# TDC del documento de prueba

## Contenido



## Introducción

Este documento describe la metodología general para resolver problemas de una experiencia de GUI de APIC lenta.

## Inicio rápido

Se observa con frecuencia que los problemas de la GUI de APIC lenta son el resultado de una alta tasa de solicitudes de API originadas en un script, integración o aplicación. El archivo access.log de un APIC registra cada solicitud de API procesada. El access.log de un APIC se puede analizar rápidamente con el script Access Log Analyzer dentro del proyecto Github Datacenter group aci-tac-scripts.

## Antecedentes

### APIC como servidor web: NGINX

NGINX es el DME responsable de los terminales API disponibles en cada APIC. Si NGINX no funciona, no se pueden gestionar las solicitudes de API. Si NGINX está congestionado, la API también lo está. Cada APIC ejecuta su propio proceso NGINX, por lo que es posible que solo un

APIC pueda tener problemas de NGINX si solo ese APIC es el objetivo de cualquier consultor agresivo.

La interfaz de usuario de APIC realiza varias solicitudes de API para rellenar cada página. De manera similar, todos los comandos show de APIC (CLI de estilo NXOS) son contenedores para scripts de Python que realizan varias solicitudes de API, controlan la respuesta y, a continuación, se la suministran al usuario.

### Registros relevantes

| Nombre de archivo de registro | Ubicación | ¿En qué soporte técnico se encuentra? | Comentarios |
|---|---|---|---|
| access.log | /var/log/dme/log | APIC 3de3 | Independiente de ACI, ofrece 1 línea por solicitud de API |
| error.log | /var/log/dme/log | APIC 3de3 | Independiente de ACI, muestra errores nginx (limitación incluida) |
| nginx.bin.log | /var/log/dme/log | APIC 3de3 | específico de ACI, registra las transacciones de DME |
| nginx.bin.warnplus.log | /var/log/dme/log | APIC 3de3 | Las opciones específicas de ACI contienen registros con advertencia+ gravedad |

## Metodología

### Aislar desencadenador inicial

¿Qué se ve afectado?

- Qué APIC se ven afectados; ¿uno, varios o todos los APIC?
- ¿Dónde se ve la lentitud? mediante la interfaz de usuario, comandos CLI o ambos?
- ¿Qué páginas o comandos específicos de la interfaz de usuario son lentos?

¿Cómo se experimenta la lentitud?

- ¿Se ve esto en varios navegadores para un solo usuario?
- ¿Varios usuarios informan de lentitud o solo un único/subconjunto de usuarios?
- ¿Los usuarios afectados comparten una ubicación geográfica o ruta de red similar desde el

navegador al APIC?

¿Cuándo se notó por primera vez la lentitud?

- ¿Se ha agregado recientemente una secuencia de comandos o integración de ACI?
- ¿Se ha habilitado recientemente una extensión de explorador?
- ¿Ha habido algún cambio reciente en la configuración de ACI?

## Comprobar el uso y el estado de NGINX

Formato de entrada de Access.log

access.log es una función de NGINX y, por lo tanto, es independiente de APIC. Cada línea representa 1 solicitud HTTP que el APIC recibió. Consulte este registro para comprender el uso de NGINX de un APIC.

El formato predeterminado de access.log en ACI versión 5.2+:

```
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 línea representa una entrada access.log cuando se realiza un moquery -c fvTenant:

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

Mapa de la entrada access.log de ejemplo a log_format:

| campo log_format | Contenido del ejemplo | Comentarios |
|---|---|---|
| $remote_addr | 127.0.0.1 | IP del host que envió esta solicitud |
| $http_x_real_ip | - | IP del último solicitante si hay proxies en uso |
| $remote_user | - | Generalmente no se utiliza. Marque nginx.bin.log para realizar un seguimiento del usuario que ha iniciado sesión para realizar solicitudes |

| | | |
|---|---|---|
| $time_local | 07/Abr/2022:20:10:59 +0000 | Cuándo se procesó la solicitud |
| $request | GET /api/class/fvTenant.xml HTTP/1.1 | Método Http (GET, POST, DELETE) y URI |
| $status | 200 | Código de estado de respuesta HTTP |
| $body_bytes_sent | 1586 | tamaño de carga útil de respuesta |
| $http_referer | - | - |
| $http_user_agent | Python-urllib | Qué tipo de cliente envió la solicitud |

Comportamientos de Access.log

Ráfagas de solicitudes de alta velocidad durante un período de tiempo prolongado:

- Las ráfagas continuas de más de 40 solicitudes por segundo pueden causar lentitud en la interfaz de usuario
- Identificar qué host(s) son responsables de las consultas
- Reduzca o desactive el origen de las consultas para ver si esto mejora el tiempo de respuesta de APIC.

Respuestas 4xx o 5xx coherentes:

- Si se encuentra, identifique el mensaje de error de nginx.bin.log

El access.log de un APIC se puede analizar rápidamente con el script Access Log Analyzer dentro del proyecto Github Datacenter group aci-tac-scripts.

Comprobación del uso de recursos NGINX

El uso de memoria y CPU de NGINX se puede verificar con el comando top del 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
```

```
  2.8   759:05.78
nginx.bin
```

Un uso elevado de los recursos NGINX puede estar directamente relacionado con una alta tasa de solicitudes procesadas.

## Comprobar si hay núcleos

Una caída de NGINX no es típica para problemas de GUI de Slow APIC. Sin embargo, si se encuentran núcleos NGINX, adjúntelos a un TAC SR para su análisis. Consulte la guía de soporte técnico de ACI para conocer los pasos para comprobar los núcleos.

## Comprobar latencia cliente a servidor

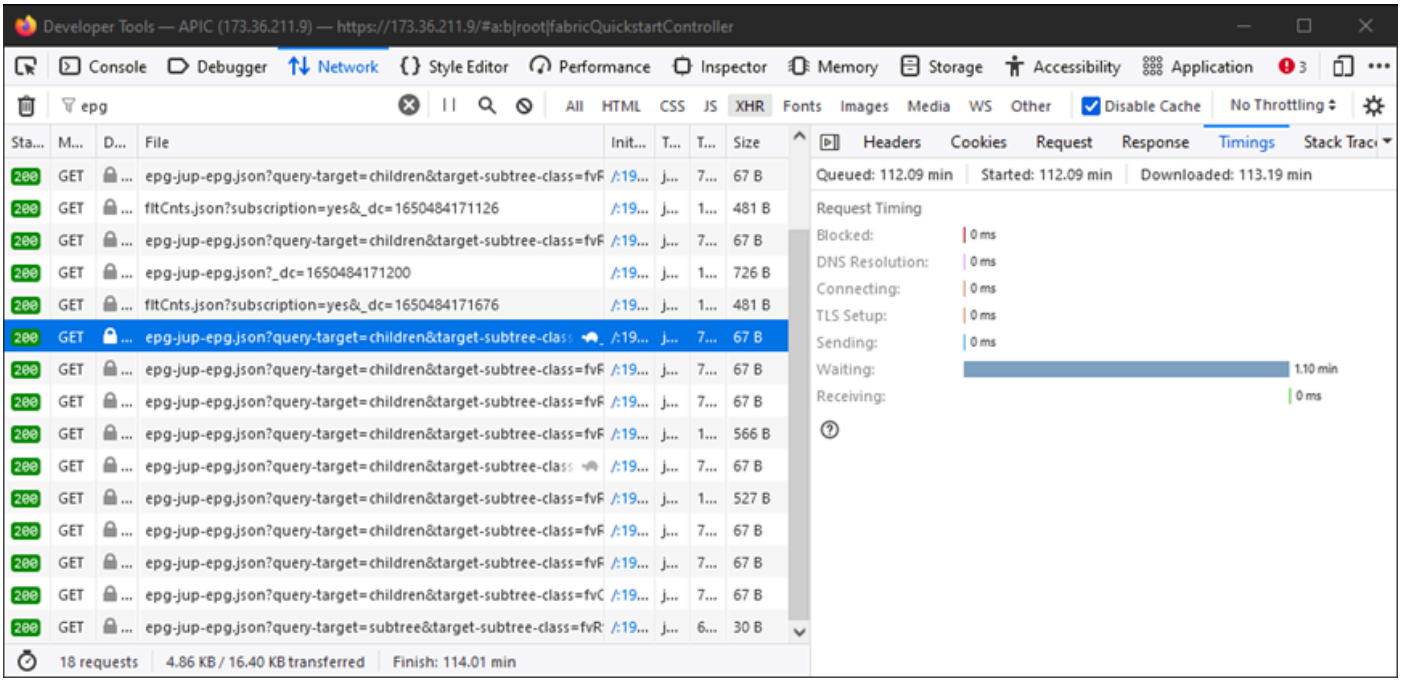
Si no se encuentran solicitudes rápidas pero un usuario sigue mostrando lentitud en la interfaz de usuario, el problema puede ser la latencia de cliente (navegador) a servidor (APIC).

En estos casos, valide la ruta de datos desde el navegador al APIC (distancia geográfica, VPN, etc.). Si es posible, implemente y pruebe el acceso desde un servidor de acceso directo ubicado en la misma región geográfica o Data Center que los APIC para aislar. Validar si otros usuarios presentan una latencia similar.

Ficha Red de herramientas de desarrollo del explorador

Todos los exploradores tienen la capacidad de validar solicitudes y respuestas HTTP mediante su kit de herramientas Desarrollo de explorador, normalmente en una ficha Red.

Esta herramienta se puede utilizar para validar la cantidad de tiempo que se tarda en cada etapa de las solicitudes originadas en el navegador, como se muestra en la imagen.



Ejemplo de un navegador que espera 1,1 minutos para que el APIC responda

Mejoras para Páginas Específicas de la IU

Página Grupo de Políticas:

La GUI de Cisco bug ID CSCvx14621 - APIC se carga lentamente en las políticas IPG en la pestaña Fabric.

Interfaz en la página Inventario:

ID de bug de Cisco CSCvx90048 - La carga inicial de la ficha operativa "Configuración de interfaz física de capa 1" es larga/induce la "congelación".

Recomendaciones generales para cliente > Latencia del servidor

Ciertos navegadores, como Firefox, permiten más conexiones web por host de forma predeterminada.

- Compruebe si este parámetro se puede configurar en la versión del explorador que se utilice
- Esto es más importante para las páginas de múltiples consultas, como la página Grupo de Políticas

La VPN y la distancia al APIC aumentan la lentitud general de la interfaz de usuario, dadas las solicitudes del navegador del cliente y el tiempo de viaje de respuesta del APIC. Un cuadro de salto geográficamente local a los APIC reduce significativamente el tiempo de viaje del navegador a APIC.

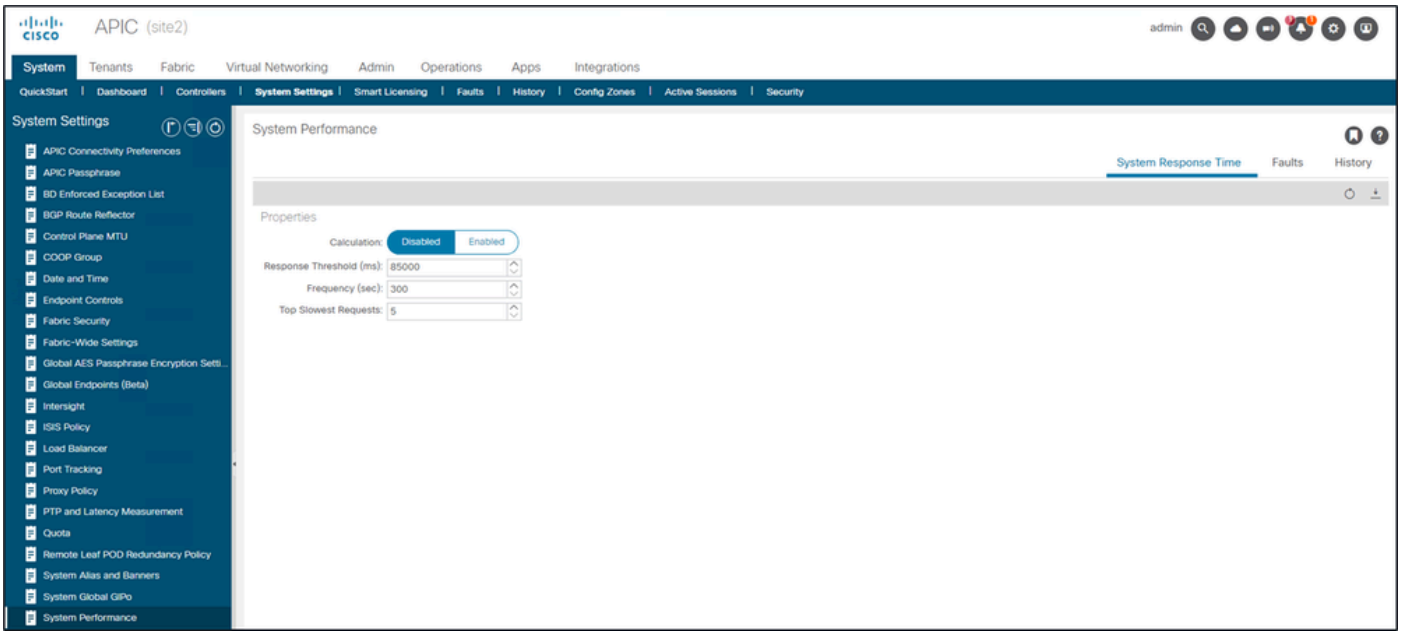
## Comprobar solicitudes Long-Web

Si un servidor web (NGINX en APIC) gestiona un gran volumen de solicitudes web largas, esto puede afectar al rendimiento de otras solicitudes recibidas en paralelo.

Esto es especialmente cierto en el caso de los sistemas que tienen bases de datos distribuidas, como los APIC. Una única solicitud de API puede requerir solicitudes y búsquedas adicionales enviadas a otros nodos del fabric, lo que puede dar lugar a tiempos de respuesta esperadamente más largos. Una ráfaga de estas solicitudes web largas en un período de tiempo reducido puede aumentar la cantidad de recursos necesarios y provocar tiempos de respuesta inesperadamente más largos. Además, las solicitudes recibidas pueden agotar el tiempo de espera (90 segundos), lo que se traduce en un comportamiento inesperado del sistema desde la perspectiva del usuario.

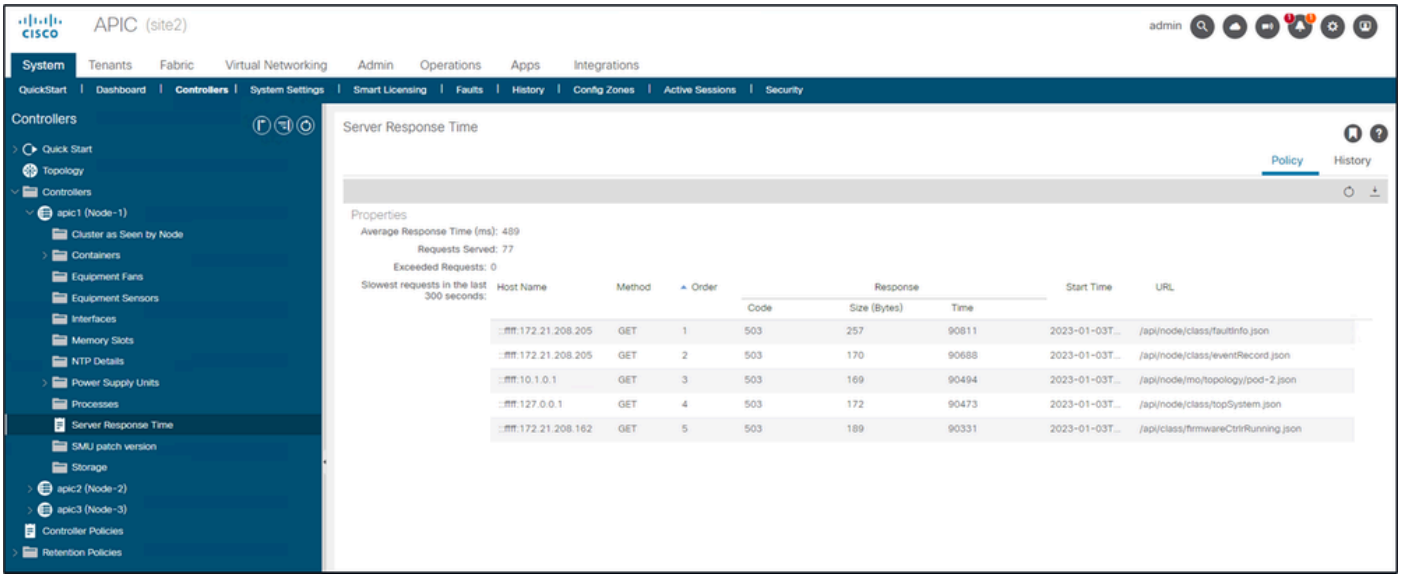
Tiempo de respuesta del sistema - Habilitar cálculo para el tiempo de respuesta del servidor

En 4.2(1)+, un usuario puede habilitar el "Cálculo del rendimiento del sistema", que realiza un seguimiento de las solicitudes de API y las resalta, que tardaron mucho tiempo en gestionarse.




El cálculo se puede habilitar desde Sistema - Configuración del sistema - Rendimiento del sistema

Una vez que se habilita "Cálculo", un usuario puede navegar a APIC específicos bajo Controladores para ver las solicitudes de API más lentas en los últimos 300 segundos.




Sistema - Controladores - Carpeta de controladores - APIC x - Tiempo de respuesta del servidor

# Consideraciones sobre el uso de API APIC

## Punteros generales para garantizar que un script no dañe a Nginx

- Cada APIC ejecuta su propio NGINX DME.
  - Solo el NGINX de APIC 1 procesa las solicitudes al APIC 1. El NGINX de APIC 2 y 3 no procesa esas solicitudes.
- En general, más de 40 solicitudes de API por segundo durante un largo período de tiempo debilita al NGINX.
  - Si lo encuentra, reduzca la agresividad de las solicitudes.

- Si el host de solicitudes no se puede modificar, considere Límites de velocidad NGINX en el APIC.

## Abordar las ineficiencias de scripts

- No inicie/cierre sesión antes de cada solicitud de API.
  - El tiempo de espera predeterminado para un inicio de sesión es de 10 minutos. Esta misma sesión se puede utilizar para varias solicitudes y se puede actualizar para ampliar el tiempo de validez.
  - Consulte Guía de configuración de la API REST de Cisco APIC: acceso a la API REST: autenticación y mantenimiento de una sesión de API.
- Si la secuencia de comandos consulta muchos DN que comparten un elemento primario, en lugar de contraer las consultas en una única consulta primaria lógica con filtros de consulta.
  - Consulte Guía de configuración de la API REST de Cisco APIC - Redacción de consultas de la API REST - Aplicación de filtros de ámbito de consulta.
- Si necesita actualizaciones de un objeto o clase de objeto, considere las suscripciones de websocket en lugar de las solicitudes rápidas de API.

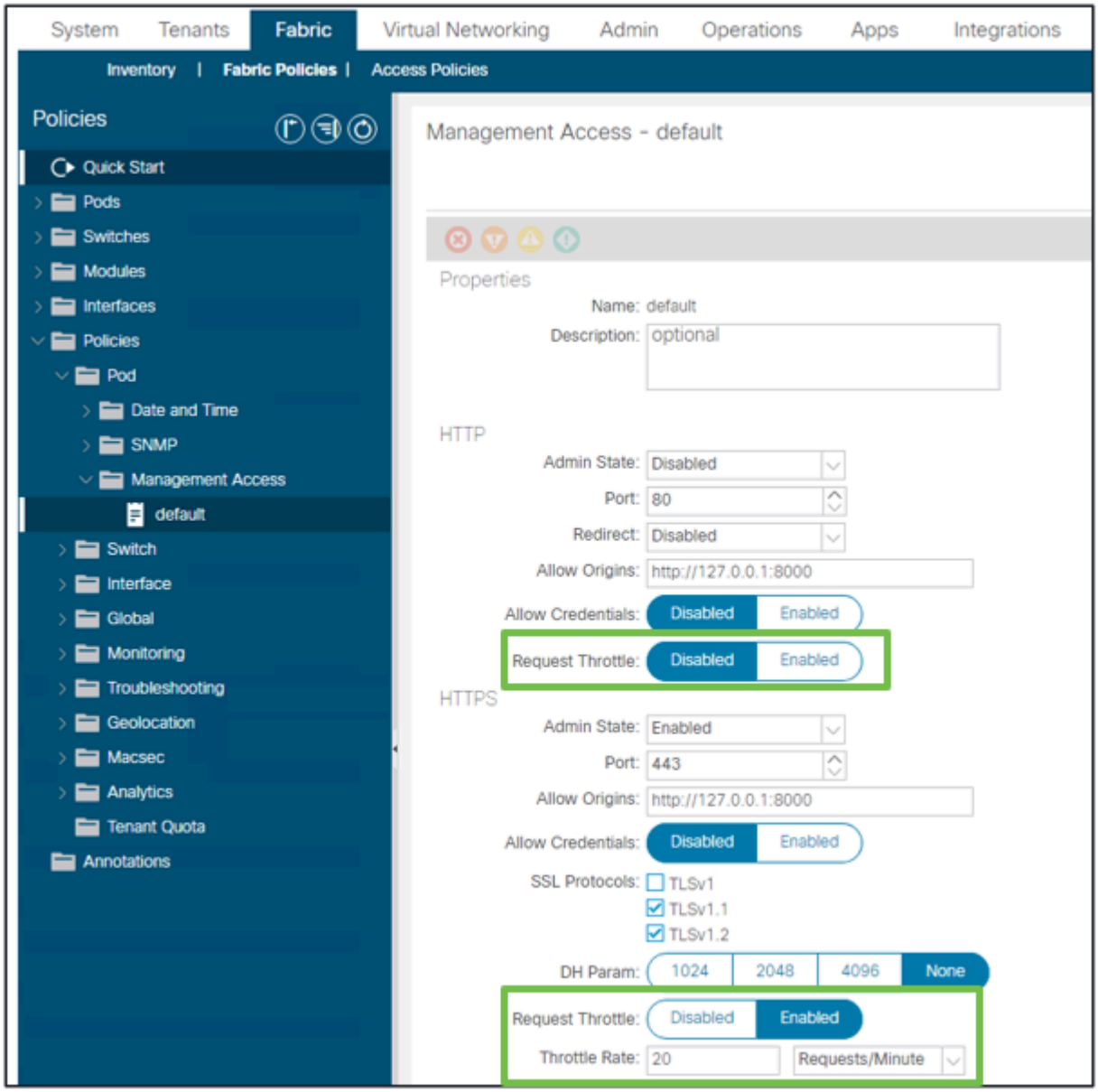
## Acelerador de solicitud NGINX

Disponible en 4.2(1)+, un usuario puede habilitar el acelerador de solicitudes contra HTTP y HTTPS de forma independiente.

---

Nota: a partir de la versión 6.1(2) de ACI, la velocidad máxima admitida para esta función se redujo a 40 solicitudes por segundo (r/s) o a 2400 solicitudes por minuto (r/m) desde 10 000 r/m.

---




Fabric - Políticas de fabric - Carpeta de políticas - Carpeta de acceso a la gestión - predeterminada

Cuando está habilitado:

- NGINX se reinicia para aplicar los cambios del archivo de configuración
  - Una nueva zona, httpsClientTagZone, se escribe en nginx config
- La velocidad del acelerador se puede establecer en Solicitudes por minuto (r/m) o Solicitudes por segundo (r/s).
- El acelerador de solicitudes se basa en la [implementación de límite de velocidad incluida en NGINX]
  - Las solicitudes de API contra /api/ URI utilizan la velocidad del acelerador definida por el usuario + burst= (velocidad del acelerador x 2) + nodelay
    - Hay un acelerador no configurable (zone aaaApiHttps) para /api/aaaLogin y /api/aaaRefresh que limita la velocidad a 2r/s + ráfaga=4 + nodelay
  - El seguimiento del acelerador de solicitudes se realiza por dirección IP de cliente
  - Las solicitudes de API que se originan en el APIC de IP automática (UI + CLI) omiten el acelerador
  - Cualquier dirección IP de cliente que cruce el umbral de velocidad + ráfaga definido

por el usuario recibe una respuesta 503 del APIC
- Estos 503 se pueden correlacionar dentro de los registros de acceso
- error.log tiene entradas que indican cuándo se ha activado la regulación (zona httpsClientTagZone) y contra qué hosts del 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 regla general, el acelerador de solicitudes solo sirve para proteger el servidor (APIC) de los síntomas similares a DDOS inducidos por clientes que realizan consultas agresivas. Comprender y aislar el cliente agresivo en la solicitud para las soluciones finales en la lógica de aplicación/script.

# Recomendaciones

Estas recomendaciones están diseñadas para ayudar a reducir la carga y el estrés operativo en el APIC, especialmente en situaciones en las que ningún origen es responsable de un gran volumen de llamadas de API. Al implementar estas prácticas recomendadas, puede minimizar el procesamiento, el registro y la generación de eventos innecesarios en el fabric, lo que se traduce en una mejora del rendimiento y la estabilidad del sistema. Estas sugerencias son especialmente relevantes en entornos en los que los comportamientos agregados en lugar de los incidentes aislados contribuyen a la tensión del APIC.

## Desactivar registro de ACL

Asegúrese de que el registro de ACL esté apagado durante las operaciones normales. Actívela sólo durante las ventanas de mantenimiento programado para la solución de problemas o la depuración. El registro continuo puede generar demasiados mensajes de información,

especialmente con caídas de tráfico de gran volumen en varios switches, lo que aumenta la carga de trabajo de APIC.

Para obtener más información, consulte la Guía de configuración de seguridad de Cisco APIC (enlace de la guía 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 la conversión de Syslog a eventos críticos

Configure el sistema para que sólo los mensajes de syslog de gravedad ALERT se conviertan en eventRecords. Evite convertir el nivel de INFORMACIÓN (que incluye ACL.logging) para evitar que los eventos ruidosos saturen el APIC:

1. Vaya a Fabric → Fabric Policies → Policies → Monitoring → Common Policy → Syslog Message Policies → Default.
2. Ajuste el filtro de recursos para establecer la gravedad de syslog en alerta.

## Códigos de eventos no esenciales de Squelch

Suprima (silencie) los códigos de evento que no sean relevantes para sus necesidades de monitorización para reducir el ruido.
Para silenciar el código de evento E4204939, utilice este comando en cualquier APIC CLI:

```
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 -
```

Alternativamente, verifíquelo a través de la interfaz:
Fabric > Políticas de fabric > Políticas > Supervisión > Política común > Política de asignación de gravedad de evento

## Optimización de actualizaciones de suscripción ND

En el caso de los fabrics administrados por versiones ND anteriores a 3.2.2m o 4.1.1g, actualice a una de estas versiones o posterior para optimizar los intervalos de actualización de suscripciones. Las versiones anteriores se actualizan cada 45 segundos por MO, lo que, a escala, puede dar lugar a más de 300 000 solicitudes APIC al día. Las versiones actualizadas aumentan el tiempo de espera de la suscripción a 3600 segundos (1 hora), lo que reduce las actualizaciones a aproximadamente 5000 al día.

## Supervisar consultas relacionadas con la información

Los fabrics habilitados para intersight generan consultas periódicas del sistema topdesde el conector DC (cada 15 segundos), lo que se suma a la carga de APIC.
En la versión 6.1.2 y posteriores, esta consulta se ha optimizado para reducir la sobrecarga.

## Ajustar políticas de retención para registros

Establezca la directiva de retención para eventRecord, faultRecord y healthRecord en 1.000 para evitar una acumulación excesiva de registros. Esto es especialmente útil cuando extrae estos registros de forma regular para cualquier actividad operativa específica. Evalúe siempre el impacto de la reducción de la granularidad de la supervisión en relación con sus requisitos operativos y de resolución de problemas.

Acerca de esta traducción
Cisco ha traducido este documento combinando la traducción automática y los recursos humanos a fin de ofrecer a nuestros usuarios en todo el mundo contenido en su propio idioma.

Tenga en cuenta que incluso la mejor traducción automática podría no ser tan precisa como la proporcionada por un traductor profesional.

Cisco Systems, Inc. no asume ninguna responsabilidad por la precisión de estas traducciones y recomienda remitirse siempre al documento original escrito en inglés (insertar vínculo URL).