

使用CPS vDRA进行主动式静态和动态速率限制

目录

[简介](#)

[先决条件](#)

[要求](#)

[使用的组件](#)

[背景信息](#)

[问题](#)

[解决方案](#)

[负载均衡器的静态速率限制](#)

[入口速率限制](#)

[出口速率限制](#)

[动态速率限制](#)

简介

本文档介绍DRA中的速率限制选项，DRA是路由Diameter消息和管理网络流量的电信组件。

先决条件

要求

Cisco 建议您了解以下主题：

- 思科策略套件(CPS)Diameter路由代理(vDRA)
- Diameter路由代理基础知识和规范

使用的组件

本文档中的信息基于Cisco Policy Suite DRA。

本文档中的信息都是基于特定实验室环境中的设备编写的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您的网络处于活动状态，请确保您了解所有命令的潜在影响。

背景信息

DRA是电信网络中的一个组件，特别是在基于Diameter协议的网络环境中。DRA可在不同的网络元素（例如策略服务器、计费系统和其他支持Diameter的设备）之间有效路由Diameter消息。速率限制是一种网络流量管理技术，用于控制进出网络元素的流量量。它有助于确保网络资源不被耗尽，保持服务质量，并防止网络被误用或滥用。

问题

网络中的每个组件都可以根据其额定容量处理流量负载，但实时出现生成的流量大于系统可处理流量的情况。其中包括：

- 用户行为 — 在短时间内生成大量数据的活动，如流传输事件或软件更新。通常从网关(Gw)发送到DRA。
- 网络拥塞 — 在网络使用率较高时，拥塞可能会增加，导致排队的数据在容量可用时以突发形式发送。
- 网络弹性机制 — 在中断或维护期间重新路由流量，导致临时高峰。这可能会影响没有任何网络问题的匹配站点中的流量。
- 网络元素行为 — 如果出现过载和拥塞，您可能会开始看到一个或多个网络元素没有响应/超时，从而导致重新连接导致系统进一步过载。
- 网关刷新 — 由于策略更改、拓扑更改或任何维护或故障排除活动，网关可以刷新现有会话。在这些情况下，会话会被清除，并且您可以接收一组Gx信用控制请求(CCR)-T请求。

解决方案

DRA可以在多个Diameter服务器之间分配负载，以确保有效处理请求并避免单个服务器过载。如果服务器发生故障，DRA可以将消息重定向到备用服务器，从而确保网络服务的高可用性和可靠性。

对DRA进行速率限制，不仅可保护DRA，还可通过确保受控的报文流保护其他实体。速率限制的主要优势包括：

- 服务连续性 — 通过确保关键网络组件不会超载并防止停机，保持连续的服务可用性。
- 可扩展性 — 允许网络处理各种负载，而不会降低性能。
- 符合服务级别协议(SLA) — 通过保持一致的性能和可靠性级别确保网络满足其SLA。

负载均衡器的静态速率限制

这是一种直接转发方法，根据DRA/数据包网关(pGW)和其他网络元素的额定容量设置固定阈值，不会根据网络条件或系统资源而更改。通过限制传入请求的速率，您可以对DRA处理的流量量产生可预测的结果。

静态速率限制的配置取决于应用该限制的用例。

入口速率限制

场景:从pGW突发

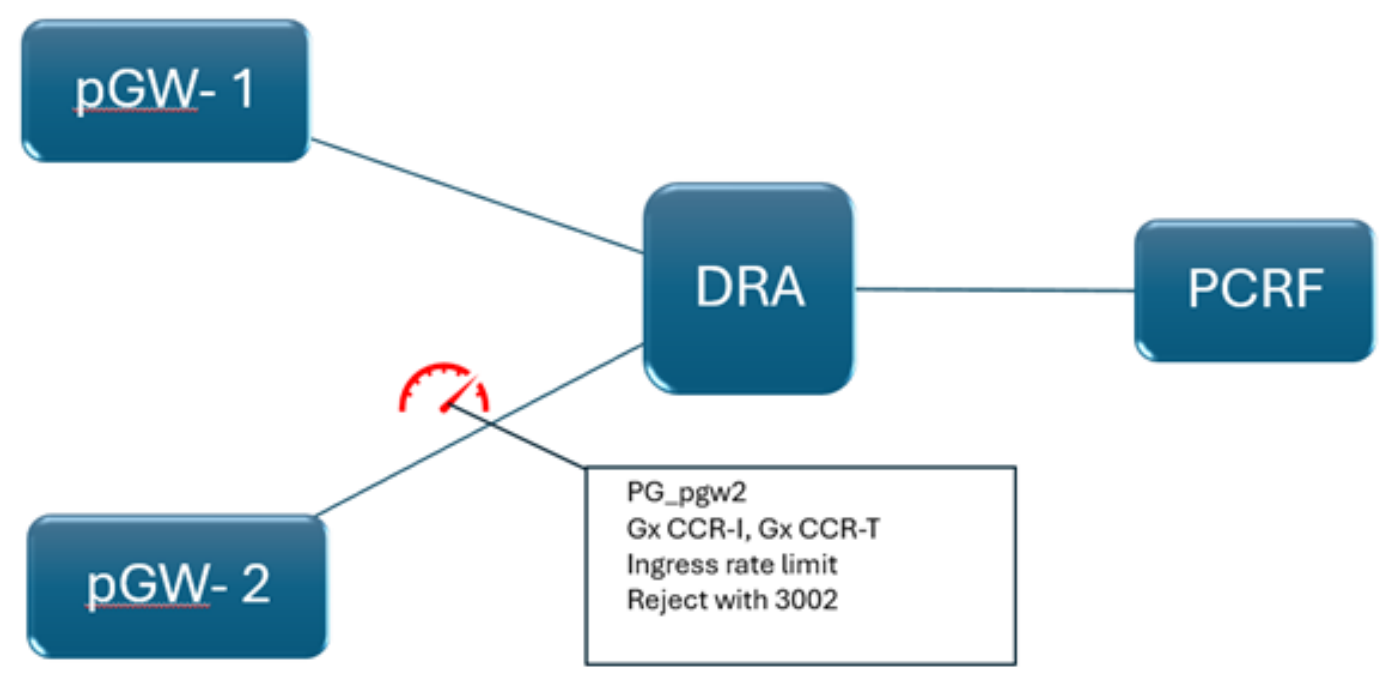
配置特定于pGW且易受这些流量突发影响的阈值。该值必须基于在这些突发事件期间可看到的常规流量/峰值流量数得出。

可以针对每种消息类型定义阈值编号，以确保仅限制突发流量，例如，必须限制来自GW的Gx CCR-I和Gx CCR-T请求，但转发Gx CCR-U流量或Gy流量。

在这种情况下，您可以在入口端进行调节，即DRA在收到消息后立即对其进行调节，因为此处的目

的是基于接收请求的网元拒绝消息，并避免处理比DRA可处理的请求数更多的请求。

限制行为可以是拒绝带有特定错误代码和错误消息的消息，也可以丢弃该消息。



通过配置自定义参考数据(CRD)表“对等体速率限制配置文件”和“消息速率限制配置文件”，可以在 CPS vDRA中启用此行为。在这些CRD表中，您需要配置以下值：

对等组	对等体组是基于Diameter节点的领域和主机的逻辑分组。您需要配置需要限制的对等体组。
对等体完全限定域名(FQDN)	对等组中需要速率限制的对等体的FQDN（完全匹配或正则表达式匹配）。
消息方向	限制的方向 — 入口或出口。在本例中，为Ingress。
速率Limit配置文件	用于定义需要限制的消息类型的消息速率限制配置文件名称。
对等体速率限制	允许此对等组的请求的速率。这包括来自该对等组的所有消息类型。
丢弃行为	您可以选择丢弃请求或拒绝带有错误代码的请求。
结果代码	结果代码值，以防您拒绝消息。不适用于丢弃消息的情况。

错误字符串	被拒绝的请求响应消息中使用的错误字符串。不适用于邮件被丢弃的情况。
应用标识符	要限制的邮件的应用程序ID。
命令代码	要限制的的消息的命令代码。
消息/请求类型	需要限制请求的应用ID和请求类型。
消息速率限制	由DRA处理的该消息类型请求的网真服务器(TPS)。此TPS之外的请求将被限制。此值按对等体组中的每个对等体显示。

这里是一个示例，其中为pGW2配置了1000条消息的总速率限制，并配置了200 Gx CCR-I和200 Gx CCR-T的速率限制。超过此速率的所有请求都会被3002拒绝，并出现一则错误消息，指示速率限制已被违反。

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- xxxxxx	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

>

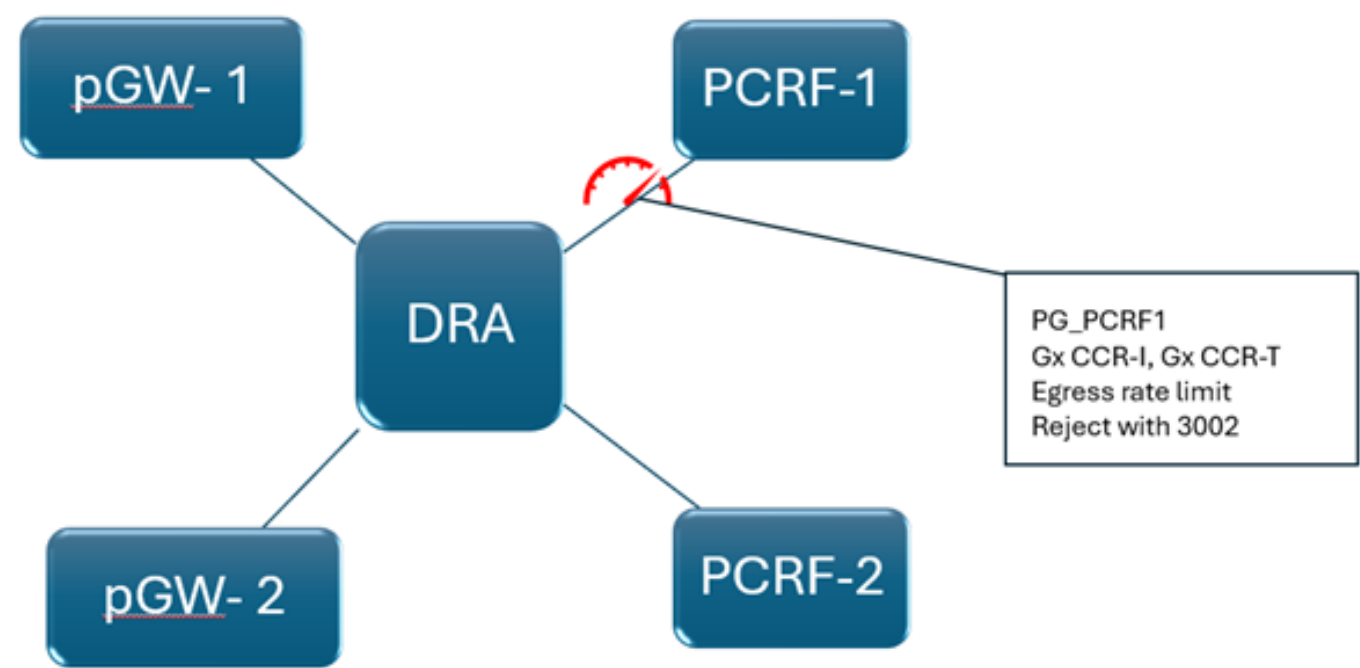
出口速率限制

场景:接收请求的网络元素的保护

考虑一个Gx事务示例，其中从pGW接收请求并将其转发到策略和计费规则功能(PCRF)。如果PCRF可以处理的数据量存在限制，则即使DRA可以处理传入流量，也可以使用DRA在DRA限制消息，而不是将请求转发到PCRF并使其过载。

此处您需要在出口端进行调节，即DRA根据基于DRA路由逻辑识别的PCRF对等组在消息转发到PCRF之前调节消息。

限制行为可以是拒绝带有特定错误代码和错误消息的消息，也可以丢弃该消息。



通过配置CRD表“对等速率限制配置文件”和“消息速率限制配置文件”，可以在CPS vDRA中启用此行

为。 在以下CRD表中，您需要配置以下值：

对等组	对等体组是基于Diameter节点的领域和主机的逻辑分组。您需要配置需要限制的对等体组。
对等FQDN	对等组中需要速率限制的对等体的FQDN（完全匹配或正则表达式匹配）。
消息方向	限制的方向 — 入口或出口。在本例中为Egress。
速率Limit配置文件	用于定义需要限制的消息类型的消息速率限制配置文件名称。
对等体速率限制	允许此对等组的请求的速率。这包括来自该对等组的所有消息类型。
丢弃行为	您可以选择丢弃请求或拒绝带有错误代码的请求。
结果代码	结果代码值，以防您拒绝消息。不适用于丢弃消息的情况。
错误字符串	被拒绝的请求响应消息中使用的错误字符串。不适用于邮件被丢弃的情况。
应用标识符	要限制的邮件的应用程序ID。
命令代码	要限制的消息的命令代码。
消息/请求类型	需要限制请求的应用ID和请求类型。
消息速率限制	DRA处理的该消息类型请求的TPS。此TPS之外的请求将被限制。此值按对等体组中的每个对等体显示。

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-...	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

场景:网络缓慢导致流量拥塞，导致DRA完全/部分故障

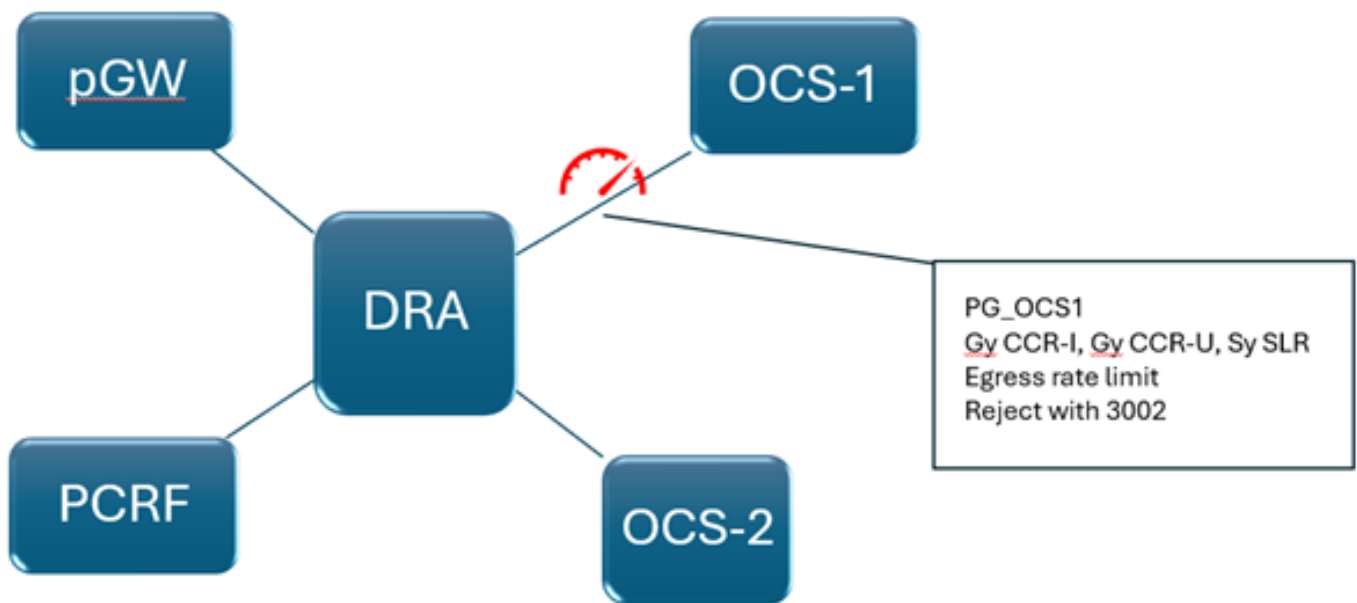
以pGW和在线计费系统(OCS)之间交换的Gy交易为例。如果DRA-OCS信道上的网络速度缓慢（由于来自pGW的高流量或由于任何其他网络问题），则由于SLA违规，请求将超时。这些超时不仅会影响DRA，也会影响整个网络。

DRA资源在尝试通过慢速网络向OCS发送请求时受阻，导致其资源耗尽。这会导致DRA拒绝多个请求，但DRA的额定容量并未遭到破坏。

这也会影响不在DRA-OCS信道上的流量。这些拒绝/超时和丢弃会触发多个网络元素上的重新连接。

在这种情况下，需要在出口端进行限制 — DRA根据具有容量限制或网络问题的OCS对等体组在消息转发到OCS之前限制消息)。

限制行为可以是拒绝带有特定错误代码和错误消息的消息，也可以丢弃该消息。



通过配置CRD表“对等速率限制配置文件”和“消息速率限制配置文件”，可以在CPS vDRA中启用此行为。在以下CRD表中，您需要配置以下值：

对等组	对等体组是基于Diameter节点的领域和主机的逻辑分组。您需要配置需要限制的对等体组。
对等FQDN	对等组中需要速率限制的对等体的FQDN（完全匹配或正则表达式匹配）。
消息方向	限制的方向 — 入口或出口。在本例中为Egress。
速率限制配置文件	用于定义需要限制的消息类型的消息速率限制配置文件名称。
对等体速率限制	允许此对等组的请求的速率。这包括来自该对等组的所有消息类型。
丢弃行为	您可以选择丢弃请求或拒绝带有错误代码的请求。
结果代码	结果代码值，以防您拒绝消息。不适用于丢弃消息的情况。
错误字符串	被拒绝的请求响应消息中使用的错误字符串。不适用于邮件

	被丢弃的情况。
应用标识符	要限制的邮件的应用程序ID。
命令代码	要限制的消息的命令代码。
消息/请求类型	需要限制请求的应用ID和请求类型。
消息速率限制	DRA处理的该消息类型请求的TPS。此TPS之外的请求将被限制。此值按对等体组中的每个对等体显示。

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-	Egress	Gy_Sy_Limit	1000	Send Error Answer	3002	DRA rate limit breached	

Showing 1 out of 1

Message Rate Limit Profile

Filter by All Visible Columns

Gy_Sy_Limit

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

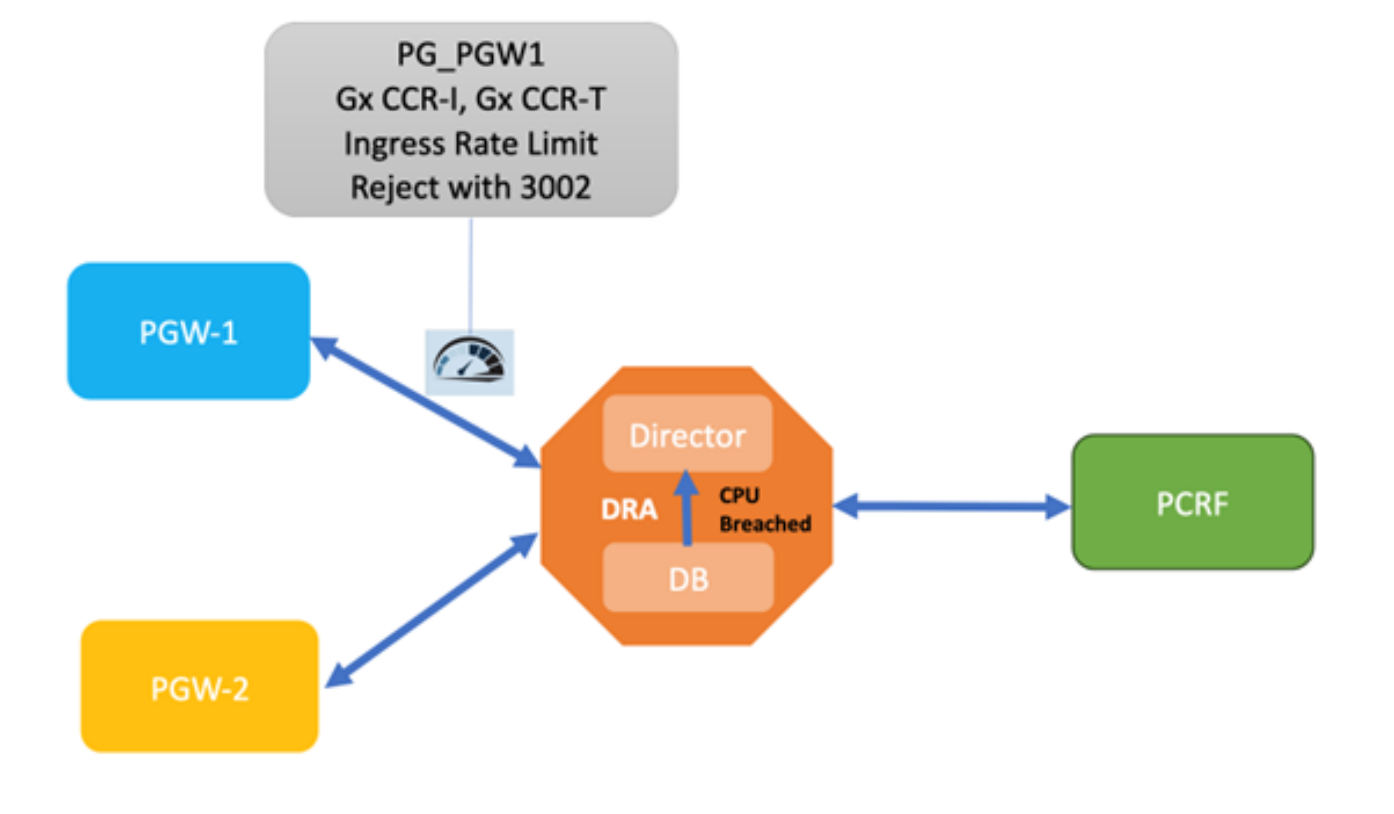
Show 50 rows 1 out of 1

动态速率限制

当发生CCR-I或CCR-T突发时，数据库(DB)上可能会出现过载，从而导致系统不稳定。为了克服这一点，DRA支持基于可用数据库容量的动态速率限制（仅适用于Gx CCR-I和Gx CCR-T）。

DRA监视数据库CPU利用率，每当超过阈值时，它就会限制传入请求。用于限制的CPU阈值和要限制的传入流量是可配置的。

可以配置具有相应节流百分比的不同CPU阈值。DRA根据当前DB CPU使用情况调整限制级别。当CPU使用率变为稳定时，限制会逐渐停止。



通过配置CRD表“对等速率限制配置文件”、“消息速率限制配置文件”、“动态对等速率限制配置文件”和“动态限制数据库CPU配置文件”，可以在CPS vDRA中启用此行为。在以下CRD表中，您需要配置以下值：

对等组	对等体组是基于Diameter节点的领域和主机的逻辑分组。在本示例中，您将配置pGW的对等组。
对等FQDN	对等组中需要速率限制的对等体的FQDN（完全匹配或正则表达式匹配）。
消息/请求类型	需要限制请求的应用ID和请求类型。在本例中，Gx CCR-I、Gx CCR-T。
消息方向	限制的方向 — 入口或出口。在本例中，为Ingress。
速率限制配置文件	用于定义需要限制的消息类型的消息速率限制配置文件名称。

对等体速率限制	允许此对等组的请求的速率。这包括来自该对等组的所有消息类型。
丢弃行为	您可以选择丢弃请求或拒绝带有错误代码的请求。
结果代码	结果代码值，以防您拒绝消息。不适用于丢弃消息的情况。
错误字符串	被拒绝的请求响应消息中使用的错误字符串。不适用于邮件被丢弃的情况。
应用标识符	要限制的邮件的应用程序ID。
命令代码	要限制的消息的命令代码。
消息速率限制	DRA处理的该消息类型请求的TPS。此TPS之外的请求将被限制。此值按对等体组中的每个对等体显示。
动态限制DB CPU配置文件	这是指CPU配置文件名称，用于定义不同CPU范围的限制百分比。
数据库CPU使用率阈值	您可以根据部署中的流量模式选择配置为漏洞限制的CPU级别的正确值。
限制百分比	超过相应CPU级别时应用的速率限制百分比。

Peer Rate Limit Profile 

Filter by

All Visible Columns

DRA rate



 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

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

DynRateLimit			
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

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

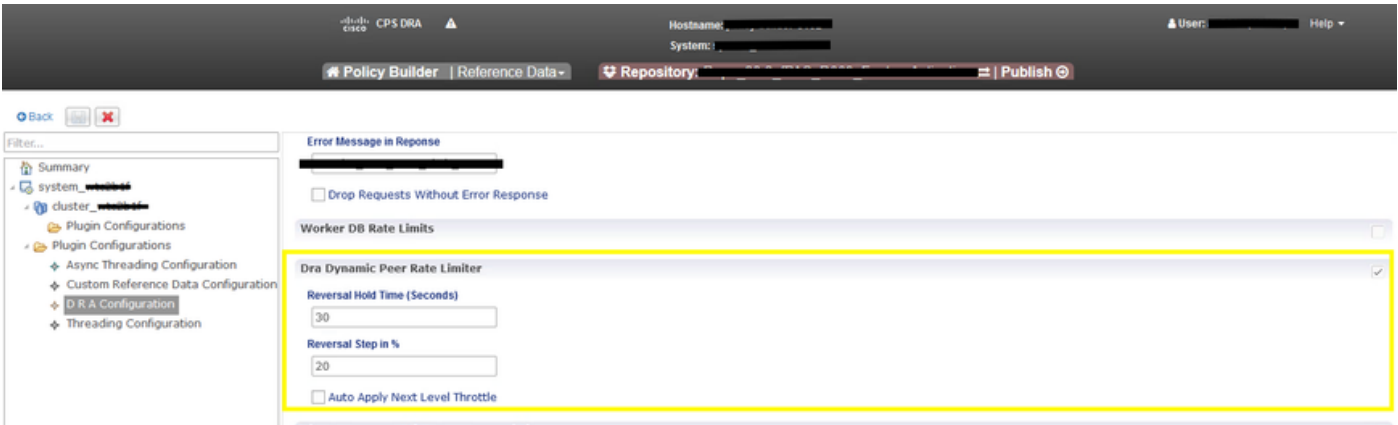
Showing 4 out of 4

Show 50 rows 1 out of 1

此外，必须启用此行为，方法是选中“DRA配置插件”部分下“DRA动态对等体速率限制器”下的策略生成器中的复选框。

Reversal Hold Time — 在应用冲销之前监控CPU使用率的时间段。

倒置步骤% — 倒置限制百分比。



场景:基于CPU利用率的动态速率限制

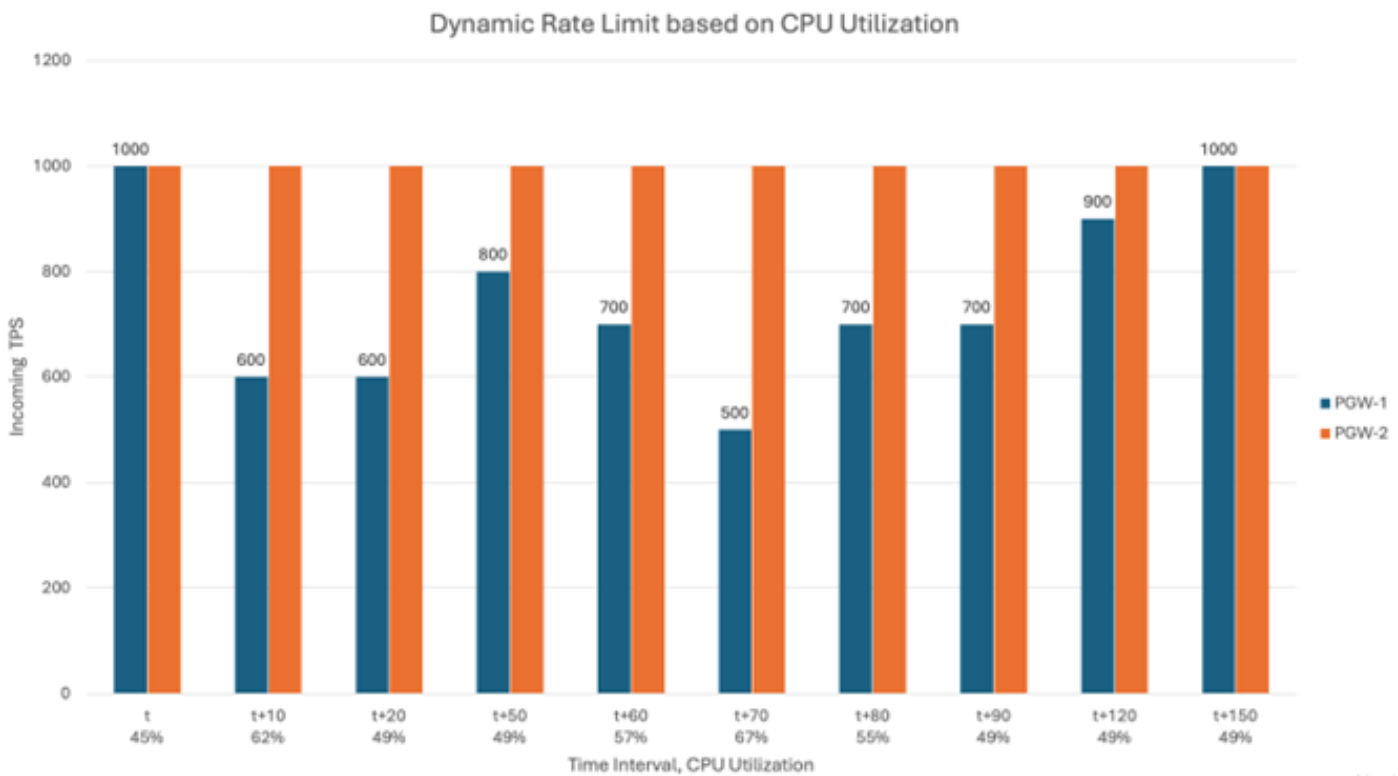
请考虑以下在DRA上的配置：

静态消息速率限制：1000（因此它是传入TPS的值）

冲销暂挂时间：30年代

冲销步骤(%):20%

当DB CPU利用率超过阈值时，它会引用“动态限制DB CPU配置文件”配置，并通过通知导向器相应地限制传入TPS。由于它是基于不断变化的CPU使用率值进行限制，因此您可以将其动态速率限制流量。



- 最初，数据库CPU利用率低于限制，因此不会发生限制。此外，PGW-2没有动态速率限制配置，因此无论CPU使用率如何，都不会发生限制。
- 当数据库CPU使用率为62%时，流量将节流40%，有效速率限制为600（传入TPS为1000,DRA仅允许600）。
- 如果CPU使用率保持在60-65%之间，则在配置的速率限制1000上继续应用40%的调节，并且有效速率限制为600（传入TPS为1000,DRA仅允许600。）
- CPU利用率降至49%，带宽限制逆转开始于pGW-1。
- 如果CPU利用率在30秒内保持在49%或更低，则限制会减少20%至20%。现在有效速率限制为800（传入TPS为1000,DRA仅允许800。）根据配置，进行反向操作时，其步骤为20%。
- 当DB CPU利用率增加到57%时，流量将节流30%，有效速率限制为700（传入TPS为1000,DRA仅允许700。）
- 当DB CPU利用率增加到67%时，流量将节流50%，有效速率限制为500（传入TPS为1000,DRA仅允许500。）
- 当DB CPU利用率降至55%时，流量将节流30%，有效速率限制为700（传入TPS为1000,DRA仅允许700。）
- 如果CPU在接下来的30秒间隔内降至49%或更低，则限制减少20%至10%，有效速率限制为900（传入TPS为1000,DRA仅允许900。）
- 如果CPU在接下来的30秒间隔内进一步保持在49%或更低，则限制会减少20%至0，并且当反向完成时不会应用速率限制（传入TPS为1000,DRA允许1000。）

关于此翻译

思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。