

ASR5x00 MME過載保護功能

目錄

[簡介](#)

[MME保護](#)

[網路過載保護：連線速率限制](#)

[網路過載保護：呼叫限制](#)

[示例配置](#)

[網路過載保護：DDN限制（服務網關功能，保護MME）](#)

[網路過載保護：EGTP路徑故障限制](#)

[示例配置](#)

[增強型擁塞控制](#)

[擁塞條件閾值](#)

[閾值和容差級別](#)

[服務控制CPU閾值](#)

[系統CPU閾值](#)

[系統記憶體閾值](#)

[相關資訊](#)

簡介

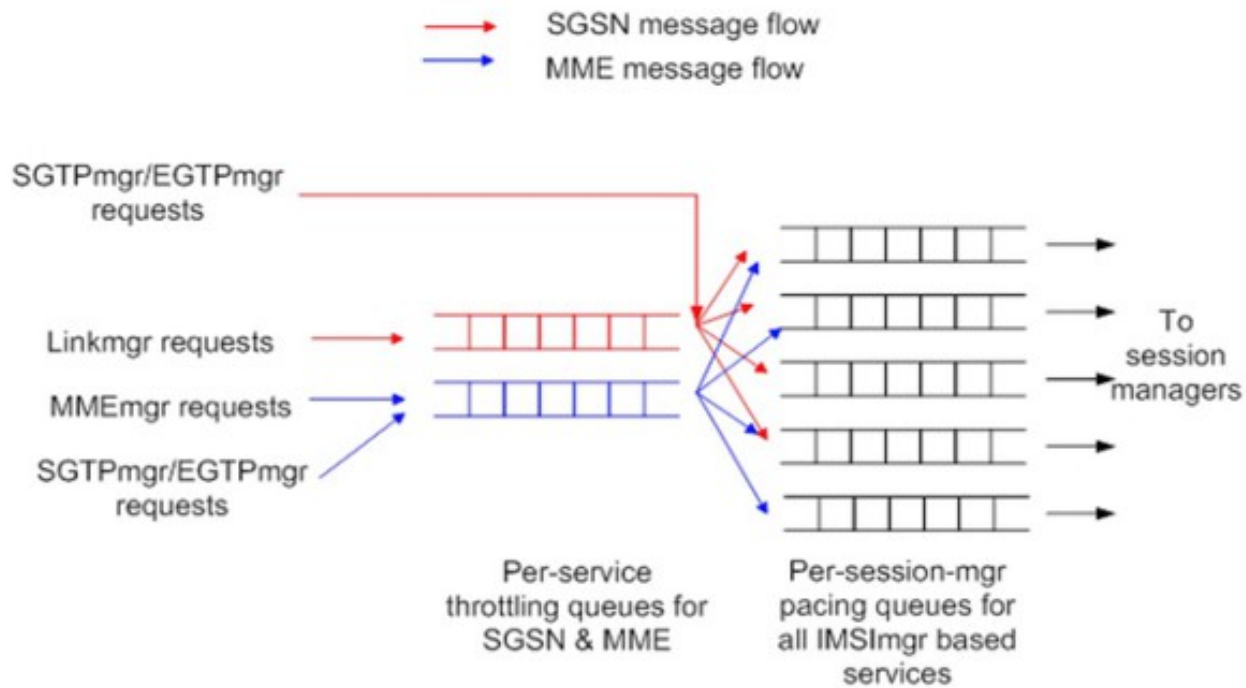
本文檔重點介紹思科聚合服務路由器(ASR)5000系列上可用的各種移動管理實體(MME)過載保護方法和功能。在ASR 5000系列中，思科為客戶提供了各種實現控制的方法，本文說明了功能和相關CLI命令。

MME保護

網路過載保護：連線速率限制

連線速率限制可保護鄰居網路元素(如歸屬使用者伺服器(HSS)、策略和計費規則功能(PCRF)和線上計費伺服器(OCS))以及內部MME資源(如imsimgr和sessmgr)。連線速率限制處理到達imsimgr的新呼叫，例如連線和MME/服務GPRS支援節點(SGSN)跟蹤區域更新(TAU)。

此圖片顯示了呼叫和限制隊列的消息流。



為了保護MME (imsimgr和sessmgr以上)，應定義限制率、隊列等待時間和隊列大小時間。限制速率取決於MME呼叫模型，因為MME容量取決於呼叫模型。

對於MME來說，限制速率計算相對簡單，請採用網路中的標準每秒呼叫事件(CEPS)加上容差。此外，如果需要HSS保護，您還需要考慮HSS資料庫容量。

範例

在繁忙時間內，MME每秒處理多達170到200個呼叫(Attaches+ Inter TAU)。如果有一個站點發生故障，則最多可能每秒會有350到370個呼叫到達一個MME。在此呼叫速率下，MME利用率上升近80%，400 calls per second是限制限制速率的最佳級別，以便避免MME盒內出現過多的信令負載。

*queue wait time*預設為5秒。這是客戶的最佳選擇。*queue size*預設為2500。這是客戶的最佳選擇。

配置命令如下。

```
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>}
```

new_connections

定義每秒要接受的新MME連線數。必須是介於50和5000之間的整數。預設值為500。

動作

定義步調隊列已滿時要執行的操作。每當在MME接收到新連線時，它們就會在調步隊列中排隊，而imsimgr會以配置的速率處理來自隊列的消息。當隊列溢位 (由於高傳入速率) 時，根據配置的「操作」，資料包將被丟棄或拒絕。

queue-size

定義用於緩衝資料包的調步隊列的最大大小。必須是介於250和25000之間的整數。預設值為2500。

示例配置

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

現在，每秒呼叫速率設定為400，並且此操作是智慧拒絕，因為這會導致使用者裝置(UE)重新連線到不同的無線接入技術(RAT)#15。等待時間設定為預設值（5秒），隊列大小為2500。

附註：以EMM cause #15 "no-suitable-cell-in-tracking-area"作為動作的「reject」是優先選項，因為以#15被拒絕的呼叫大多不會返回到MME並將轉到不同的RAT層(3G、2G)。用於服務無線網路子系統(SRNS)重新定位的操作「丟棄」用於將來使用，並將阻止在拒絕後快速重新連線到MME。

網路過載保護：呼叫限制

尋呼限制將內部MME資源(memgr)保護為eNodeB/無線電資源（如果需要）。對於給定ASR 5000機箱，此速率限制閾值應適用於與MME關聯的所有eNodeB。S1對eNodeB的尋呼請求應被速率限制在此閾值。S1超過此閾值的對eNodeB的尋呼請求將被丟棄。

對於MME，限制速率計算相對簡單，取網路內的標準出口尋呼速率加上容差。（這完全基於設計團隊的決定。）

範例

在繁忙時間內，每個MME每秒最多處理35000條尋呼消息。如果有一個站點發生故障，則一個MME每秒可能最多會傳送70000頁。在此尋呼速率下，MME利用率(memgr)將上升近80%，70000到80000 pages per second將是限制限制限制速率的最佳級別，以避免超過memgr的S1信令過多。

但是，每個平均eNodeB的速率是有限的。每個eNodeB（對於6500 eNodeB）的平均速率是每秒10頁。但是，跟蹤區域(TA)在使用者數量上不相等，並且不同的TA/成員eNodeB以不同的方式載入分頁。如果TA大小與每個TA的平均訂戶數之差為兩倍，則每個eNodeB的速率將為20。如果TA大小與每個TA的平均訂戶數之差為20倍，則每個eNodeB的速率將為200。這意味著當TA（訂戶數）均勻載入時，此功能將變得最有效。

應該並行執行的另一項操作是啟用智慧尋呼。請參閱《ASR 5000 MME管理指南》中的「TAI mgmt db和LTE尋呼」部分。

組態指令如下：

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

- network-overload-protection標識網路過載保護
- mme-tx-msg-rate-control enb標識每個平均eNodeB的MME消息速率控制
- s1-paging標識S1尋呼的消息速率控制
- <rate>指定速率閾值，以每秒eNodeB消息數為單位 — 範圍(1到65535)

示例配置

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

附註：

- 速率限制是需要進一步調整的主題，朝減小的方向。調整的基礎是TA上的使用者數（分頁數）（需要TA級別的統計資訊）。
- 當TA（按每個TA的使用者數/尋呼數計算）均勻載入時，該功能將變得最有效。

網路過載保護：DDN限制（服務網關功能，保護MME）

下行鏈路資料通知(DDN)節流是一種功能，用於控制從服務網關(SGW)端向MME發出的DDN請求的速率。它可保護MME資源（例如mmgr和sessmgr）免受DDN（即入口尋呼請求）浪湧的影響。

此功能包括兩個部分，一個用於Rel-10相容MME，另一個用於Rel-10不相容MME：

- 對於符合Rel-10的MME，請在SGW服務中設定DDN限制分配和保留優先順序(ARP)水線以啟用此功能。
- 對於Rel-10非相容MME，需要在SGW服務中設定一些其他引數以及ARP水印（如限制因子、限制時間、穩定時間、輪詢間隔等）。

在SGW上啟用此功能後，它會將DDN請求中的ARP水印傳送到MME。作為回覆，MME將傳送限制延遲單元、限制延遲值和限制因子。延遲值和延遲單位的組合將計算限制時間。收到這些值後，SGW將丟棄特定ARP的DDN請求，直到限制計時器過期。

對於使用本地配置的非Rel-10相容MME，SGW會使用特定ARP水印限制DDN請求。

Cisco ASR5x00 MME版本16和17不支援自動DDN限制，因此在DDN限制方面，它的工作方式與非Rel 10相容。

附註：DDN限制在入口端(S11)而不是出口端(S1)的MME尋呼限制上提供了進一步的粒度。如果已配置尋呼限制，思科不會要求您實施DDN限制，但它提供了更早的過載檢測和消除。

技術規範(TS)23.401,MME參考：

限制DDN請求

在異常情況下（例如，當MME負載超過運營商配置的閾值時），MME可能會限制其SGW在其上生成的信令負載（如果配置為這樣做）。

MME可以拒絕對於處於空閒模式的UE的低優先順序流量的DDN請求，或者進一步減輕MME的負荷。MME可以請求SGW根據在DDN Ack消息中指定的節流係數和DDN Ack消息中指定的節流延遲，選擇性地減少它傳送的DDN請求的數量，該請求用於為處於空閒模式的UE接收的下行鏈路低優先順序業務。

SGW基於承載的ARP優先順序和運營商策略（即，運營商在SGW中將被視為優先順序或非優先順序流量的運營商配置）來確定承載是否用於低優先順序業務。MME根據從SGW和運營商策略接收的ARP優先順序來確定DDN請求是否針對低優先順序流量。

如果UE的空閒狀態信令減少(ISR)未啟用，則在限制延遲期間，SGW丟棄在其所有低優先順序承載上接收到的下行鏈路分組，該低優先順序承載用於被稱作非連線使用者平面的UE(即，SGW上下文資料指示該MME服務的下行鏈路使用者平面隧道端識別符號(TEID)，該TEID與限制因子成比例，並且只針對非限制承載向MME傳送DDN消息。

如果ISR在限制延遲期間對UE處於活動狀態，則SGW不會將DDN傳送到MME，而只將DDN傳送到SGSN。如果MME和SGSN都請求減少負載，則SGW丟棄在其所有低優先順序承載上接收到的下行鏈路分組，該下行鏈路分組用於已知未連線的使用者平面(即，SGW上下文資料指示無下行鏈路使用者平面TEID)的UE，與限制因子成比例。

SGW在限制延遲到期後恢復正常操作。限制係數和限制延遲的上次接收值將取代從該MME接收的任何以前的值。接收限制延遲將重新啟動與該MME關聯的SGW計時器。

對於SGW與MME，節流速率計算相對簡單。採用允許的最大輸入尋呼率，即每個MME框每秒鐘傳送1100條消息。

配置命令如下：

```
#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

如果配置了ARP水印，並且MME/SGSN在DDN ACK消息中傳送了限制因子和延遲，則所有ARP值大於配置值的DDN都將由指定延遲的限制因子限制。

*arp_value*是一個介於1和15之間的整數。

速率限制

配置速率限制(僅當MME是非版本10 MME時，才使用此令牌和後續令牌進行速率限制)。

*limit*是一個介於1和999999999之間的整數。

時間因子秒

配置SGW做出限制決定的持續時間。

*seconds*是一個介於1和300之間的整數。

節流係數百分比

配置DDN限制係數。輸入在檢測到DDN浪湧時要丟棄的DDN的百分比。

*百分比*是一個介於1和100之間的整數。

增量因子百分比

配置DDN限制增量因子。輸入應增加DDN限制的百分比。

*百分比*是一個介於1和100之間的整數。

poll-interval seconds

配置DDN限制中的輪詢間隔。

*seconds*是介於2和999999999之間的整數。

throttle-time-sec秒

配置DDN限制時間 (以秒為單位)。輸入在SGW上限制DDN的時間段 (以秒為單位)。
*seconds*是介於0和59之間的整數。

throttle-time-min minutes

配置DDN限制時間 (以分鐘為單位)。輸入DDN在SGW處受限的時間段 (以分鐘為單位)。
*minutes*是從0到59的整數。

throttle-time-hour hour

配置DDN限制時間 (小時)。輸入DDN在SGW被限制的時段 (小時)。
*hour*是從0到310的整數。

stab-time-sec秒

配置DDN限制穩定時間 (以秒為單位)。輸入時間段 (以秒為單位)，如果系統已穩定，將禁用限制。
*seconds*是介於0和59之間的整數。

stab-time-min minutes

配置DDN限制穩定時間 (以分鐘為單位)。輸入時間段 (以分鐘為單位)，如果系統穩定下來，將禁用限制。
*minutes*是從0到59的整數。

插入時間 — 小時

配置DDN限制穩定時間 (小時)。輸入時間段 (以小時為單位)，如果系統已穩定，將禁用限制。
*hour*是從0到310的整數。

示例配置

```
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頁/秒是允許的最大輸入速率 (包括DDN)
- 1100頁/秒，如果DDN激增對應於1100 DDN/秒
- 每個MME站點具有4xSGW的區域>最大允許的速率= 275 DDN/秒
- 每個MME站點具有3xSGW的區域>最大允許的速率= 366 DDN/秒
- 每個MME站點具有2xSGW的區域>最大允許的速率= 550 DDN/秒
- 每個MME站點具有1xSGW的區域>最大允許的速率= 1100 DDN/秒

網路過載保護：EGTP路徑故障限制

此功能可保護MME資源(*sessmgr*、*mmgr*)以及4G資源免受增強型GPRS隧道協定(EGTP)路徑故障浪潮的影響，以防IP骨幹網和IP回程中的傳輸失敗，並且在S1尋呼限制之外，側網元 *failures/restarts*。The功能可根據*sessmgr*對檢測到的EGTP路徑故障事件進行限制，並為使用者管理定義進一步的粒度。根據空閒使用者和已連線使用者的劃分，設定限制。它非常特定於網路，需要根據eUTRAN和UE狀態進行調整。

範例

使用者大約在80:20 IDLE到CONNECTED。最糟糕的情況是，IDLE使用者的EGTP PF會導致尋呼激增，這可能導致鏈路中最窄的瓶頸*mmgr*過載。這種EGTP尋呼因子(PF)突發 (對於IDLE使用者) 首先會導致尋呼突發，並且此突增會遇到*mmgr*瓶頸，因此您需要首先保護*mmgr*以避免此情況。

因此，IDLE的EGTP PF可能被視為意外入口尋呼突發，其最大值被允許為1100頁/秒。

- 對於空閒訂戶，建議的限制限制為1000消息/秒。
- CONNECTED子節點的數量大約是空閒子節點的5到7倍。
- 已連線的訂閱者不會發生尋呼突發，因此建議對已連線的訂閱者安全應用2000消息/秒。

附註：EGTP PF限制在入口端(S11,Sv)而不是出口端(S1)的MME尋呼限制上提供了進一步的粒度。如果配置了呼叫限制，思科不要求您實施EGTP PF限制，但它提供了更早的過載檢測和消除。

此配置適用於介面型別為「interface-mme」的EGTP服務。

組態指令如下：

```
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 >
```

- network-overload-protection標識網路過載保護
- mme-tx-msg-rate-control標識MME消息速率控制
- egtp-pathfail識別EGTP路徑故障的消息速率控制
- ecm-idle標識ECM-Idle模式下MME UE會話的速率
- ecm-connected識別ECM-Connected模式下MME UE會話的速率
- <速率 (以會話數每秒為單位) >指定速率閾值 (以會話數每秒為單位)，範圍為1到5000

示例配置

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

增強型擁塞控制

使用增強型擁塞控制功能，MME可以向其所連線的eNodeB發出訊號，以便將流量重定向到MME池中的其他MME。這通過S1介面過載程式 (TS 36.300和TS 36.413) 完成。

當配置了過載控制並且達到擁塞閾值時，可以將MME配置為將S1AP介面過載開始消息傳送到與MME連線的eNodeB的某個百分比。為了反映MME希望減少的負載量，此百分比是可以配置的。在傳送到eNodeB的超負荷響應資訊元素(IE)中，MME可以請求eNodeB拒絕或允許特定型別的會話，其中包括：

- 拒絕非緊急會話
- 拒絕新會話
- 允許緊急會話
- 允許高優先順序會話和移動端終止服務
- 拒絕容遲訪問

擁塞控制功能允許您設定策略和閾值，並指定系統在遇到重負載情況時的反應方式。擁塞控制會監控系統是否有可能會在系統負載較重時降低效能的條件。通常，這些情況是暫時的 (例如，高CPU或記憶體利用率)，並且會快速得到解決。但是，在特定時間間隔內連續或大量出現這些情況可能會影響系統服務使用者會話的能力。擁塞控制有助於識別這些情況並呼叫策略來解決這種情況。

。

擁塞條件閾值

- 系統CPU使用情況
- 系統服務CPU使用率 (解複用卡CPU使用率)
- 系統記憶體使用情況
- 許可證使用情況
- 每個服務的最大會話數

閾值和容差級別

為嚴重、主要和次要擁塞級別配置閾值和容差時，閾值級別和容差決不能重疊。請考慮以下閾值級別不重疊的示例配置：

- 嚴重擁塞觸發率為95%，清除率為90%
- 主要擁塞觸發率為90%，清除率為85%
- 輕微擁塞觸發率為85%，清除率為80%

服務控制CPU閾值

此閾值由系統的demux CPU計算。閾值是根據五分鐘的平均CPU使用率計算的。

考慮解複用器CPU的兩個CPU核心的最大CPU使用值。例如，如果CPU核心0的五分鐘CPU使用率為40%，而CPU核心1的五分鐘CPU使用率為80%，則考慮CPU核心1進行閾值計算。

系統CPU閾值

此閾值使用所有CPU (備用CPU和SMC CPU除外) 的五分鐘CPU使用平均值計算。

考慮所有CPU中兩個CPU核心的最大CPU使用值。

系統記憶體閾值

此閾值是根據所有CPU (備用CPU和SMC CPU除外) 的五分鐘記憶體使用平均值計算的。

配置擁塞操作配置檔案

擁塞操作配置檔案定義了一組操作，可在超過相應閾值後執行。

將擁塞操作配置檔案與擁塞控制策略相關聯

每個擁塞控制策略 (關鍵、主要、次要) 必須與擁塞控制配置檔案相關聯。

配置過載控制

當在MME上檢測到過載情況時，系統可以配置為將過載情況報告給eNodeB的指定百分比，並對傳入會話執行已配置的操作。

以下過載操作也可用 (除了reject-new-sessions)：

- permit-emergency-sessions-and-mobile-terminated-services
- permit-high-priority-sessions-and-mobile-terminated-services
- reject-delay-tolerant-access
- 拒絕非緊急會話

配置說明示例

這樣會啟用擁塞控制功能：

```
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
```

定義擁塞操作配置檔案 (嚴重、主要和次要)

```
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
```

應用擁塞策略

```
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
```

相關資訊

- [Cisco ASR 5000移動管理實體管理指南](#)
- [技術支援與文件 - Cisco Systems](#)