Este é o teste de publicação do artigo com Licenciamento como

Introdução

Este documento descreve a metodologia geral para solucionar problemas de uma experiência de GUI do APIC lenta.

Início rápido

É comum perceber que problemas lentos de GUI do APIC são o resultado de uma alta taxa de solicitações de API originadas de um script, integração ou aplicativo. O access.log de um APIC registra cada solicitação de API processada. O access.log de um APIC pode ser rapidamente analisado com o script Access Log Analyzer no projeto do grupo Github Datacenter aci-tac-scripts.

Informações de Apoio

APIC como um servidor Web - NGINX

O NGINX é o DME responsável pelos endpoints de API disponíveis em cada APIC. Se NGINX estiver inoperante, as solicitações de API não poderão ser tratadas. Se NGINX estiver congestionado, a API está congestionada. Cada APIC executa seu próprio processo NGINX, portanto, é possível que apenas um único APIC possa ter problemas de NGINX se apenas esse APIC for direcionado por qualquer consultante agressivo.

A IU do APIC executa várias solicitações de API para preencher cada página. Da mesma forma, todos os comandos show do APIC (CLI do estilo NXOS) são empacotadores para scripts python que executam várias solicitações de API, manipulam a resposta e a enviam ao usuário.

Registros relevantes

Nome do arquivo de log	Local	Em qual suporte técnico ele está	Comentários
access.log	/var/log/dme/log	APIC 3de3	Independentemente da ACI, oferece 1 linha por solicitação de API
error.log	/var/log/dme/log	APIC 3de3	Independente da ACI, mostra erros nginx (limitação incluída)

nginx.bin.log	/var/log/dme/log	APIC 3de3	Específico da ACI, registra transações DME
nginx.bin.warnplus.log	/var/log/dme/log	APIC 3de3	Específico da ACI contém logs que são de aviso+ gravidade

Metologia

Isolar disparador inicial

O que é afetado?

- Quais APIC são afetados; um, muitos ou todos os APICs?
- Onde se observa lentidão; via interface do usuário, comandos CLI ou ambos?
- Quais páginas ou comandos específicos da interface do usuário são lentos?

Como ocorre a lentidão?

- Isso é visto em vários navegadores para um único usuário?
- Vários usuários relatam lentidão ou apenas um único/subconjunto de usuários?
- Os usuários afetados compartilham uma localização geográfica ou um caminho de rede semelhante do navegador para o APIC?

Quando a lentidão foi notada pela primeira vez?

- Foi adicionada recentemente uma integração ou script da ACI?
- Uma extensão do navegador foi habilitada recentemente?
- Houve uma alteração recente na configuração da ACI?

Verificar o uso e a integridade do NGINX

Formato de Entrada Access.log

O access.log é um recurso do NGINX e, portanto, não depende do APIC. Cada linha representa 1 solicitação HTTP recebida pelo APIC. Consulte esse registro para entender o uso do NGINX de um APIC.

O formato access.log padrão na versão 5.2+ do ACI:

```
log_format proxy_ip '$remote_addr ($http_x_real_ip) - $remote_user [$time_local]'
                    '"$request" $status $body_bytes_sent '
                    '"$http_referer" "$http_user_agent"';
```

Esta linha representa uma entrada access.log quando um moquery -c fvTenant é executado:

```
127.0.0.1 (-) - - [07/Apr/2022:20:10:59 +0000]"GET /api/class/fvTenant.xml HTTP/1.1" 200 15863 "-" "Pyt
```

Mapa da entrada access.log de exemplo para log_format:

campo log_format	Conteúdo do exemplo	Comentários
$remote_addr	127.0.0.1	IP do host que enviou esta solicitação
$http_x_real_ip	-	IP do último solicitante se proxies estiverem em uso
$remote_user	-	Não é geralmente usado. Marque nginx.bin.log para rastrear qual usuário efetuou login para executar solicitações
$time_local	07/abr/2022:20:10:59 +0000	Quando a solicitação foi processada
$request	OBTENHA /api/class/fvTenant.xml HTTP/1.1	Método Http (GET, POST, DELETE) e URI
$status	200	Código de Status da Resposta HTTP
$body_bytes_sent	1586	tamanho de payload de resposta
$http_referer	-	-
$http_user_agent	Python- urllib	Que tipo de cliente enviou a solicitação

Comportamentos do Access.log

Intermitências de solicitação de alta taxa durante um grande período de tempo:

- As intermitências contínuas de mais de 40 solicitações por segundo podem causar lentidão na interface do usuário
- Identificar quais hosts são responsáveis pelas consultas

- Reduza ou desative a origem de consultas para ver se isso melhora o tempo de resposta do APIC.

Respostas 4xx ou 5xx consistentes:

- Se encontrado, identifique a mensagem de erro de nginx.bin.log

O access.log de um APIC pode ser rapidamente analisado com o script Access Log Analyzer no projeto do grupo Github Datacenter aci-tac-scripts.

Verificar uso de recurso NGINX

A utilização da CPU e da memória do NGINX pode ser verificada com o comando top do APIC:

<#root>

```
top - 13:19:47 up 29 days, 2:08, 11 users, load average: 12.24, 11.79, 12.72
Tasks: 785 total, 1 running, 383 sleeping, 0 stopped, 0 zombie
%Cpu(s): 3.5 us, 2.0 sy, 0.0 ni, 94.2 id, 0.1 wa, 0.0 hi, 0.1 si, 0.0 st
KiB Mem : 13141363+total, 50360320 free, 31109680 used, 49943636 buff/cache
KiB Swap: 0 total, 0 free, 0 used. 98279904 avail Mem

PID   USER PR NI VIRT    RES   SHR   S %CPU %MEM TIME+     COMMAND
21495 root 20 0  4393916 3.5g 217624 S
```

2.6

2.8 759:05.78

nginx.bin

O alto uso de recursos NGINX pode se correlacionar diretamente a uma alta taxa de solicitações processadas.

Verificar núcleos

Um travamento de NGINX não é típico para problemas de GUI do Slow APIC. No entanto, se houver núcleos NGINX, anexe-os a um TAC SR para análise. Consulte o guia de suporte técnico da ACI para obter as etapas para verificar os núcleos.

Verificar Latência de Cliente para Servidor

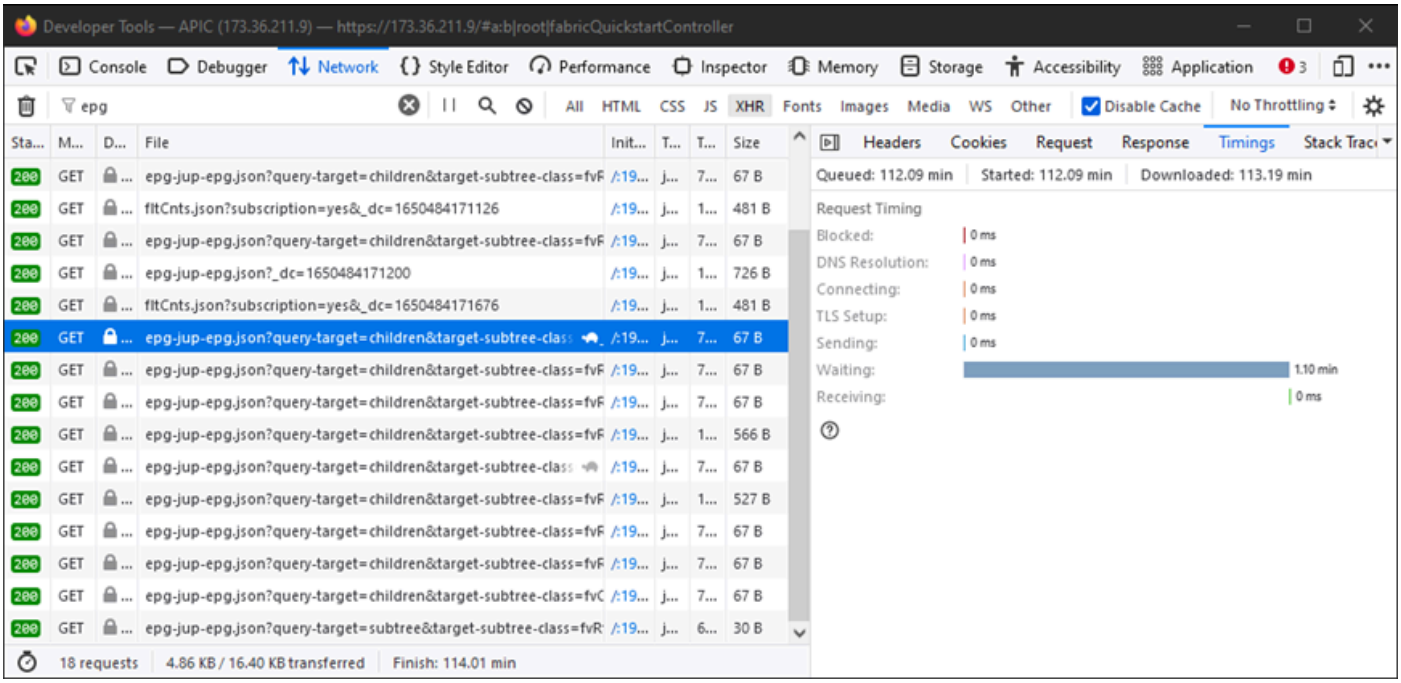
Se solicitações rápidas não forem encontradas, mas um usuário continuar a exibir lentidão da interface do usuário, o problema pode ser a latência de Cliente (navegador) para Servidor (APIC).

Nesses cenários, valide o caminho de dados do navegador para o APIC (distância geográfica, VPN etc.). Se possível, implante e teste o acesso de um servidor de salto localizado na mesma região geográfica ou no mesmo data center que os APICs para isolar. Valide se outros usuários exibem uma quantidade similar de latência.

Guia Rede de Ferramentas de Desenvolvimento de Navegador

Todos os navegadores têm a capacidade de validar solicitações e respostas HTTP por meio do kit de ferramentas Desenvolvimento de navegador, normalmente em uma guia Rede.

Essa ferramenta pode ser usada para validar o tempo necessário para cada estágio de solicitações originadas no navegador, conforme mostrado na imagem.



Exemplo do navegador aguardando 1,1 minuto para que o APIC responda

Aprimoramentos para Páginas de IU Específicas

Página Grupo de Políticas:

ID de bug Cisco CSCvx14621 - GUI do APIC carrega lentamente nas políticas de IPG na guia Estrutura.

Interface na página Inventário:

ID de bug Cisco CSCvx90048 - A carga inicial da guia operacional "Configuração de interface física da camada 1" é longa/induz 'congelamento'.

Recomendações gerais para cliente > Latência do servidor

Certos navegadores, como o Firefox, permitem mais conexões da Web por host por padrão.

- Verifique se essa configuração pode ser definida na versão do navegador usada
- Isso é mais importante para páginas de várias consultas, como a página Grupo de Políticas

A VPN e a distância para o APIC aumentam a lentidão geral da interface do usuário, considerando as solicitações do navegador do cliente e o tempo de viagem de resposta do APIC. Uma caixa de salto geograficamente local para os APICs reduz significativamente o tempo de viagem do navegador para o APIC.

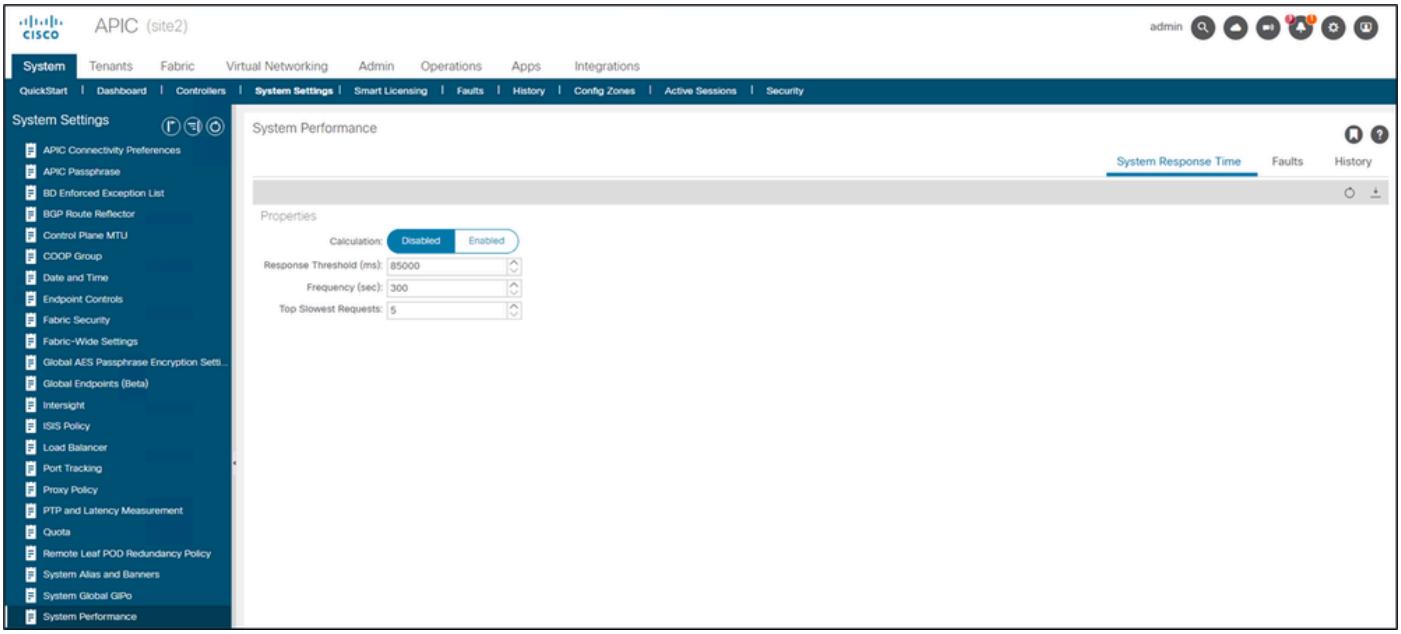
Verificar Solicitações Longas da Web

Se um servidor da Web (NGINX no APIC) lidar com um grande volume de solicitações da Web longas, isso pode afetar o desempenho de outras solicitações recebidas em paralelo.

Isso é especialmente verdadeiro para sistemas que têm bancos de dados distribuídos, como APICs. Uma única solicitação de API pode exigir solicitações e pesquisas adicionais enviadas a outros nós na malha, o que pode resultar em tempos de resposta esperadamente mais longos. Um pico dessas Solicitações da Web Longa em um pequeno intervalo de tempo pode compor a quantidade de recursos necessários e levar a tempos de resposta inesperadamente mais longos. Além disso, as solicitações recebidas podem expirar (90 segundos), o que resulta em um comportamento inesperado do sistema da perspectiva do usuário.

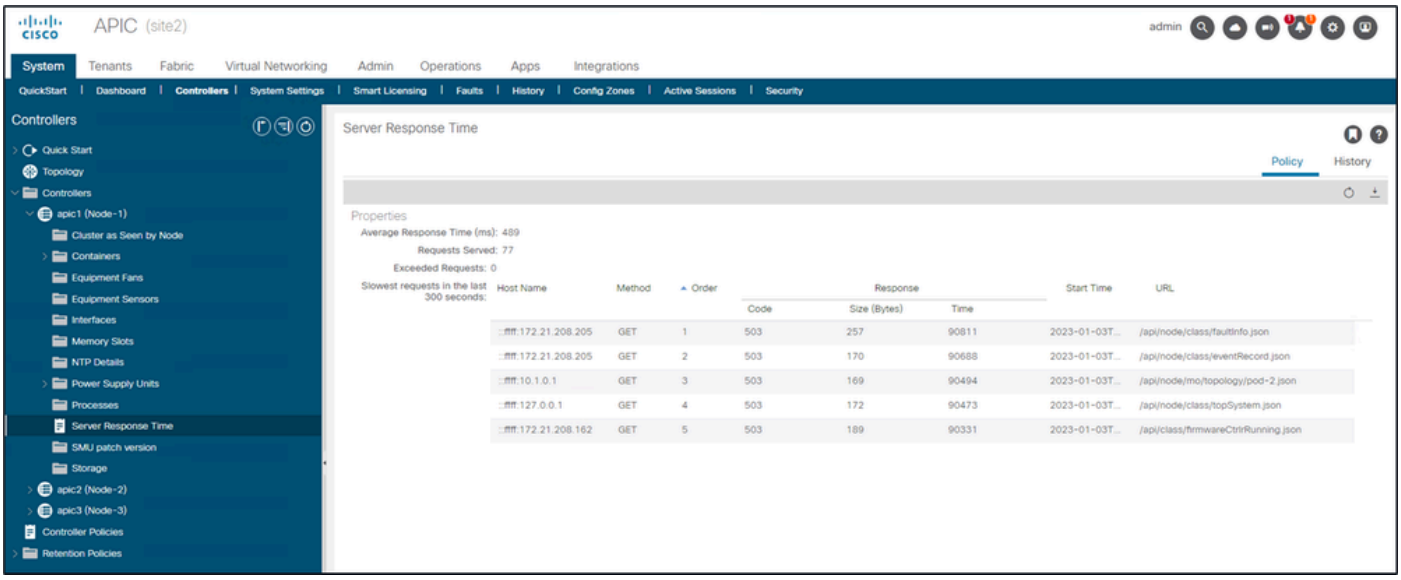
Tempo de Resposta do Sistema - Habilitar Cálculo para Tempo de Resposta do Servidor

No 4.2(1)+, um usuário pode habilitar o "Cálculo de desempenho do sistema", que rastreia e destaca solicitações de API que levaram tempo para serem tratadas.



O cálculo pode ser ativado em Sistema - Configurações do sistema - Desempenho do sistema

Quando o "Cálculo" estiver habilitado, um usuário poderá navegar para APICs específicos em Controladores para visualizar as Solicitações de API mais lentas nos últimos 300 segundos.



Sistema - Controladores - Pasta Controladores - APIC x - Tempo de resposta do servidor

Considerações sobre o uso da API do APIC

Ponteiros gerais para garantir que um script não prejudique o Nginx

- Cada APIC executa seu próprio NGINX DME.
  - Somente o NGINX do APIC 1 processa solicitações para o APIC 1. O NGINX do APIC 2 e do APIC 3 não processa essas solicitações.
- Em geral, mais de 40 solicitações de API por segundo durante um longo período debilita o NGINX.
  - Se encontrados, reduza a agressividade das solicitações.
  - Se o host Requests não puder ser modificado, considere NGINX Rate Limits no APIC.

Ineficiências do script de endereço

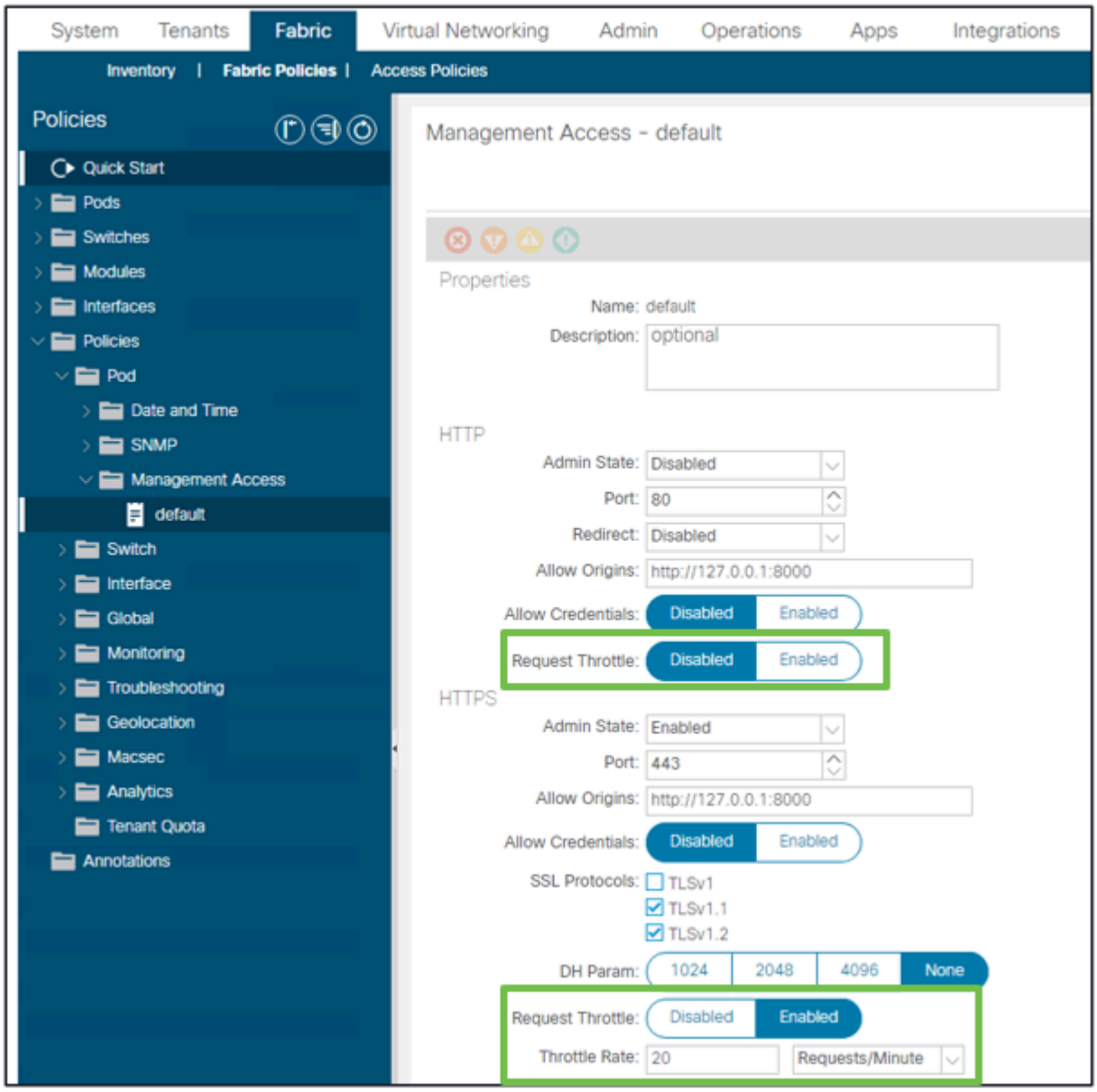
- Não faça logon/logoff antes de cada solicitação de API.
  - O tempo limite padrão para uma sessão de logon é de 10 minutos. Essa mesma sessão pode ser usada para várias solicitações e pode ser atualizada para estender o tempo de validade.
  - Consulte o Guia de configuração da API REST do Cisco APIC - Acesso à API REST - Autenticação e manutenção de uma sessão de API.
- Se o seu script consultar muitos DNs que compartilham um pai, em vez de recolher as consultas em uma única consulta pai lógica com Filtros de Consulta.
  - Consulte o Guia de configuração da API REST do Cisco APIC - Redigindo consultas da API REST - Aplicando filtros de escopo de consulta.

- Se precisar de atualizações de um objeto ou classe de objeto, considere assinaturas de websocket em vez de solicitações rápidas de API.

Acelerador de Solicitação NGINX

Disponível no 4.2(1)+, um usuário pode habilitar o acelerador de solicitações em HTTP e HTTPS independentemente.

Observação: a partir da versão 6.1(2) da ACI, a taxa máxima suportada para esse recurso foi reduzida para 40 solicitações por segundo (r/s) ou 2.400 solicitações por minuto (r/m) de 10.000 r/m.



Malha - Políticas de malha - Pastas de políticas - Pasta de acesso de gerenciamento - padrão

Quando habilitado:

- O NGINX é reiniciado para aplicar as alterações do arquivo de configuração
  - Uma nova região, httpsClientTagZone, é gravada na configuração nginx
- A taxa de aceleração pode ser definida em Solicitações por minuto (r/m) ou Solicitações por segundo (r/s).
- O Acelerador de Solicitação depende da Implementação de Limite de Taxa incluída no NGINX

- As Solicitações de API em relação ao /api/ URI usam a Taxa de Aceleração definida pelo usuário + intermitência= (Taxa de Aceleração x 2) + nodelay
  - Há um acelerador não configurável (zone aaaApiHttps) para /api/aaaLogin e /api/aaaRefresh que limita a taxa em 2r/s + burst=4 + nodelay
- O Acelerador de Solicitação é rastreado por cliente-endereço-IP
- As solicitações de API originadas do autoip (UI + CLI) do APIC ignoram o acelerador
- Qualquer endereço IP do cliente que cruze a taxa de aceleração definida pelo usuário + limite de intermitência recebe uma resposta 503 do APIC
- Esses 503s podem ser correlacionados nos registros de acesso
- error.log tem entradas que indicam quando a limitação foi ativada (zona httpsClientTagZone) e em quais hosts do cliente

<#root>

apic#

less /var/log/dme/log/error.log

...
2023/04/17 20:19:14 [error] ...

limiting requests

, excess: 40.292 by zone "

httpsClientTagZone

", client: h.o.s.t, ... request: "GET /api/class/...", host: "a.p.i.c"
2023/04/17 20:19:14 [error] ...

limiting requests

, excess: 40.292 by zone "

httpsClientTagZone

", client: h.o.s.t, ... request: "GET /api/node/...", host: "a.p.i.c"

Como regra geral, o Request Throttle serve apenas para proteger o servidor (APIC) de sintomas do tipo DDOS induzidos por Clientes agressivos de consulta. Entender e isolar o cliente agressivo de solicitação para soluções finais na lógica de aplicativo/script.

Recomendações

Essas recomendações foram projetadas para ajudar a reduzir a carga e o estresse operacional no APIC, particularmente em cenários em que nenhuma fonte é responsável por um grande volume de chamadas de API. Implementando essas práticas recomendadas, você pode minimizar o processamento, o registro e a geração de eventos desnecessários em sua malha, resultando em melhor estabilidade e desempenho do sistema. Essas sugestões são especialmente relevantes em ambientes em que comportamentos agregados em vez de incidentes isolados contribuem para a tensão do APIC.

Desativar registro de ACL

Certifique-se de que o registro da ACL esteja DESATIVADO durante as operações normais. Habilite-o somente durante janelas de manutenção agendadas para solução de problemas ou depuração. O registro contínuo pode gerar mensagens informativas excessivas, especialmente com quedas de tráfego de alto volume em vários switches, aumentando a carga de trabalho do APIC.

Para obter mais detalhes, consulte o Guia de configuração de segurança do Cisco APIC (link do guia 5.2.x):
https://www.cisco.com/c/en/us/td/docs/dcn/aci/apic/5x/security-configuration/cisco-apic-security-configuration-guide-release-52x/security-policies-52x.html

Limitar a conversão de syslog a eventos críticos

Configure o sistema de modo que somente as mensagens de syslog de severidade ALERT sejam convertidas em eventRecords. Evite converter o nível INFORMATION (que inclui ACL.logging) para evitar que eventos ruidosos sobrecarreguem o APIC:

1. Navegue até Fabric → Fabric Policies → Policies → Monitoring → Common Policy → Syslog Message Policies → Default.
2. Ajuste o filtro de instalação para definir a gravidade do syslog como alerta.

Códigos de Evento Não Essenciais de Squelch

Para reduzir o ruído, suprima os códigos de evento (squelch) que não são relevantes para suas necessidades de monitoramento.
Para silenciar o código de evento E4204939, use este comando em qualquer CLI do APIC:

```
bash
icurl -k -sX POST -d '<fabricInst><monCommonPol><eventSevAsnP code="E4204939" sev="squelched"/></monCom
```

Para verificar:

```
bash
icurl -k -sX GET 'https://localhost/api/node/class/eventSevAsnP.xml' | xmllint --format -
```

Como alternativa, verifique via interface do usuário:
Fabric > Fabric Policies > Policies > Monitoring > Common Policy > Event Severity Assignment Policy (Estrutura > Políticas de estrutura > Políticas de monitoramento > Política comum > Política de atribuição de severidade de evento)

Otimizar Atualizações de Assinatura ND

Para malhas gerenciadas por versões ND anteriores a 3.2.2m ou 4.1.1g, faça a atualização para uma dessas versões ou posterior para otimizar os intervalos de atualização da assinatura. As versões anteriores são atualizadas a cada 45 segundos por MO, o que, em escala, pode resultar em mais de 300.000 solicitações APIC por dia. As versões atualizadas aumentam o tempo limite da assinatura para 3600 segundos (1 hora), reduzindo as atualizações para cerca de 5.000 por dia.

Consultas Relacionadas à Interceptação do Monitor

As malhas ativadas pela Intersight geram consultas periódicas ao topsystem a partir do conector DC (a cada 15 segundos), adicionando à carga do APIC.
Na versão 6.1.2 e posterior, essa consulta foi otimizada para reduzir a sobrecarga.

Ajustar políticas de retenção para registros

Defina a política de retenção para eventRecord, faultRecord e healthRecord como 1.000 para evitar o acúmulo excessivo de registros. Isso é especialmente útil quando você extrai esses registros regularmente para qualquer atividade operacional específica. Sempre avalie o impacto da redução da granularidade do monitoramento em relação aos requisitos operacionais e de solução de problemas.

Sobre esta tradução
A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês (link fornecido) seja sempre consultado.