Implementar proteção contra sobrecarga para gateways e elementos de rede vizinhos no ASR5x00 Series

Contents

Introduction

Este documento descreve como implementar os recursos de proteção disponíveis para Gateways (GWs) e elementos de rede vizinhos no Cisco Aggregated Services Router (ASR) 5x00 Series para proteger o desempenho geral da rede.

Controle de congestionamento para GWs

Controle de congestionamento é um recurso genérico de autoproteção. Ele é usado para proteger o sistema contra surtos de utilização desses recursos:

- Uso da CPU em placas de processamento
- Uso de memória em placas de processamento

Quando a utilização excede os limites predefinidos, todas as novas chamadas (ativações do Protocolo de Dados de Pacotes (PDP - Packet Data Protocol), ativações de sessão da Rede de Dados de Pacotes (PDN - Packet Data Network) são *descartadas* ou *rejeitadas*, dependendo da configuração.

Este é um exemplo que mostra como monitorar a utilização geral da placa de processamento de dados (DPC):

```
congestion-control threshold system-cpu-utilization 85

congestion-control threshold system-memory-utilization 85

congestion-control policy ggsn-service action drop

congestion-control policy sgw-service action drop

congestion-control policy pgw-service action drop
```

> **Note**: O limite de engenharia do sistema é 80% da utilização da CPU, que é definida como o limite de engenharia recomendado que não deve ser excedido para garantir a operação regular do sistema. A carga além do valor pode afetar as operações da plataforma, como sua estabilidade e previsibilidade, e deve ser evitada com o planejamento adequado da capacidade.

> **Note**: A Cisco recomenda que você use a ação *drop* em vez da ação *reject*, pois as chamadas rejeitadas causam tentativas imediatas de reconexão repetidas do equipamento do usuário (UE). No caso de uma ação de queda, a UE espera alguns segundos antes de fazer tentativas repetidas de reconexão, portanto, a taxa de chamada é reduzida.

Proteção contra sobrecarga de rede para limitação de mensagens GTP-C de entrada

Esse recurso protege os processos de GW de pacote (P-GW)/GPRS Support Node (GGSN) de surtos de transmissão e falhas de elementos de rede. Em um nó de suporte GPRS P-GW/Serving GPRS (SGSN), o principal gargalo está relacionado ao processamento de dados do usuário, como a utilização do gerenciador de sessão e a utilização geral da CPU e memória DPC.

Um *valor Nenhum* é configurado no SGSN/Mobility Management Entity (MME) para limitar as mensagens GPRS Tunneling Protocol-Control (GTP-C) de entrada quando a proteção de sobrecarga de rede é ativada.

> **Note**: O uso de GTP e limitação de interface de diâmetro exige a instalação de uma chave de licença válida.

Esse recurso ajuda a controlar a taxa de mensagens de entrada/saída no P-GW/GGSN, o que ajuda a garantir que o P-GW/GGSN não seja sobrecarregado pelas mensagens do plano de controle GTP. Além disso, ajuda a garantir que o P-GW/GGSN não sobrecarregue o peer GTP-C com as mensagens do plano de controle GTP. Este recurso exige que as mensagens de controle GTP (Versão 1 (v1) e Versão 2 (v2)) sejam moldadas/policiadas sobre as interfaces Gn/Gp e S5/S8. Este recurso cobre a proteção contra sobrecarga dos nós P-GW/GGSN e dos outros nós externos com os quais ele se comunica. A limitação é feita somente para mensagens de controle no nível da sessão, de modo que as mensagens de gerenciamento de caminho não são limitadas por taxa.

A sobrecarga de nó externo pode ocorrer em um cenário em que o P-GW/GGSN gera solicitações de sinalização em uma taxa mais alta do que os outros nós podem lidar. Além disso, se a taxa de entrada for alta no nó P-GW/GGSN, ela poderá inundar o nó externo. Por esse motivo, o controle

de fluxo das mensagens de controle de entrada e de saída é necessário. Para proteger os nós externos de uma sobrecarga devido à sinalização de controle P-GW/GSN, uma estrutura é usada para modelar e policiar as mensagens de controle de saída para as interfaces externas.

Configurar a limitação de mensagens GTP-C de entrada

Insira este comando para configurar o controle de limitação de mensagem GTP-C de entrada:

```
gtpc overload-protection Ingress
```

Isso configura a proteção contra sobrecarga da interface GGSN/PGW, limitando as mensagens de controle GTPv1 e GTPv2 de entrada sobre a interface Gn/Gp (GTPv1) ou S5/S8 (GTPv2) com os outros parâmetros para os serviços que são configurados em um contexto e aplicados à GGSN e PGW.

Quando você digita o comando anterior, este prompt é gerado:

```
[context_name]host_name(config-ctx)# gtpc overload-protection ingress
{msg-rate msg_rate} [delay-tolerance dur] [queue-size size]
[no] gtpc overload-protection Ingress
```

Aqui estão algumas notas sobre esta sintaxe:

- **não:** Esse parâmetro desabilita o controle de entrada GTP de limitação de mensagem para os serviços GGSN/PGW neste contexto.
- **msg-rate msg_rate**: Esse parâmetro define o número de mensagens de entrada GTP que podem ser processadas por segundo. O *msg_rate* é um inteiro que varia de cem a 12.000.
- **tolerância a retardo**: Esse parâmetro define o número máximo de segundos em que uma mensagem GTP de entrada pode ser enfileirada antes de ser processada. Após essa tolerância ser excedida, a mensagem será removida. O *dur* é um inteiro que varia de um a dez.
- **tamanho da fila**: Esse parâmetro define o tamanho máximo da fila para as mensagens GTP-C de entrada. Se a fila exceder o tamanho definido, todas as novas mensagens de entrada serão descartadas. O *tamanho* é um inteiro que varia de cem a 10.000.

Você pode usar esse comando para ativar o controle de entrada de mensagens GTP para os serviços GGSN/PGW configurados no mesmo contexto. Como exemplo, este comando ativa as mensagens de controle GTP de entrada em um contexto com uma taxa de mensagem de *1.000* por segundo, um tamanho de fila de mensagens de *10.000* e um atraso de *um* segundo:

```
gtpc overload-protection ingress msg-rate 1000 delay-tolerance 1 queue-size 10000
```

Proteção do Elemento de Rede Vizinho

Muitos elementos de rede vizinhos usam seus próprios mecanismos para se proteger, e a proteção adicional contra sobrecarga de rede no lado ASR5x00 pode não ser necessária. A proteção dos elementos da rede vizinha pode ser necessária nos casos em que a estabilidade

geral da rede pode ser alcançada somente quando o controle de mensagens é aplicado no lado de saída.

Proteção contra sobrecarga de rede com limitação de diâmetro em uma interface S6a

Esse recurso protege as interfaces S6a e S13 na direção de saída. Ele protege o Home Subscriber Server (HSS), o Diameter Routing Agent (DRA) e o Equipment Identity Register (EIR). O recurso usa a RLF (Rate Limiting Function, função de limitação de taxa).

Considere estas notas importantes ao aplicar a configuração de ponto final de diâmetro:

- Um modelo RLF deve ser associado ao peer.
- Um RLF é anexado somente por peer (individualmente).

Configurar a limitação de diâmetro em uma interface S6a

Esta é a sintaxe de comando que é usada para configurar a limitação de diâmetro em uma interface S6a:

```
[context_na>me]host_name(config-ctx-diameter)#>peer [*] peer_name [*]
[ realm realm_name ] { address ipv4/ipv6_address [ [ port port_number ]
[connect-on-application-access] [ send-dpr-before-disconnect disconnect-cause
disconnect_cause ] [ sctp ] ] + | fqdn fqdn [ [ port port_number ]
[ send-dpr-before-disconnect disconnect-cause disconnect_cause ]
[ rlf-template rlf_template_name ] ] }

no peer peer_name [ realm realm_name ]
```

Aqui estão algumas notas sobre esta sintaxe:

- **não:** Este parâmetro remove a configuração de peer especificada.
- **[*] peer_name [*]**: Esse parâmetro especifica o nome do peer como uma string alfanumérica que varia de um a 63 caracteres (caracteres de pontuação são permitidos).**Note**: O ponto final do servidor de diâmetro agora pode ser um nome de peer cardado (com o caractere * como um caractere curinga válido). Os peers clientes que satisfazem o padrão cardado pela rede são tratados como peers válidos, e a conexão é aceita. O token cardado como selvagem indica que o nome do peer é cardado como selvagem e qualquer caractere * na string que precede é tratado como curinga.
- **realm_name**: Esse parâmetro especifica o território desse peer como uma string alfanumérica que varia de um a 127 caracteres. O nome do território pode ser uma empresa ou um nome de serviço.
- **endereço ipv4/ipv6_address**: Esse parâmetro especifica o endereço IP do peer de diâmetro em notação decimal pontuada IPv4 ou notação hexadecimal separada por dois pontos IPv6. Esse endereço deve ser o endereço IP do dispositivo com o qual o chassi se comunica.
- **fqdn fqdn**: Esse parâmetro especifica o par de diâmetro FQDN (Fully Qualified Domain Name,

nome de domínio totalmente qualificado) como uma sequência alfanumérica que varia de um a 127 caracteres.

- **port_number**: Este parâmetro especifica o número da porta para este par de diâmetro. O número da porta deve ser um inteiro que varie de um a 65.535.
- **connect-on-application-access**: Esse parâmetro ativa o peer no acesso inicial do aplicativo.
- **send-dpr-before-disconnect**: Esse parâmetro envia o DPR (Disconnect-Peer-Request).
- **causa da desconexão**: Esse parâmetro encerra o DPR para o peer especificado, com o motivo de desconexão especificado. A causa da desconexão deve ser um inteiro que varia de zero a dois, correspondendo às seguintes causas:

  0Â Â REINICIALIZAÇÃO

  1 Â Â OCUPADO

  2 â Â DO_NOT_WANT_TO_TALK_TO_VOCÊ

- **rlf-template rlf_template_name**: Esse parâmetro especifica o modelo de RLF a ser associado a esse par de diâmetro. O *rlf_template_name* deve ser uma cadeia de caracteres alfanumérica que vai de um a 127 caracteres.

  Note: Uma licença RLF é necessária para configurar um modelo RLF.

Proteção contra sobrecarga de rede com limitação de diâmetro em uma interface Gx/Gy

Este recurso protege as interfaces Gx e Gy na direção de saída. Ele protege a função de política e regras de cobrança (PCRF) e o sistema de cobrança online (OCS) e usa RLF.

Considere estas notas importantes ao aplicar a configuração de ponto final de diâmetro:

- Um modelo RLF deve ser associado ao peer.
- Um RLF é anexado somente por peer (individualmente).

Este comando é usado para configurar a proteção contra sobrecarga de rede:

```
[context_name]host_name(config-ctx-diameter)# rlf-template rlf_template_name
```

Note: É necessária uma licença RLF para configurar um modelo RLF

Configurar a limitação de diâmetro em uma interface Gx/Gy

Você pode considerar o uso do RLF para interfaces de diâmetro. Aqui está um exemplo de

configuração:

```
rlf-template rlf1

msg-rate 1000 burst-size 100

threshold upper 80 lower 60

delay-tolerance 4

#exit


diameter endpoint Gy

use-proxy

origin host Gy address 10.55.22.3

rlf-template rlf1

peer peer1 realm foo.com address 10.55.22.1 port 3867 rlf-template rlf2

peer peer2 realm fo.com address 10.55.22.1 port 3870

#exit
```

Aqui estão algumas observações sobre esta configuração:

- O peer chamado *peer1* está vinculado a *RFL2*, e o restante dos peers sob o endpoint está associado ao *RLF1*.
- O modelo de RLF de nível de peer tem precedência sobre o modelo de nível de endpoint.
- O número de mensagens é enviado a uma taxa máxima de 1.000 por segundo.(msg-rate). Estas considerações aplicam-se igualmente:

  Apenas cem mensagens (tamanho de intermitência) são enviadas a cada cem milissegundos (para alcançar as 1.000 mensagens por segundo).

  Se o número de mensagens na fila RLF exceder 80% da taxa de mensagens (80% de 1.000 = 800), o RLF passará para o estado *OVER_THRESHOLD*.

  Se o número de mensagens na fila RLF exceder a taxa de mensagens (1.000), o RLF passará para o estado *OVER_LIMIT*.

  Se o número de mensagens na fila de RLF diminuir abaixo de 60% da taxa de mensagem (60% de 1.000 = 600), o RLF muda de volta para o estado *READY*.

  O número máximo de mensagens que podem ser enfileiradas é igual à taxa de mensagem multiplicada pela tolerância de atraso (1.000 x 4 = 4.000).

  Se o aplicativo enviar mais de 4.000 mensagens para o RLF, os primeiros 4.000 serão enfileirados e o restante serão descartados.

As mensagens que são descartadas são repetidas vezes/reenviadas pelo aplicativo para o RLF em um período de tempo apropriado.

O número de novas tentativas é de responsabilidade do aplicativo.

- O modelo pode ser desvinculado do ponto de extremidade com o parâmetro *no rlf-template*. Por exemplo, ele desvincularia o *RLF1* do *peer2*.

- Não use o parâmetro *no rlf-template rlf1* no modo de *configuração de ponto final*, pois a CLI tenta excluir o *RLF1* do modelo RLF. Esse comando CLI faz parte da configuração global, não da configuração de ponto de extremidade.

- O modelo pode ser associado a peers individuais através de um destes comandos:

```
no peer peer2 realm foo.com

peer peer2 realm foo.com address 10.55.22.1 port 3867
```

- O RLF só pode ser usado para endpoints de diâmetro em que diamproxy é usado.

- A taxa de mensagem configurada é implementada por proxy de diamante. Por exemplo, se a taxa de mensagens for 1.000 e 12 diamproxies estiverem ativos (chassi totalmente preenchido = 12 Packet Services Card (PSC) ativo + 1 Demux + 1 PSC em standby), as transmissões efetivas por segundo (TPS) serão 12.000. Você pode inserir um destes comandos para exibir as estatísticas de contexto de RLF:

```
show rlf-context-statistics diamproxy

show rlf-context-statistics diamproxy verbose
```

Proteção contra sobrecarga de rede por meio da limitação de página com RLF

O recurso de limitação de página limita o número de mensagens de paginação que são enviadas do SGSN. Ele oferece flexibilidade e controle ao operador, que agora pode reduzir o número de mensagens de paging que são enviadas do SGSN com base nas condições da rede. Em alguns locais, a quantidade de mensagens de paging iniciadas pelo SGSN é muito alta devido a condições de rádio ruins. Um número maior de mensagens de paginação resulta no consumo de largura de banda na rede. Este recurso fornece um limite de taxa configurável, no qual a mensagem de paginação é limitada nestes níveis:

- O nível global para acesso 2G e 3G

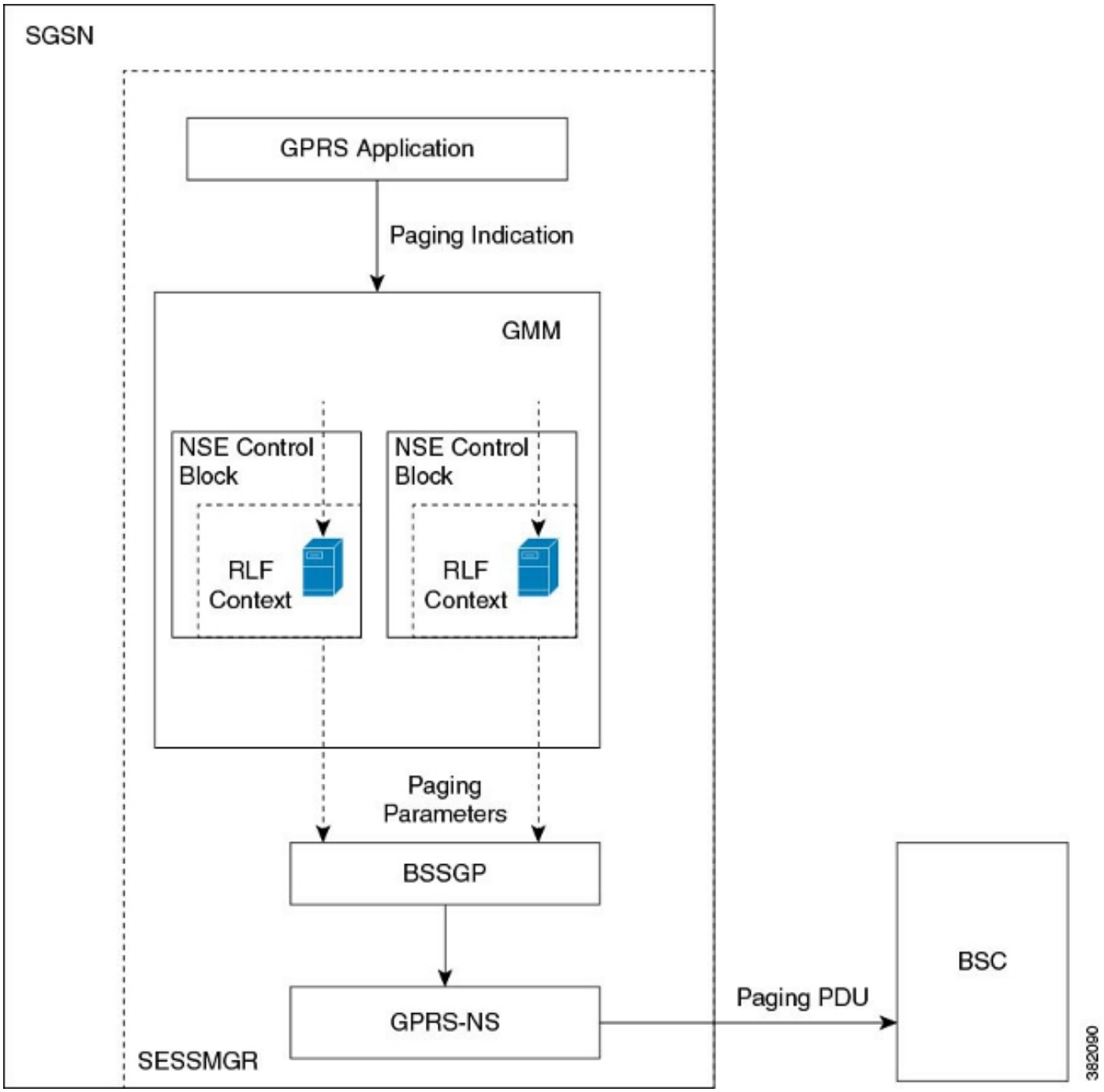
- O nível NSE (Network Service Entity, entidade de serviço de rede) somente para acesso 2G

- O nível RNC (Radio Network Controller, Controlador de Rede de Rádio) somente para acesso 3G

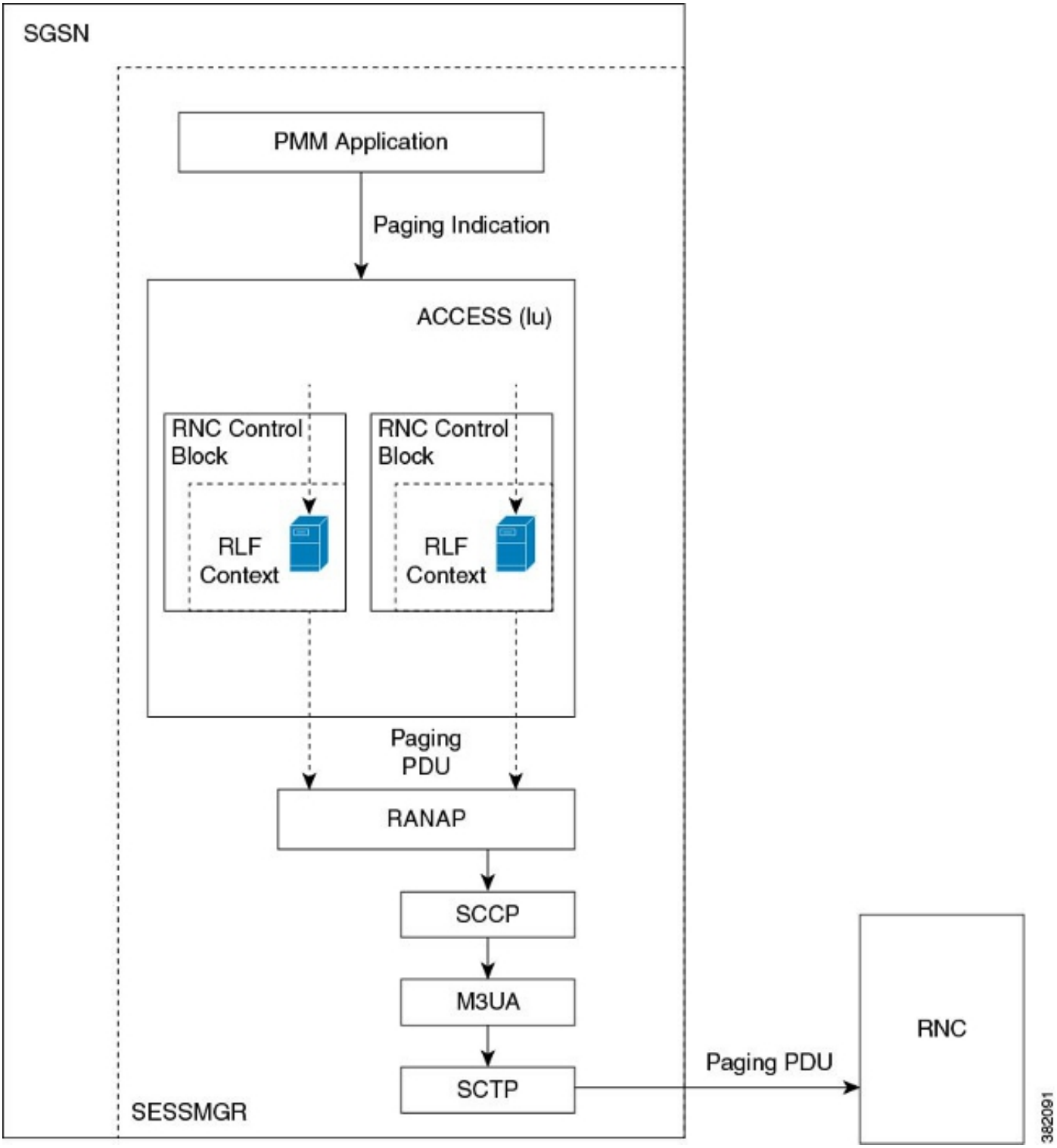
Esse recurso melhora o consumo de largura de banda na interface de rádio.

Note: Uma licença RLF é necessária para configurar um modelo RLF.

Aqui está um exemplo do processo de paginação com acesso 2G e limitação de taxa:



Aqui está um exemplo do processo de paginação com acesso 3G e limitação de taxa:



Configurar a limitação de página com RLF

Os comandos descritos nesta seção são usados para configurar o recurso de limitação de página. Esses comandos CLI são usados para associar/remover o modelo RLF para limitação de página no nível global, no nível NSE e no nível RNC no SGSN.

Mapeie o nome RNC para o identificador RNC

O comando **interface** é usado para configurar o mapeamento entre o Identificador RNC (ID) e o nome RNC. Você pode configurar o *paging-rlf-template* por nome RNC ou ID RNC. Esta é a sintaxe usada:

```
config
sgsn-global
interface-management
[ no ] interface {gb
peer-nsei | iu peer-rnc} {name <value> | id <value>}
exit
```

Note: A forma *no* do comando remove o mapeamento e outra configuração associada à *paging-rlf-template do* RNC da SGSN e redefine o comportamento para o padrão desse RNC.

Aqui está um exemplo de configuração:

```
[local]asr5000# configure
[local]asr5000(config)# sgsn-global
[local]asr5000(config-sgsn-global)# interface-management
[local]asr5000(config-sgsn-interface-mgmt)# interface
iu peer-rnc id 250 name bng_rnc1
[local]asr5000(config-sgsn-interface-mgmt)# end
[local]asr5000#
```

Associar um modelo de RLF de paginação

Esse comando permite que o SGSN associe um modelo de RLF no nível global, o que limita as mensagens de paging iniciadas no acesso 2G (nível de NSE) e 3G (nível de RNC), ou no nível por entidade, que está no nível de RNC para acesso 3G ou no nível de NSE para acesso 2G. Esta é a sintaxe usada:

```
config
sgsn-global
interface-management
[no] paging-rlf-template {template-name <template-name>} {gb
peer-nsei | iu peer-rnc} {name <value> | id <value>}
exit
```

Note: Se não houver nenhum modelo de RLF associado a um NSE/RNC específico, a carga de paginação será limitada com base no modelo global de RLF associado (se houver). Se nenhum modelo global de RLF estiver associado, nenhum limite de taxa será aplicado na carga de paginação.

Aqui está um exemplo de configuração:

```
[local]asr5000(config)# sgsn-global
[local]asr5000(config-sgsn-global)# interface-management
[local]asr5000(config-sgsn-interface-mgmt)# paging-rlf-template
template-name rlf1
[local]asr5000(config-sgsn-interface-mgmt)# end
[local]asr5000#
[local]asr5000# configure
[local]asr5000(config)# sgsn-global
[local]asr5000(config-sgsn-global)# interface-management
[local]asr5000(config-sgsn-interface-mgmt)# paging-rlf-template
template-name rlf2 gb peer-nsei id 1
[local]asr5000(config-sgsn-interface-mgmt)# end
[local]asr5000#
```

```
[local]asr5000# configure
[local]asr5000(config)# sgsn-global
[local]asr5000(config-sgsn-global)# interface-management
[local]asr5000(config-sgsn-interface-mgmt)# paging-rlf-template
template-name rlf2 iu peer-rnc name bng_rnc1
[local]asr5000(config-sgsn-interface-mgmt)# end
[local]asr5000#
```