

Solucionar problemas do recurso de limitação AAA

Contents

[Introduction](#)

[Prerequisites](#)

[Requirements](#)

[Componentes Utilizados](#)

[Informações de Apoio](#)

[Mecanismo de trabalho](#)

[Filas AAAMGR](#)

[Limitações](#)

[Discussões relacionadas da comunidade de suporte da Cisco](#)

Introduction

Este documento descreve o recurso de limitação de registros AAA (RADIUS) que suporta limitação de acesso (autenticação e autorização) e registros de contabilidade que são enviados ao servidor RADIUS.

Esse recurso permite que um usuário configure a taxa de limitação apropriada para evitar congestionamento e instabilidade da rede quando não houver largura de banda suficiente para acomodar uma intermitência repentina de registros gerados do roteador Cisco para o servidor RADIUS.

Prerequisites

Requirements

Não existem requisitos específicos para este documento.

Componentes Utilizados

As informações neste documento são baseadas na plataforma ASR5k.

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, make sure that you understand the potential impact of any command.

Informações de Apoio

Quando o aaamgr envia as mensagens de raio para o servidor RADIUS a uma taxa alta (por exemplo, quando um grande número de sessões é desativado ao mesmo tempo, as mensagens de parada de contabilidade para todas as sessões são geradas ao mesmo tempo), o servidor

RADIUS pode não conseguir receber as mensagens em taxas tão altas. Para lidar com essa condição, precisamos de um mecanismo de controle de taxa efetivo no aaamgr, de modo que o aaamgr envie mensagens em uma taxa ótima, de modo que o servidor RADIUS seja capaz de receber todas as mensagens e assegure que nenhuma mensagem seja liberada devido à sobrecarga no servidor RADIUS.

Mecanismo de trabalho

Quando o aaamgr envia mensagens na taxa configurada para o servidor RADIUS, ele envia mensagens uniformemente por todo o segundo em vez de enviar todas as mensagens em uma única intermitência. Dependendo da configuração, cada segundo é dividido em vários slots de tempo igual (com um período de tempo específico por slot). O período mínimo de um slot pode ser 50 milissegundos.

A taxa deve ser configurada levando-os em conta

- A taxa de chamadas recebidas,
- Número de instâncias do aaamgr
- A taxa na qual o servidor RADIUS pode receber as mensagens e
- Intervalo de intervalo (para configuração contábil)
- Algoritmo usado para seleção de servidor

Se o valor configurado para servidores de autenticação for muito baixo, haverá um gargalo de garrafa que levará a congestionamento, o que pode levar ao descarte das chamadas devido ao tempo limite de configuração da sessão. Se um valor baixo for configurado para servidores de contabilidade, muita limpeza de mensagens de contabilidade será observada devido ao estouro da fila.

Quando o recurso é configurado, o número de slots de tempo em um segundo e o período de um segundo são computados e armazenados no nível de raio. Quando uma mensagem está pronta para ser enviada ao servidor RADIUS, verifica-se se a cota (número de mensagens para este intervalo de tempo) foi atingida. Se o limite não for atingido, a mensagem será enviada, se for, então a mensagem será colocada na fila de nível de servidor a ser enviada em intervalos de tempo futuros. Cada servidor RADIUS mantém detalhes sobre o número de mensagens enviadas no intervalo de tempo atual e a hora em que o intervalo de tempo expira. Quando as mensagens em fila são selecionadas da fila de nível de servidor, elas são colocadas no cabeçalho da fila de nível de instância, garantindo preferência por mensagens mais antigas que qualquer outra nova mensagem. As mensagens da fila de nível de instância são selecionadas para manutenção.

Filas AAAMGR

Há dois tipos de filas na AAMGR para mensagens:

1. Filas de nível de instância
2. Filas de nível de servidor

Quando uma mensagem é gerada, ela é inicialmente colocada na fila de nível de instância para manutenção.

A fila de nível de instância é processada por 25 milissegundos para cada 50 milissegundos. Qualquer mensagem removida da fila de nível de instância será tentada para ser enviada ao servidor RADIUS. Sob algumas condições, talvez não possamos enviar as mensagens (sem largura de banda disponível ou sem IDs disponíveis). Nesses casos, as mensagens que falharam na tentativa serão enfileiradas nas filas de nível de servidor. Para cada 50 milissegundos, você seleciona quantas mensagens têm IDs disponíveis e também largura de banda disponível e as coloca no cabeçalho da fila de nível de instância (essas mensagens são mais antigas que qualquer outra mensagem presente na fila de nível de instância).

Quando há um controle de taxa para mensagens de relatório e se há muitas mensagens de relatório na fila de nível de instância, qualquer nova mensagem de autenticação vai para o fim da fila de nível de instância. Para ser processado, ele deve aguardar que toda a mensagem contábil (antes da nova mensagem de autenticação) seja enviada ao servidor RADIUS ou movida para a fila de nível de servidor. É um comportamento existente e não é modificado. Isso pode causar um pequeno atraso para que a nova mensagem de autenticação seja processada.

Exemplo

Com base na taxa máxima com o valor de 5, você pode enviar cinco mensagens em 1 segundo e ter 256 mensagens de autenticação radius pendentes (configuração máxima padrão pendente) não atendidas por aaamgr para o servidor de autenticação Radius. Caso haja mais de 5 mensagens, em 1 segundo as mensagens são colocadas na fila até que o servidor AAA responda às solicitações existentes.

Caso você alcance 256 mensagens de autenticação Radius enviadas de um aaamgr para o servidor, as solicitações restantes serão colocadas em fila até que o servidor AAA responda às solicitações existentes. Ele voltará a entrar na mesma fila que a taxa máxima. A mensagem é retirada da fila somente quando você tem um slot livre. O slot livre entra quando você recebe uma resposta para a mensagem ou quando ele expira.

Limitações

Como o Cisco ASR5K é um sistema distribuído com pares independentes sessmgr/aaamgr processando as chamadas, o controle de taxa pode ser implementado somente para instâncias independentes do aaamgr. É teórico estender a taxa de uma única instância para toda a caixa do Cisco ASR5K como um todo simplesmente multiplicando o número total de instâncias pela taxa máxima de cada instância.

Esse número é apenas o limite máximo absoluto em um cenário de dia ensolarado. Você não pode tratar o Cisco ASR5K como uma caixa preta e não pode supor que todas as chamadas terão êxito se o valor calculado visto no sistema não tiver ultrapassado o limite superior.

A taxa máxima do rádio está vinculada a outros parâmetros internos e externos relacionados ao sistema. Veja o impacto esperado se uma das condições não for atendida.

Condições

Distribuição uniforme de chamadas do demuxmgr para todos os sessmgrs

Impacto se não atendido

Se a distribuição da chamada não for uniforme, as mensagens de raio podem ser enfileirado para algumas instâncias. Assim, mesmo que o limite teórico de taxa máxima não seja atingido, as chamadas serão descartadas para instâncias onde as mensagens s

Distribuição uniforme de IMSI (isso é apenas um caso de contabilidade de mediação de rodízio)

Nenhuma intermitência repentina de chamadas recebidas

Os servidores Radius devem responder a tempo

colocadas em fila.

O rodízio de relatório de mediação é baseado no roteamento baseado em IMSI.

Nesse caso, com base na distribuição IMSI, alguns conjuntos de servidores podem ser preferidos em vez de outros com base na lógica de roteamento; a fila pode ser criada para servidores que levam à queda de chamada.

Se houver uma intermitência de novas chamadas, as mensagens de raio recém-geradas serão enfileiradas no sistema. Quando as novas solicitações radius forem processadas. O tempo de configuração da sessão pode ter expirado, levando a quedas de chamada.

Quando as solicitações radius expirarem devido a problemas de servidor, haverá novamente a criação da fila, pois novas solicitações não serão enviadas a menos que a atual que espera uma resposta seja removida do sistema. A taxa na qual as mensagens de tempo limite serão removidas do sistema depende do máximo de configurações pendentes e de tempo limite.

Em muitos casos, podemos ver que as solicitações de acesso não são processadas por todas as tarefas ativas do aaamgr. Isso significa que temos distribuição desigual de chamadas dentro das tarefas do sessmgr e mais adiante, nem todas as instâncias do aaamgr estão envolvidas no processamento de chamadas.

A distribuição de chamadas não se baseia no mecanismo de rodízio estrito, ou seja, se houver 10 chamadas recebidas, elas irão para 10 sms em um algoritmo monotônico.

A distribuição de chamadas se baseia nesses quatro fatores principais

- **ative_session_count**
- **cpu_load**
- **Round_trip_delay** (demuxmgr - sessmgr - demuxmgr)
- **notável_add_request** (demux para sessmgr)

Esta é a implementação atual. A taxa máxima é apenas um limite superior, mas devido à natureza distribuída de nossa arquitetura, você não pode extrapolá-la diretamente para a carga do chassi. O comportamento depende da carga em um AAAMgr específico em um determinado momento.

A fila de taxa máxima de rádio deve ser usada para **monitorar o status** do sistema. Se houver uma **compilação de fila**,

significa que uma dessas condições 4 (consulte a tabela) não foi atendida e você deve identificar a causa raiz para a mesma.

**o limite máximo da fila pode ser implementado e monitorado constantemente.