

ASR5x00 MME Überlastungsschutz - Funktionen

Inhalt

[Einführung](#)

[MME-Schutz](#)

[Schutz vor Netzwerküberlastungen: Drosselung der Attach-Rate](#)

[Schutz vor Netzwerküberlastungen: Paging-Throttling](#)

[Beispielkonfiguration](#)

[Schutz vor Netzwerküberlastungen: DDN Throttling \(Serving GateWay Funcality, Protected MME\)](#)

[Schutz vor Netzwerküberlastungen: Fehlerdrosselung beim EGTP-Pfad](#)

[Beispielkonfiguration](#)

[Verbesserte Überlastungskontrolle](#)

[Grenzwerte für Überlastungsbedingungen](#)

[Grenzwerte und Toleranzwerte](#)

[Servicesteuerung CPU-Grenzwerte](#)

[System-CPU-Grenzwerte](#)

[Grenzwerte des Systemspeichers](#)

[Zugehörige Informationen](#)

Einführung

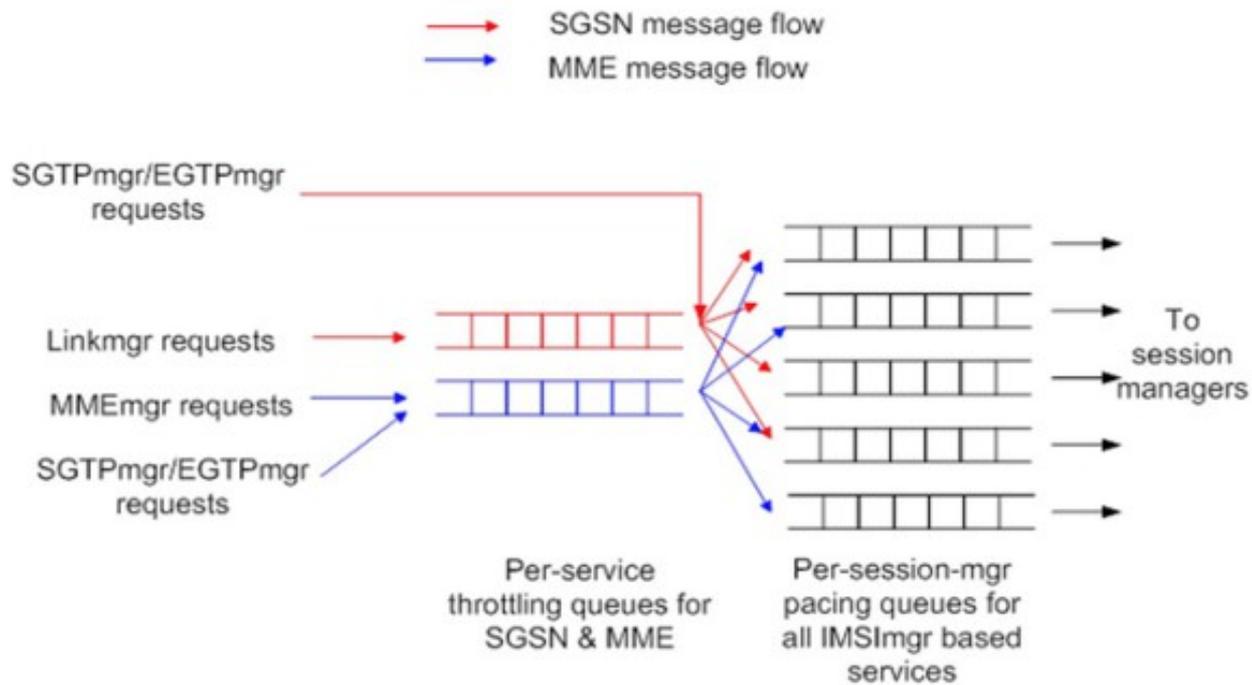
In diesem Dokument werden die verschiedenen Überlastungsschutzmethoden und -funktionen der Mobility Management Entity (MME) für die Cisco Aggregation Services Router (ASR) der Serie 5000 vorgestellt. In der Serie ASR 5000 bietet Cisco dem Kunden verschiedene Möglichkeiten zur Steuerung. In diesem Artikel werden die Funktionen und zugehörigen CLI-Befehle erläutert.

MME-Schutz

Schutz vor Netzwerküberlastungen: Drosselung der Attach-Rate

Die Attach Rate Throttling schützt benachbarte Netzwerkelemente wie Home Subscriber Server (HSS), Policy and Charging Rules Function (PCRF) und Online Charging Server (OCS) sowie interne MME-Ressourcen wie `imsimgr` und `sessmgr`. Attach Rate Throttling verarbeitet die neuen Anrufe, die an `imsimgr` eingehen, z. B. bei Attachs und Inter-MME/Serving GPRS Supporting Node (SGSN) Tracking Area Update (TAUs).

Dieses Bild zeigt den Nachrichtenfluss für Anrufe und Warteschlangen zum Beenden.



Um die MME zu schützen (imsimgr und sessmgr onwards), sollten die *Drosselungsrate*, die *Wartezeit in der Warteschlange* und die *Zeit in der Warteschlangengröße* definiert werden. Die *Drosselungsrate* hängt vom MME-Anrufmodell ab, da die MME-Kapazität vom Anrufmodell abhängt.

Bei MME ist die Berechnung der Drosselungsrate relativ einfach. Verwenden Sie dazu die standardmäßigen Call Events Per Second (CEPS) im Netzwerk plus Toleranz. Möglicherweise müssen Sie auch die HSS-Datenbankkapazität berücksichtigen, wenn HSS-Schutz erforderlich ist.

Beispiel

In Spitzenzeiten verarbeitet die MME bis zu 170 bis 200 Anrufe pro Sekunde (Attaches+ Inter-TAU). Bei Ausfall eines Standorts werden bis zu 350 bis 370 Anrufe pro Sekunde an eine MME weitergeleitet. Bei dieser Anrufrate steigt die MME-Auslastung um fast 80 %, und 400 Anrufe pro Sekunde sind ein optimales Niveau, um die Drosselungsrate zu begrenzen, um eine übermäßige Signallast innerhalb der MME-Box zu vermeiden.

Die *Wartezeit in der Warteschlange* beträgt standardmäßig 5 Sekunden. Es ist optimal für KUNDEN. Die *Warteschlangengröße* ist standardmäßig 2500. Es ist optimal für KUNDEN.

Der Konfigurationsbefehl lautet wie folgt.

```
asr5k(config)#network-overload-protection mme-new-connections-per-second
new_connections action attach { drop | reject-with-emm-cause
{ congestion | network-failure | no-suitable-cell-in-tracking-area}
tau { drop | reject-with-emm-cause { congestion | network-failure
| no-suitable-cells-in-tracking-area | no-sec-ctxt-in-nw} fwd-reloc
{ drop | reject} }{wait-time <wait-time>} {queue-size <queue-size>}
```

Neue_Verbindungen

Definiert die Anzahl der neuen MME-Verbindungen, die pro Sekunde akzeptiert werden. Muss eine ganze Zahl zwischen 50 und 5000 sein. Der Standardwert ist 500.

Aktion

Definiert die Aktion, die ausgeführt wird, wenn die Warteschlange voll ist. Wenn neue Verbindungen an der MME empfangen werden, werden sie in die Warteschlange für das Paging gestellt, und die Nachrichten aus der Warteschlange werden mit der konfigurierten Geschwindigkeit verarbeitet. Wenn die Warteschlange (aufgrund einer hohen eingehenden Rate) aufgrund der konfigurierten Aktion überlastet wird, werden Pakete entweder verworfen oder abgelehnt.

Warteschlangengröße

Definiert die maximale Größe der Paging-Warteschlange, die zum Puffern der Pakete verwendet wird. Muss eine ganze Zahl zwischen 250 und 25.000 sein. Der Standardwert ist 2500.

Beispielkonfiguration

```
network-overload-protection mme-new-connections-per-second 400 action attach  
reject-with-cause no-suitable-cell-in-tracking-area tau  
reject-with-cause no-suitable-cell-in-tracking-area fwd-reloc drop
```

Jetzt ist die Anrufrate pro Sekunde auf 400 festgelegt, und die Aktion ist intelligent Ablehnen, mit der Ursache 15, dass die Benutzergeräte (UE) wieder eine Verbindung zu verschiedenen RATs herstellen. Die Wartezeit ist auf den Standardwert (5 Sekunden) und die Warteschlangengröße auf 2500 festgelegt.

Hinweis: Die Aktion "Ablehnen" mit der EMM-Ursache Nr. 15 "No-usable-cell-in-tracking-area" wird bevorzugt, da die mit Nr. 15 abgelehnten Anrufe meist nicht wieder bei MME eintreffen und auf verschiedene RAT-Ebenen (3G, 2G) übertragen werden. Die Aktion "Drop" für die Umverteilung des SRNS-Subsystems (Serving Radio Network Subsystem) dient der zukünftigen Verwendung und verhindert eine schnelle Neuanschlusssuche an MME nach der Ablehnung.

Schutz vor Netzwerküberlastungen: Paging-Throttling

Paging Throttling schützt interne MME-Ressourcen (Memgr) als eNodeB-/Funkressourcen (falls erforderlich). Dieser Grenzwert für die Durchsatzbegrenzung gilt für alle eNodeB, die mit MME für ein bestimmtes ASR 5000-Chassis verknüpft sind. S1 Paging-Anfragen an eNodeB werden bei diesem Schwellenwert begrenzt. S1 Paging-Anfragen an einen eNodeB, die diesen Grenzwert überschreiten, werden verworfen.

Bei MME ist die Berechnung der Drosselungsrate relativ einfach: Verwenden Sie die standardmäßige Ausgangs-Paging-Rate im Netzwerk plus die Toleranz. (Dies basiert ausschließlich auf der Entscheidung des Designteams.)

Beispiel

In Spitzenzeiten verarbeitet jede MME bis zu 35.000 Paging-Nachrichten pro Sekunde. Bei Ausfall eines Standorts können bis zu 70.000 Seiten pro Sekunde von einem MME übertragen werden. Bei dieser Paging-Rate wäre die MME-Nutzung (Memgr) fast 80 % und die *70000 bis 80000* Seiten pro Sekunde ein optimales Niveau, um die Drosselungsrate zu begrenzen, um eine übermäßige S1-Signalisierung über MGR zu vermeiden.

Die Rate ist jedoch pro durchschnittlichem eNodeB begrenzt. Die durchschnittliche Rate pro

eNodeB (bei 6500 eNodeB) beträgt 10 Seiten pro Sekunde. Die Tracking Areas (TAs) sind jedoch in der Anzahl der Teilnehmer nicht gleich, und verschiedene TA/Mitglieder eNodeB werden unterschiedlich mit Paging geladen. Bei einer doppelten Differenz zwischen der TA-Größe und der durchschnittlichen Anzahl von Abonnenten pro TA beträgt die Rate pro eNodeB 20. Bei einer 20-fachen Differenz zwischen der TA-Größe und der durchschnittlichen Anzahl von Abonnenten pro TA würde die Rate pro eNodeB 200 betragen. Dies bedeutet, dass die Funktion am effizientesten ist, wenn TA (in der Anzahl der Teilnehmer) gleichmäßig geladen wird.

Eine weitere Aktion, die parallel durchgeführt werden sollte, ist die Aktivierung des Intelligent Paging. Weitere Informationen finden Sie im Abschnitt "TAI mgmt db and LTE Paging" im ASR 5000 MME Administration Guide.

Der Konfigurationsbefehl lautet wie folgt:

```
asr5000(config)# network-overload-protection mme-tx-msg-rate-control enb s1-paging
```

- Netzwerküberlastungsschutz zur Identifizierung von Netzwerküberlastungsschutz
- ME-tx-msg-rate-control enb identifiziert die MME-Nachrichtenratenkontrolle pro durchschnittlichem eNodeB
- s1-Paging identifiziert die Nachrichtenratenkontrolle für S1-Paging.
- <rate> gibt einen Durchsatzschwellenwert in Nachrichten pro Sekunde pro eNodeB an - Bereich (1 bis 65.535)

Beispielkonfiguration

```
network-overload-protection mme-tx-msg-rate-control enb s1-paging 200
```

Hinweise:

- Die Ratenbeschränkung ist Gegenstand einer weiteren Abstimmung in absteigender Richtung. Die Grundlage für die Abstimmung ist die Anzahl der Teilnehmer (Anzahl der Paging) über TAs (Statistiken auf TA-Ebene sind erforderlich).

- Diese Funktion wird am effizientesten, wenn TAs (in der Anzahl der Abonnenten/Paging pro TA) gleichmäßig geladen werden.

Schutz vor Netzwerküberlastungen: DDN Throttling (Serving GateWay Funcality, Protected MME)

Die Downlink Data Notification (DDN)-Drosselung ist eine Funktion, mit der die Geschwindigkeit von DDN-Anfragen an MME von der Serving GateWay (SGW)-Seite gesteuert wird. Sie schützt MME-Ressourcen wie Memgr und sessmgr vor DDN (d. h. Eingangs-Paging-Anfragen)-Spitzen.

Diese Funktion besteht aus zwei Teilen: einer für Rel-10-konforme MMEs und einer für Rel-10-nicht konforme MMEs:

- Legen Sie für Rel-10-konforme MMEs im SGW-Dienst das Wasserzeichen ARP (DDN Throttling Allocation and Retention Priority) fest, um die Funktion zu aktivieren.
- Bei Rel-10-nicht konformen MMEs müssen einige andere Parameter zusammen mit dem

ARP-Wasserzeichen (wie Drosselungsfaktor, Drosselungszeit, Stabilisierungszeit, Abfrageintervall usw.) im SGW-Dienst festgelegt werden.

Wenn diese Funktion auf dem SGW aktiviert ist, sendet sie ein ARP-Wasserzeichen im DDN Req an MME. Als Antwort sendet MME eine Throttling Delay Unit, Throttling Delay Value und Throttling Factor. Die Kombination aus Delay Value and Delay Unit berechnet die Throttling Time. Nach Erhalt dieser Werte verwirft das SGW DDN Req für einen bestimmten ARP, bis der Throttling-Timer abläuft.

Bei nicht Rel-10-konformen MMEs, die die lokale Konfiguration verwenden, wird DDN Req mit einem bestimmten ARP-Wasserzeichen gedrosselt.

Die Cisco ASR5x00 MME-Versionen 16 und 17 unterstützen keine automatische DDN-Drosselung, sodass sie hinsichtlich der DDN-Drosselung als nicht Rel 10-konform fungieren.

Hinweis: Die DDN-Drosselung bietet eine weitere Detailgenauigkeit auf der MME-Paging-Drosselung (S11) statt auf der Ausgangs-Seite (S1). Bei Cisco *müssen* Sie die DDN-Drosselung nicht implementieren, wenn die Paging-Drosselung konfiguriert ist. Sie ermöglicht jedoch eine frühere Erkennung und Beseitigung von Überlastungen.

Technische Spezifikationen (TS) 23.401, Referenz für MME:

Drosselung von DDN-Anfragen

Unter ungewöhnlichen Umständen (z. B. wenn die MME-Last einen vom Operator konfigurierten Grenzwert überschreitet) kann die MME die Signalisierungslast einschränken, die ihre SGWs darauf generieren, wenn dies konfiguriert ist.

Die MME kann DDN-Anfragen für Datenverkehr mit niedriger Priorität für UEs im Inaktivitätsmodus ablehnen oder die MME weiter entlasten. Die MME kann die SGWs auffordern, die Anzahl der DDN-Anfragen, die sie für Downlink-Datenverkehr mit niedriger Priorität, der für UEs im Inaktivitätsmodus empfangen wird, gemäß einem Drosselungsfaktor und einer in der DDN-Ack-Nachricht angegebenen Drosselungsverzögerung selektiv zu reduzieren.

Der SGW bestimmt anhand der ARP-Prioritätsstufe und der Operatorrichtlinie des Inhabers, ob ein Träger für Datenverkehr mit niedriger Priorität zuständig ist oder nicht (d. h. die Konfiguration der ARP-Prioritätsstufen im SGW als Prioritäts- oder Nicht-Prioritätsdatenverkehr gilt). Die MME bestimmt, ob eine DDN-Anforderung für Datenverkehr mit niedriger Priorität oder nicht, basierend auf der ARP-Prioritätsebene, die von der SGW- und Operatorrichtlinie empfangen wurde.

Wenn die Inaktivitätsreduzierung (ISR) für die UE nicht aktiv ist, verwirft das SGW während der Drosselungsverzögerung Downlink-Pakete, die auf allen Trägern mit niedriger Priorität für UEs empfangen wurden, die als nicht mit der Benutzerebene verbunden bezeichnet werden (d. h. die SGW-Kontextdaten weisen auf keinen TEID (Downlink User Plane Tunnel End Identifier) hin, der von diesem MME im Verhältnis zum Drosselungsfaktor bereitgestellt wird, und sendet eine DDN-Nachricht nur für die nicht gedrosselten Träger an die MME übertragen werden.

Wenn der ISR während der Drosselungsverzögerung für den UE-Router aktiv ist, sendet der SGW kein DDN an den MME und sendet nur den DDN an den SGSN. Wenn sowohl MME- als auch SGSN-Anforderungslasten verringert werden, verwirft das SGW Downlink-Pakete, die auf allen Trägern mit niedriger Priorität für UEs empfangen wurden, die als nicht mit der Benutzerebene verbunden bezeichnet werden (d. h., die SGW-Kontextdaten weisen im Verhältnis zu den

Drosselungsfaktoren auf keine TEID der Benutzerebene für Downlink hin).

Der SGW nimmt nach Ablauf der Drosselungsverzögerung den normalen Betrieb wieder auf. Der zuletzt erhaltene Wert des Drosselungsfaktors und der Drosselungsverzögerung setzt alle zuvor von dieser MME empfangenen Werte außer Kraft. Beim Empfang einer Drosselungsverzögerung wird der diesem MME zugeordnete SGW-Timer neu gestartet.

Für SGW und MME ist die Berechnung der Drosselungsrate relativ einfach. Nehmen Sie die maximal zulässige *Eingangs*-Paging-Rate von 1100 Nachrichten pro Sekunde pro MME-Box.

Die Konfigurationsbefehle lauten wie folgt:

```
#configure
```

```
#context saegw-gtp
```

```
#sgw-service sgw-svc
```

```
#ddn throttle arp-watermark <arp_value> rate-limit <limit> time-factor <seconds>  
throttle-factor <percent> increment-factor <percent> poll-interval <second>  
throttle-time-sec <seconds> throttle-time-min <minutes> throttle-time-hour <hour>  
stab-time-sec <seconds> stab-time-min <minutes> stab-time-hour <hour>
```

throttle arp-watermark arp_value

Wenn das ARP-Wasserzeichen konfiguriert ist und ein MME/SGSN den Drosselungsfaktor und die Verzögerung in einer DDN ACK-Nachricht sendet, werden alle DDNs, deren ARP-Wert größer als der konfigurierte Wert ist, durch den Drosselungsfaktor für die angegebene Verzögerung ausgelöst.

arp_value ist eine ganze Zahl zwischen 1 und 15.

Grenzwert

Konfiguriert die Ratenbeschränkung (verwenden Sie diese und nachfolgende Token, um die Ratenbeschränkung nur dann zu begrenzen, wenn es sich bei MME um eine Non-Release 10 MME handelt).

limit ist eine ganze Zahl zwischen 1 und 999999999.

Zeitfaktor-Sekunden

Konfiguriert die Zeitdauer, während der der SGW Entscheidungen zur Drosselung trifft.

Sekunden ist eine ganze Zahl zwischen 1 und 300.

Drosselfaktor Prozent

Konfiguriert den DDN-Drosselungsfaktor. Geben Sie den Prozentsatz des DDN ein, der bei Erkennung einer DDN-Überspannung fallen soll.

% ist eine Ganzzahl zwischen 1 und 100.

Inkrement-Faktor-Prozent

Konfiguriert den DDN-Throttling-Zuwachsfaktor. Geben Sie den Prozentsatz ein, um den die DDN-Drosselung erhöht werden soll.

% ist eine Ganzzahl zwischen 1 und 100.

Abfrageintervall-Sekunden

Konfiguriert das Abfrageintervall in der DDN-Drosselung.

Sekunden ist eine ganze Zahl zwischen 2 und 999999999.

Drosselungszeit Sek.

Konfiguriert die DDN-Drosselungszeit in Sekunden. Geben Sie einen Zeitraum in Sekunden ein, über den DDN am SGW gedrosselt werden.

Sekunden ist eine ganze Zahl zwischen 0 und 59.

Drosselungszeit min

Konfiguriert die DDN-Drosselungszeit in Minuten. Geben Sie einen Zeitraum in Minuten ein, über den DDN am SGW gedrosselt werden.

minutes ist eine ganze Zahl zwischen 0 und 59.

Drosselstundenzeit

Konfiguriert die DDN-Drosselungszeit in Stunden. Geben Sie einen Zeitraum in Stunden ein, über den DDN am SGW gedrosselt werden.

Stunde ist eine ganze Zahl zwischen 0 und 310.

Stichtag-Zeit-Sek.

Konfiguriert die DDN-Drosselungsstabilisierungszeit in Sekunden. Geben Sie einen Zeitraum in Sekunden ein, in dem die Drosselung deaktiviert wird, wenn das System stabilisiert ist.

Sekunden ist eine ganze Zahl zwischen 0 und 59.

Stichtag-min Minuten

Konfiguriert die DDN-Drosselungsstabilisierungszeit in Minuten. Geben Sie einen Zeitraum in Minuten ein, innerhalb dessen die Drosselung deaktiviert wird, wenn das System stabilisiert ist.

minutes ist eine ganze Zahl zwischen 0 und 59.

Arbeitsstunde

Konfiguriert die DDN-Drosselungsstabilisierungszeit in Stunden. Geben Sie einen Zeitraum in Stunden ein, in dem die Drosselung deaktiviert wird, wenn das System stabilisiert ist.

Stunde ist eine ganze Zahl zwischen 0 und 310.

Beispielkonfiguration

```
ddn throttle arp-watermark 1 rate-limit RATE time-factor 120 throttle-factor 50
increment-factor 10 poll-interval 30 throttle-time-sec 0 throttle-time-min 1
throttle-time-hour 0 stab-time-sec 0 stab-time-min 2 stab-time-hour 0
```

- 1100 Seiten/Sekunden ist die maximal zulässige Eingangsrate (einschließlich DDN).
- 1100 Seiten/Sekunden bei einer DDN-Überspannung entspricht 1100 DDN/Sekunden
- Regionen mit 4 x SGW pro MME-Standort > RATE = **275** DDN/Sekunde pro SGW maximal zulässig
- Regionen mit 3 x SGW pro MME-Standort > RATE = **366** DDN/Sekunde pro SGW maximal zulässig
- Regionen mit 2 x SGW pro MME-Standort > RATE = **550** DDN/Sekunde pro SGW maximal zulässig
- Regionen mit 1 x SGW pro MME-Standort > RATE = **1100** DDN/Sekunde pro SGW maximal zulässig

Schutz vor Netzwerküberlastungen: Fehlerdrosselung beim EGTP-Pfad

Diese Funktion schützt MME-Ressourcen (sessmgr, mmemgr) sowie 4G-Ressourcen gegen den Ausfall des Enhanced GPRS Tunneling Protocol (EGTP)-Pfads bei Übertragungsfehlern im IP-Backbone und IP-BackHaul. Die Funktion failures/restarts.The ermöglicht die Begrenzung der

erkannten EGTP-Pfadfehler pro Sitzung und definiert eine weitere Detailliertheit des Teilnehmermanagements, zusätzlich zum S1 Paging. Abspaltung. Abhängig von der Aufteilung auf inaktive und verbundene Teilnehmer werden die Grenzwerte festgelegt. Es ist sehr netzwerkspezifisch und erfordert eine Abstimmung in Bezug auf den eUTRAN- und den UE-Status.

Beispiel

Die Teilnehmer werden etwa 80:20 IDLE-to-CONNECTED aufgeteilt. Im schlimmsten Fall verursacht der EGTP-PF für IDLE-Teilnehmer eine Flut von Paging, die zu einer Überlastung der Mitglieder führen könnte, dem kleinsten Engpass in der Kette. Eine derartige Zunahme des EGTP-Paging-Faktors (PF) (für IDLE-Abonnenten) verursacht zunächst einen Paging-Anstieg, der den Engpass des Mitglieds erreicht. Daher müssen Sie den MGM zuerst vor diesem Problem schützen. Daher kann der EGTP PF für IDLE als unerwartete Paging-Überspannung betrachtet werden, die maximal 1100 Seiten pro Sekunde betragen darf.

- Die empfohlene Drosselungsgrenze beträgt 1000 msg/Sekunde für IDLE-Teilnehmer.
- Die Anzahl der VERNETZTEN Teilnehmer ist ca. 5- bis 7-mal geringer als bei IDLE.
- Paging-Spitzen treten bei CONNECTED-Abonnenten nicht auf, daher wird empfohlen, 2000 msg/sec sicher auf CONNECTED-Teilnehmer anzuwenden.

Hinweis: Die EGTP-PF-Drosselung ermöglicht eine weitere Feinabstimmung auf der MME-Paging-Drosselung auf der Eingangsseite (S11, Sv) und nicht auf der Ausgangsseite (S1). Bei Cisco *müssen* Sie *keine* EGTP-PF-Drosselung implementieren, wenn die Paging-Drosselung konfiguriert ist. Sie ermöglicht jedoch eine frühere Erkennung und Beseitigung von Überlastungen.

Diese Konfiguration gilt für einen EGTP-Dienst mit dem Schnittstellentyp "interface-mme".

Der Konfigurationsbefehl lautet wie folgt:

```
asr5000(config)# network-overload-protection mme-tx-msg-rate-control egtp-pathfail ecm-idle  
< rate in sessions per second > ecm-connected < rate in sessions per second >
```

- Netzwerküberlastungsschutz zur Identifizierung von Netzwerküberlastungsschutz
- mme-tx-msg rate control identifiziert MME-Nachrichtenratensteuerung
- egtp-pathfail identifiziert die Nachrichtenratenkontrolle für EGTP-Pfadfehler.
- ecm-idle identifiziert die Rate für MME UE-Sitzungen im ECM-Leerlauf-Modus.
- Identifizierungsrate für MME UE-Sitzungen im ECM-Connected-Modus
- <Rate in Sitzungen pro Sekunde> gibt einen Durchsatzschwellenwert in Sitzungen pro Sekunde an, der Bereich liegt zwischen 1 und 5.000

Beispielkonfiguration

```
network-overload-protection mme-tx-msg-rate-control egtp-pathfail ecm-idle  
1000 ecm-connected 2000
```

Verbesserte Überlastungskontrolle

Mithilfe der erweiterten Überlastungskontrolle kann die MME Signale an die eNodeBs senden, an

die sie angeschlossen ist, um Datenverkehr an andere MMEs im MME-Pool umzuleiten. Dies wird mithilfe des S1-Schnittstellenüberlastungsverfahrens (TS 36.300 und TS 36.413) erreicht.

Wenn eine Überlastungskontrolle konfiguriert und ein Überlastungsschwellenwert erreicht wird, kann die MME so konfiguriert werden, dass sie eine S1AP-Schnittstelle Overload Start-Nachricht an einen Prozentsatz der eNodeBs sendet, mit denen die MME verbunden ist. Dieser Prozentsatz ist konfigurierbar, um die von der MME zu reduzierende Last widerzuspiegeln. In an eNodeBs gesendeten Overload Response Information Element (IE) kann die MME eNodeB auffordern, bestimmte Arten von Sitzungen abzulehnen oder zuzulassen, darunter:

- Nichtnotfallsitzungen ablehnen
- neue Sitzungen ablehnen
- Notfallsitzungen zulassen
- Zulassen von Sitzungen mit hoher Priorität und Diensten mit mobiler Beendigung
- Ablehnen des verzögerungstoleranten Zugriffs

Mit der Funktion zur Überlastungskontrolle können Sie Richtlinien und Schwellenwerte festlegen und angeben, wie das System bei starker Auslastung reagiert. Die Überlastungskontrolle überwacht das System auf Bedingungen, die die Leistung beeinträchtigen könnten, wenn das System unter hoher Belastung ist. In der Regel handelt es sich um temporäre Bedingungen (z. B. hohe CPU- oder Speichernutzung), die schnell behoben werden. Jedoch kann eine große Anzahl dieser Bedingungen innerhalb eines bestimmten Zeitintervalls die Fähigkeit des Systems beeinträchtigen, Teilnehmersitzungen zu bedienen. Mithilfe der Überlastungskontrolle können solche Bedingungen identifiziert und Richtlinien aufgerufen werden, um die Situation zu beheben.

Grenzwerte für Überlastungsbedingungen

- CPU-Auslastung des Systems
- CPU-Auslastung des Systemdiensts (CPU-Auslastung der Demux-Card)
- Systemspeicher-Nutzung
- Lizenznutzung
- Maximale Anzahl Sitzungen pro Service

Grenzwerte und Toleranzwerte

Wenn Sie Grenzwerte und Toleranzen für kritische, schwerwiegende und geringfügige Engpässe konfigurieren, sollten sich die Schwellenwerte und Toleranzen niemals überschneiden. Betrachten Sie die folgenden Beispielfiguren, bei denen sich die Schwellenwerte nicht überschneiden:

- Staus verursachen 95 % und sind um 90 % überflüssig
- Staus verursachen eine hohe Belastung von 90 % und sind bei 85 % zu hören
- Geringfügige Engpässe verursachen 85 % und gehen bei 80 % auf

Servicesteuerung CPU-Grenzwerte

Dieser Grenzwert wird aus der demux-CPU des Systems berechnet. Der Grenzwert wird basierend auf einer durchschnittlichen CPU-Auslastung von fünf Minuten berechnet.

Der höchste CPU-Nutzwert von zwei CPU-Kernen der demux-CPU wird berücksichtigt. Wenn

beispielsweise der CPU-Core 0 eine CPU-Auslastung von 5 Minuten von 40 % und der CPU-Core 1 eine CPU-Auslastung von 5 Minuten von 80 % hat, wird der CPU-Core 1 für die Berechnung des Schwellenwerts herangezogen.

System-CPU-Grenzwerte

Dieser Grenzwert wird mithilfe des CPU-Nutzungs-Durchschnitts von fünf Minuten aller CPUs (außer Standby-CPU und SMC-CPU) berechnet.

Der höchste CPU-Nutzwert von zwei CPU-Kernen aller CPUs wird berücksichtigt.

Grenzwerte des Systemspeichers

Dieser Grenzwert wird mit dem 5-Minuten-Arbeitsspeichernutzungsdurchschnitt aller CPUs (außer Standby-CPU und SMC-CPU) berechnet.

Konfigurieren eines Überlastungsaktionsprofils

Überlastungs-Aktionsprofile definieren eine Reihe von Aktionen, die ausgeführt werden können, wenn der entsprechende Grenzwert überschritten wird.

Zuordnen eines Überlastungsaktionsprofils zu Überlastungskontrollrichtlinien

Jede Richtlinie zur Kontrolle von Überlastungen (kritisch, schwerwiegend, geringfügig) muss einem Profil zur Kontrolle von Überlastungen zugeordnet werden.

Konfigurieren der Überlastungskontrolle

Wenn eine Überlastungsbedingung für eine MME erkannt wird, kann das System so konfiguriert werden, dass die Bedingung einem angegebenen Prozentsatz von eNodeBs gemeldet und die konfigurierte Aktion für eingehende Sitzungen ausgeführt wird.

Diese Überlastungsaktionen sind auch verfügbar (zusätzlich zu Ablehnen-Neusitzungen):

- permit-Emergency-Sitzungen und mobile-terminierte Dienste
- permit-high-priority-sessions-and-mobile-terminierte Dienste
- Nur mit Verzögerung toleranter Zugriff
- Nicht-Notrufsitzungen ablehnen

Beispielkonfiguration - Erläuterung

Dies ermöglicht die Kontrolle von Überlastungen:

```
congestion-control
```

```
This monitors the overall CPU Utilization including the sessmgr and demux mgrs
```

```
congestion-control threshold system-cpu-utilization critical 90
```

```
congestion-control threshold system-cpu-utilization major 85
```

congestion-control threshold system-cpu-utilization minor 80

Memory utilization thresholds:

congestion-control threshold system-memory-utilization critical 85

congestion-control threshold system-memory-utilization major 75

congestion-control threshold system-memory-utilization minor 70

CPU utilization on DEMUX card:

congestion-control threshold service-control-cpu-utilization critical 85

congestion-control threshold service-control-cpu-utilization major 75

congestion-control threshold service-control-cpu-utilization minor 70

Defining tolerance margins:

congestion-control threshold tolerance critical 5

congestion-control threshold tolerance major 5

congestion-control threshold tolerance minor 5

Definition von Aktionsprofilen für Überlastungen (kritisch, groß und klein)

lte-policy

congestion-action-profile criticalCogestionProfile

reject s1-setup time-to-wait 60

drop handovers

drop combined-attaches

drop service-request

drop addn-brr-requests

drop addn-pdn-connects

exclude-voice-events

exclude-emergency-events

report-overload permit-emergency-sessions-and-mobile-terminated-service enodeb-percentage 50

congestion-action-profile majorCogestionProfile

report-overload permit-emergency-sessions-and-mobile-terminated-service enodeb-percentage 50

congestion-action-profile minorCogestionProfile

report-overload permit-emergency-sessions-and-mobile-terminated-service enodeb-percentage 30

end

Überlastungsrichtlinien anwenden

configure

```
congestion-control policy critical mme-service action-profile criticalCogestionProfile
```

```
congestion-control policy major mme-service action-profile majorCogestionProfile
```

```
congestion-control policy minor mme-service action-profile minorCogestionProfile
```

end

.

Zugehörige Informationen

- [Cisco ASR 5000 Mobility Management Entity Administration Guide](#)
- [Technischer Support und Dokumentation - Cisco Systems](#)