

Präventive Begrenzung statischer und dynamischer Übertragungsraten mit CPS vDRA

Inhalt

[Einleitung](#)

[Voraussetzungen](#)

[Anforderungen](#)

[Verwendete Komponenten](#)

[Hintergrundinformationen](#)

[Problem](#)

[Lösung](#)

[Statische Durchsatzbegrenzung am Load Balancer](#)

[Grenzwert für Eingangs-Übertragungsrate](#)

[Grenzwert für Ausgangs-Übertragungsrate](#)

[Dynamischer Ratengrenzwert](#)

Einleitung

In diesem Dokument werden die Optionen für die Durchsatzbegrenzung in DRA beschrieben, einer Telekommunikationskomponente, die Nachrichten mit dem Durchmesser weiterleitet und den Netzwerkverkehr verwaltet.

Voraussetzungen

Anforderungen

Cisco empfiehlt, dass Sie über Kenntnisse in folgenden Bereichen verfügen:

- Cisco Policy Suite (CPS) Diameter Routing Agent (vDRA)
- Diameter Routing Agent - Grundlagen und Spezifikationen

Verwendete Komponenten

Die Informationen in diesem Dokument basieren auf Cisco Policy Suite DRA.

Die Informationen in diesem Dokument beziehen sich auf Geräte in einer speziell eingerichteten Testumgebung. Alle Geräte, die in diesem Dokument benutzt wurden, begannen mit einer gelöschten (Nichterfüllungs) Konfiguration. Wenn Ihr Netzwerk in Betrieb ist, stellen Sie sicher, dass Sie die möglichen Auswirkungen aller Befehle kennen.

Hintergrundinformationen

DRA ist ein Bestandteil von Telekommunikationsnetzen, insbesondere im Rahmen von auf dem Diameter-Protokoll basierenden Netzen. Die DRA leitet Diameter-Nachrichten effizient zwischen verschiedenen Netzwerkelementen weiter, z. B. Richtlinienserver, Abrechnungssysteme und andere Geräte mit aktiviertem Durchmesser. Bei der Ratenbegrenzung handelt es sich um eine Technik zur Verwaltung des Netzwerkverkehrs, mit der die Menge des von einem Netzwerkelement ein- oder ausgehenden Datenverkehrs gesteuert wird. Sie trägt dazu bei, dass Netzwerkressourcen nicht erschöpft sind, gewährleistet die Quality of Service und verhindert Missbrauch oder Missbrauch des Netzwerks.

Problem

Jede Komponente des Netzwerks kann die Datenverkehrslast auf Basis der Nennkapazität verarbeiten, aber in Echtzeit kann es Szenarien geben, bei denen der generierte Datenverkehr mehr ist, als das System verarbeiten kann. Einige davon sind:

- Benutzerverhalten - Aktivitäten wie das Streaming von Ereignissen oder Software-Updates, die innerhalb kurzer Zeit große Datenmengen generieren. Wird in der Regel vom Gateway (Gw) an DRA gesendet.
- Netzwerküberlastung - In Zeiten hoher Netzwerkauslastung kann sich eine Überlastung aufbauen, die zu Daten in der Warteschlange führt, die dann in Spitzenzeiten gesendet werden, wenn die Kapazität verfügbar ist.
- Mechanismen zur Ausfallsicherheit des Netzwerks - Datenverkehr wird bei Ausfällen oder Wartungsarbeiten umgeleitet, was zu vorübergehenden Spitzen führt. Dies kann sich auf den Datenverkehrsfluss an verbundenen Standorten auswirken, bei denen keine Netzwerkprobleme auftreten.
- Verhalten von Netzwerkelementen - Bei Überlastung und Überlastung können Sie anfangen, keine Antworten/Zeitüberschreitungen von einem oder mehreren Netzwerkelementen zu sehen, die eine erneute Verbindung verursachen können, die zu einer weiteren Überlastung des Systems beiträgt.
- Gateway-Leerung - Gateway kann vorhandene Sitzungen aufgrund von Richtlinienänderungen, Topologieänderungen oder Wartungs- oder Fehlerbehebungsaktivitäten leeren. Während dieser Szenarien werden die Sitzungen gelöscht, und Sie können eine Häufung von CCR-T-Anforderungen (Global Credit Control Request) erhalten.

Lösung

DRA kann die Last auf mehrere Diameter-Server verteilen, um eine effiziente Verarbeitung der Anfragen zu gewährleisten und eine Überlastung eines einzelnen Servers zu vermeiden. Bei einem Serverausfall kann das DRA Nachrichten an alternative Server umleiten und so eine hohe Verfügbarkeit und Zuverlässigkeit der Netzwerkdienste gewährleisten.

Die Ratenbegrenzung auf der DRA schützt nicht nur die DRA, sondern auch andere Einheiten, indem sie einen kontrollierten Nachrichtenfluss sicherstellt. Die wichtigsten Vorteile der Ratenbegrenzung sind:

- Servicekontinuität - Kontinuierliche Serviceverfügbarkeit durch die Vermeidung von Überlastung kritischer Netzwerkkomponenten und Vermeidung von Ausfällen.
- Skalierbarkeit - Ermöglicht dem Netzwerk die Verarbeitung unterschiedlicher Lasten ohne Leistungseinbußen.
- Einhaltung von Service Level Agreements (SLAs) - Sicherstellen, dass das Netzwerk seine SLAs erfüllt, indem es ein konsistentes Leistungs- und Zuverlässigkeitsniveau aufrechterhält.

Statische Durchsatzbegrenzung am Load Balancer

Hierbei handelt es sich um einen unkomplizierten Ansatz, bei dem ein fester Schwellenwert auf der Grundlage der Nennkapazität von DRA/Packet Gateway (pGW) und anderen Netzwerkelementen festgelegt wird und sich nicht aufgrund der Netzwerkbedingungen oder Systemressourcen ändert. Durch die Begrenzung der eingehenden Anfragen erhalten Sie ein vorhersehbares Ergebnis für den von DRA verarbeiteten Datenverkehr.

Die Konfigurationen für die statische Ratenbegrenzung hängen vom jeweiligen Anwendungsfall ab.

Grenzwert für Eingangs-Übertragungsrate

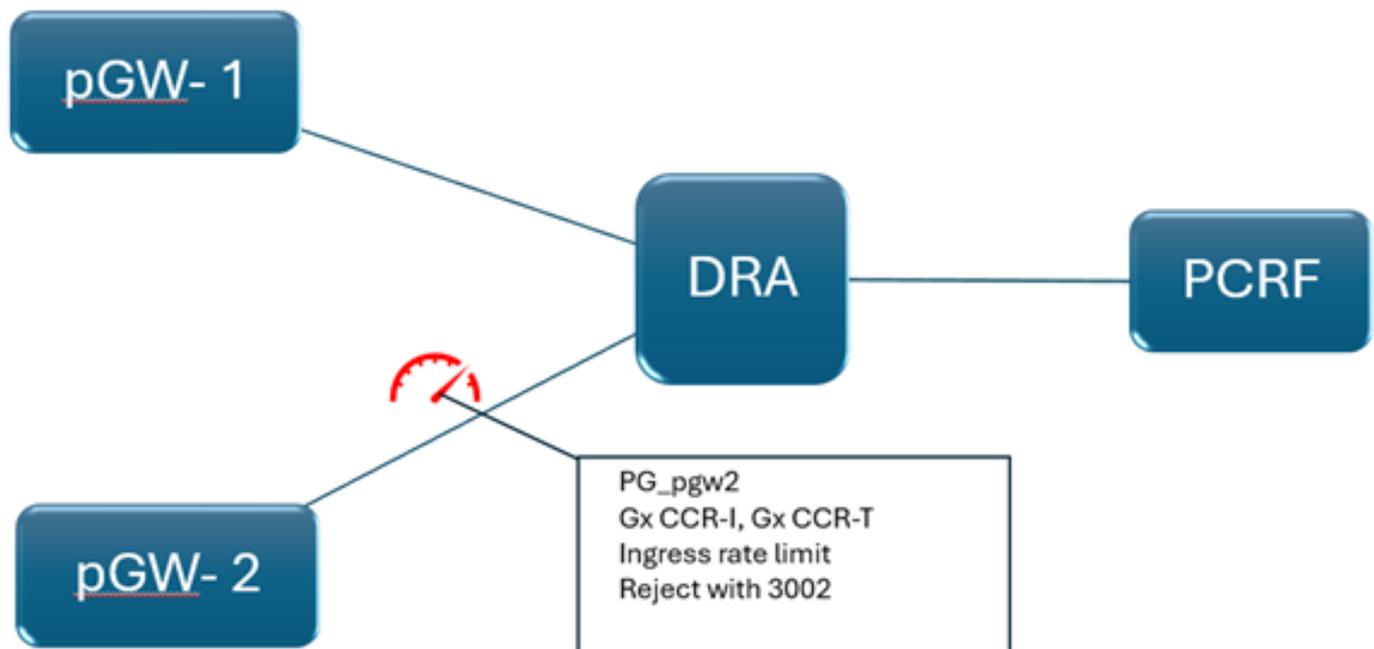
Szenario: Bursts vom pGW

Es werden Grenzwerte für pGWs konfiguriert, die für diese Datenverkehrsspitzen anfällig sind. Der Wert muss auf Basis der regulären Verkehrszahlen bzw. Verkehrsspitzen ermittelt werden, die während dieser Spitzen zu sehen sind.

Die Schwellenwerte können für jeden Nachrichtentyp definiert werden, um sicherzustellen, dass nur der Burst-Datenverkehr gedrosselt wird. Beispielsweise müssen nur Gx CCR-I- und Gx CCR-T-Anforderungen von einem GW gedrosselt werden, aber Gx CCR-U-Datenverkehr oder Gy-Datenverkehr wird wie empfangen weitergeleitet.

In diesem Fall können Sie einseitig drosseln, d. h., DRA drosselt die Nachricht, sobald sie eingeht, da es hier darum geht, sie aufgrund des Netzwerkelements, von dem sie die Anforderung empfängt, abzulehnen und eine höhere Anzahl von Anforderungen zu vermeiden, als die DRA verarbeiten kann.

Das Drosselungsverhalten kann darin bestehen, die Nachricht mit einem bestimmten Fehlercode und einer Fehlermeldung entweder abzulehnen oder zu löschen.



Dieses Verhalten kann in CPS vDRA durch Konfigurieren der benutzerdefinierten Referenzdaten (CRD)-Tabellen "Peer Rate Limit Profile" und "Message Rate Limit Profile" aktiviert werden. In diesen CRD-Tabellen müssen Sie die folgenden Werte konfigurieren:

Peer-Gruppe	Eine Peer-Gruppe ist eine logische Gruppierung von Diameter-Knoten, basierend auf ihrem Bereich und Host. Sie müssen die Peer-Gruppe konfigurieren, die gedrosselt werden muss.
FQDN (Peer Fully Qualified Domain Name)	FQDN (exakte Übereinstimmung oder Übereinstimmung mit regulären Ausdrücken) für die Peers in der Peer-Gruppe, für die Sie eine Durchsatzbegrenzung benötigen.
Nachrichtenrichtung	Richtung der Drosselung - Eingang oder Ausgang. In diesem Fall - Ingress.
Ratenlimit-Profil	Der Name des Profils für die Beschränkung der Nachrichtenrate, mit dem der zu drosselnde Nachrichtentyp definiert wurde.
Peer-Rate-Limit	Die Rate der Anforderungen, die für diese Peer-Gruppe zulässig sind. Dies schließt alle Nachrichtentypen dieser Peer Group ein.
Verhalten verwerfen	Sie können die Anforderung löschen oder mit einem

	Fehlercode ablehnen.
Ergebniscode	Ergebniscodewert, falls Sie die Nachrichten ablehnen. Nicht zutreffend, wenn die Nachrichten verworfen werden.
Fehlerzeichenfolge	Die Fehlerzeichenfolge, die in der Antwortnachricht der Anforderung verwendet wird und abgelehnt wird. Nicht zutreffend, wenn die Nachricht verworfen wird.
Anwendungs-ID	Anwendungs-ID der zu drosselnden Nachricht.
Befehlscode	Befehlscode der zu drosselnden Nachricht.
Nachrichten-/Anforderungstyp	Anwendungs-ID und Anforderungstyp der Anforderungen, die gedrosselt werden müssen
Übertragungsratenlimit für Nachrichten	TelePresence Server (TPS) der Anforderung dieses Nachrichtentyps, die von DRA verarbeitet wird. Anforderungen, die über diesen TPS hinausgehen, werden gedrosselt. Dieser Wert gilt pro Peer in der Peer-Gruppe.

Hier ist ein Beispiel, bei dem Sie eine allgemeine Durchsatzgrenze von 1000 Nachrichten für pGW2 und eine Durchsatzgrenze von 200 Gx CCR-I und 200 Gx CCR-T konfigurieren. Jede Anforderung, die über diese Rate hinausgeht, wird mit einem Wert von 3002 abgelehnt. Außerdem wird eine Fehlermeldung angezeigt, die besagt, dass die Durchsatzgrenze überschritten wird.

Peer Rate Limit Profile ✖

Filter by All Visible Columns ▾

CCR_I_T_Limit 🔍 🔒

Peer Group *	Peer FQDN *	Message Direction *	Rate Limit Profile	Peer Rate Limit	Discard Behavior *	Result Code	Error String	Actions
PG_pGW2	match=peer- ...	Ingress	CCR_I_T_Limit	1000	Send Error Answer	3002	DRA rate limit breached	✎ 🗑️

Showing 1 out of 1

Show 50 rows ⏪ < 1 > ⏩ out of 1

Message Rate Limit Profile



Filter by All Visible Columns

Rate Limit Profile Name *	Application Identifier *	Command Code *	Message/Request Type *	Message Rate Limit *	Actions
CCR_I_T_Limit	16777238	272	1	200	
CCR_I_T_Limit	16777238	272	3	200	

Showing 2 out of 2

Show 50 rows 1 out of 1

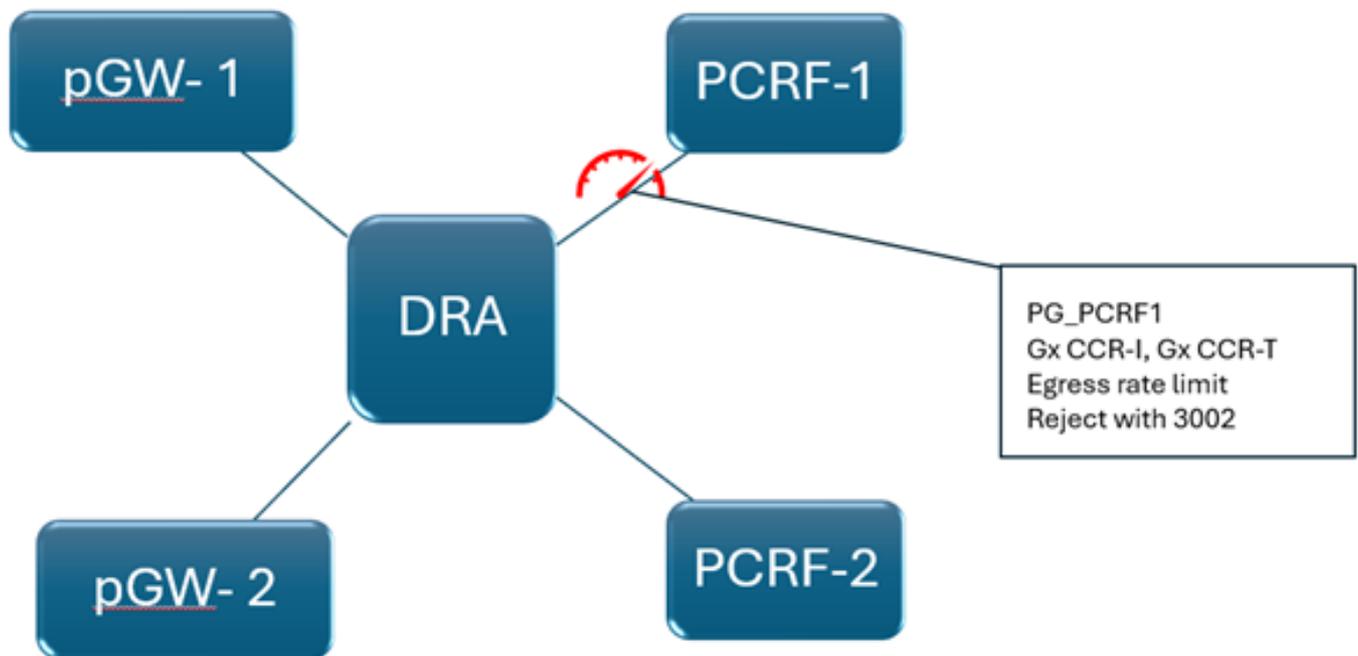
Grenzwert für Ausgangs-Übertragungsrate

Szenario: Schutz von Netzwerkelementen, die Anfragen empfangen

Nehmen wir als Beispiel eine Gx-Transaktion, bei der die Anforderung vom pGW empfangen und an die Policy and Charging Rules Function (PCRF) weitergeleitet wird. Wenn die Datenmenge, die PCRF verarbeiten kann, begrenzt ist, können Sie DRA verwenden, um die Nachricht bei DRA zu drosseln, anstatt die Anfrage an PCRF weiterzuleiten und zu überladen, obwohl DRA den eingehenden Datenverkehr verarbeiten kann.

Hier muss die Drosselung auf der Ausgangsseite erfolgen, d. h., DRA drosselt die Nachricht direkt bevor sie an die PCRF weitergeleitet wird, basierend auf der PCRF-Peer-Gruppe, die basierend auf der DRA-Routing-Logik identifiziert wird.

Das Drosselungsverhalten kann darin bestehen, die Nachricht mit einem bestimmten Fehlercode und einer Fehlermeldung entweder abzulehnen oder zu löschen.



Dieses Verhalten kann in CPS vDRA durch Konfigurieren der CRD-Tabellen 'Peer Rate Limit Profile' und 'Message Rate Limit Profile' aktiviert werden. In diesen CRD-Tabellen müssen Sie die folgenden Werte konfigurieren:

Peer-Gruppe	Eine Peer-Gruppe ist eine logische Gruppierung von Diameter-Knoten, basierend auf ihrem Bereich und Host. Sie müssen die Peer-Gruppe konfigurieren, die gedrosselt werden muss.
Peer-FQDN	FQDN (exakte Übereinstimmung oder Übereinstimmung mit regulären Ausdrücken) für die Peers in der Peer-Gruppe, für die Sie eine Durchsatzbegrenzung benötigen.
Nachrichtenrichtung	Richtung der Drosselung - Eingang oder Ausgang. In diesem Fall - Egress.
Ratenlimit-Profil	Der Name des Profils für die Beschränkung der Nachrichtenrate, mit dem der zu drosselnde Nachrichtentyp definiert wurde.
Peer-Rate-Limit	Die Rate der Anforderungen, die für diese Peer-Gruppe zulässig ist. Dies schließt alle Nachrichtentypen dieser Peer Group ein.

Verhalten verwerfen	Sie können die Anforderung löschen oder mit einem Fehlercode ablehnen.
Ergebniscode	Ergebniscodewert, falls Sie die Nachrichten ablehnen. Nicht zutreffend, wenn die Nachrichten verworfen werden.
Fehlerzeichenfolge	Die Fehlerzeichenfolge, die in der Antwortnachricht der Anforderung verwendet wird und abgelehnt wird. Nicht zutreffend, wenn die Nachricht verworfen wird.
Anwendungs-ID	Anwendungs-ID der zu drosselnden Nachricht.
Befehlscode	Befehlscode der zu drosselnden Nachricht.
Nachrichten- /Anforderungstyp	Anwendungs-ID und Anforderungstyp der Anforderungen, die gedrosselt werden müssen
Übertragungsratenlimit für Nachrichten	TPS der Anforderung dieses Nachrichtentyps, die von DRA verarbeitet wird. Anforderungen, die über diesen TPS hinausgehen, werden gedrosselt. Dieser Wert gilt pro Peer in der Peer-Gruppe.

Peer Rate Limit Profile 🗨️ ✕

Filter by All Visible Columns ▾

CCR_I_T 🔍 🗑️

Peer Group *	Peer FQDN *	Message Direction *	Rate Limit Profile	Peer Rate Limit	Discard Behavior *	Result Code	Error String	Actions
PG_PCRF1	match=peer- XXXXXXXXXX	Egress	CCR_I_T_Limit	1000	Send Error Answer	3002	DRA rate limit breached	✎ 🗑️

Showing 1 out of 1

Show 50 rows ⏪ < 1 > ⏩ out of 1

Message Rate Limit Profile



Filter by All Visible Columns

CCR_I_T_Limit



Rate Limit Profile Name *	Application Identifier *	Command Code *	Message/Request Type *	Message Rate Limit *	Actions
CCR_I_T_Limit	16777238	272	1	200	
CCR_I_T_Limit	16777238	272	3	200	

Showing 2 out of 2

Show 50 rows 1 out of 1

Szenario: Langsame Netzwerkverbindungen, die zu Datenverkehrsengepässen führen und DRA zu einem vollständigen oder teilweisen Ausfall führen

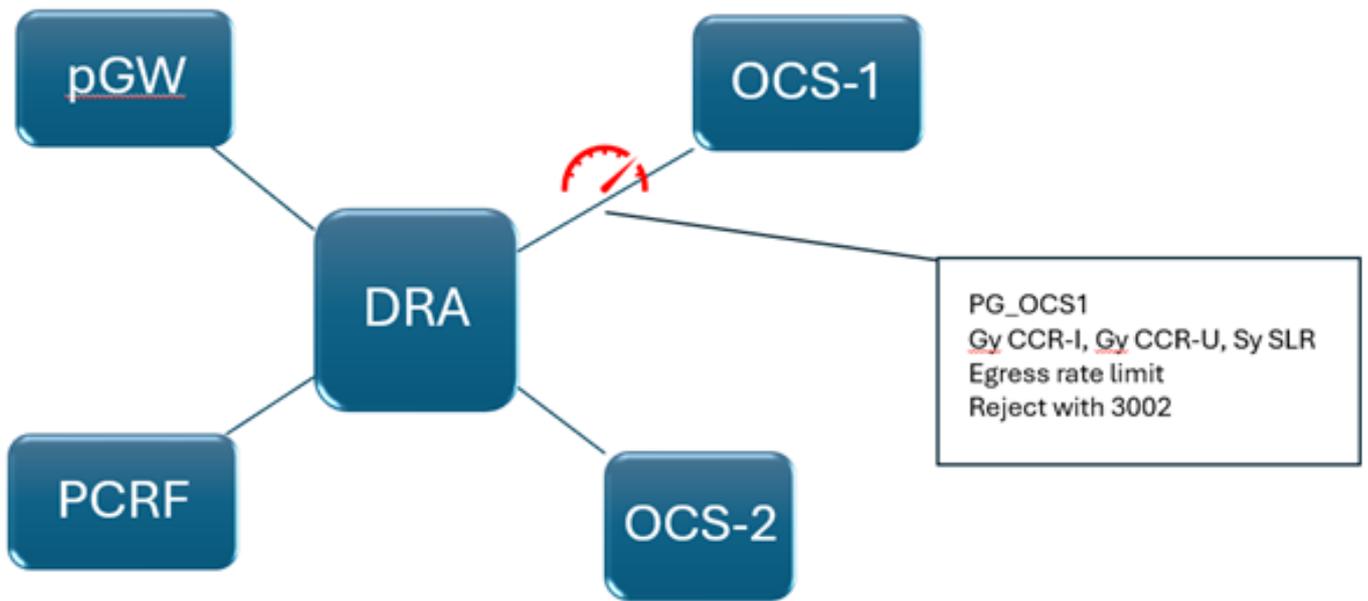
Nehmen wir als Beispiel eine Gy-Transaktion, die zwischen dem pGW und dem Online Charging System (OCS) ausgetauscht wird. Bei einer Netzwerkverlangsamung auf dem DRA-OCS-Kanal (aufgrund eines hohen Datenverkehrs vom pGW oder aufgrund eines anderen Netzwerkproblems) wird die Anforderung aufgrund einer SLA-Verletzung abgelaufen. Diese Zeitüberschreitungen wirken sich nicht nur auf DRA, sondern auf das gesamte Netzwerk aus.

Die DRA-Ressourcen werden bei dem Versuch, die Anfrage über das langsame Netzwerk an OCS zu senden, aufgehalten, was dazu führt, dass die Ressourcen erschöpft werden. Dies führt dazu, dass mehrere Anfragen von DRA abgelehnt werden, obwohl die Nennkapazität von DRA nicht überschritten wird.

Dies wirkt sich auch auf Datenverkehr aus, der sich nicht auf dem DRA-OCS-Kanal befindet. Diese Ablehnungen/Zeitüberschreitungen und Unterbrechungen lösen eine erneute Verbindung bei mehreren Netzwerkelementen aus.

In diesem Fall müssen Sie den Ausgang drosseln. DRA drosselt die Nachricht, bevor sie an den OCS weitergeleitet wird, basierend auf der OCS-Peer-Gruppe, die Kapazitätsbeschränkungen oder Netzwerkprobleme aufweist.

Das Drosselungsverhalten kann darin bestehen, die Nachricht mit einem bestimmten Fehlercode und einer Fehlermeldung entweder abzulehnen oder zu löschen.



Dieses Verhalten kann in CPS vDRA durch Konfigurieren der CRD-Tabellen 'Peer Rate Limit Profile' und 'Message Rate Limit Profile' aktiviert werden. In diesen CRD-Tabellen müssen Sie die folgenden Werte konfigurieren:

Peer-Gruppe	Eine Peer-Gruppe ist eine logische Gruppierung von Diameter-Knoten, basierend auf ihrem Bereich und Host. Sie müssen die Peer-Gruppe konfigurieren, die gedrosselt werden muss.
Peer-FQDN	FQDN (exakte Übereinstimmung oder Übereinstimmung mit regulären Ausdrücken) für die Peers in der Peer-Gruppe, für die Sie eine Durchsatzbegrenzung benötigen.
Nachrichtenrichtung	Richtung der Drosselung - Eingang oder Ausgang. In diesem Fall - Egress.
Profil für Ratenlimit	Der Name des Profils für die Beschränkung der Nachrichtenrate, mit dem der zu drosselnde Nachrichtentyp definiert wurde.
Peer-Rate-Limit	Die für diese Peer-Gruppe zulässige Rate von Anforderungen. Dies schließt alle Nachrichtentypen dieser Peer Group ein.

Verhalten verwerfen	Sie können die Anforderung löschen oder mit einem Fehlercode ablehnen.
Ergebniscode	Ergebniscodewert, falls Sie die Nachrichten ablehnen. Nicht zutreffend, wenn die Nachrichten verworfen werden.
Fehlerzeichenfolge	Die Fehlerzeichenfolge, die in der Antwortnachricht der Anforderung verwendet wird und abgelehnt wird. Nicht zutreffend, wenn die Nachricht verworfen wird.
Anwendungs-ID	Anwendungs-ID der zu drosselnden Nachricht.
Befehlscode	Befehlscode der zu drosselnden Nachricht.
Nachrichten- /Anforderungstyp	Anwendungs-ID und Anforderungstyp der Anforderungen, die gedrosselt werden müssen
Übertragungsratenlimit für Nachrichten	TPS der Anforderung dieses Nachrichtentyps, die von DRA verarbeitet wird. Anforderungen, die über diesen TPS hinausgehen, werden gedrosselt. Dieser Wert gilt pro Peer in der Peer-Gruppe.

Peer Rate Limit Profile 🗖️ ✕

Filter by All Visible Columns ▾

gy_sy 🔍 🔒

Peer Group *	Peer FQDN *	Message Direction *	Rate Limit Profile	Peer Rate Limit	Discard Behavior *	Result Code	Error String	Actions
PG_OCS_1	match=peer- XXXXXXXXXX	Egress	Gy_Sy_Limit	1000	Send Error Answer	3002	DRA rate limit breached	✎ 🗑️

Showing 1 out of 1

Message Rate Limit Profile



Filter by

Rate Limit Profile Name *	Application Identifier *	Command Code *	Message/Request Type *	Message Rate Limit *	Actions
Gy_Sy_Limit	16777302	8388635	1	300	
Gy_Sy_Limit	4	272	1	500	

Showing 2 out of 2

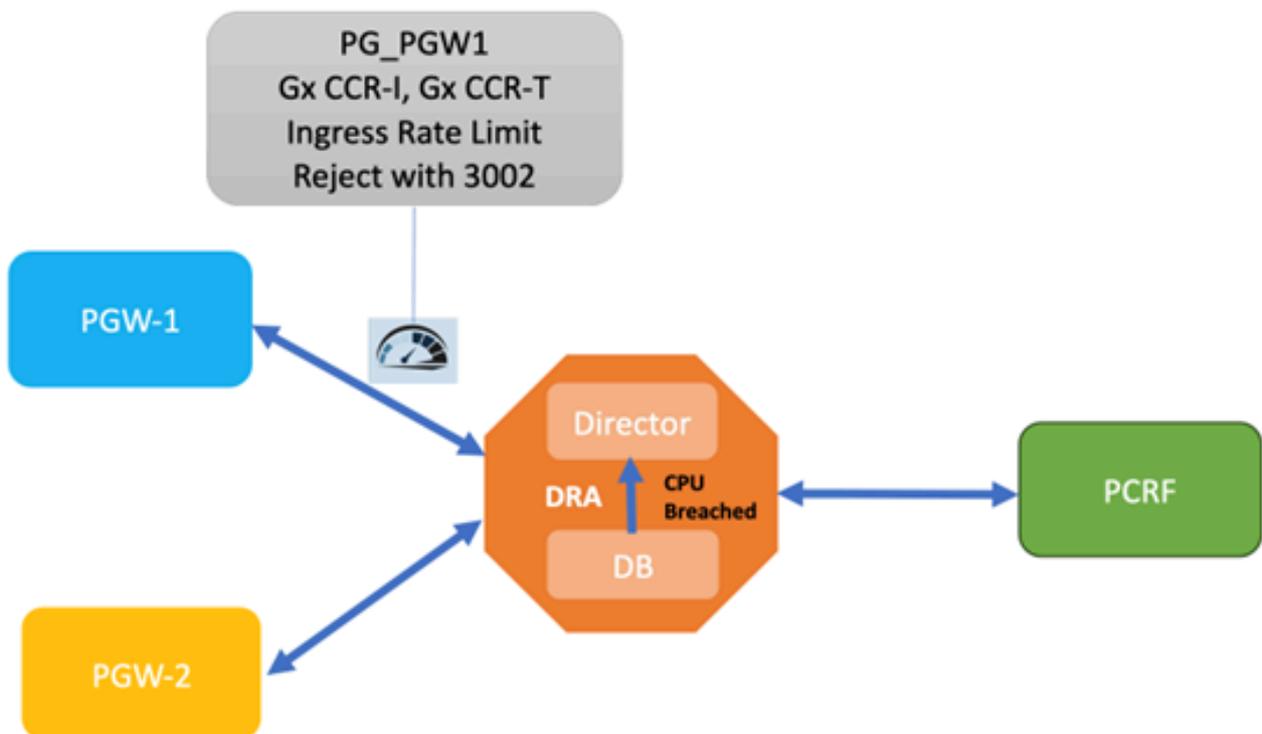
Show rows out of 1

Dynamischer Ratengrenzwert

Wenn CCR-I- oder CCR-T-Bursts auftreten, kann es zu einer Überlastung der Datenbank (DB) kommen, die das System destabilisieren kann. Um dieses Problem zu beheben, unterstützt DRA eine dynamische Ratenbegrenzung (nur für Gx CCR-I und Gx CCR-T) basierend auf der verfügbaren DB-Kapazität.

DRA überwacht die CPU-Auslastung der DB und drosselt bei Überschreiten des Grenzwerts die eingehenden Anfragen. Die CPU-Grenzwerte für die Drosselung und der zu drosselnde eingehende Datenverkehr können konfiguriert werden.

Es können verschiedene CPU-Schwellenwerte mit den entsprechenden Drosselungsprozentsätzen konfiguriert werden. DRA passt die Drosselungsstufe basierend auf der aktuellen CPU-Auslastung der DB an. Wenn die CPU-Auslastung stabil wird, wird die Drosselung schrittweise angehalten.



Dieses Verhalten kann in CPS vDRA durch Konfigurieren der CRD-Tabellen 'Peer Rate Limit Profile', 'Message Rate Limit Profile', 'Dynamic Peer Rate Limit Profile' und 'Dynamic Throttling DB CPU Profile' aktiviert werden. In diesen CRD-Tabellen müssen Sie die folgenden Werte konfigurieren:

Peer-Gruppe	Eine Peer-Gruppe ist eine logische Gruppierung von Diameter-Knoten, basierend auf ihrem Bereich und Host. In diesem Beispiel konfigurieren Sie die Peer-Gruppe des pGW.
Peer-FQDN	FQDN (exakte Übereinstimmung oder Übereinstimmung mit regulären Ausdrücken) für die Peers in der Peer-Gruppe, für die Sie eine Durchsatzbegrenzung benötigen.
Nachrichten-/Anforderungstyp	Anwendungs-ID und Anforderungstyp der Anforderungen, die gedrosselt werden müssen In diesem Beispiel Gx CCR-I, Gx CCR-T.
Nachrichtenrichtung	Richtung der Drosselung - Eingang oder Ausgang. In diesem Fall - Ingress.

Profil für Ratenlimit	Der Name des Profils für die Beschränkung der Nachrichtenrate, mit dem der zu drosselnde Nachrichtentyp definiert wurde.
Peer-Rate-Limit	Die für diese Peer-Gruppe zulässige Rate von Anforderungen. Dies schließt alle Nachrichtentypen dieser Peer Group ein.
Verhalten verwerfen	Sie können die Anforderung löschen oder mit einem Fehlercode ablehnen.
Ergebniscode	Ergebniscodewert, falls Sie die Nachrichten ablehnen. Nicht zutreffend, wenn die Nachrichten verworfen werden.
Fehlerzeichenfolge	Die Fehlerzeichenfolge, die in der Antwortnachricht der Anforderung verwendet wird und abgelehnt wird. Nicht zutreffend, wenn die Nachricht verworfen wird.
Anwendungs-ID	Anwendungs-ID der zu drosselnden Nachricht.
Befehlscode	Befehlscode der zu drosselnden Nachricht.
Übertragungsratenlimit für Nachrichten	TPS der Anforderung dieses Nachrichtentyps, die von DRA verarbeitet wird. Anforderungen, die über diesen TPS hinausgehen, werden gedrosselt. Dieser Wert gilt pro Peer in der Peer-Gruppe.
Dynamische Einschränkung des DB-CPU-Profiles	Dies bezieht sich auf den Namen des CPU-Profiles, das verwendet wird, um den Drosselungsprozentsatz für verschiedene CPU-Bereiche zu definieren.
Schwellenwert für DB-CPU-Auslastung	Sie können den korrekten Wert der CPU-Ebenen auswählen, die gemäß dem Datenverkehrsmuster in Ihrer Bereitstellung als Bedrohungsgrenzwerte konfiguriert sind.
Drosselungsprozentsatz	Der Prozentsatz des Ratenlimits, der angewendet wird, wenn die entsprechende CPU-Ebene überschritten wird.

Peer Rate Limit Profile



Filter by All Visible Columns

Peer Group *	Peer FQDN *	Message Direction *	Rate Limit Profile	Peer Rate Limit	Discard Behavior *	Result Code	Error String	Actions
PG_pGW_1	match=peer-*	Ingress	CCR_LT	1000	Send Error Answer	3002	DRA rate limit breached	

Showing 1 out of 1

Message Rate Limit Profile



Filter by All Visible Columns

Rate Limit Profile Name *	Application Identifier *	Command Code *	Message/Request Type *	Message Rate Limit *	Actions
CCR_LT	16777238	272	1	1000	
CCR_LT	16777238	272	3	1000	

Showing 2 out of 2

Show 50 rows 1 out of 1

Dynamic Peer Rate Limit Profile



Filter by All Visible Columns

Peer Group *	Peer FQDN *	Dynamic Throttling DB CPU Profile	Actions
PG_pGW_1	*	DynRateLimit	

Showing 1 out of 1

Show 50 rows 1 out of 1

Dynamic Throttling DB CPU Profile



Filter by

All Visible Columns

CPU Profile Name *	DB CPU Utilization Threshold *	Throttle Percentage	Actions
DynRateLimit	50	20	
DynRateLimit	55	30	
DynRateLimit	60	40	
DynRateLimit	65	50	

Showing 4 out of 4

Show 50 rows 1 out of 1

Außerdem muss dieses Verhalten aktiviert werden, indem das Kontrollkästchen im Policy Builder unter "DRA Configuration Plugin" im Abschnitt "DRA Dynamic Peer Rate Limiter" aktiviert wird.

Umgekehrte Haltezeit - Der Zeitraum, für den die CPU-Auslastung überwacht wird, bevor die Umkehrung angewendet wird.

Umkehrrschritt in % - Der Prozentsatz der Drosselung, der umgekehrt wird.

The screenshot shows the Policy Builder interface for the DRA Configuration Plugin. The configuration for the DRA Dynamic Peer Rate Limiter is highlighted in yellow. The configuration includes a Reversal Hold Time of 30 seconds and a Reversal Step of 20%. The checkbox for 'Auto Apply Next Level Throttle' is checked.

Szenario: Dynamische Ratenbegrenzung basierend auf der CPU-Auslastung

Sehen Sie sich diese Konfiguration bei DRA an:

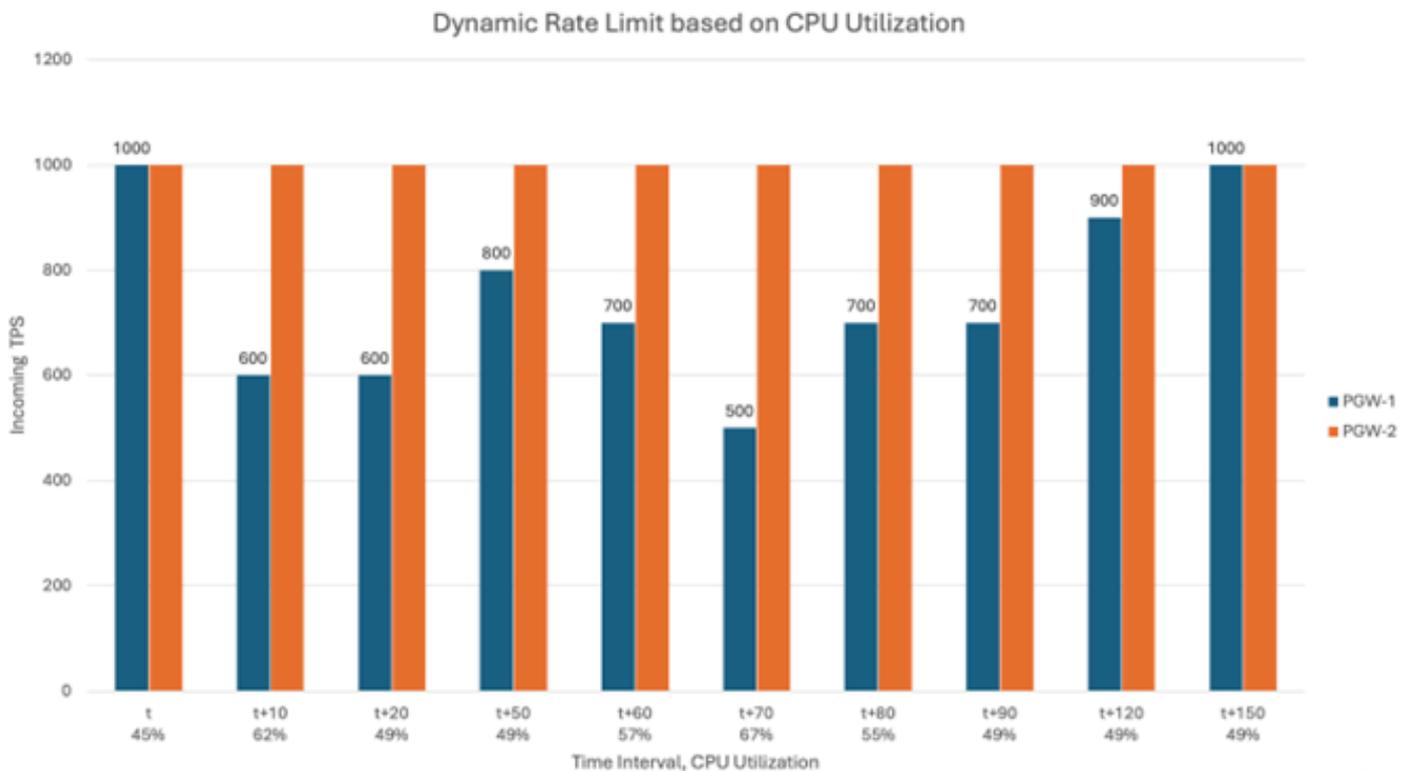
Übertragungsratenlimit statischer Nachrichten: 1000 (ist somit der Wert des eingehenden TPS)

Zurückstellungszeit: 30 s

Umkehrrschritt in %: 20%

Wenn die CPU-Auslastung der DB den Schwellenwert überschreitet, verweist sie auf die Konfiguration "Dynamic Throttling DB CPU Profile" und drosselt den eingehenden TPS entsprechend, indem sie den Director benachrichtigt. Da die Drosselung auf sich ständig

ändernden CPU-Auslastungswerten basiert, kann man sagen, dass die dynamische Rate den Datenverkehr einschränkt.



- Die CPU-Auslastung der DB liegt anfänglich unter dem Grenzwert, sodass keine Drosselung erfolgt. Außerdem verfügt der PGW-2 nicht über die Konfiguration einer dynamischen Ratenbegrenzung, sodass dort keine Drosselung unabhängig von der CPU-Nutzung stattfindet.
- Wenn die CPU-Auslastung der DB 62 % beträgt, wird der Datenverkehr um 40 % gedrosselt, und die effektive Durchsatzratengrenze beträgt 600 (eingehender TPS ist 1000, DRA erlaubt nur 600).
- Wenn die CPU-Auslastung zwischen 60 und 65 % bleibt, wird die Drosselung von 40 % bei konfigurierterem Ratenlimit von 1.000 und einem effektiven Ratenlimit von 600 weiterhin angewendet (eingehendes TPS ist 1.000, DRA erlaubt nur 600).
- Die CPU-Auslastung wird auf 49 % reduziert, die Umkehr der Drosselung beginnt bei pGW-1.
- Wenn die CPU-Auslastung 30 Sekunden lang bei 49 % oder weniger bleibt, wird die Drosselung um 20 % auf 20 % reduziert. Die effektive Durchsatzgrenze liegt jetzt bei 800 (eingehender TPS bei 1000, DRA nur bei 800). Die Umkehrung erfolgt gemäß der Konfiguration in Schritten von 20 %.
- Wenn die CPU-Auslastung der DB auf 57 % zunimmt, wird der Datenverkehr um 30 % gedrosselt, und die effektive Durchsatzratengrenze beträgt 700 (eingehender TPS ist 1000, DRA erlaubt nur 700).
- Wenn die CPU-Auslastung der DB auf 67 % zunimmt, wird der Datenverkehr um 50 % gedrosselt, und das effektive Durchsatzlimit beträgt 500 (eingehender TPS ist 1000, DRA erlaubt nur 500).
- Wenn die CPU-Auslastung der DB auf 55 % sinkt, wird der Datenverkehr um 30 % gedrosselt, und das effektive Durchsatzlimit beträgt 700 (eingehender TPS ist 1000, DRA

erlaubt nur 700).

- Wenn die CPU im nächsten 30-Sekunden-Intervall auf 49 % oder weniger fällt, wird die Drosselung um 20 % auf 10 % verringert, und die effektive Durchsatzgrenze beträgt 900 (eingehender TPS ist 1000, DRA erlaubt nur 900).
- Wenn die CPU während des nächsten 30-Sekunden-Intervalls weiterhin bei 49 % oder weniger bleibt, wird die Drosselung um 20 % auf 0 verringert, und es wird keine Ratenbeschränkung angewendet, da die Umkehr abgeschlossen ist (eingehender TPS ist 1000, DRA erlaubt 1000).

Informationen zu dieser Übersetzung

Cisco hat dieses Dokument maschinell übersetzen und von einem menschlichen Übersetzer editieren und korrigieren lassen, um unseren Benutzern auf der ganzen Welt Support-Inhalte in ihrer eigenen Sprache zu bieten. Bitte beachten Sie, dass selbst die beste maschinelle Übersetzung nicht so genau ist wie eine von einem professionellen Übersetzer angefertigte. Cisco Systems, Inc. übernimmt keine Haftung für die Richtigkeit dieser Übersetzungen und empfiehlt, immer das englische Originaldokument (siehe bereitgestellter Link) heranzuziehen.