

基於CPS vDRA的搶佔式靜態和動態速率限制

目錄

[簡介](#)

[必要條件](#)

[需求](#)

[採用元件](#)

[背景資訊](#)

[問題](#)

[解決方案](#)

[負載平衡器上的靜態速率限制](#)

[輸入速率限制](#)

[輸出速率限制](#)

[動態速率限制](#)

簡介

本文檔介紹DRA中的速率限制選項，DRA是路由Diameter消息和管理網路流量的電信元件。

必要條件

需求

思科建議您瞭解以下主題：

- Cisco Policy Suite(CPS)Diameter Routing Agent(vDRA)
- Diameter路由代理基礎知識和規範

採用元件

本檔案中的資訊是根據Cisco Policy Suite DRA。

本文中的資訊是根據特定實驗室環境內的裝置所建立。文中使用到的所有裝置皆從已清除（預設）的組態來啟動。如果您的網路運作中，請確保您瞭解任何指令可能造成的影響。

背景資訊

DRA是電信網路中的一個元件，特別是在基於Diameter協定的網路環境中。DRA可在不同的網路元素（例如策略伺服器、計費系統和其它支援Diameter的裝置）之間有效地路由Diameter消息。速率限制是一種網路流量管理技術，用於控制進出網路元素的流量量。它有助於確保網路資源不被耗盡，保持服務品質，並防止網路被濫用或濫用。

問題

網路中的每個元件都可以根據其額定容量處理流量負載，但即時情況下，生成的流量可能高於系統可處理的流量。其中一些是：

- 使用者行為 — 在短時間內生成大量資料的活動，如流事件或軟體更新。通常從網關(Gw)傳送到DRA。
- 網路擁塞 — 在網路使用率較高的時期，擁塞可能會累積，導致排隊的資料在容量可用時以突發形式傳送。
- 網路恢復機制 — 在中斷或維護期間重新路由流量，導致臨時高峰。這可能會影響沒有任何網路問題的匹配站點中的流量。
- 網路元素行為 — 如果出現過載和擁塞，您可能會開始看到一個或多個網路元素沒有響應/超時，從而導致重新連線導致系統進一步過載。
- 網關刷新 — 由於策略更改、拓撲更改或任何維護或故障排查活動，網關可以刷新現有會話。在這些情況下，會話會被清除，並且您可以接收一組Gx信用控制請求(CCR)-T請求。

解決方案

DRA可以在多個Diameter伺服器之間分配負載，以確保有效處理請求並避免單個伺服器過載。在伺服器發生故障時，DRA可以將消息重定向到備用伺服器，確保網路服務的高可用性和可靠性。

對DRA進行速率限制，不僅可保護DRA，還可通過確保受控報文流保護其他實體。速率限制的主要優勢包括：

- 服務連續性 — 通過確保關鍵網路元件不會超載並防止停機，保持連續的服務可用性。
- 可擴充性 — 允許網路處理各種負載，而不會降低效能。
- 符合服務級別協定(SLA) — 通過保持一致的效能和可靠性級別確保網路滿足其SLA。

負載平衡器上的靜態速率限制

這是一種直接轉發方法，根據DRA/資料包網關(pGW)和其他網元的額定容量設定固定閾值，並且不會根據網路條件或系統資源而更改。通過限制傳入請求的速率，您可以對DRA處理的流量產生可預測的結果。

靜態速率限制的配置取決於應用此限制的使用情形。

輸入速率限制

案例：從pGW猝發

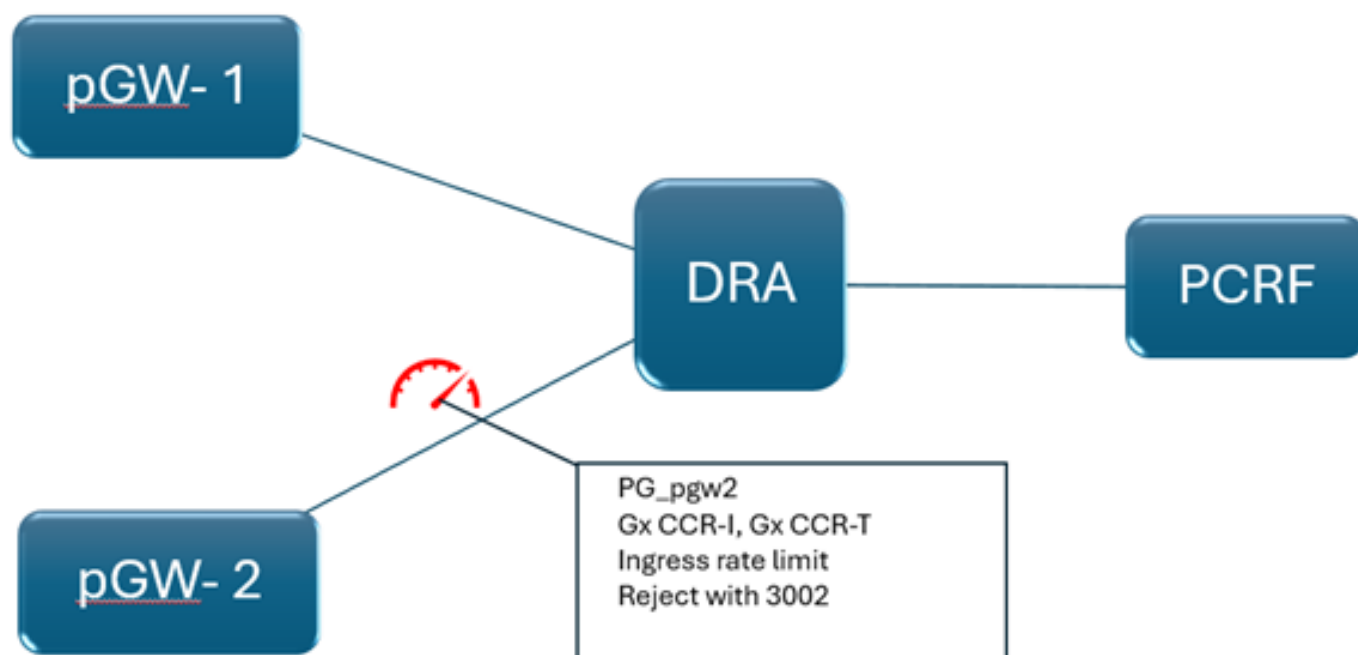
配置了對易受這些流量突發影響的pGW特定的閾值。該值必須基於在這些突發事件中可看到的常規流量/峰值流量數來獲得。

可以針對每種消息型別定義閾值編號，以確保只有突發流量被限制，例如，必須限制來自GW的Gx CCR-I和Gx CCR-T請求，但轉發Gx CCR-U流量或Gy流量作為接收流量。

在這種情況下，可以在入口端進行調節，即DRA在收到消息後立即對其進行調節，因為這裡的目的

是基於接收請求的網元拒絕消息，並避免處理比DRA可以處理的請求數更多的請求。

限制行為可以是拒絕帶有特定錯誤代碼和錯誤消息的消息，也可以丟棄該消息。



通過配置自定義引用資料(CRD)表「對等體速率限制配置檔案」和「消息速率限制配置檔案」，可以在CPS vDRA中啟用此行為。在這些CRD表中，需要配置以下值：

對等體組	對等體組是基於Diameter節點的領域和主機的邏輯分組。您需要設定需要節流的對等體組。
對等完全限定域名(FQDN)	需要速率限制的對等組中對等體的FQDN (完全匹配或正規表示式匹配)。
消息方向	限制的方向 — Ingress或Egress。在本例中 — Ingress。
速率Limit配置檔案	用於定義需要限制的消息型別的消息速率限制配置檔名稱。
對等速率限制	允許此對等組的請求的速率。這包括來自該對等組的所有消息型別。
丟棄行為	您可以選擇丟棄請求或拒絕帶有錯誤代碼的請求。
結果代碼	結果代碼值，以防您拒絕消息。不適用於丟棄消息的情況。

錯誤字串	被拒絕的請求響應消息中使用的錯誤字串。在郵件被丟棄的情況下不適用。
應用程式識別符號	要限制的郵件的應用程式ID。
命令代碼	要限制的訊息的命令代碼。
消息/請求型別	需要限制請求的應用程式ID和請求型別。
消息速率限制	由DRA處理的該消息型別請求的TelePresence Server(TPS)。此TPS以外的請求將被限制。此值是對等組中的每個對等體。

此處提供了一個示例，其中為pGW2配置1000條消息的總速率限制以及200 Gx CCR-I和200 Gx CCR-T的速率限制。超過此速率的任何請求都會被3002拒絕，並出現一條錯誤消息，指示速率限制已被違反。

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- 10.10.10.10	Ingress	CCR_I_T_Limit	1000	Send Error Answer	3002	DRA rate limit breached	Edit Delete

Showing 1 out of 1

Show 50 rows
 << < 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

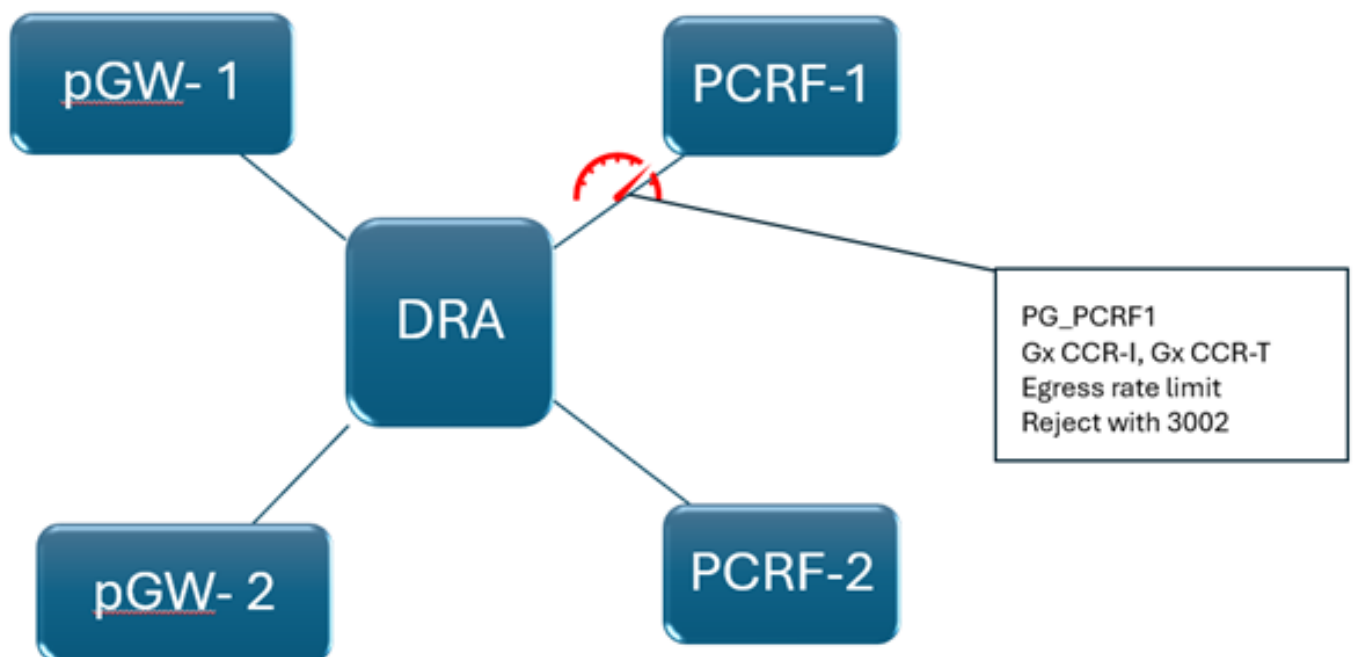
輸出速率限制

案例：對接收請求的網元的保護

考慮一個Gx事務的示例，其中從pGW接收請求並將其轉發到策略和計費規則功能(PCRF)。如果PCRF可以處理的資料量存在限制，則即使DRA可以處理傳入流量，也可以使用DRA在DRA限制消息，而不是將請求轉發到PCRF並將其過載。

此處您需要在出口端進行限制，即DRA根據基於DRA路由邏輯識別的PCRF對等組，在消息轉發到PCRF之前立即限制消息。

限制行為可以是拒絕帶有特定錯誤代碼和錯誤消息的消息，也可以丟棄該消息。



通過配置CRD表「對等速率限制配置檔案」和「消息速率限制配置檔案」，可以在CPS vDRA中啟

用此行為。 在以下CRD表中，您需要配置以下值：

對等體組	對等體組是基於Diameter節點的領域和主機的逻辑分組。您需要設定需要節流的對等體組。
對等FQDN	需要速率限制的對等組中對等體的FQDN (完全匹配或正規表示式匹配)。
消息方向	限制的方向 — Ingress或Egress。本例中為Egress。
速率Limit配置檔案	用於定義需要限制的消息型別的消息速率限制配置檔名稱。
對等速率限制	允許此對等組的請求的速率。這包括來自該對等組的所有消息型別。
丟棄行為	您可以選擇丟棄請求或拒絕帶有錯誤代碼的請求。
結果代碼	結果代碼值，以防您拒絕消息。不適用於丟棄消息的情況。
錯誤字串	被拒絕的請求響應消息中使用的錯誤字串。在郵件被丟棄的情況下不適用。
應用程式識別符號	要限制的郵件的應用程式ID。
命令代碼	要限制的消息的命令代碼。
消息/請求型別	需要限制請求的應用程式ID和請求型別。
消息速率限制	由DRA處理的該消息型別請求的TPS。此TPS以外的請求將被限制。此值是對等組中的每個對等體。

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

案例：網路速度慢導致流量擁塞，導致DRA完全/部分故障

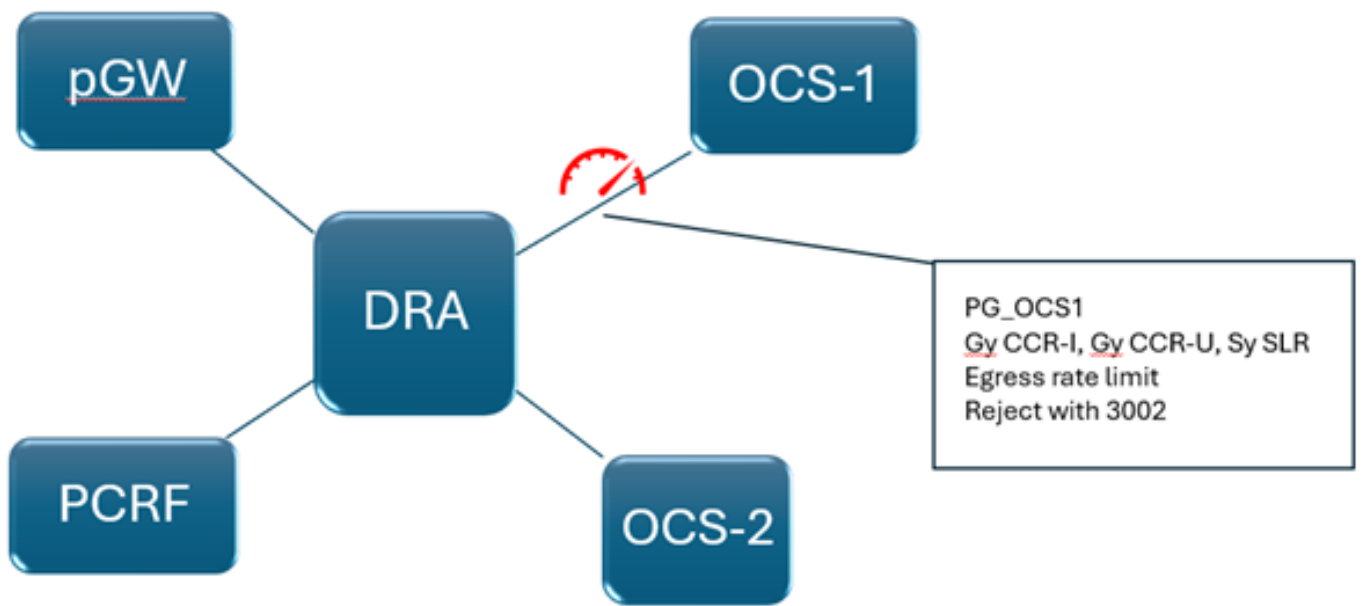
以pGW和線上計費系統(OCS)之間交換的Gy交易為例。如果DRA-OCS通道上的網路速度緩慢（由於來自pGW的高流量或由於任何其他網路問題），由於SLA違規，請求將超時。這些超時不僅會影響DRA，也會影響整個網路。

DRA資源在嘗試通過慢速網路向OCS傳送請求時受阻，導致其資源耗盡。這導致多個請求被DRA拒絕，但不會違反DRA的額定容量。

這也影響不在DRA-OCS通道上的流量。這些拒絕/超時和丟棄觸發多個網路元素上的重新連線。

在這種情況下，您需要在出口端進行限制 — DRA根據具有容量限制或網路問題的OCS對等體組，在消息轉發到OCS之前立即限制消息。

限制行為可以是拒絕帶有特定錯誤代碼和錯誤消息的消息，也可以丟棄該消息。



通過配置CRD表「對等速率限制配置檔案」和「消息速率限制配置檔案」，可以在CPS vDRA中啟用此行為。在以下CRD表中，您需要配置以下值：

對等體組	對等體組是基於Diameter節點的領域和主機的邏輯分組。您需要設定需要節流的對等體組。
對等FQDN	需要速率限制的對等組中對等體的FQDN (完全匹配或正規表示式匹配)。
消息方向	限制的方向 — Ingress或Egress。本例中為Egress。
速率限制配置檔案	用於定義需要限制的消息型別的消息速率限制配置檔名稱。
對等速率限制	允許此對等組的請求的速率。這包括來自該對等組的所有消息型別。
丟棄行為	您可以選擇丟棄請求或拒絕帶有錯誤代碼的請求。
結果代碼	結果代碼值，以防您拒絕消息。不適用於丟棄消息的情況。
錯誤字串	被拒絕的請求響應消息中使用的錯誤字串。在郵件被丟棄的情況下不適用。

應用程式識別符號	要限制的郵件的應用程式ID。
命令代碼	要限制的消息的命令代碼。
消息/請求型別	需要限制請求的應用程式ID和請求型別。
消息速率限制	由DRA處理的該消息型別請求的TPS。此TPS以外的請求將被限制。此值是對等組中的每個對等體。

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

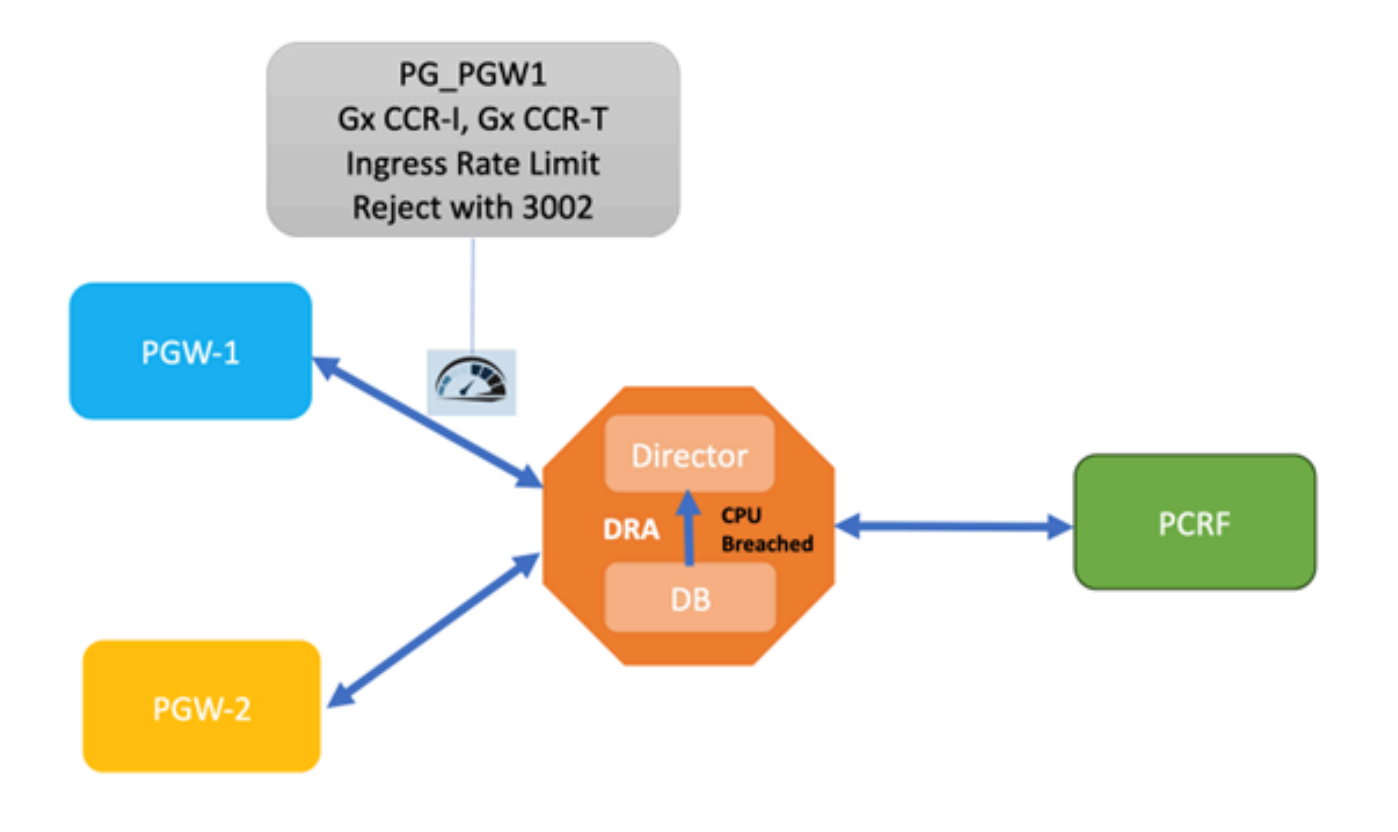
動態速率限制

當發生CCR-I或CCR-T突發事件時，資料庫(DB)上可能會出現過載，從而導致系統不穩定。為了克服這一點，DRA支援基於可用資料庫容量的動態速率限制（僅適用於Gx CCR-I和Gx CCR-T）。

DRA監視資料庫CPU利用率，並且一旦超過閾值，就會限制傳入請求。用於限制的CPU閾值和要限制的傳入流量是可配置的。

可以配置具有不同限制百分比的CPU閾值。DRA根據當前DB CPU使用情況調整限制級別。當

CPU使用率穩定時，限制會逐漸停止。



通過配置CRD表「對等體速率限制配置檔案」、「消息速率限制配置檔案」、「動態對等體速率限制配置檔案」和「動態限制資料庫CPU配置檔案」，可以在CPS vDRA中啟用此行為。在以下CRD表中，您需要配置以下值：

對等體組	對等體組是基於Diameter節點的領域和主機的邏輯分組。在此示例中，您配置pGW的對等組。
對等FQDN	需要速率限制的對等組中對等體的FQDN（完全匹配或正規表示式匹配）。
消息/請求型別	需要限制請求的應用程式ID和請求型別。在本例中，Gx CCR-I、Gx CCR-T。
消息方向	限制的方向 — Ingress或Egress。在本例中 — Ingress。
速率限制配置檔案	用於定義需要限制的消息型別的消息速率限制配置檔名稱。
對等速率限制	允許此對等組的請求的速率。這包括來自該對等組的所有消息

	型別。
丟棄行為	您可以選擇丟棄請求或拒絕帶有錯誤代碼的請求。
結果代碼	結果代碼值，以防您拒絕消息。不適用於丟棄消息的情況。
錯誤字串	被拒絕的請求響應消息中使用的錯誤字串。在郵件被丟棄的情況下不適用。
應用程式識別符號	要限制的郵件的應用程式ID。
命令代碼	要限制的消息的命令代碼。
消息速率限制	由DRA處理的該消息型別請求的TPS。此TPS以外的請求將被限制。此值是對等組中的每個對等體。
動態限制DB CPU配置檔案	這是指CPU配置檔名稱，用於定義不同CPU範圍的限制百分比。
DB CPU利用率閾值	根據部署中的流量模式，您可以選擇配置為漏洞限制的CPU級別的正确值。
限制百分比	突破相應的CPU級別時應用的速率限制的百分比。

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_I_T	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	

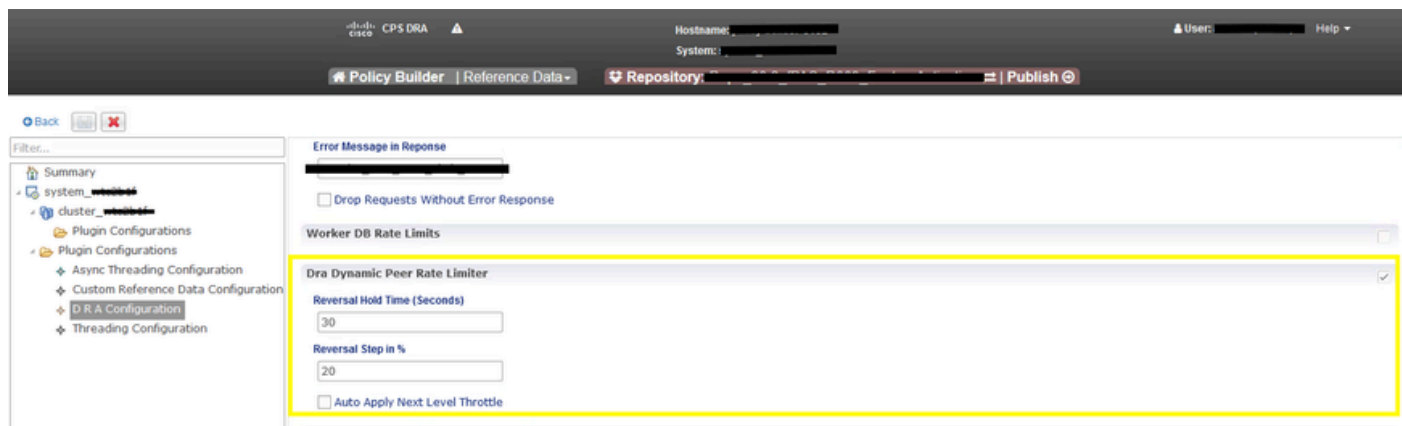
Showing 4 out of 4

Show 50 rows 1 out of 1

此外，必須通過選擇「DRA動態對等速率限制器」一節中「DRA配置外掛」下的策略生成器中的覈取方塊來啟用此行為。

Reversal Hold Time — 在應用沖銷之前監控CPU使用率的時間段。

倒轉步驟(%) — 倒轉的限制百分比。



案例：基於CPU使用率的動態速率限制

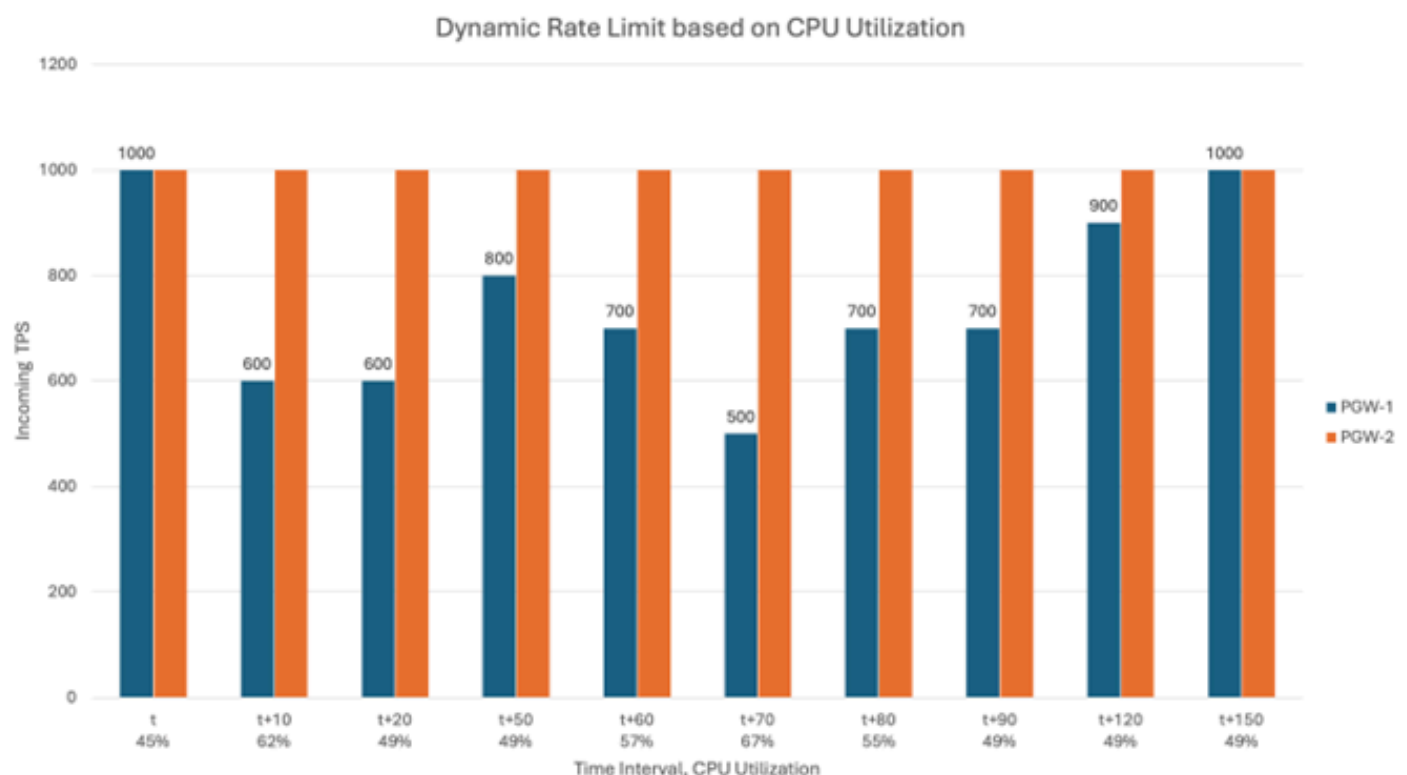
在DRA考慮此配置：

靜態郵件速率限制：1000（因此它是傳入TPS的值）

沖銷暫掛時間：30年代

沖銷步驟%:20%

每當DB CPU利用率超過閾值時，它會引用「動態限制DB CPU配置檔案」配置，並通過通知導向器相應地限制傳入TPS。由於它是基於不斷變化的CPU使用率值進行限制，因此可以將其動態速率限制流量稱為限制速率。



- 最初，資料庫CPU利用率低於限制，因此不會發生限制。此外，PGW-2沒有動態速率限制配置，因此無論CPU使用率如何，都不會出現頻寬限制。
- 當DB CPU使用率為62%時，流量會被40%限制，而有效速率限制為600（傳入的TPS為1000,DRA僅允許600）。
- 如果CPU使用率保持在60-65%之間，則繼續將40%的限制應用到配置的速率限制1000，並且有效速率限制為600（傳入TPS為1000,DRA僅允許600。）
- CPU利用率降低到49%，頻寬限制逆轉開始於pGW-1。
- 如果CPU利用率在30秒內保持在49%或更低，則限制會減少20%到20%。現在有效速率限制為800（傳入TPS為1000,DRA僅允許800。）根據配置，反轉的步驟是20%。
- 當DB CPU利用率增加到57%時，流量將節流30%，有效速率限制為700（傳入的TPS為1000,DRA僅允許700。）
- 當DB CPU利用率增加到67%時，流量將節流50%，有效速率限制為500（傳入的TPS為1000,DRA僅允許500。）
- 當DB CPU利用率降低到55%時，流量將節流30%，並且有效速率限制為700（傳入的TPS為1000,DRA僅允許700。）
- 如果在接下來的30秒間隔內CPU使用率下降至49%或更低，則限制會減少20%至10%，有效速率限制為900（傳入TPS為1000,DRA僅允許900。）
- 如果在接下來的30秒間隔內CPU進一步保持在49%或更低，則限制會減少20%到0，並且當反向完成時，不會應用速率限制（傳入的TPS為1000,DRA允許1000）。

關於此翻譯

思科已使用電腦和人工技術翻譯本文件，讓全世界的使用者能夠以自己的語言理解支援內容。請注意，即使是最佳機器翻譯，也不如專業譯者翻譯的內容準確。Cisco Systems, Inc. 對這些翻譯的準確度概不負責，並建議一律查看原始英文文件（提供連結）。