

CPS vDRA를 통한 선점적 정적 및 동적 속도 제한

목차

[소개](#)

[사전 요구 사항](#)

[요구 사항](#)

[사용되는 구성 요소](#)

[배경 정보](#)

[문제](#)

[솔루션](#)

[로드 밸런서의 고정 속도 제한](#)

[인그레스 속도 제한](#)

[이그레스 속도 제한](#)

[동적 속도 제한](#)

소개

이 문서에서는 Diameter 메시지를 라우팅하고 네트워크 트래픽을 관리하는 통신 구성 요소인 DRA의 속도 제한 옵션에 대해 설명합니다.

사전 요구 사항

요구 사항

다음 주제에 대한 지식을 보유하고 있으면 유용합니다.

- CPS(Cisco Policy Suite) vDRA(Diameter Routing Agent)
- Diameter Routing Agent 기본 및 사양

사용되는 구성 요소

이 문서의 정보는 Cisco Policy Suite DRA를 기반으로 합니다.

이 문서의 정보는 특정 랩 환경의 디바이스를 토대로 작성되었습니다. 이 문서에 사용된 모든 디바이스는 초기화된(기본) 컨피그레이션으로 시작되었습니다. 현재 네트워크가 작동 중인 경우 모든 명령의 잠재적인 영향을 미리 숙지하시기 바랍니다.

배경 정보

DRA는 통신 네트워크의 구성 요소이며, 특히 Diameter 프로토콜 기반 네트워크의 경우 그렇습니다.

. DRA는 정책 서버, 충전 시스템 및 기타 Diameter 지원 디바이스와 같은 서로 다른 네트워크 요소 간에 Diameter 메시지를 효율적으로 라우팅합니다. 속도 제한은 네트워크 요소를 오가는 트래픽의 양을 제어하는 데 사용되는 네트워크 트래픽 관리 기술입니다. 네트워크 리소스가 소진되지 않도록 하고, 서비스 품질을 유지하며, 네트워크의 오용이나 남용을 방지합니다.

문제

네트워크의 각 구성 요소는 정격 용량에 따라 트래픽 로드를 처리할 수 있지만, 실시간으로 생성되는 트래픽이 시스템에서 처리할 수 있는 것보다 많은 경우가 발생할 수 있습니다. 그 중 일부는 다음과 같습니다.

- 사용자 행동 - 단기간에 대량의 데이터를 생성하는 스트리밍 이벤트 또는 소프트웨어 업데이트와 같은 활동 일반적으로 게이트웨이(Gw)에서 DRA로 전송됩니다.
- 네트워크 혼잡 - 네트워크 사용량이 많은 시간에는 혼잡이 가중될 수 있으며, 대기 중인 데이터는 용량이 사용 가능해지면 버스트(burst)로 전송됩니다.
- 네트워크 탄력성 메커니즘 - 중단 또는 유지 관리 중에 트래픽을 다시 라우팅하여 일시적인 스파이크를 유발합니다. 이는 네트워크 문제가 없는 연결된 사이트의 트래픽 흐름에 영향을 줄 수 있습니다.
- 네트워크 요소 동작 - 오버로드 및 혼잡의 경우, 하나 이상의 네트워크 요소에서 응답/시간 초과가 표시되지 않기 시작할 수 있으며, 이로 인해 재연결이 시스템의 추가 오버로드를 초래할 수 있습니다.
- 게이트웨이 플래싱 - 게이트웨이는 정책 변경, 토폴로지 변경 또는 유지 보수 또는 문제 해결 활동으로 인해 기존 세션을 플래싱할 수 있습니다. 이러한 시나리오 세션 중에 세션이 지워지며 Gx CCR(Credit Control Request)-T 요청을 버스트할 수 있습니다.

솔루션

DRA는 요청을 효율적으로 처리하고 단일 서버의 오버로드를 방지하기 위해 여러 Diameter 서버 간에 로드를 분산할 수 있습니다. 서버 장애 시 DRA는 메시지를 대체 서버로 리디렉션하여 네트워크 서비스의 고가용성과 안정성을 보장할 수 있습니다.

DRA에 대한 속도 제한 - DRA뿐만 아니라 메시지의 제어 흐름을 보장하여 다른 엔터티를 보호합니다. 속도 제한의 주요 이점은 다음과 같습니다.

- 서비스 연속성 - 중요한 네트워크 구성 요소가 과부하 상태가 되지 않도록 보장하고 가동 중단을 방지함으로써 지속적인 서비스 가용성을 유지합니다.
- 확장성 - 성능 저하 없이 네트워크에서 다양한 부하를 처리할 수 있습니다.
- SLA(Service Level Agreements) 준수 - 일관된 성능 및 안정성 수준을 유지하여 네트워크가 SLA를 충족하도록 보장

로드 밸런서의 고정 속도 제한

이는 straight forward 접근 방식입니다. 고정 임계값은 pGW(DRA/Packet Gateway) 및 기타 네트워크 요소의 정격 용량에 따라 설정되며 네트워크 조건이나 시스템 리소스에 따라 변경되지 않습니다. 수신 요청 속도를 제한하면 DRA에서 처리하는 트래픽 양에 대해 예측 가능한 결과를 얻을 수 있습니다.

고정 속도 제한에 대한 컨피그레이션은 적용되는 활용 사례에 따라 달라집니다.

인그레스 속도 제한

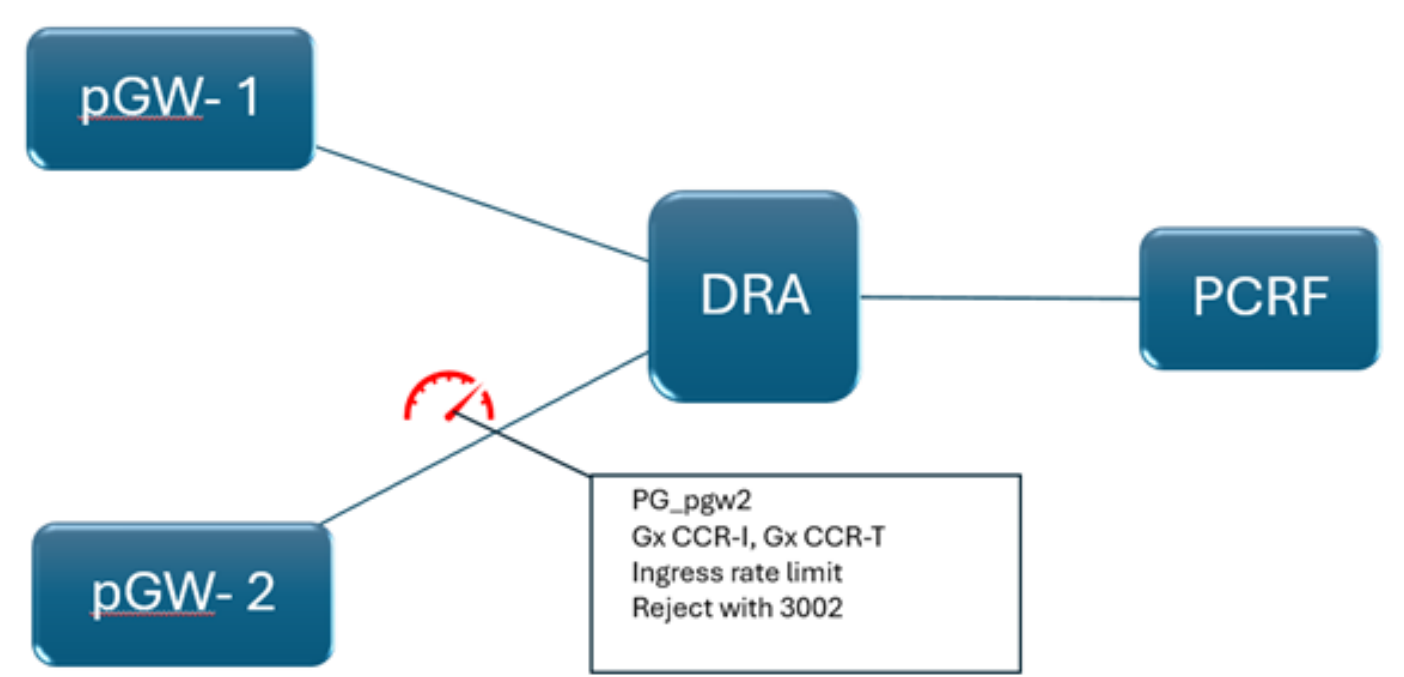
시나리오: pGW에서 버스트

이러한 트래픽 버스트에 영향을 받기 쉬운 pGW에 특정한 임계값이 구성됩니다. 이 값은 이러한 버스트 중에 표시되는 정규 트래픽/피크 트래픽 수를 기반으로 도달해야 합니다.

GW의 Gx CCR-I 및 Gx CCR-T 요청만 제한해야 하지만 Gx CCR-U 트래픽 또는 Gy 트래픽은 수신된 대로 전달되어야 하는 등 버스트 트래픽만 제한하도록 각 메시지 유형에 따라 임계값 번호를 정의할 수 있습니다.

이 경우 인그레스 측에서 스로틀(throttle)할 수 있습니다. 즉, DRA는 메시지를 수신하는 즉시 메시지를 스로틀(throttle)할 수 있습니다. 이 경우 요청을 수신하는 네트워크 요소를 기준으로 메시지를 거부하고 DRA가 처리할 수 있는 것보다 많은 수의 요청을 처리하지 않도록 하기 위한 목적이 있습니다.

스로틀 동작은 특정 Error-Code 및 Error-Message가 있는 메시지를 거부하거나 삭제하는 것일 수 있습니다.



이 동작은 CPS vDRA에서 CRD(사용자 지정 참조 데이터) 테이블 '피어 속도 제한 프로필' 및 '메시지 속도 제한 프로필'을 구성하여 활성화할 수 있습니다. 이러한 CRD 테이블에서 다음 값을 구성해야 합니다.

피어 그룹	피어 그룹은 영역 및 호스트를 기반으로 하는 Diameter 노드의 논리적 그룹입니다. 조절해야 하는 피어 그룹을 구성해야 합니다.
-------	---

피어 FQDN(정규화된 도메인 이름)	속도 제한이 필요한 피어 그룹의 피어에 대한 FQDN(정확한 또는 regex 일치).
메시지 방향	스로틀링의 방향 - 인그레스 또는 이그레스 이 경우 - 인그레스(Ingress).
Rate Limit 프로파일	제한해야 하는 메시지 유형을 정의하는 데 사용된 메시지 속도 제한 프로파일 이름입니다.
피어 속도 제한	이 피어 그룹에 허용된 요청 비율입니다. 여기에는 해당 피어 그룹의 모든 메시지 유형이 포함됩니다.
버리기 동작	요청을 삭제하거나 오류 코드로 거부하도록 선택할 수 있습니다.
결과 코드	메시지를 거부하는 경우의 결과 코드 값입니다. 메시지가 삭제되는 경우에는 적용되지 않습니다.
오류 문자열	거부된 요청의 응답 메시지에 사용되는 오류 문자열입니다. 메시지가 삭제되는 경우에는 적용되지 않습니다.
애플리케이션 식별자	조절할 메시지의 애플리케이션 ID입니다.
명령 코드	조절할 메시지의 명령 코드입니다.
메시지/요청 유형	제한해야 하는 요청의 애플리케이션 ID 및 요청 유형입니다.
메시지 속도 제한	DRA에서 처리하는 해당 메시지 유형의 요청에 대한 TPS(TelePresence Server)입니다. 이 TPS를 초과하는 요청은 제한됩니다. 이 값은 피어 그룹의 피어당입니다.

여기에서는 pGW2에 대한 전체 속도 제한 1000개 메시지, 속도 제한 200Gx CCR-I 및 200Gx CCR-T를 구성하는 예를 보여줍니다. 이 속도를 초과하는 요청은 3002로 거부되고 속도 제한이 위반되었음을 나타내는 오류 메시지가 표시됩니다.

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- gwgw.com	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

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

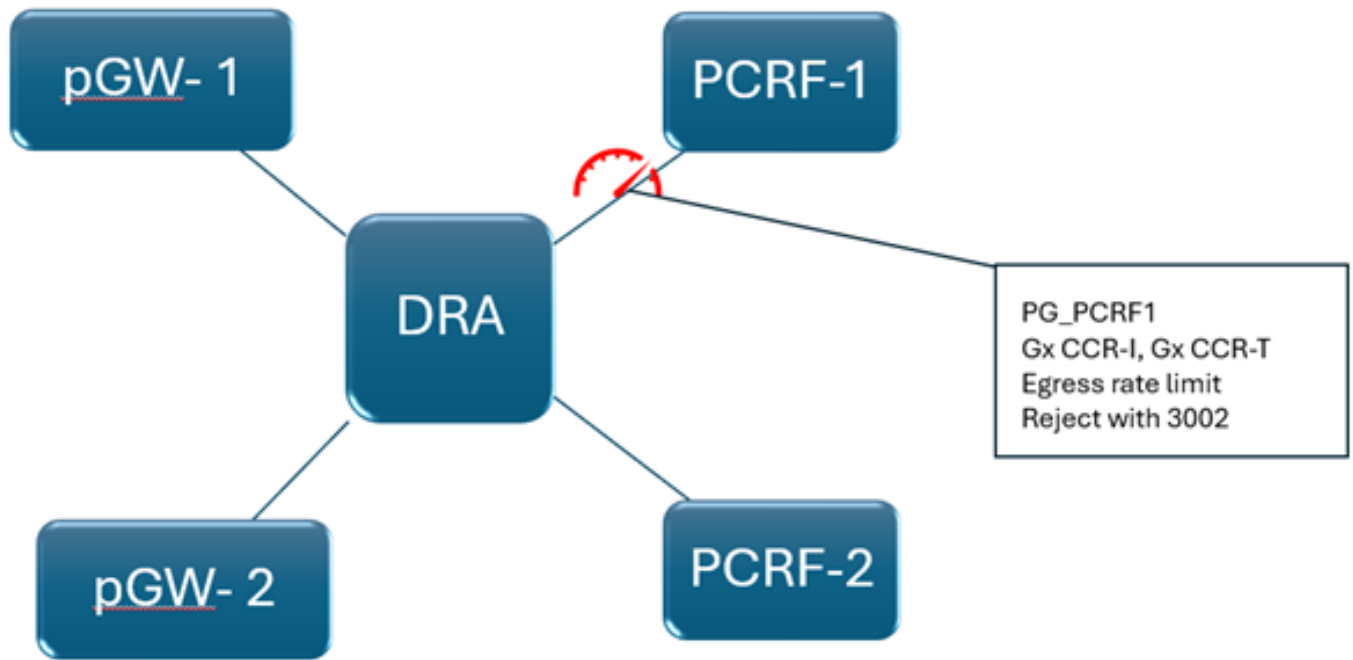
이그레스 속도 제한

시나리오: 요청을 받는 네트워크 요소의 보호

pGW로부터 요청이 수신되고 Policy and Charging Rules Function (PCRF)으로 전달되는 Gx 트랜잭션의 예를 고려하십시오. PCRF가 처리할 수 있는 데이터 양에 제한이 있는 경우, DRA가 수신 트래픽을 처리할 수 있더라도 DRA를 사용하여 요청을 PCRF에 전달하고 오버로드하는 대신 DRA에서 메시지를 조절할 수 있습니다.

여기서 이그레스 측에서 스로틀링해야 합니다. 즉, DRA는 DRA 라우팅 논리에 기초하여 식별된 PCRF 피어 그룹에 기초하여 메시지가 PCRF로 전달되기 직전에 메시지를 스로틀링합니다.

스로틀 동작은 특정 Error-Code 및 Error-Message가 있는 메시지를 거부하거나 삭제하는 것일 수 있습니다.



이 동작은 CPS vDRA에서 CRD 테이블 'Peer Rate Limit Profile'(피어 속도 제한 프로파일) 및 'Message Rate Limit Profile'(메시지 속도 제한 프로파일)을 구성하여 활성화할 수 있습니다. 이러한 CRD 테이블에서 다음 값을 구성해야 합니다.

피어 그룹	피어 그룹은 영역 및 호스트를 기반으로 하는 Diameter 노드의 논리적 그룹입니다. 조절해야 하는 피어 그룹을 구성해야 합니다.
피어 FQDN	속도 제한이 필요한 피어 그룹의 피어에 대한 FQDN(정확한 또는 regex 일치).
메시지 방향	스로틀링의 방향 - 인그레스 또는 이그레스 이 경우 - 이그레스(Egress).
Rate Limit 프로파일	제한해야 하는 메시지 유형을 정의하는 데 사용된 메시지 속도 제한 프로파일 이름입니다.
피어 속도 제한	이 피어 그룹에 대해 허용되는 요청 비율입니다. 여기에는 해당 피어 그룹의 모든 메시지 유형이 포함됩니다.
버리기 동작	요청을 삭제하거나 오류 코드로 거부하도록 선택할 수 있습니다.
결과 코드	메시지를 거부하는 경우의 결과 코드 값입니다. 메시지가 삭

	제되는 경우에는 적용되지 않습니다.
오류 문자열	거부된 요청의 응답 메시지에 사용되는 오류 문자열입니다. 메시지가 삭제되는 경우에는 적용되지 않습니다.
애플리케이션 식별자	조절할 메시지의 애플리케이션 ID입니다.
명령 코드	조절할 메시지의 명령 코드입니다.
메시지/요청 유형	제한해야 하는 요청의 애플리케이션 ID 및 요청 유형입니다.
메시지 속도 제한	DRA에서 처리하는 해당 메시지 유형의 요청 TPS입니다. 이 TPS를 초과하는 요청은 제한됩니다. 이 값은 피어 그룹의 피어당 입니다.

Peer Rate Limit Profile

Filter by

All Visible Columns

CCR_I_T

Q Peer Group *

Q Peer FQDN *

Q 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

Q Rate Limit Profile Name *

Q Application Identifier *

Q Command Code *

Q 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

시나리오: 네트워크 속도가 느려 트래픽 정체가 발생하고 DRA에 전체/부분 장애 발생

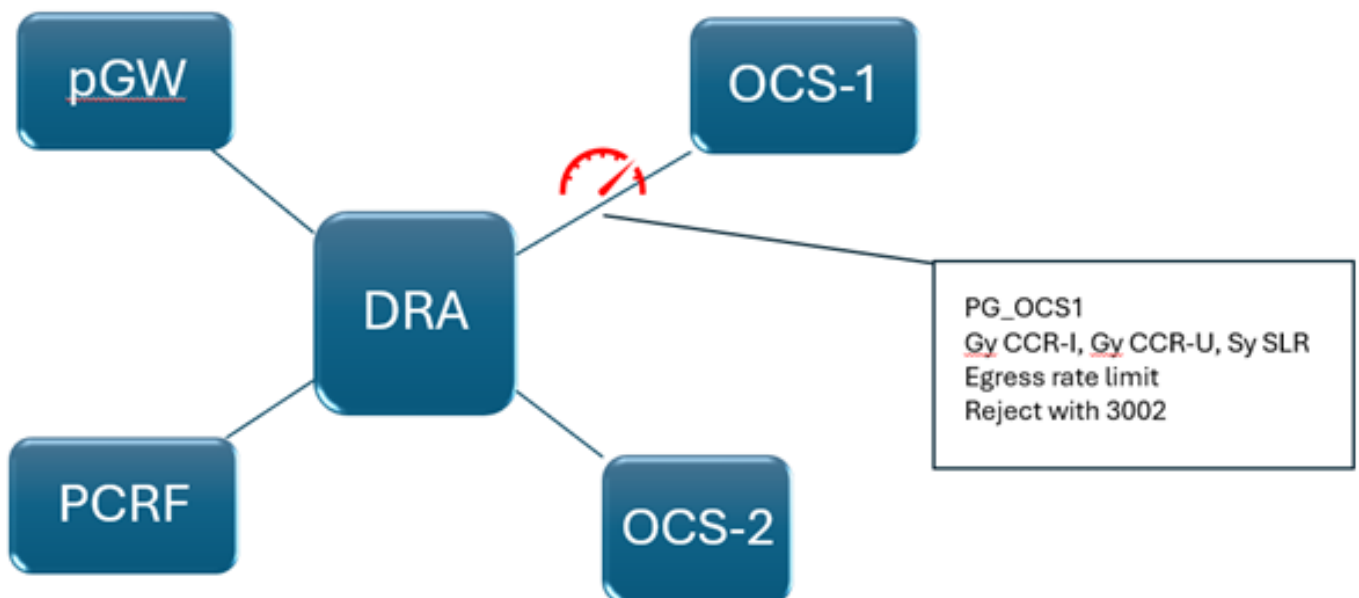
pGW와 OCS(Online Charging System) 간에 교환되는 Gy 트랜잭션의 예를 고려하십시오. DRA-OCS 채널의 네트워크가 느려지는 경우(pGW의 트래픽이 높거나 기타 네트워크 문제로 인해), SLA 위반으로 인해 요청이 시간 초과됩니다. 이러한 시간 제한은 DRA뿐만 아니라 전체 네트워크에 영향을 미칩니다.

DRA 리소스는 느린 네트워크를 통해 OCS에 요청을 보내려고 하다가 지연되어 리소스가 모두 소모됩니다. 따라서 DRA의 정격 용량이 위반되지 않더라도 여러 요청이 DRA에 의해 거부됩니다.

이는 DRA-OCS 채널에 없는 트래픽에도 영향을 줍니다. 이러한 거부/시간 초과 및 삭제는 여러 네트워크 요소에 대한 재연결을 트리거합니다.

이 경우 이그레스 측에서 스로틀해야 합니다. DRA는 용량 제한 또는 네트워크 문제가 있는 OCS 피어 그룹을 기준으로 메시지가 OCS로 전달되기 직전에 메시지를 스로틀합니다.

스로틀 동작은 특정 Error-Code 및 Error-Message가 있는 메시지를 거부하거나 삭제하는 것일 수 있습니다.



이 동작은 CPS vDRA에서 CRD 테이블 'Peer Rate Limit Profile'(피어 속도 제한 프로파일) 및 'Message Rate Limit Profile'(메시지 속도 제한 프로파일)을 구성하여 활성화할 수 있습니다. 이러한 CRD 테이블에서 다음 값을 구성해야 합니다.

피어 그룹	피어 그룹은 영역 및 호스트를 기반으로 하는 Diameter 노드의 논리적 그룹입니다. 조절해야 하는 피어 그룹을 구성해야 합니다.
피어 FQDN	속도 제한이 필요한 피어 그룹의 피어에 대한 FQDN(정확한 또는 regex 일치).

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

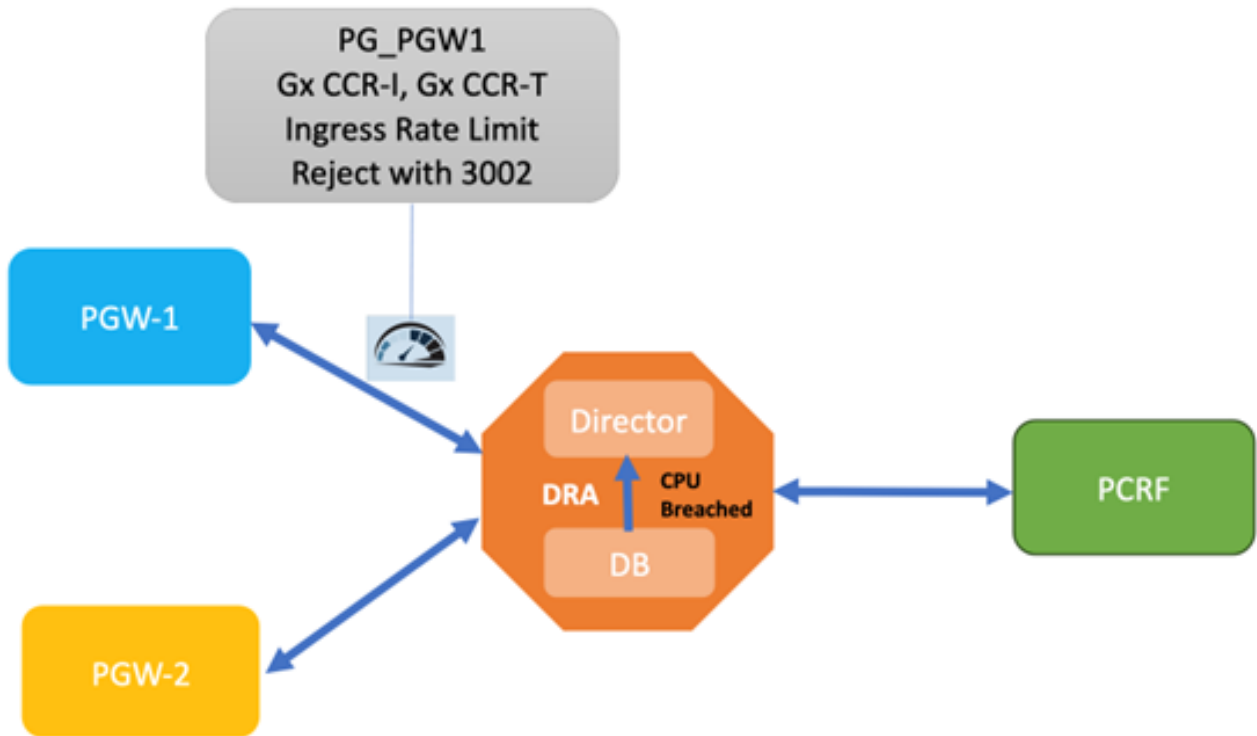
Show 50 rows 1 out of 1

동적 속도 제한

CCR-I 또는 CCR-T 버스트가 발생하면 데이터베이스(DB)에 오버로드가 발생하여 시스템이 불안정해질 수 있습니다. 이를 극복하기 위해 DRA는 사용 가능한 DB 용량을 기반으로 동적 속도 제한(Gx CCR-I 및 Gx CCR-T에만 해당)을 지원합니다.

DRA는 DB CPU 사용률을 모니터링하며, 임계값에 도달할 때마다 수신 요청을 제한합니다. 스로틀링에 대한 CPU 임계값과 스로틀링할 수신 트래픽을 구성할 수 있습니다.

해당 스로틀 백분율과 함께 서로 다른 CPU 임계값을 구성할 수 있습니다. DRA는 현재 DB CPU 사용률을 기반으로 조절 레벨을 조정합니다. CPU 사용률이 안정적이면 스로틀링이 점진적으로 중지됩니다.



이 동작은 CPS vDRA에서 CRD 테이블 'Peer Rate Limit Profile', 'Message Rate Limit Profile', 'Dynamic Peer Rate Limit Profile' 및 'Dynamic Throttling DB CPU Profile'을 구성하여 활성화할 수 있습니다. 이러한 CRD 테이블에서 다음 값을 구성해야 합니다.

피어 그룹	피어 그룹은 영역 및 호스트를 기반으로 하는 Diameter 노드의 논리적 그룹입니다. 이 예에서는 pGW의 피어 그룹을 구성합니다.
피어 FQDN	속도 제한이 필요한 피어 그룹의 피어에 대한 FQDN(정확한 또는 regex 일치).
메시지/요청 유형	제한해야 하는 요청의 애플리케이션 ID 및 요청 유형입니다. 이 예에서는 Gx CCR-I, Gx CCR-T를 보여 줍니다.
메시지 방향	스로틀링의 방향 - 인그레스 또는 이그레스 이 경우 - 인그레스(Ingress).
속도 제한 프로파일	제한해야 하는 메시지 유형을 정의하는 데 사용된 메시지 속도 제한 프로파일 이름입니다.

피어 속도 제한	이 피어 그룹에 허용된 요청 비율입니다. 여기에는 해당 피어 그룹의 모든 메시지 유형이 포함됩니다.
버리기 동작	요청을 삭제하거나 오류 코드로 거부하도록 선택할 수 있습니다.
결과 코드	메시지를 거부하는 경우의 결과 코드 값입니다. 메시지가 삭제되는 경우에는 적용되지 않습니다.
오류 문자열	거부된 요청의 응답 메시지에 사용되는 오류 문자열입니다. 메시지가 삭제되는 경우에는 적용되지 않습니다.
애플리케이션 식별자	조절할 메시지의 애플리케이션 ID입니다.
명령 코드	조절할 메시지의 명령 코드입니다.
메시지 속도 제한	DRA에서 처리하는 해당 메시지 유형의 요청 TPS입니다. 이 TPS를 초과하는 요청은 제한됩니다. 이 값은 피어 그룹의 피어당입니다.
동적 조절 DB CPU 프로파일	이는 CPU 프로파일 이름을 나타내며, 서로 다른 CPU 범위에 대한 스로틀 백분율을 정의하는 데 사용됩니다.
DB CPU 사용률 임계값	구축의 트래픽 패턴에 따라 보안 침해 제한으로 구성된 CPU 레벨의 올바른 값을 선택할 수 있습니다.
조절 비율	해당 CPU 수준이 침해될 때 적용되는 비율 제한의 백분율.

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	

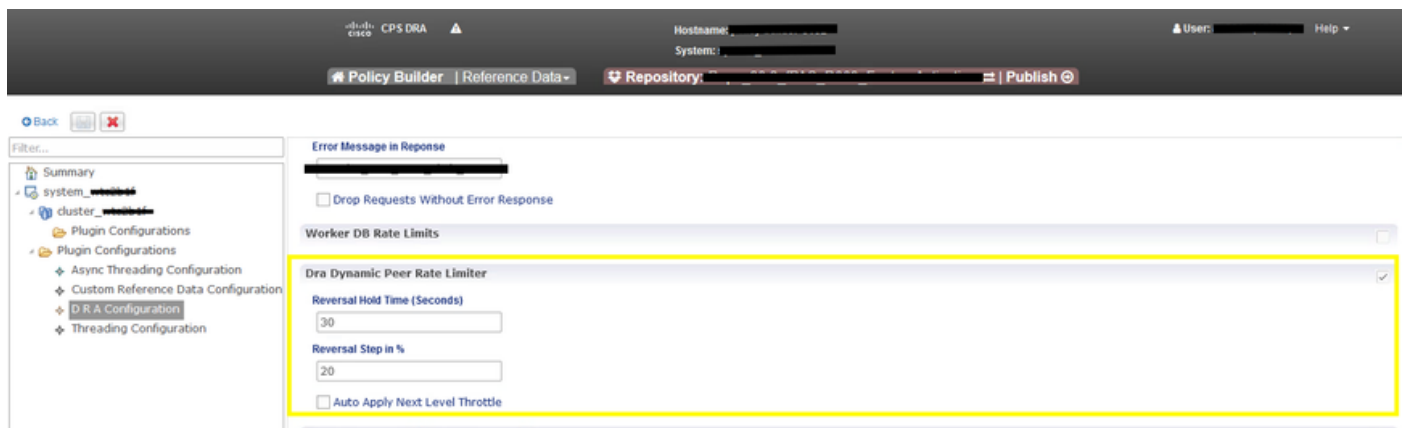
Showing 4 out of 4

Show 50 rows 1 out of 1

또한 정책 구성기의 'DRA Configuration Plugin' 아래의 'DRA Dynamic Peer Rate Limiter' 섹션 아래에서 확인란을 선택하여 이 동작을 활성화해야 합니다.

취소 보류 시간 - 취소가 적용되기 전에 CPU 사용률이 모니터링되는 기간입니다.

취소 단계(%) - 취소된 조절 백분율입니다.



시나리오: CPU 사용률을 기반으로 한 동적 속도 제한

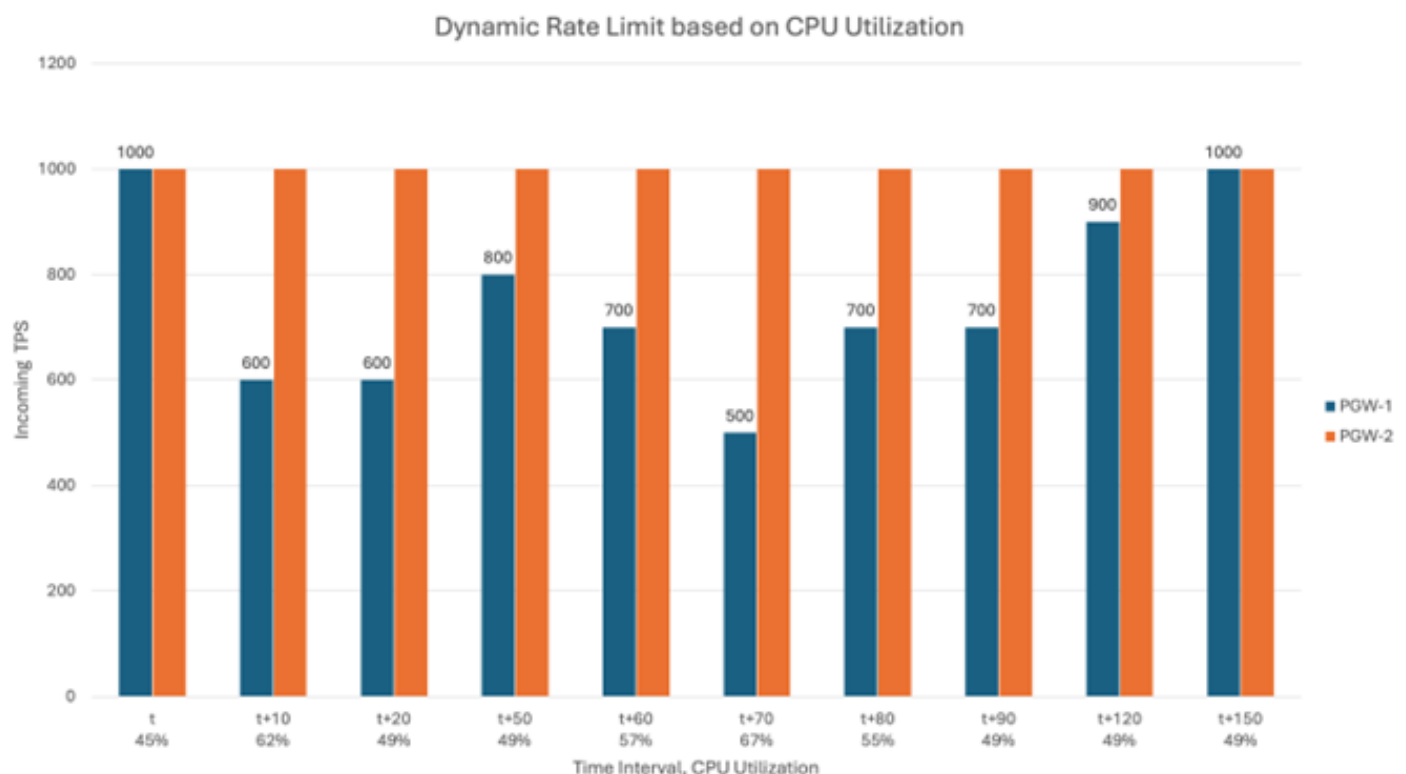
DRA에서 이 컨피그레이션을 고려하십시오.

정적 메시지 속도 제한: 1000(따라서 수신 TPS의 값)

취소 보류 시간: 30년대

취소 단계(%): 20%

DB CPU Utilization(DB CPU 사용률)이 임계값을 초과할 때마다 'Dynamic Throttling DB CPU Profile(동적 조절 DB CPU 프로파일)' 컨피그레이션을 참조하고, 이를 디렉터에 알림으로써 그에 따라 수신 TPS를 조절합니다. 끊임없이 변화하는 CPU 사용률 값을 기반으로 조절(throttling)되므로 트래픽 제한을 동적으로 속도 조절한다고 말할 수 있습니다.



- 초기에는 DB CPU 사용률이 한계에 도달하므로 제한이 발생하지 않습니다. 또한 PGW-2에는 Dynamic Rate Limiting의 컨피그레이션이 없으므로 CPU 사용률과 상관없이 throttling이 발생하지 않습니다.
- DB CPU 사용률이 62%이면 트래픽이 40% 감소하고 유효 속도 제한이 600입니다(수신 TPS는 1000이고 DRA는 600만 허용함).
- CPU 사용률이 60-65% 사이를 유지할 경우 구성된 속도 제한인 1000에 40%의 속도 제한이 계속 적용되고 유효 속도 제한은 600입니다(수신 TPS는 1000이고 DRA는 600만 허용함).
- CPU 사용률이 49%로 감소하며, 속도 제한의 역전이 pGW-1부터 시작됩니다.
- CPU 사용률이 30초 동안 49% 이하를 유지할 경우 스로틀링이 20%에서 20%로 감소합니다. 현재 유효 속도 제한은 800입니다(수신 TPS는 1000이고, DRA는 800만 허용함). 반면 컨피그레이션에 따라 20% 단계로 수행됩니다.
- DB CPU 사용률이 57%로 증가하면 트래픽은 30% 감소하고 유효 속도 제한은 700입니다(수신 TPS는 1000이고 DRA는 700만 허용함).
- DB CPU 사용률이 67%로 증가하면 트래픽은 50% 감소하고 유효 속도 제한은 500입니다(수신 TPS는 1000이고 DRA는 500만 허용함).
- DB CPU 사용률이 55%로 감소하면 트래픽이 30% 감소하고 유효 속도 제한이 700입니다(수신 TPS는 1000이고 DRA는 700만 허용함).
- 다음 30초 간격 동안 CPU가 49% 이하로 떨어질 경우 스로틀링이 20%에서 10%로 감소하고 유효 속도 제한이 900입니다(수신 TPS는 1000이고 DRA는 900만 허용함).
- CPU가 다음 30초 간격 동안 49% 이하의 상태를 유지하면 스로틀링이 0으로 20% 감소하며, 전환이 완료될 때 속도 제한이 적용되지 않습니다(수신 TPS는 1000이고 DRA는 1000을 허용함).

이 번역에 관하여

Cisco는 전 세계 사용자에게 다양한 언어로 지원 콘텐츠를 제공하기 위해 기계 번역 기술과 수작업 번역을 병행하여 이 문서를 번역했습니다. 아무리 품질이 높은 기계 번역이라도 전문 번역가의 번역 결과물만큼 정확하지는 않습니다. Cisco Systems, Inc.는 이 같은 번역에 대해 어떠한 책임도 지지 않으며 항상 원본 영문 문서(링크 제공됨)를 참조할 것을 권장합니다.