# Contents



# RESSALVA

Este documento fornece um resumo de alto nível de algumas práticas recomendadas estabelecidas para o roteamento OSPF/IS-IS e BGP. Essas recomendações não representam um projeto validado da Cisco e são necessários o devido

cuidado e atenção para a implantação em qualquer ambiente operacional específico. Eles devem ser lidos em conjunto com os guias de configuração e a documentação técnica dos produtos relevantes que descrevem em mais detalhes como essas práticas recomendadas podem ser implementadas. As referências neste documento a guias de configuração e documentação técnica de produtos específicos são apenas exemplos. Consulte os guias de configuração e a documentação técnica de seus produtos específicos.

# Introduction

Este documento descreve algumas práticas recomendadas estabelecidas e recomendações para criar redes simplificadas, eficientes e escaláveis alimentadas por plataformas de roteamento IOS XR. Este documento concentra-se em técnicas de implementação específicas e opções de suporte de recursos disponíveis no IOS XR para ajudar a personalizar as implantações de OSPF/IS-IS e BGP.

# Implementação do OSPF

O protocolo OSPF, definido no RFC 2328, é um IGP usado para distribuir informações de roteamento em um único sistema autônomo. O OSPF oferece vários benefícios em relação a outros protocolos, mas é necessário um projeto adequado para criar uma rede escalável e tolerante a falhas.

Para obter mais informações sobre o OSPF, consulte:

- Nota técnica sobre OSPF: https://www.cisco.com/c/en/us/support/docs/ip/open-shortest-path-first-ospf/7039-1.html#anc13
- Guia de configuração do OSPF: https://www.cisco.com/c/en/us/td/docs/routers/asr9000/software/asr9k-r7-6/routing/configuration/guide/b-routing-cg-asr9000-76x/implementing-ospf.html
- Referência de comando: https://www.cisco.com/c/en/us/td/docs/routers/asr9000/software/asr9k-r7-5/routing/command/reference/b-routing-cr-asr9000-75x/ospf-commands.html#wp2421918195

# Principais conceitos

- Hierarquia: Um modelo de rede hierárquica é uma ferramenta de alto nível útil para projetar infraestruturas de rede confiáveis e ajuda a dividir problemas complexos de projeto de rede em áreas menores e mais gerenciáveis.
- Modularidade: Dividindo várias funções em uma rede em módulos, a rede é muito mais fácil de projetar. A Cisco identificou vários módulos, incluindo campus empresarial, bloco de serviços, data center e borda da Internet.
- Resiliência: A rede está disponível em condições normais e anormais. As condições normais incluem fluxos de tráfego esperados, padrões e eventos programados, como janelas de manutenção. As condições anormais incluem falhas de hardware ou software, cargas de tráfego extremas, padrões de tráfego incomuns, eventos de negação de serviço (DoS) e outros eventos planejados ou não.
- Flexibilidade: A capacidade de modificar partes da rede, adicionar novos serviços ou aumentar a capacidade sem passar por uma atualização em larga escala (ou seja, substituir os principais dispositivos de hardware).

Como melhor prática geral, a implantação de rede deve considerar o "alcance" da rede para conter as rotas dentro de um limite específico e as rotas que são relevantes e exigidas pelos roteadores dentro de um domínio para encaminhamento. O uso eficaz de áreas OSPF ajuda a reduzir o número de anúncios de link-state (LSAs) e outro tráfego de sobrecarga enviado através da rede. Uma das vantagens de criar uma hierarquia é que essa abordagem ajuda a garantir que o tamanho do banco de dados de topologia que cada roteador precisará manter seja gerenciável e esteja em conformidade com o perfil de memória do roteador.

## Domínio OSPF e redistribuição de BGP

O OSPF foi projetado para transportar apenas alguns milhares de rotas. Em um alto nível, as "áreas" do OSPF são seções de uma rede onde qualquer roteador sabe sobre a capacidade de roteamento de todos os outros roteadores na área. Isso permite uma convergência rápida quando qualquer dispositivo tem um problema, mas ao custo de escalabilidade reduzida. Como tal, o OSPF é usado em um núcleo de Provedor de Serviços para fornecer a conectividade de nível básico entre todos os dispositivos de núcleo, e todos os dispositivos de núcleo são configurados dentro da mesma área OSPF. Este é um projeto padrão de uma rede "subjacente".

Por outro lado, o BGP é projetado para transportar significativamente mais rotas do que a maioria dos IGPs, como o OSPF. Riscos associados à redistribuição de rotas BGP em um IGP como o OSPF. Se um provedor de serviços exigir que as rotas BGP sejam redistribuídas no domínio IGP para qualquer caso de uso, isso precisará ser gerenciado pelo provedor de serviços com a filtragem adequada nos roteadores de borda de sistema autônomo (ASBRs) e com a proteção contra sobrecarga configurada no roteador receptor. Se a redistribuição do BGP não for filtrada em um OSPF, cada dispositivo OSPF no ASBR começará a receber rotas muito além de sua capacidade de lidar ao mesmo tempo. Os roteadores Cisco IOS XR, por exemplo, permitirão que apenas 10.000 rotas BGP sejam redistribuídas no OSPF por padrão. Quando as rotas BGP são redistribuídas no IGP, é possível que todos os roteadores dentro do domínio IGP possam receber essas rotas, dependendo do design do IGP. De acordo com o protocolo RFC do OSPF, qualquer rota externa que esteja sendo redistribuída no OSPF deve ser distribuída para todos os roteadores na área OSPF.

## Gerenciamento da redistribuição no IGP

Como melhor prática geral, a redistribuição só deve ser feita de maneira cuidadosa e planejada quando não houver outras opções para aprender as rotas de alcance que uma função de redistribuição fornecerá.

Como prática geral, você deve:

- Evitar redistribuição
- Evitar transportar rotas em um domínio IGP
- Implementar o BGP para acessibilidade externa
- Usar IGP para transportar somente informações do próximo salto; por exemplo, Loopback 0

## Limitações de redistribuição de rota OSPF

A escala de prefixos redistribuídos do BGP no OSPF é gerenciada com a configuração de proteção contra sobrecarga (max-lsa). Essa é a única proteção contra vazamento de um grande número de rotas no domínio OSPF. Em caso de redistribuição em uma única área OSPF, você deve implementar várias camadas de proteção contra redistribuição de rota.

Aqui estão algumas das opções disponíveis para proteção contra redistribuição de rota:

- Filtragem de redistribuição usando ACL
- Limite de redistribuição - configuração global para impedir que mais de um número específico de rotas seja redistribuído. Se o filtro for removido, o limite de redistribuição global será a segunda linha de defesa e protegerá os núcleos.
- Configurações LSA máximo em todos os dispositivos na área OSPF - se as proteções mencionadas nos marcadores acima falharem, force os roteadores receptores a recusar os LSAs excessivos de entrada.

## Proteção contra sobrecarga do banco de dados link-state do OSPF

O recurso de proteção contra sobrecarga do banco de dados link-state do OSPF fornece um mecanismo no nível do OSPF para limitar o número de LSAs não gerados automaticamente para um determinado processo OSPF. Se outros roteadores na rede tiverem sido configurados incorretamente, eles poderão gerar um grande volume de LSAs, por exemplo, para redistribuir um grande número de prefixos no OSPF. Esse mecanismo de proteção ajuda a evitar que os roteadores recebam muitos LSAs e, portanto, sofram escassez de CPU e memória.

### Comportamento do recurso

Veja como o recurso se comporta:

- Quando esse recurso está habilitado, o roteador mantém uma contagem do número de LSAs recebidos (não gerados automaticamente).
- Quando o valor de limite configurado é atingido, uma mensagem de erro é registrada.
- Quando o número máximo configurado de LSAs recebidos é excedido, o roteador pára de aceitar novos LSAs.

```
max-lsa <max-lsa-count> <%-threshold-to-log-warning> ignore-count <ignore-count-value> ignore-time
<ignore-time-in-minutes> reset-time <time-to-reset-ignore-count-in-minutes>
```

## Estados do OSPF

Se a contagem de LSAs recebidos for maior que o número máximo configurado após um minuto, o processo OSPF desativará todas as adjacências e limpará o banco de dados OSPF. Esse estado é chamado de estado ignorar. Nesse estado, todos os pacotes OSPF recebidos em todas as interfaces pertencentes à instância OSPF são ignorados e nenhum pacote OSPF é gerado nas interfaces. O processo OSPF permanece no estado ignorar durante o tempo de

ignorar configurado (o padrão é 5 minutos). Quando o tempo de ignorar expira, o processo OSPF retorna à operação normal e cria adjacências em todas as suas interfaces.

Se a contagem LSA exceder o número máximo assim que a instância OSPF retornar do estado ignorar, a instância OSPF poderá oscilar infinitamente entre seu estado normal e o estado ignorar. Para evitar essa oscilação infinita, a instância OSPF conta quantas vezes esteve no estado ignorar. Esse contador é chamado de ignore-count. Se ignore-count (o padrão ignore-count é 5) excede seu valor configurado, a instância do OSPF permanece permanentemente no estado ignore.

Você deve emitir o comando clear ospf para retornar a instância OSPF ao seu estado normal. O comando ignore-count será redefinido como zero se a contagem de LSA não exceder o número máximo novamente durante o tempo configurado pela palavra-chave reset-time.

Se você usar a palavra-chave warning-only, a instância OSPF nunca entrará no estado ignorar. Quando a contagem de LSA excede o número máximo, o processo OSPF registra uma mensagem de erro e a instância OSPF continua em sua operação de estado normal.

Você deve emitir o comando clear ospf para retornar a instância OSPF ao seu estado normal. O ignore-count será redefinido para zero se a contagem de LSA não exceder o número máximo novamente durante o tempo configurado pela palavra-chave reset-time.

Se você usar a palavra-chave warning-only, a instância OSPF nunca entrará no estado ignorar. Quando a contagem de LSA excede o número máximo, o processo OSPF registra uma mensagem de erro e a instância OSPF continua em sua operação de estado normal.

Não há valor padrão para max-lsa. O limite será verificado apenas se estiver configurado especificamente.

Depois que max-lsa é configurado, outros parâmetros podem ter valores padrão:

- %-threshold-to-log-warning padrão - 75%
- valor padrão de ignore-count-value - 5
- default ignore-time-in-minutes - 5 minutos
- default time-to-reset-ignore-count - 10 minutos

Aqui está um exemplo da implementação que mostra como configurar a instância do OSPF para aceitar 12000 LSAs não gerados automaticamente e 1000 LSAs não gerados automaticamente no VRF V1.

```
RP/0/RSP0/CPU0:router# configure
RP/0/RSP0/CPU0:router(config)# router ospf 0
RP/0/RSP0/CPU0:router(config-ospf)# max-lsa 12000
RP/0/RSP0/CPU0:router(config-ospf)# vrf V1
RP/0/RSP0/CPU0:router(config-ospf)# max-lsa 1000
```

O exemplo a seguir mostra como exibir o status atual da instância OSPF.

```
RP/0/RSP0/CPU0:router# show ospf 0
    Processo de roteamento "ospf 0" com ID 10.0.0.2
    NSR (roteamento ininterrupto) está desativado
    Suporta apenas rotas TOS(TOS0) únicas
    Suporta LSA opaco
    É um roteador de borda de área
    Número máximo de LSAs não gerados automaticamente permitidos 12000
       Número atual de LSA 1 não gerado automaticamente
       Limite para mensagem de aviso 75%
       Ignore-time 5 minutes, reset-time 10 minutes
       Ignore-count allowed 5, current ignore-count 0
```

# Implementando o BGP

As famílias de endereços BGP fazem do BGP um protocolo de roteamento "multiprotocolo". É altamente recomendável que você entenda como as famílias de endereços são usadas para criar topologias escaláveis fáceis de implementar e gerenciar. Usando famílias de endereços, o operador pode criar topologias diferentes para tecnologias diferentes, por exemplo, EVPN, Multicast e assim por diante.

Para obter mais informações sobre o BGP, consulte o guia de configuração do BGP: https://www.cisco.com/c/en/us/td/docs/iosxr/ncs5500/bgp/76x/b-bgp-cg-ncs5500-76x/implementing-bgp.html

# BGP e BFD

A convergência de BGP em uma rede de provedores de serviços é importante para atender às expectativas do cliente para criar redes resilientes e tolerantes a falhas. Por padrão, o BGP tem um temporizador de Keepalive de 60 segundos e um temporizador de espera de 180 segundos. Tudo isso significa que o BGP será muito lento para convergir a menos que haja ajuda disponível dos protocolos de suporte. O BFD (Bi-directional Forwarding, encaminhamento bidirecional) é um protocolo desse tipo projetado para ajudar os protocolos clientes a convergirem mais rapidamente. Com o BFD, os protocolos podem convergir em segundos.

## Additional Information

- Este guia fornece informações conceituais e de configuração para BFD: https://www.cisco.com/c/en/us/td/docs/iosxr/ncs5500/routing/76x/b-routing-cg-ncs5500-76x/implementing-bfd.html
- Este white paper apresenta uma visão centrada no provedor de serviços sobre a convergência rápida usando BFD nos roteadores Cisco NCS 5500 e Cisco Network Convergence System 500 Series: https://xrdocs.io/ncs5500/tutorials/bfd-architecture-on-ncs5500-and-ncs500/
- Para obter informações mais detalhadas sobre o uso de BFD em interfaces de pacote e a implementação de BFD de vários caminhos e multi-hop, consulte o repositório https://xrdocs.io/.

# Detecção de peer lento de BGP

Um peer lento é um peer que não consegue acompanhar a taxa na qual o roteador está gerando mensagens de atualização durante um período prolongado (na ordem de minutos) em um grupo de atualização. Quando um peer lento está presente em um grupo de atualização, o número de atualizações formatadas com transmissão pendente aumenta. Quando o limite do cache é atingido, o grupo não tem mais cotas para formatar novas mensagens. Para que uma nova mensagem seja formatada, algumas mensagens existentes devem ser transmitidas com o peer lento e removidas do cache. O restante dos membros do grupo que são mais rápidos que o peer lento e completaram a transmissão das mensagens formatadas não terão nada de novo para enviar, mesmo que possa haver redes BGP recém-modificadas esperando para serem anunciadas ou retiradas. Esse efeito de bloquear a formatação de todos os pares em um grupo quando um dos pares estiver lento no consumo de atualizações é o problema do "par lento".

Eventos que causam uma variação significativa na tabela BGP (como redefinições de conexão) podem causar um breve pico na taxa de geração de atualização. Um peer que fica temporariamente para trás durante tais eventos, mas se recupera rapidamente após o evento, não é considerado um peer lento. Para que um peer seja marcado como lento, ele deve ser incapaz de acompanhar a taxa média de atualizações geradas durante um período mais longo (na ordem de alguns minutos).

O peer BGP Slow pode ser causado por:

- Perda de pacotes ou alto tráfego no link para o peer.
- Um peer de BGP pode estar muito carregado em termos de CPU e, portanto, não pode atender à conexão TCP na velocidade necessária.
- Nesse caso, a capacidade de hardware da plataforma e a carga oferecida devem ser verificadas.
- Problemas de throughput com a conexão BGP
- Para obter mais informações sobre a detecção de peer BGP Slow, consulte: https://www.cisco.com/c/en/us/td/docs/iosxr/ncs5500/bgp/76x/b-bgp-cg-ncs5500-76x/implementing-bgp.html#concept_ir5_j4w_p4b

Aqui estão algumas atenuações e práticas recomendadas para gerenciar colegas lentos:

- QoS fim-a-fim, que reserva largura de banda para tráfego plano de controle BGP durante o congestionamento.
- Uso de valores corretos e apropriados de MSS/MTU usando as configurações de BGP PMTUD e/ou TCP MSS.
- Use o hardware correto e minimize o número de rotas em relação ao hardware.

A detecção de ponto lento é habilitada por padrão no Cisco IOS XR a partir da versão 7.1.2. Os pontos lentos são aqueles que demoram a receber e processar as atualizações de BGP de entrada e confirmam as atualizações ao remetente. Se o peer lento estiver participando do mesmo grupo de atualização que outros peers, isso poderá retardar o processo de atualização para todos os peers. Nesta versão, quando o IOS XR detectar um peer lento, ele criará um syslog que tenha os detalhes sobre o peer específico.

## Convergência Rápida usando a Convergência Independente de Prefixo BGP

Para prefixos BGP, a convergência rápida é alcançada usando o BGP Prefix Independent Convergence (PIC), no qual o BGP calcula um melhor caminho alternativo e o melhor caminho principal e instala ambos os caminhos na tabela de roteamento como caminhos principais e de backup.

Se o remoto do próximo salto do BGP se tornar inalcançável, o BGP imediatamente alterna para o caminho alternativo usando o BGP PIC em vez de recalcular o caminho após a falha.

Se o PE remoto do próximo salto do BGP estiver ativo, mas houver uma falha de caminho, o IGP TI-LFA FRR processará a reconvergência rápida para o caminho alternativo e o BGP atualizará o próximo salto do IGP para o PE remoto.

O BGP PIC é configurado na família de endereços VRF para convergência rápida de prefixos VPN se um PE remoto se tornar inalcançável.

Para obter mais informações sobre a convergência independente de prefixo BGP, consulte: https://www.cisco.com/c/en/us/td/docs/iosxr/ncs5500/bgp/76x/b-bgp-cg-ncs5500-76x/bgp-pic.html

## Segurança BGP com BGP Flowspec

O BGP Flowspec, em poucas palavras, é um recurso que permite que você receba especificações de fluxo de tráfego IPv4/IPv6 (origem X, destino Y, protocolo UDP, porta origem A e assim por diante) e ações que precisam ser tomadas nesse tráfego (como queda, policiamento ou redirecionamento) através da atualização BGP.
Dentro da atualização BGP, os critérios de correspondência Flowspec são representados por BGP NLRI, e as comunidades estendidas BGP representam as ações.

Esse recurso é baseado no RFC 5575 e pode ser usado para ajudar a atenuar ataques DDoS. Quando um determinado host dentro de uma rede está sendo atacado, podemos enviar uma atualização Flowspec para roteadores de borda de forma que o tráfego de ataque possa ser policiado ou descartado, ou até mesmo redirecionado para outro lugar, talvez para um dispositivo que possa limpar o tráfego (filtre o tráfego "ruim" e encaminhe apenas o tráfego "bom" para o host afetado).

Depois que as especificações de fluxo são recebidas por um roteador e programadas nas placas de linha aplicáveis, todas as portas L3 ativas nessas placas de linha começarão a processar o tráfego de entrada de acordo com as regras de especificação de fluxo.

Para obter mais informações sobre a implementação do BGP FlowSpec, consulte:

- White paper que inclui links para o canal do Youtube do Cisco IOS XR, consulte https://xrdocs.io/ncs5500/tutorials/bgp-flowspec-on-ncs5500/
- Manual de configuração de BGP: https://www.cisco.com/c/en/us/td/docs/iosxr/ncs5500/bgp/76x/b-bgp-cg-ncs5500-76x/implementing-bgp.html#concept_ugv_bxq_h2b

# Práticas recomendadas e recomendações

A lista a seguir fornece uma visão geral das melhores práticas e recomendações gerais, listadas em uma ordem específica:

- Auditoria de rede para a integridade geral do sistema. Comece com uma auditoria de configuração e mude sequencialmente de configurações de interface para roteamento e serviços.
- Ter uma estratégia de monitoramento em vigor. Embora o SNMP seja uma prática padrão, considere a implantação de técnicas mais robustas e descritivas usando a Telemetria de Streaming. Consulte o whitepaper a seguir para obter as melhores práticas de implementação de telemetria em um IOS XR Router: https://xrdocs.io/telemetry/

## OSPF

Estas são as melhores práticas e recomendações gerais para o OSPF:

- Implemente a sumarização de rotas para rotas intra-área para OSPF.

- Configure o ID do roteador explicitamente dentro do OSPF como um dos endereços de loopback habilitados para OSPF.
- Projete uma rede hierárquica para limitar os LSAs dentro de uma área para OSPF. Mantenha o número de ABRs para uma área dentro de um intervalo razoável (~3 a 4).
- Implementar a configuração "max-lsa" do OSPF para OSPF, ou equivalente, para limitar os LSAs no banco de dados para usar efetivamente a memória do sistema.
- Limite o número máximo de rotas que podem ser distribuídas do BGP ao OSPF. No IOS-XR, o limite padrão é 10K.
- Use a política de rota (RPL) para redistribuir as rotas no OSPF.
- Resuma a rota entre áreas e as rotas externas tipo 5 sempre que aplicável.
- Uso de autenticação quando necessário.
- Sempre use NSF e NSR.
- Configure a filtragem de redistribuição na origem em vez do destino.
- Use a interface passiva onde aplicável.
- O OSPF deve transportar apenas rotas de Loopback e de Interface do Roteador - remova qualquer outra redistribuição de BGP para OSPF.
- Considere mover cada hub principal para sua própria área (NSSA).
- Use BFD para detecção rápida de falhas em comparação com os temporizadores agressivos do protocolo de roteamento.
- Não use o comando mtu-ignore o máximo possível.
- Considere o uso da sincronização IGP-LDP em um ambiente MPLS para evitar o envio de tráfego em um caminho sem rótulo.
- Considere a escala dentro dos limites da plataforma suportada (número de prefixos, número de rótulos, ECMP, número de áreas, etc.).
- Evite a redistribuição mútua em vários pontos.
- Configure a distância administrativa de modo que cada prefixo nativo de cada protocolo ou processo seja alcançado por meio do protocolo ou processo do domínio correspondente.
- Controle os prefixos (usando a combinação de distância ou lista de prefixos) para que o mesmo prefixo não seja anunciado de volta ao domínio de origem.
- Embora o ID do processo OSPF tenha significado local para o roteador, é recomendável ter o mesmo ID de processo para todos os roteadores no mesmo domínio OSPF. Isso melhora a consistência da configuração e facilita as tarefas de configuração automáticas.
- Ao configurar o OSPF para ambientes hub-and-spoke, projete as áreas OSPF com um número menor de roteadores.
- Configure a largura de banda de referência de custo automático do OSPF em todo o domínio OSPF para o link de largura de banda mais alta na rede.

- De uma perspectiva de projeto, recomendamos que você implemente o peering IGP com domínios sob os mesmos controles administrativos ou operacionais para ajudar a evitar que uma atualização IGP não planejada ou não autorizada se propague pela rede. Isso deve permitir melhor capacidade de serviço e facilidade de solução de problemas em caso de erros. Caso um domínio IGP grande seja uma necessidade comercial, planeje o uso do BGP nesses casos para limitar o número de rotas no domínio de rede IGP.
- Se precisar de conectividade MPLS de ponta a ponta, continue usando hierarquia/segmentação e use opções como RFC3107 BGP-LU ou computação de caminho entre domínios via PCE, ou selecione redistribuição/vazamento com política como último recurso.
- O recurso Shortest Path First Throttling do OSPF pode ser usado para configurar a programação do SPF em intervalos de milissegundos e potencialmente atrasar os cálculos do SPF durante a instabilidade da rede.
- O recurso OSPF SPF Prefix Prioritization permite que um administrador converja prefixos importantes mais rapidamente durante a instalação da rota.

## IS-IS

Aqui estão as melhores práticas e recomendações gerais para IS-IS:

- Se você tem uma rede linear de nível único, pense na escala. Configure todos os roteadores apenas como L2. Por padrão, o roteador é L1-L2 e o vazamento de informações de roteamento de L1 para L2 é ativado por padrão. Isso pode fazer com que todos os roteadores vazem todas as rotas L1 para L2, sobrecarregando o banco de dados de link-state.
- Se você executar uma rede de vários níveis (várias áreas), certifique-se de que a topologia da camada 3 siga a hierarquia ISIS. Não crie links de backdoor entre áreas L1.
- Se você executar uma rede multinível (várias áreas), verifique se os roteadores L1 e L2 estão conectados por meio de áreas L1 e L2. Isso não exige várias conexões físicas ou virtuais entre eles; execute o link entre os roteadores L1 e L2 como um circuito L1/L2.
- Se você executar uma rede multinível (várias áreas), resuma o que pode ser resumido - por exemplo, no caso de MPLS, o loopback de roteadores PE precisa ser propagado entre áreas, mas os endereços de link de infraestrutura não.
- Crie e siga o plano de endereçamento adequado, se possível. Isso permite o resumo e ajuda a escalar.
- Defina o tempo de vida do LSP para um máximo de 18 horas.
- Evite a redistribuição por qualquer meio. A redistribuição é complexa e precisa ser gerenciada manualmente para evitar loops de roteamento. Use o design multiárea/nível, se possível.
- Se você precisar usar a redistribuição, use a rotulação durante a redistribuição e a filtragem "distribute-list in" com base em tags para gerenciá-la. Resuma durante a redistribuição, se possível.
- Configure as interfaces como "ponto-a-ponto" sempre que possível. Isso melhora o desempenho e a escalabilidade do protocolo.
- Não use ISIS em topologia altamente em malha. Os protocolos link-state se comportam mal em ambientes altamente em malha.

- Configure uma métrica padrão alta no submodo da família de endereços ISIS. Isso impede que links recém-adicionados atraiam tráfego se eles forem configurados inadvertidamente sem uma métrica.
- Configure "alterações de adjacência de log" para ajudar na solução de problemas de conexão.
- Use "metric-style wide" no submodo ipv4 da família de endereços ISIS. Métricas estreitas não são muito úteis e não dão suporte a recursos como roteamento de segmentos ou flex-algo.
- Se estiver usando o TI-LFA de SR-MPLS, lembre-se de adicionar "ipv4 unnumbered mpls traffic-eng Loopback0" à configuração para permitir que o ISIS aloque túneis TE quando necessário.
- Deixe o padrão das configurações "lsp-gen-interval" e "spf-interval", a menos que você tenha certeza de que é necessária uma convergência nativa mais rápida. Com TI-LFA, a convergência nativa não é tão importante, já que o redirecionamento rápido lidará com alterações de topologia única em 50 ms ou menos.
- Se você modificar "lsp-gen-interval" ou "spf-interval", não use um tempo de atraso inicial menor que 50 ms.
- Na maioria dos casos, "set-overload-bit" é uma escolha melhor do que "max-metric", pois é uma alteração atômica que é suportada pelo fast-reroute.
- Use a autenticação criptográfica para Hello (hello-password) e LSPs (lsp-password). Os chaveiros proporcionam a maior flexibilidade e podem
  acomodar substituições de teclas sem interrupções.
- Configure "nsf cisco" para autenticação sem interrupções de reinicializações de processos ISIS e instalação de SMU. Apesar do nome, isso fornece melhor interoperabilidade com outros fornecedores do que "nsf ietf".
- Em uma plataforma com RPs duplos, configure TAMBÉM "nsr" para lidar com switchovers RP.
- Use os modelos "group" e "apply-group" para configurar seções de configuração repetidas. Isso é menos propenso a erros e mais fácil de alterar, se necessário.
- Em uma rede de vários níveis, considere cuidadosamente se você precisa usar "propagar" para vazar prefixos do Nível 2 para o Nível 1. Isso pode limitar a escalabilidade e, muitas vezes, a rota padrão de nível 1 fornecida pelo bit Anexado é suficiente.
- Se estiver usando várias instâncias ISIS no mesmo VRF, considere configurar valores exclusivos de "distância" para elas. Isso tornará a instalação de rotas na RIB mais determinística se cada uma tiver uma rota para o mesmo prefixo.
- Use BFD para detecção rápida de link desativado. Com o BFD fornecendo essa função, o intervalo de hello do ISIS pode ser aumentado com segurança para melhorar a escalabilidade.

## BGP

Aqui estão as melhores práticas gerais e recomendações para BGP:

- Use o NSR e o NSF / reinicialização suave com temporizadores cuidadosamente ajustados, dependendo da escala esperada.
- Configure o BGP usando a interface de loopback ‘sempre UP’, não a interface física para o peering IBGP.
- Não redistribua rotas BGP (alto volume) no IGP (volume comparativamente baixo) e vice-versa sem RPL adequado, restringindo o número de rotas redistribuídas do BGP para um IGP (OSPF/ISIS).

- A redistribuição de BGP para IGP sem uma política (ACL) adequada e bem testada pode causar esgotamento de recursos (memória) no roteador.
- Uso de rotas de sumarização no BGP para diminuir o tamanho da tabela de roteamento e o uso da memória. Agregar rotas com sumarização apenas onde fizer sentido
- Use a filtragem de rota para anunciar e receber rotas de forma eficiente, especialmente no BGP.
- Recomendamos o uso de Refletor de Rota (RR) e confederação para aumentar a escala da rede.
- Algumas das considerações de design do Refletor de rota são:
- A escala de caminho aumenta com base no número de clientes/não clientes.
- Em RRs hierárquicos, use o mesmo cluster-id no mesmo nível (RR redundante) para prevenção de loop e escala.
- Controle a MTU no caminho BGP ou use o protocolo PMTUD para ajustar o BGP MSS automaticamente.
- Use BFD ou ajuste temporizadores BGP para detecções de falhas mais rápidas.
- A escala BGP é conforme a configuração e o caso de uso, e nenhum tamanho único serve para todos. Você precisa ter uma boa ideia sobre:
- escala de rota
  - escala de caminho (com reconfiguração suave, aumentará)
  - escala de atributo
- Se o add-path estiver configurado, ele consumirá mais memória.
- Compreensão cuidadosa das políticas de vizinhança de BGP:
  - passar tudo (especialmente em um roteador de limite) pode causar estragos à medida que a escala de memória for aumentando.
  - Use construções de política que evitarão correspondências de expressões regulares em RPL.
- Com o NSR, o RP em espera usará cerca de 30% mais memória virtual do que o ativo. Tenha isso em consideração se houver um RP em espera.
- Observe a rotatividade contínua em um número significativo de rotas (bumps de versão). Isso pode manter a memória de geração de atualização em marca d'água alta.
- Proteja os pares com o botão max-prefix.
- Use os parâmetros de retardo de disparo do próximo salto de acordo com as metas de escala e convergência.
- No projeto de rede, tente evitar novos atributos. Atributos exclusivos levam a empacotamento ineficiente e resultam em mais atualizações de BGP.
- Configurar vários caminhos na rede pode levar a loops de encaminhamento. Use com cuidado.
- Use a política de tabela para evitar a instalação da rota para rib se o RR não for inline-RR (no next-hop-self)

## Monitore a memória do sistema para processos de roteamento

Nenhum dispositivo tem recursos infinitos - se enviarmos um número infinito de rotas a um dispositivo, o dispositivo deverá escolher como falhará. Os roteadores tentarão atender todas as rotas até que os limites de memória se esgotem, e isso pode fazer com que todos os protocolos e processos de roteamento falhem.

Cada processo no roteador central tem um "RLIMIT" definido. O "RLIMIT" é a quantidade de memória do sistema que cada processo pode consumir.

Esta seção descreve algumas técnicas padrão para monitorar e verificar a memória do sistema usada pelo processo BGP.

## Memória de processo

Mostra a quantidade de memória consumida por um processo.

```
RP/0/RP0/CPU0:NCS-5501#show proc memory
Texto JID (KB) Dados (KB) Pilha (KB) Processo dinâmico (KB)
------ ---------- ---------- ---------- ----------- ------------------------------
1150 896 368300 136 33462 lspv_server
380 316 1877872 136 32775 parser_server
1084 2092 2425220 136 31703 bgp
1260 1056 1566272 160 31691 ipv4_rib
1262 1304 1161960 152 28962 ipv6_rib
1277 4276 1479984 136 21555 pim6
1301 80 227388 136 21372 schema_server
1276 4272 1677244 136 20743 pim
250 124 692436 136 20647 invmgr_proxy
1294 4540 2072976 136 20133 l2vpn_mgr
211 212 692476 136 19408 sdr_invmgr
1257 4 679752 136 17454 statsd_manager_g
```

Cada processo recebe uma quantidade máxima de memória que pode consumir. Isso é definido como o limite.

```
RP/0/RP0/CPU0:NCS-5501#show proc memory detail
Pilha de dados de texto JID Dyn-Limit dinâmico Shm-Tot Phy-Tot  Processo
=============================================================================================================
=======
1150 896K 359M 136K 32M 1024M 18M 24M    lspv_server
1084 2M 2368M 136K 30M 7447M 43M 69M  bgp
1260 1M 1529M 160K 30M 8192M 38M 52M  ipv4_rib
380 316K 1833M 136K 29M 2048M 25M 94M    parser_server
1262 1M 1134M 152K 28M 8192M 22M 31M  ipv6_rib
1277 4M 1445M 136K 21M 1024M 18M 41M  PIM6
1301 80K 222M 136K 20M 300M 5M 33M    servidor_esquema
1276 4M 1637M 136K 20M 1024M 19M 41M  pim
250 124K 676M 136K 20M 1024M 9M 31M   invmgr_proxy
1294 4M 2024M 136K 19M 1861M 48M 66M  l2vpn_mgr
211 212K 676M 136K 18M 300M 9M 29M    sdr_invmgr
1257 4K 663M 136K 17M 2048M 20M 39M   statsd_manager_g
288 4K 534M 136K 16M 2048M 15M 33M    statsd_manager_l
...
```

## Principais consumidores de memória

```
RP/0/RP0/CPU0:NCS-5501#show memory-top-consumer
######################################################################
 Principais consumidores de memória em 0/0/CPU0 (em 2022/Apr/13/15:54:12)
######################################################################
    PID Processo Total (MB) Heap (MB) Shared (MB)
   3469 fia_driver 826 492,82 321
   4091 fib_mgr 175 1094,43 155
   3456 spp 130 9,68 124
   4063 dpa_port_mapper 108 1.12 105
   3457 pacote 104 1.36 101
   5097 l2fib_mgr 86 52.01 71
   4147 bfd_agent 78 6,66 66
   4958 eth_intf_ea 66 4,76 61
   4131 optics_driver 62 141,23 22
   4090 ipv6_nd 55 4.13 49
######################################################################
 Principais consumidores de memória em 0/RP0/CPU0 (em 2022/Apr/13/15:54:12)
######################################################################
    PID Processo Total (MB) Heap (MB) Shared (MB)
   3581 spp 119 9,62 114
   4352 dpa_port_mapper 106 2.75 102
   4494 fib_mgr 99 7,71 90
   3582 pacote 96 1,48 94
   3684 parser_server 95 64.27 25
   8144 te_control 71 15,06 55
   8980 bgp 70 27,61 44
   7674 l2vpn_mgr 67 23,64 48
   8376 mibd_interface 65 35.28 28
   3608 gsp 65 15,75 48
```

## Memória Total - Usada e Disponível

Os componentes do sistema têm uma quantidade fixa de memória disponível.

```
RP/0/RP0/CPU0:NCS-5501#show memory summary location all
node: node0_0_CPU0
------------------------------------------------------------------
 Memória física: Total de 8192 M (6172 M disponíveis)
 Memória do Aplicativo: 8192M (6172M disponíveis)
 Imagem: 4M (botram: 0 M)
 Reservado: 0M, IOMem: 0M, flashfsys: 0 M
 Janela total compartilhada: 226 M
node: node0_RP0_CPU0
------------------------------------------------------------------
 Memória física: Total de 18432M (15344M disponíveis)
 Memória do Aplicativo: 18432M (15344M disponíveis)
 Imagem: 4M (botram: 0 M)
 Reservado: 0M, IOMem: 0M, flashfsys: 0 M
 Janela total compartilhada: 181 M
```

A janela de memória compartilhada fornece informações sobre as alocações de memória compartilhada no sistema.

```
RP/0/RP0/CPU0:NCS-5501#show memory summary detail location 0/RP0/CPU0
```

```
node: node0_RP0_CPU0
------------------------------------------------------------------
 Memória física: Total de 18432M (15344M disponíveis)
 Memória do Aplicativo: 18432M (15344M disponíveis)
 Imagem: 4M (botram: 0 M)
 Reservado: 0M, IOMem: 0M, flashfsys: 0 M
 Janela compartilhada soasync-app-1: 243,328K
 Janela compartilhada soasync-12: 3,328 K
 ...
 Janela compartilhada rewrite-db: 272,164K
 Janela compartilhada l2fib_brg_shm: 139,758 K
 _Regras de im da janela compartilhada: 384,211K
 Janela compartilhada grid_svr_shm: 44,272M
 Janela compartilhada spp: 86,387M
 im_db de janela compartilhada: 1,306 M
 Janela total compartilhada: 180,969 M
 Memória Alocada: 2,337G
 Texto do programa: 127,993T
 Dados do programa: 64,479G
 Pilha de programas: 2,034G
RAM do sistema: 18432M ( 19327352832)
Total usado: 3088M ( 3238002688)
 Privado usado: 0M ( 0)
 Usado compartilhado: 3088M ( 3238002688)
```

Você pode verificar os processos do participante com uma janela de memória compartilhada.

```
RP/0/RP0/CPU0:NCS-5501#sh shmwin spp lista de participantes
Dados relativos ao Window "spp":
-----------------------------
Lista dos participantes atuais:-
NAME PID JID INDEX
3581 113 0
pacote 3582 345 1
ncd 4362 432 2
netio 4354 234 3
nsr_ping_reply 4371 291 4
aib 4423 296 5
ipv6_io 4497 430 6
ipv4_io 4484 438 7
fib_mgr 4494 293 8
...
snmpd 8171 1002 44
ospf 8417 1030 45
mpls_ldp 7678 1292 46
bgp 8980 1084 47
cdp 9295 337 48
RP/0/RP0/CPU0:BRU-SPCORE-PE6#sh shmwin soasync-1 lista de participantes
Dados para a janela "soasync-1":
-----------------------------
Lista dos participantes atuais:-
NAME PID JID INDEX
tcp 5584 168 0
bgp 8980 1084
```

## Monitoramento de recursos e vigilantes

A utilização da memória é monitorada através de um watchdog do sistema no cXR e com Resmon no eXR.

```
RP/0/RP0/CPU0:NCS-5501#show watchdog memory-state
---- node0_RP0_CPU0 ----
Informações de memória:
    Memória física: 18432,0 MB
    Memória livre: 15348,0 MB
    Estado da Memória: Normal
RP/0/RP0/CPU0:NCS-5501#
RP/0/RP0/CPU0:NCS-5501#show watchdog threshold memory defaults location 0/RP0/CPU0
---- node0_RP0_CPU0 ----
 Limites de memória padrão:
 Secundária: 1843 MB ß — 10%
 Grave: 1474 MB ß — 8%
 Crítica: 921,599 MB ß — 5%
Informações de memória:
    Memória física: 18432,0 MB
    Memória livre: 15340,0 MB
    Estado da Memória: Normal
RP/0/RP0/CPU0:NCS-5501#
RP/0/RP0/CPU0:NCS-5501(config)#watchdog threshold memory minor ?
  <5-40> consumo de memória em porcentagem
```

Um aviso será impresso se os limites forem ultrapassados.

```
RP/0/RP0/CPU0:Fev 17 23:30:21.663 UTC: resmon[425]: %HA-HA_WD-4-MEMORY_ALARM : Limite de memória
ultrapassado: Secundário com 1840.000MB livres. Estado anterior: Normal
RP/0/RP0/CPU0:Fev 17 23:30:21.664 UTC: resmon[425]: %HA-HA_WD-6-TOP_MEMORY_USERS_INFO : Os 5
principais consumidores de memória do sistema (1884160 Kbytes livres):
RP/0/RP0/CPU0:Fev 17 23:30:21.664 UTC: resmon[425]: %HA-HA_WD-6-TOP_MEMORY_USER_INFO : 0: Nome do
processo: bgp[0], pid: 7861, Uso de pilha: 12207392 kbytes.
RP/0/RP0/CPU0:Fev 17 23:30:21.664 UTC: resmon[425]: %HA-HA_WD-6-TOP_MEMORY_USER_INFO : 1: Nome do
processo: ipv4_rib[0], pid: 4726, Uso de pilha: 708784 kbytes.
RP/0/RP0/CPU0:Fev 17 23:30:21.664 UTC: resmon[425]: %HA-HA_WD-6-TOP_MEMORY_USER_INFO : 2: Nome do
processo: fib_mgr[0], pid: 3870, Uso de pilha: 584072 kbytes.
RP/0/RP0/CPU0:Fev 17 23:30:21.664 UTC: resmon[425]: %HA-HA_WD-6-TOP_MEMORY_USER_INFO : 3: Nome do
processo: netconf[0], pid: 9260, Uso de pilha: 553352 kbytes.
RP/0/RP0/CPU0:Fev 17 23:30:21.664 UTC: resmon[425]: %HA-HA_WD-6-TOP_MEMORY_USER_INFO : 4: Nome do
processo: netio[0], pid: 3655, Uso de pilha: 253556 kbytes.
LC/0/3/CPU0:08 de março 05:48:58.414 PST: resmon[172]: %HA-HA_WD-4-MEMORY_ALARM : Limite de memória
ultrapassado: Grave, com 600,182MB livres. Estado anterior: Normal
LC/0/3/CPU0:08 de março 05:48:58.435 PST: resmon[172]: %HA-HA_WD-4-TOP_MEMORY_USERS_WARNING : Os 5
principais consumidores de memória do sistema (624654 Kbytes livres):
LC/0/3/CPU0:08 de março 05:48:58.435 PST: resmon[172]: %HA-HA_WD-4-TOP_MEMORY_USER_WARNING : 0: Nome
do processo: fib_mgr[0], pid: 5375, Uso de pilha 1014064 Kbytes.
LC/0/3/CPU0:08 de março 05:48:58.435 PST: resmon[172]: %HA-HA_WD-4-TOP_MEMORY_USER_WARNING : 1: Nome
do processo: ipv4_mfwd_partner[0], pid: 5324, Uso de pilha 185596 Kbytes.
LC/0/3/CPU0:08 de março 05:48:58.435 PST: resmon[172]: %HA-HA_WD-4-TOP_MEMORY_USER_WARNING : 2: Nome
do processo: nfsvr[0], pid: 8357, Uso de pilha 183692 Kbytes.
LC/0/3/CPU0:08 de março 05:48:58.435 PST: resmon[172]: %HA-HA_WD-4-TOP_MEMORY_USER_WARNING : 3: Nome
do processo: fia_driver[0], pid: 3542, Uso de pilha 177552 Kbytes.
LC/0/3/CPU0:08 de março 05:48:58.435 PST: resmon[172]: %HA-HA_WD-4-TOP_MEMORY_USER_WARNING : 4: Nome
do processo: npu_driver[0], pid: 3525, Uso de pilha 177156 Kbytes.
```

Alguns processos podem executar ações específicas com base no estado da memória do watchdog. Por exemplo, o BGP faz o seguinte:

- no estado secundário, o BGP pára de ativar novos peers
- no estado severo, o BGP gradualmente derruba alguns peers.
- em um estado crítico, o processo BGP é desligado.

Os processos podem ser configurados para registrar notificações de estado de memória.

```
Mostrar watchdog ou processo de reconhecimento
```

Os usuários podem desativar o desligamento automático do processo devido ao tempo limite do watchdog.

```
watchdog restart memory-hog disable
```

# Onde encontrar mais informações?

- Repositório de blogs e white papers do Cisco IOS XR (xrdocs.io)
  - Design de malha central: https://xrdocs.io/design/blogs/latest-core-fabric-hld: Este whitepaper discute as tendências e a evolução recentes em redes de backbone central.
  - Design de malha de peering: https://xrdocs.io/design/blogs/latest-peering-fabric-hld : Este whitepaper fornece uma visão geral abrangente dos desafios e das práticas recomendadas para o projeto de peering com foco na simplificação da rede.
- Guia de configuração: Este guia fornece informações sobre o BGP: https://www.cisco.com/c/en/us/td/docs/iosxr/ncs5500/bgp/76x/b-bgp-cg-ncs5500-76x/implementing-bgp.html
- Guia de referência de comandos: Este guia descreve os comandos usados para configurar e monitorar o BGP em Cisco NCS 5500 Series Routers usando o software Cisco IOS XR: https://www.cisco.com/c/en/us/td/docs/iosxr/ncs5500/bgp/b-ncs5500-bgp-cli-reference.html