada que nas redes IP. Isso difere do comportamento das redes IP pela aplicação adicional do protocolo de sinalização LDP, também descentralizado, quando são atribuidos e distribuídos os labels para os diferentes links da rede MPLS, do que resultam os LSPs constituidos nessa rede.

O roteamento IGP no MPLS-TE e no MPLS-TP obedece a diferentes paradigmas de roteamento e de sinalização, que embora apresentem um certo grau de roteamento centralizado, mantêm um número de processos descentralizados.

No roteamento IGP, as condicionantes (*constraints*) aplicáveis aos links da rede são configuradas descentralizadamente nos LSRs de toda a rede, onde são captadas e transportadas nas LSAs opacas dos protocolos de roteamento para os PEs da rede. É nos PEs, na qualidade de *head end LSRs*, que ocorrem então funções centralizadas, como o armazenamento das informações de roteamento nas TEDs e realização, ou o controle, da função PCE responsável pelo cálculo do caminho mais indicado para um dado LSP.

Quando da sinalização, cabe ao *head end LSR* iniciar o processo de sinalização, emitindo mensagens *RSVP Path* para a constituição dos LSPs, mensagens essas onde é inserido o objeto ERO de cada um dos LSPs.

O objeto ERO, embora especifique o caminho a ser utilizado pelo LSP, não é suficiente para o encaminhamento de quadros de dados nesse caminho, servindo apenas como um atributo para o processo de sinalização.

Esse processo de sinalização opera descentralizadamente como no LDP, sendo os labels também atribuidos e distribuídos por todos os LSRs da rede MPLS.

A grande diferença dos RSVP-TE e GMPLS RSVP-TE para o LDP é a possibilidade de utilização de engenharia de tráfego (TE) no cálculo de caminhos, do que resulta a possibilidade de explicitação de caminhos (pelo objeto ERO) que melhor atendam aos requisitos estabelecidos para cada um dos LSPs.

O algoritmo normalmente utilizado para o cálculo de caminhos no MPLS-TE e no MPLS-TP (controlado pelo GMPLS) é o CSPF (*Constrained SPF*), que consiste basicamente na aplicação sucessiva das condicionantes válidas para o LSP sobre o melhor caminho inicialmente calculado.

Uma análise cuidadosa das alternativas de tecnologia de rede acima apresentadas, que, em maior ou menor proporção utilizam paradigmas de controle descentralizado, evidencia os seguintes pontos negativos:

- Verifica-se uma enorme carga operacional na rede, com a utilização de múltiplos protocolos de controle;

- Verifica-se a utilização de uma multiplicidade de tabelas no interior da rede, com o registro de diferentes estados, representando certamente um grau mais elevado de sofisticação operacional;

- Essas tabelas nos roteadores da rede contêm diferentes atributos, como por exemplo a disponibilidade e a utilização de labels, registros das identificações de LSPs, etc., o que certamente conduz à questão da disponibilidade desses atributos e à consequente preocupação com a escalabilidade na rede;

- As funções realizadas descentralizadamente nos roteadores no interior das redes causam maior latência na transmissão dos quadros de dados e mais ineficiência no uso dos recursos das redes. A descentralização nessas redes representa inevitavelmente um grau de dificuldade para a implementação de SDN (*Software Defined Networking*).

Surge nesse ponto a seguinte pergunta: Existe alguma possível nova tecnologia que evite esses pontos negativos em redes comutadas modo pacote? A resposta é positiva, e essa nova tecnologia se denomina *Segment Routing* (SR).

## 3 – SEGMENT ROUTING

*Segment Routing* (SR) é uma nova tecnologia de comutação em redes modo pacote, cuja arquitetura foi definida pelo IETF na RFC 8402 (*Segment Routing Architecture*), com base na essência do conceito de *source routing,* efetivado de forma flexível e escalável.

Um roteador em uma rede *Segment Routing* é capaz de selecionar qualquer caminho nessa rede e inseri-lo no cabeçalho dos quadros a percorrê-lo, caminho esse codificado sob a forma de uma lista ordenada de identificadores de segmento. Para a indicação desse caminho é suficiente a manutenção de estados definidores do fluxo exclusivamente no(s) nó(s) de ingresso ao domínio SR ou em um controlador SR centralizado (inclusive um controlador SDN).

A RFC 8402 não impõe qualquer condição quanto à forma pela qual um controlador SR central se comunica com a rede SR. As opções citadas nessa RFC são o NETCONF (*Network Configuration Protocol*), o PCEP (*Path Computation Element Communication Protocol*) ou o BGP-LS (*BGP Link State*).

A Figura 1 exibe uma configuração de uso de um controlador SDN em uma rede SR.

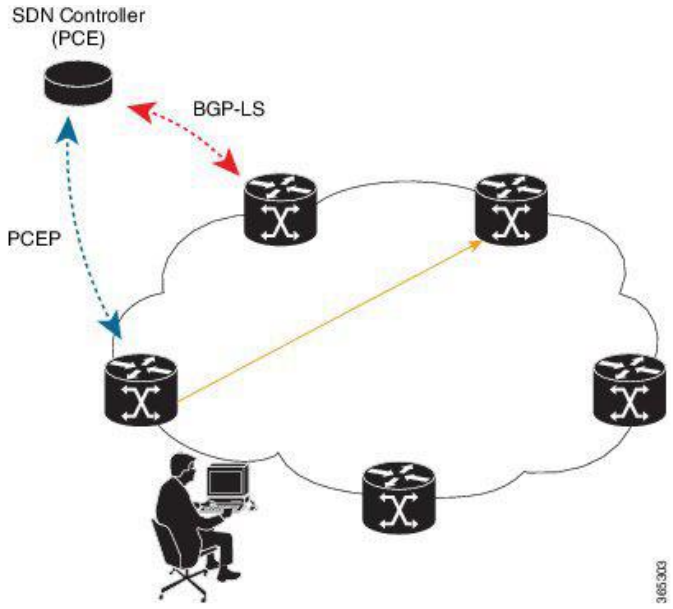


**Figura 1 – Exemplo de rede SR controlada por SDN.**

No cenário dessa figura, onde se constata o uso do PCEP ou do BGP-LS, um roteador pode solicitar ao controlador SDN um caminho com determinadas características, como, por exemplo, com uma determinada banda e com valores máximos tolerados de delay e de taxa de perda de quadros. O controlador SDN, utilizando uma função PCE, calcula um caminho otimizado e retorna a correspondente lista de segmentos para o roteador.

O roteador encontra-se então em condições de iniciar a transmissão de tráfego nesse caminho, sem a necessidade de qualquer sinalização adicional na rede.

Segmentos são qualquer tipo de instrução destinada à explicitação do caminho sendo percorrido, com base topológica ou com base em serviços. Os identificadores de segmento são referidos como SIDs (*Segment IDs*).

É importante registrar que em *Segment Routing*, assim como no IP/MPLS (MPLS Básico e MPLS-TE), os caminhos traçados são unidirecionais. Para o atendimento bidirecional é necessária a aplicação de duas instâncias de protocolo em sentidos inversos. Somente no MPLS-TP, controlado pelo GMPLS, é possível a constituição de caminhos (LSPs, no caso) bidirecionais em uma única instância de protocolo.

Registramos também que a RFC 8402 se limita à aplicação de SR para unicast, não especificando qualquer conceito ou aplicação de SR em multicast.

### 3.1 – Operação de Segment Routing

Um segmento pode possuir uma semântica com significado local para um nó SR, quando é identificado por um identificador de segmento adjacência (*Adjacency Segment ID*, referido como *Adj-SID*) ou uma semântica com significado global no domínio SR, quando é identificado por identificador de segmento nodal (*Node Segment ID,* referido como *Node-SID*) por dizer respeito às identificações dos roteadores da rede SR.

A Figura 2 ilustra os conceitos de *Adjacency Segment ID* (*Adj-SID*) e de *Node Segment ID* (*Node-SID*).

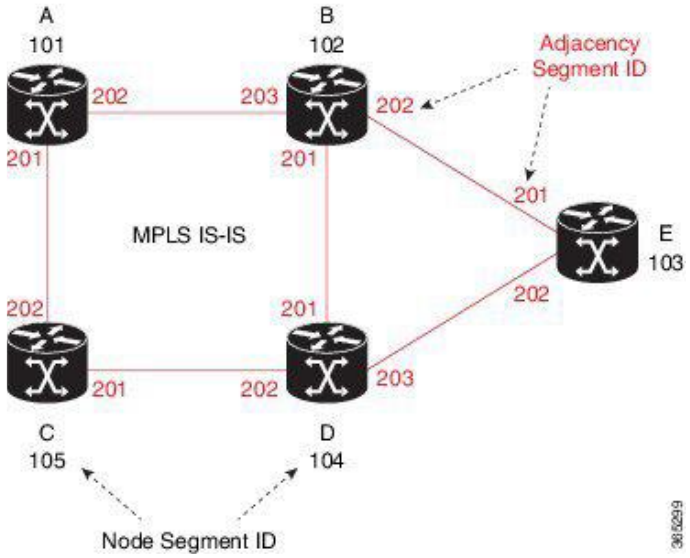


**Figura 2 – Adjacency Segment ID e Node Segment ID.**

Os valores de *Adj-SID* e de *Node-SID* são transportados em extensões do protocolo IGP ou do protocolo BGP.

A classificação acima, embora clara e didática, representa uma forma básica e parcial para classificar os segmentos e os respectivos SIDs. Apresentaremos, adiante neste tutorial, a classificação completa conforme a RFC 8402.

Os segmentos globais, como, por exemplo, os segmentos nodais, têm os seus valores de SID em um domínio SR definidos dentro de um conjunto de valores estabelecidos. Esse conjunto é referido como SRGB (*SR Global Block*).

Cada roteador possui o seu SRGB, no qual define o seu *Node-SID.* É recomendável, contudo, que os roteadores do domínio SR possuam um único SRGB.

Os segmentos adjacências, por sua vez, de significado local para cada um dos nós SR do domínio, têm os seus SIDs limitados a conjuntos locais de valores referidos como SRLBs (*SR Local Blocks*). Se um nó SR participa de múltiplos domínios SR, define-se um SRLB para cada um desses domínios SR.

*Segment Routing* fundamenta-se em um pequeno número de extensões dos protocolos IGP (OSPF e IS-IS) e do protocolo BGP-LS, destinadas essencialmente à transmissão de métricas e de SIDs. Abordaremos essas extensões, em linhas gerais, adiante neste tutorial.

Não existe a necessidade de utilização de protocolos de sinalização em *Segment Routing*, como o LDP e RSVP-TE, nem a necessidade de registro de estados ao longo da rede, sendo a comutação baseada apenas nas instruções contidas nos próprios quadros transmitidos.

Os caminhos mais curtos para alcançar cada um dos roteadores, no entanto ficam definidos e registrados ao longo da rede SR, como nos protocolos IGP tradicionais

A transmissão ocorre com base em uma lista ordenada de SIDs nodais e de SIDs locais, em sequências válidas para cada caso.

Como os caminhos mais curtos para alcançar qualquer nó da rede SR encontram-se pré-estabelecidos no domínio SR, um quadro em um nó contendo um segmento nodal como o segmento ativo (segmento a ser imediatamente processado) é transmitido pelo menor caminho para atingir o nó correspondente a esse segmento nodal. Aplica-se, quando for o caso, ECMP.

Para que um quadro tenha que ser transmitido de um nó por uma determinada interface, é necessário que o segmento ativo seja um *Adj-SID,* que indica essa interface*.* Isso significa, em outras palavras, que os *Adj-SIDs* são um instrumento para a aplicação de TE em SR.

Em configurações mais amplas de uso de *Segment Routing,* envolvendo comunicação entre ASs e normalmente utilizando TE, os SIDs podem ser distribuídos por meio do BGP.

**3.2 – Utilização de Segment Routing em Diferentes Planos de Dados**

*Segment Routing* pode ser aplicado para várias arquiteturas de rede sem qualquer alteração no plano de dados dessas redes. São exemplos os planos de dados de redes MPLS e de redes IPv6, cujas utilizações se encontram em adiantado estágio de definição no IETF.

No caso do MPLS, referido como SR-MPLS, em processo de definição no *IETF Draft* intitulado *Segment Routing with MPLS data plane*, que se encontra na 22ª versão (01/05/2019), não tendo assumido ainda, por alguma razão, a forma de RFC.

No SR-MPLS, os segmentos são codificados como labels MPLS no formato genérico. Uma lista ordenada de segmentos é codificada como uma pilha (*stack*) desses labels.

O segmento ativo em uma dada operação é aquele identificado pelo *top label* da pilha de labels, ou seja, pelo label mais próximo da camada física. Após a utilização de um segmento, o correspondente label é descartado da pilha.

No caso do IPv6, referido como SRv6, utiliza-se um novo tipo de cabeçalho de roteamento IPv6, que passa a conter a lista de segmentos, agora codificados como endereços IPv6. Esse cabeçalho é referido como *IPv6 SR Header* (IPv6 – SRH).

A utilização do IPv6-SRH, que constitui a base do SRv6, está sendo especificada no IETF Draft intitulado IPv6 Segment Routing Header, cuja última versão (26ª versão) foi emitida em 22/10/2019.

O segmento ativo no SRv6 é indicado pelo endereço IPv6 de destino do quadro. O próximo segmento ativo é indicado por um apontador no IPv6 – SRH, apontador esse denominado *SegmentsLeft* (SL) *pointer.* Quando um segmento ativo é utilizado, o

apontador SL é decrementado, e o próximo segmento é copiado no DA. Quando o próximo pacote é transmitido, o correspondente IPv6 – SRH é acrescido ao pacote.

### 3.3 – Considerações Gerais

*Segment Routing* facilita o provimento de proteção automática de tráfego sem qualquer restrição topológica e sem a necessidade de processos adicionais de sinalização, em razão de seu modo de operação. Essa proteção pode também ocorrer mediante a utilização da metodologia FRR (*Fast Rerouting*) definida originalmente para o MPLS.

Embora se trate de uma recente tecnologia, a concepção básica de *Segment Routing* remonta à da tecnologia *Source Routing Bridging* (SBR), da qual faremos uma breve recapitulação a seguir.

### 3.4 – Source Routing Bridging (SBR)

*Source Routing Bridging* (SRB) é uma tecnologia baseada em um método para a definição de caminhos para a transmissão direcionada de quadros MAC unicast entre diferentes LANs, método esse baseado no encaminhamento por rotas inseridas nos cabeçalhos desses quadros pelas estações finais de origem. Dito em outras palavras, encaminhamento por *source routing.*

Esse método foi desenvolvido pela IBM, sendo adotado pelo comitê IEEE 802.5 na especificação de LANs *Token Ring*.

No algoritmo SRB, uma estação final de uma *SRB bridged LAN* que deseja transmitir tráfego unicast para uma dada outra estação final, após se certificar que essa estação final de destino não se encontra na mesma LAN *Token Ring* que ela (exclusão essa que ocorre pelo resultado do envio de um quadro unicast para a estação final de destino), envia um quadro broadcast *Route Explorer* também endereçado à estação final de destino.

O quadro *Route Explorer* é enviado com a inserção de um campo RIF (*Routing Information Field*) inicialmente vazio, sendo que campo RIF contém uma sucessão de descritores de rota.

Como o *Route Explorer* passa por todas as bridges da rede, cada uma dessas bridges preenche o respectivo RIF, de modo que todas as estações finais das LANs, que não a LAN de origem, recebem múltiplas mensagens contendo campos RIF preenchidos com as descrições de rota registradas até alcançar cada uma delas.

Como os caminhos de ida e de volta são os mesmos nesse tipo de rede, a estação final de destino endereçada pode retornar outros quadros *Route Explorer* para a estação de origem, agora unicast, com a indicação completa de descritores de rota no RIF.

De posse desses quadros *Route Explorer* recebidos, a estação final de origem escolhe então um dos caminhos neles contidos para o envio posterior de quadros unicast. O

critério mais utilizado para essa seleção é adotar o RIF do primeiro dos quadros recebidos.

O campo RIF é composto por um subcampo *Routing Control* e por um subcampo *Route Descriptor* (que contém a sucessão de descritores de rota que indicam o caminho). Observa-se a plenitude de uso do princípio de *source routing*.

Verificamos nos parágrafos anteriores deste item que *Segment Routing*, assim como *Source Routing Bridging*, representa uma forma de utilização integral da concepção *source routing. Segment Routing,* no entanto, representa uma opção mais aprimorada de *source routing*, por diferentes razões.

Em primeiro lugar, enquanto *Source Routing Bridging* opera apenas com bridges, *Segment Routing* aplica-se sobre diferentes tipos de roteador baseados em uma variedade de protocolos. O mais importante, contudo, é que *Segment Routing* possibilita funções mais complexas, particularmente no que diz respeito à utilização de Engenharia de Tráfego (TE), inviáveis em *Source Routing Bridging*, que se limita ao uso de métricas mais simples na escolha de caminhos.

Para informação dos leitores, registramos que, subsequentemente ao SRB, foi desenvolvido o esquema referido como *Source Routing Transparent* (SRT) *bridging*, que combina o método SRB com o método *transparent bridging* desenvolvido no IEEE para *bridged Ethernet LANs.*

A opção em uso das bridges SRT é indicada por meio do bit RII (*Routing Information Indicator*), que representa o primeiro bit do endereço MAC de origem do quadro.

### 3.5 – Funcionamento de Segment Routing

Ilustraremos o funcionamento de *Segment Routing* em exemplos a seguir*,* utilizando a codificação dos segmentos por labels*,* como no MPLS.

Em *Segment Routing*, extensões do OSPF ou do IS-IS distribuem *Node-SIDs* por toda a rede SR, de forma tal que os caminhos mais curtos para que um dado roteador alcance qualquer um outro roteador ficam automaticamente estabelecidos. Assim, basta a indicação do *Node-SID* do roteador de destino como segmento ativo na lista de segmentos, para que um outro roteador transmita adequadamente os quadros pelo caminho mais curto para esse destino.

Cada roteador divulga seu label global (*Node SID*) juntamente com o seu endereço IPv4 *loopback* por meio das extensões do IGP. Todos os demais roteadores da rede SR instalam os *Node-SIDs* recebidos, no plano de dados do SR.

Adicionalmente, cada roteador aloca e anuncia *Adj-SIDs* para suas interfaces, sendo que cada um desses *Adj-SIDs* é instalado apenas pelo roteador vizinho na respectiva interface.

#### 3.5.1 – Primeiro Exemplo

Em um primeiro exemplo, a Figura 3 representa uma configuração em que um quadro, recebendo na origem (roteador A) o *Node-SID* 103 que identifica o roteador E como destino final, trafega na rede pelo caminho mais curto para esse destino. Como se observa, não se aplica ECMP nesse caso.

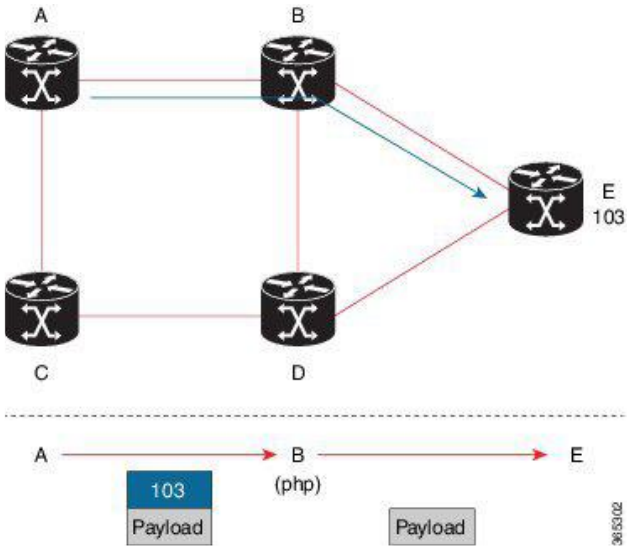


**Figura 3 – Caminho mais curto sem ECMP.**

Como se observa na figura, ocorre PHP (*Penultimate Hop Popping*) no roteador B, por ser desnecessária a transmissão de *Node-SIDs* no último passo. Apenas os *payloads* dos quadros chegam ao destino.

### 3.5.2 – Segundo Exemplo

Na Figura 4, apresentamos um outro exemplo de uso exclusivo do caminho mais curto, agora com a ocorrência de ECMP.

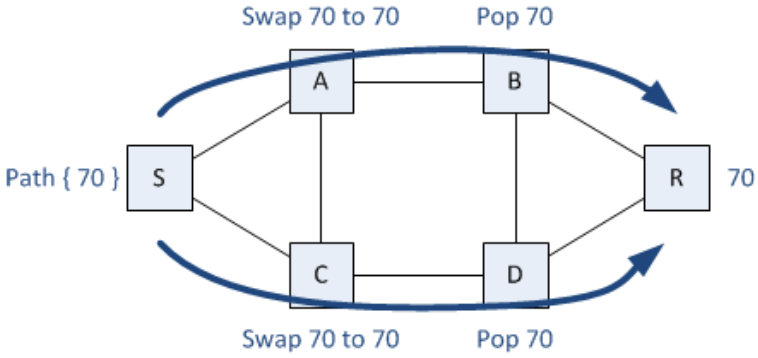


**Figura 4 – Caminho mais curto com ECMP.**

Nessa figura, os quadros transmitidos pelo roteador S para o roteador R pelo caminho mais curto, recebem apenas o *Node-SID* do roteador R (label 70). Como se observa na figura, ocorreu PHP nos roteadores B e D.

Observa-se também que, como o roteador S possui duas alternativas de caminho mais curto para o roteador R, foi aplicado ECMP no roteador S.

### 3.5.3 – Terceiro Exemplo

Na Figura 5, são utilizadas propriedades mais sofisticadas de *Segment Routing.* Foi aplicada alguma forma de TE no roteador S, o que determinou que os quadros destinados ao roteador R devem ser comutados no roteador B para o roteador D, de onde seguem para o roteador R de destino. Os quadros desviam-se do caminho mais curto no roteador B, voltando para o caminho mais curto (de D para R) no roteador D.

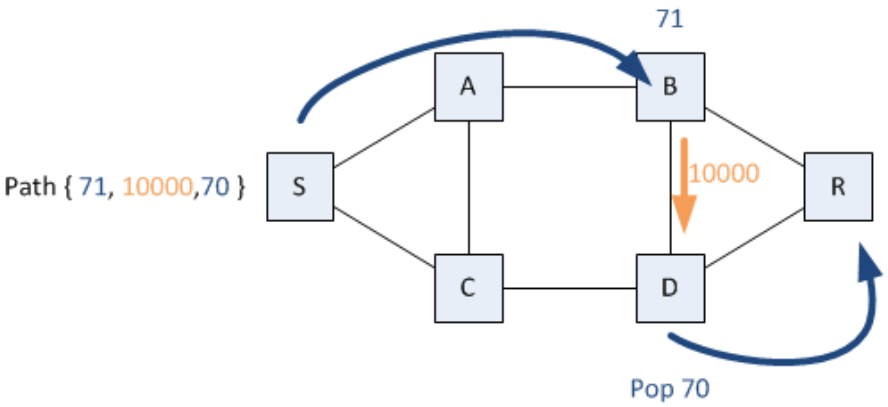


**Figura 5 – Configuração com uso de Adj-Sid.**

Com esse propósito, aplicou-se no roteador S uma lista de segmentos representada pela seguinte pilha de labels: [71,1.000,70, *Payload*]. Nessa pilha, o label 71 representa o *top-label,* ou seja, o primeiro label a ser processado (segmento ativo).

O *Node-SID* 71 faz com os quadros sejam conduzidos pelo caminho mais curto para o roteador B (*Node -SID* 71), onde esse *Node-SID* é descartado. O roteador B verifica então que o próximo SID é o *Adj-SID* 1000, e encaminha os quadros pela interface 1000, o que conduz os quadros para o roteador D.

O roteador D descarta o *Adj-SID* 1000, e verifica então que o próximo SID é o *Node-SID* 70 e sabe que deve enviar os quadros pelo caminho mais curto para o roteador R (*Node-SID* 70). O roteador D aplica então PHP antes da transmissão dos quadros para o destino final, descartando o *Node-SID* 70.

O roteador R sabe que é o destino final por receber os quadros sem qualquer label ou com apenas o seu próprio *Node-SID* (caso não houvesse PHP no roteador D).

### 3.5.4 – Quarto Exemplo

Neste quarto exemplo, onde se verifica a aplicação de TE com um pouco mais de sofisticação, foi reproduzido, com pequenas simplificações, do artigo *Segment Routing (SR) and Traffic Engineering (TE),* da série *Industry Solutions and Trends,* da *Juniper.*

A Figura 6 representa um mapa resumido dos Estados Unidos, com milhagens entre diferentes cidades, que será utilizado como ilustração deste exemplo de uso de *Segment Routing* com TE.

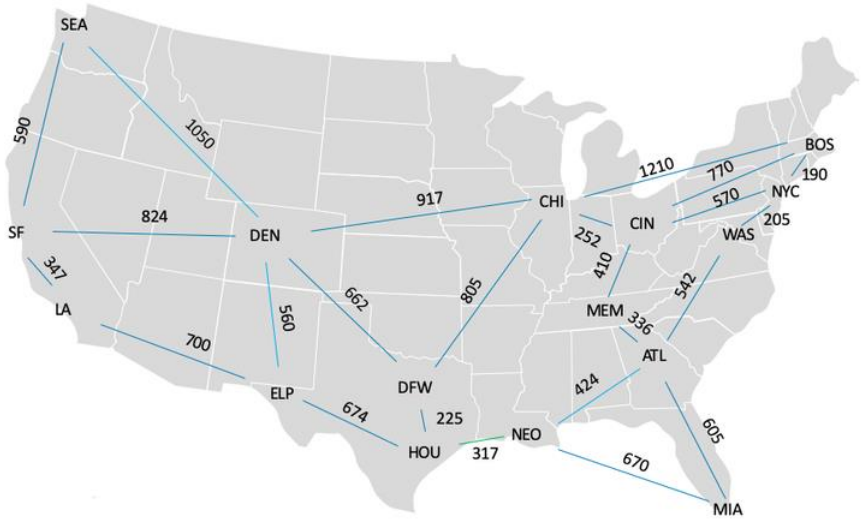


**Figura 6 – Mapa resumido dos Estados Unidos com milhagens entre cidades.**

Suponhamos que uma empresa queira transmitir tráfego Ethernet de ATL para WAS, NYC e BOS. Para isso, é requerido o uso de alguma técnica de encapsulamento, estando disponíveis as seguintes:

- VXLAN (*Virtual Extensible LAN*);

- MPLS sinalizado por LDP (MPLS Básico);

- SR-MPLS.

Em qualquer das hipóteses, considerando-se as milhagens como custo IGP por refletirem as latências dos respectivos trechos, o caminho mais curto de ATL para WAS, NYC e BOS é aquele que liga diretamente aquelas cidades (apresentam menores milhagens).

Assim, o tráfego normal de ATL para WAS, NYC e BOS deve passar pelo link ATL/WAS, enquanto o tráfego de ATL para NYC e BOS deve passar pelo link WAS/NYC.

Esse fato é motivo de preocupação para a gerência da rede, pela possibilidade de sobrecarga de tráfego nesses links de uso comum antes da consequente expansão das respectivas capacidades.

Decidiu-se então que a solução para esse problema é o desvio temporário do tráfego destinado a BOS para outro caminho, ainda que com maior custo, mantendo-se o tráfego para WAS e NYC pelo caminho mais curto.

Com essa condição, a escolha recai sobre o SR-MPLS, por ser a única das alternativas de técnicas de encapsulamento que permite a alternância de caminhos, por meio de uso recursos de TE.

Assim, o roteador de ATL deve inserir respectivamente o *prefix segment ID* de WAS, NYC ou BOS nos quadros destinados para cada uma dessas cidades.

Para o desvio temporário de tráfego conforme a decisão anterior, caso necessário, é preciso escolher entre os caminhos ATL/MEM/CIN/BOS e ATL/MEM/CIN/NYC/BOS como alternativa. É óbvio que o primeiro desses caminhos será o escolhido, por apresentar menor latência, menor risco de falhas (maior resiliência) e não utilizar o trecho NYC/BOS supostamente congestionado.

Assim, quando do uso do caminho alternativo, o roteador de ATL deve inserir uma pilha de protocolos composta por uma sucessão de *Node-SIDs* e de *Adj-SIDs*, com o seguinte conteúdo: [MEM, MEM/CIN, CIN, CIN/BOS, BOS].

### 3.6 – TIPOS DE SEGMENTO/SID

No sub-item 3.1 anterior deste tutorial, apresentamos a classificação básica simplificada de segmentos e respectivos SIDs, que engloba segmentos nodais (com significado global no domínio SR) e segmentos/SIDs adjacência (com significado local para cada roteador SR).

Vamos agora apresentar a classificação geral de segmentos/SIDs, de acordo com a RFC 8402.

Em termos gerais, foram definidas as seguintes classes de segmento:

- Segmentos *Link State IGP;*

- Segmentos BGP;

- Segmentos *Binding.*

#### 3.6.1 – Segmentos Link State IGP

Em um domínio IGP SR (Área IGP ou AS), um roteador SR identificado por um prefixo de endereço IP *loopback*, divulga segmentos com significado global (seu prefixo de endereço loopback e o *Node-SID* por ele atribuído para sua identificação no contexto de todo o domínio SR). Esses segmentos são referidos como Segmentos *Link State* IGP ou simplesmente como segmentos IGP.

A divulgação dos segmentos IGP realiza-se por extensões sendo definidas pelo IETF para cada um dos protocolos IGP utilizados. Essas extensões estão sendo especificadas pelos *IETF Drafts* OSPF-SR-EXT (*OSPF Extensions for Segment Routing*), ISIS-SR-EXT (*IS-IS Extensions for Segment Routing*) e OSPFv3-SR-EXT (*OSPFv3 Extensions for Segment Routing*).

Como essas extensões não incorporam mecanismos para TE, e dentro da filosofia de simplificação de SR, está sendo especificada, em paralelo, uma nova extensão para os protocolos IGP, definindo o que se denomina *IGP Flex-Algo.*

*IGP Flex-Algo* está em fase de definição pelo IETF, que emitiu, em 18/09/2019, a última versão do *draft* intitulado *IGP Flexible Algorithm*. Como as extensões aos

protocolos IGP previamente em análise não contemplam TE, foi definida a concepção *IGP Flex-Algo* com esse propósito.

Esse algoritmo possibilita aos próprios IGPs calcular caminhos baseados em condicionantes (*constraints)*, também por ele divulgadas.

O *draft IGP Flexible Algorithm* especifica um conjunto de extensões aos protocolos IS-IS, OSPFv2 e OSPFv3, que possibilita a um dado roteador SR enviar TLVs que:

- Identificam um tipo de cálculo;

- Especificam um tipo de métrica;

- Descrevem um conjunto de condicionantes na topologia, a serem utilizadas no cálculo dos melhores caminhos que as atendam.

O roteador SR que envia essas informações, também atribui um valor de *Flex-Algorithm* correspondente à combinação especificada dessas três informações. Os valores de SID referentes a um caminho definido são associados a um dado *Flex-Algorithm*.

Foram especificados os seguintes tipos de Segmentos IGP:

- Segmentos *IGP-Prefix,* identificados por *Prefix-SIDs;*

- Segmentos *IGP- Node*, identificados por *Node-SIDs;*

- Segmentos *IGP-Anycast,* identificados por *Anycast-SIDs;*

- Segmentos *IGP-Adjacency,* identificados por *Adj-SIDs.*

#### 3.6.1.1 - Segmentos IGP-Prefix

Um Segmento IGP-Prefix é um segmento IGP vinculado a um prefixo IGP. Esse tipo de segmento tem significado global, a menos que tenha sido explicitamente divulgado de outra forma no domínio SR.

O Segmento *IGP-Prefix* define o caminho mais curto para o roteador SR a ele associado, permitindo a ocorrência de ECMP quando aplicável.

Cada roteador SR é identificado por um único *Prefix-SID* em uma rede. Os valores de *Prefix-SID* são atribuidos da SRGB gerenciada pelo usuário.

#### 3.6.1.2 – Segmentos IGP-Node

Segmentos *IGP-Node* são um tipo de Segmento *IGP-Prefix* identificado por valores arbitrários de *Node-SID* atribuidos por cada roteador SR de destino, e divulgados juntamente com os respectivos endereços *loopback*.

O conceito e a utilização dos Segmentos IGP-Node encontram-se claramente evidenciados nos exemplos de uso de SR apresentados no sub-item 3.5 deste artigo.

### 3.6.1.3 – Segmentos IGP-Anycast

Um Segmento *IGP-Anycast* é um Segmento *IGP-Prefix* que identifica um conjunto de roteadores SR. Em um grupo anycast, todos os roteadores SR divulgam um mesmo prefixo com o mesmo valor de SID, o que facilita o balanceamento de carga de tráfego.

### 3.6.1.4 – Segmentos IGP-Adjacency

Um Segmento *IGP-Adjacency* representa uma adjacência específica em um roteador SR, como por exemplo uma interface de egresso para um roteador SR vizinho. Esse tipo se segmento é identificado por um *Adj-SID.* O roteador vizinho pode um roteador adjacente ou não adjacente, como por exemplo uma FA (*forwarding adjacency*).

O *Adj-SID* implica que, a partir do roteador SR que o divulgou, o quadro é transmitido através da adjacência, ou adjacências, por ele identificada(s), a despeito dos custos IGP/SPF. Em outras palavras, o uso de Segmentos *IGP-Adjacency* torna sem efeito as decisões de roteamento tomadas pelo algoritmo SPF.

Adj-SIDs podem ser utilizados para representar um conjunto de interfaces paralelas entre dois roteadores SR adjacentes (adjacências paralelas). Com o propósito de otimizar a função de balanceamento de carga de tráfego, um fator de peso (*weight*) pode ser associado ao *Adj-SID* divulgado com cada adjacência

## 3.6.2 – Segmentos BGP

Os Segmentos BGP são segmentos que, uma vez alocados, podem ser distribuídos por BGP.

Para possibilitar a distribuição de informações de *Segment Routing* via BGP de forma adequada, o IETF emitiu, em 27/06/2019, a última (16ª) versão do *IETF Draft* intitulado *BGP Link-State Extensions for Segment Routing.* Como se observa, trata-se de extensões ao protocolo BGP-LS, que por sua vez representa uma extensão ao BGP para habilitá-lo a operar com *Link-State* e TE**.**

O protocolo BGP-LS foi especificado na RFC 7752 (*North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP*), de março/2016.

A RFC 7752 descreve um mecanismo pelo qual informações atinentes a *link-state* e a TE, podem ser coletadas das redes e compartilhadas com componentes externos pela utilização do BGP. Isso é conseguido mediante o uso de um novo formato de codificação de NLRI (*Network Layer Reachability Information*).

Esse mecanismo é aplicável para links IGP físicos ou virtuais, e é sujeito ao controle por políticas.

Foram especificados os seguintes tipos de Segmentos BGP:

- Segmentos *BGP-Prefix,* identificados por *BGP Prefix SIDs;*

- Segmentos *BGP Peering*, identificados *BGP Peering SIDs.*

#### 3.6.2.1 – Segmentos BGP-Prefix

Um Segmento *BGP-Prefix* define um caminho para um prefixo BGP. Esse segmento é globalmente único em um domínio SR (a menos que divulgado de outra forma).

Um prefixo BGP representa o prefixo da sub-rede de destino, sub-rede essa pertencente a um AS de destino, adjacente ao AS de origem onde se situa o domínio SR (referido como *peer AS*) ou a ele distante.

Um *BGP-Prefix SID* é então um identificador do ASBR do AS de origem que conduz à sub-rede identificada pelo prefixo almejado, como por exemplo o endereço loopback desse ASBR ou um *Node-SID* a ele (e por ele) atribuído.

A presença de um *BGP-Prefix SID* como SID ativo em um quadro, encaminha esse quadro pelo caminho mais curto ao ASBR que conduz à sub-rede de destino.

#### 3.6.2.2 – Segmentos BGP Peering

Um Segment *BGP Peering* identifica um link particular em um dado ASBR de saída da rede SR, que conduz a um dado AS par (*peer AS*). Embora o link identificado se encontre no ASBR de saída da rede SR, a sua definição e a sua comunicação ocorrem no ASBR de entrada nessa rede.

O processo que efetiva os procedimentos acima é referido como BGP-EPE (*BGP Egress Peer Engineering*). A aplicação de BGP-EPE em *Segment Routing* encontra-se especificada no *IETF Draft* intitulado *Segment Routing Centralized BGP Egress Peer Engineering.*

O referido *draft* ilustra a aplicação de *Segment Routing* no atendimento ao BGP-EPE (*SR-based BGP-EPE*). Embora o *SR-based BGP-EPE* seja também utilizado de forma distribuída nos ASBRs de uma rede SR, como apresentaremos adiante, o *draft* em tela focaliza a aplicação do *SR-based BGP-EPE* utilizando controlador central (*BGP-EPE Controller*). O *BGP-EPE Controller* pode ser inclusive um controlador SDN.

Compete ao *BGP-EPE Controller* programar as políticas de BGP-EPE nos ASBRs de entrada na rede SR, políticas que são aplicadas pelos ASBRs de saída. Esses ASBRs de saída são referidos como *BGP-EPE enabled ASBRs.*

Passaremos agora a ilustrar o funcionamento do SR-based BGP-EPE, evidenciando a utilização dos Segmentos *BGP Peering.*

A Figura 7 exibe uma configuração de rede IP multi-ASes.

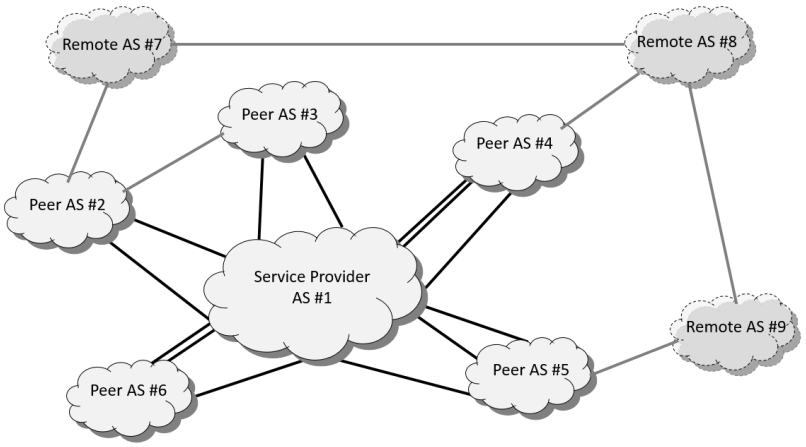


**Figura 7 – Configuração de rede IP multi-ASes.**

Como se observa nessa figura, tomando-se o AS #1 como AS central, os ASs #2, #3, #4, #5 e #6 constituem ASes pares (*peer ASes*), enquanto os ASes #7, #8 e #9 são ASes remotos.

Na Figura 8, baseada na configuração de rede da Figura 7 anterior, aplica-se o BGP tradicional (sem o uso de BGP-EPE), na transmissão de pacotes IP dos *e_ASBR1* (*egress ASBR 1*) e *e_ASBR2*, para a interrede com o prefixo de endereço IP *Prefix "A"* (situada no *Peer AS #4*).

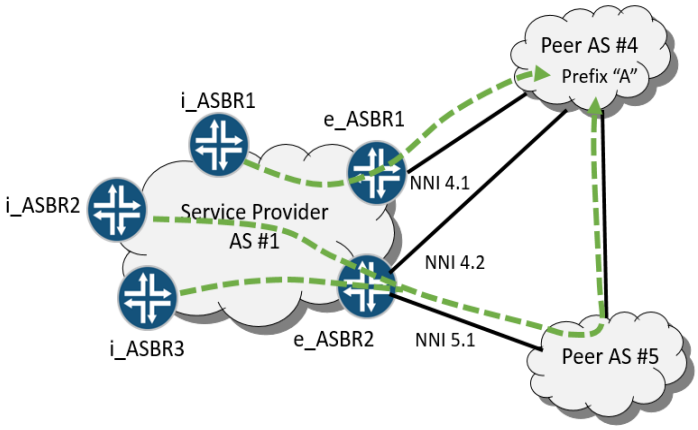


**Figura 8 – Rede da Figura 7 com o BGP tradicional (sem BGP-EPE).**

Nesse caso, todos os pacotes enviados para o *e_ASBR1 pelo i_ASBR1 (ingress ASBR 1), e* endereçados ao *Prefix "A",* egressam do AS #1 pela NNI 4.1.

Da mesma forma, todos os pacotes enviados para o *e_ASBR2*, tanto pelo *i_ASBR2* quanto pelo *i_ASBR3,* e endereçados ao *Prefix "A",* egressam do AS #1 pela NNI 5.1.

A escolha da NNI 5.1, e não da NNI 4.2, para encaminhamento de tráfego do *e_ASBR2* para o *Prefix A"* foi determinada pelo BGP.

Resulta que na solução da Figura 8, pode estar ocorrendo congestionamento no caminho percorrido pelo tráfego para o *Prefix "A"* a partir da NNI 5.1, estando o link com esse destino (via NNI 4.2) totalmente ocioso. A aplicação do BGP-EPE (*SR-based BGP-EPE,* em nosso caso), possibilita a melhoria dessa condição de desbalanceamento de tráfego entre caminhos.

Antes de abordarmos o *SR-based BGP-EPE,* é importante observarmos que no AS #1 (do Provedor de Serviço), está sendo utilizada uma tecnologia de rede com TE (como MPLS-TE, MPLS-TP ou SR), para possibilitar a constituição de dois caminhos para um mesmo destino (Prefix "A") com saídas independentes (do i_ASBR1 saindo pelo e_ASBR1 e dos i_ASBR2/i_ASBR3 saindo pelo e_ASBR2).

A Figura 9 mostra a configuração da Figura 8 com aplicação de BGP-EPE (*SR-based BGP-EPE*).

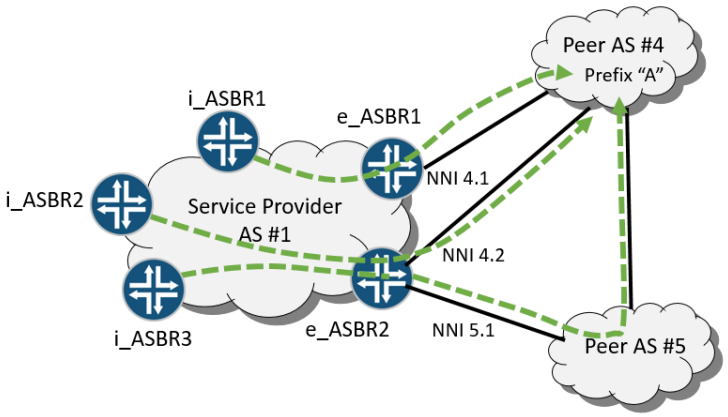


**Figura 9 – Figura 8 com aplicação de SR-based BGP-EPE.**

Como se observa nessa figura, tornou-se possível a distribuição do tráfego destinado ao *Prefix "A"* no *e_ASBR2*, em função da origem do tráfego. O tráfego originado no *i_ASBR2* egressa pela NNI 4.2 no *e_ASBR2*, enquanto o tráfego originado no *i_ASBR3* egressa pela NNI 5.1 no *e_ASBR2.*

Fica visível a ocorrência da distribuição do tráfego sainte pelo e_ASBR2, ocasionando melhoria no balanceamento de carga entre as duas alternativas de caminho para o Prefix "A".

Essa solução é possível devido à inserção de dois diferentes valores de *GBP Peering SID* nos quadros transmitidos dos *i_ASBR2* e *i_ASBR3*, por cada um deles respectivamente. O *e_ASBR2* deve ser naturalmente configurado com a associação entre cada um desses dois valores e a NNI a ser utilizada.

Finalmente, a Figura 10 apresenta, como ilustração, uma configuração de uso de *SR-based BGP-EPE* com um controlador EPE, tendo ao lado os procedimentos aplicados. Deixamos como exercício para os leitores a interpretação da figura.

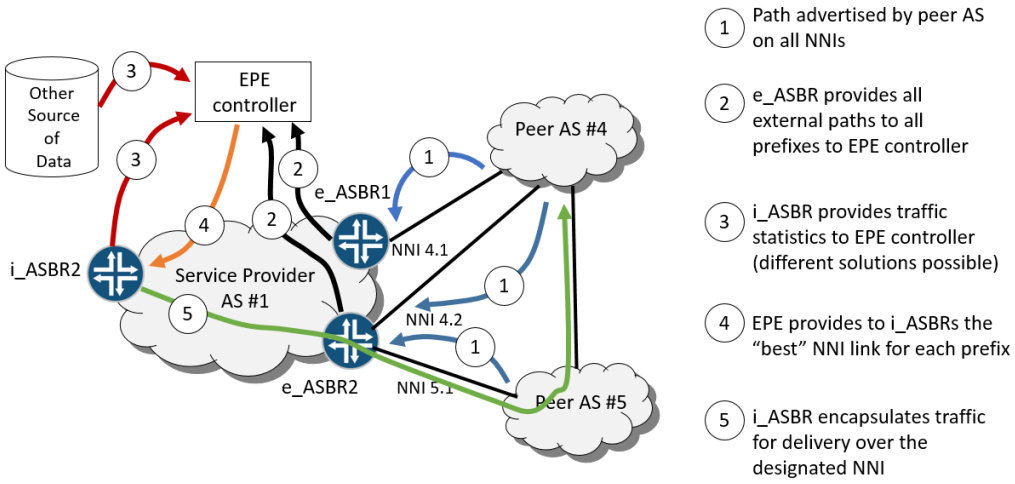


**Figura 10 – Aplicação de SR-based BGP-EPE com controlador EPE.**

## 4 – CONSIDERAÇÔES FINAIS

Apresentamos neste artigo uma abordagem conceitual de *Segment Routing* (SR). SR é uma nova tecnologia já em implementação, baseada no paradigma *Source Routing*, e que se pretende que venha a ocupar o espaço do IP/MPLS (MPLS Básico como LDP e MPLS-TE) e do modo operacional atual das redes IP. Não vimos qualquer menção ao MPLS-TP na literatura por nós obtida, embora seja teoricamente também possível a sua suplantação por SR.

*Segment Routing* tem como principal vantagem a simplificação das redes em decorrência da centralização de processos que caracteriza *Source Routing*. Com SR, torna-se desnecessária a constituição intensa de registros de estado ao longo da rede, e os processos de sinalização utilizados em todas as formas de MPLS.

Ao longo do artigo, tivemos a oportunidade de abordar algumas novas concepções em redes de telecomunicações, explicando os seus fundamentos, como por exemplo:

- BGP- LS (*BGP Link-State*);

- Aplicação do BGP-LS em *Segment Routing*;

- *IGP Flex-Algo*;

- BGP-EPE (*BGP Egress Peer Engineering*).

---