SMF Overload Protection

Feature Summary and Revision History

Summary Data

Table 1: Summary Data

Applicable Products or Functional Area	SMF
Applicable Platform(s)	SMI
Feature Default Setting	Disabled - Configuration Required
Related Changes in this Release	Not Applicable
Related Documentation	Not Applicable

Revision History

Table 2: Revision History

Revision Details	Release
Added support for message priority configuration.	2020.04.0
First introduced.	2020.03.0

Feature Description

An interface can handle only a specified number of incoming requests. When the incoming requests exceed the specified numbers, the interface overloads. For instance, an interface is overloaded when:

- There is a network element failure causing large number of re-attaches
- Multiple users perform location update or transition from idle to active mode frequently

Overloading causes the interface to either drop the requests or delay processing the request. The overall network performance degrades because of overloading at the interface. This can lead to node congestion, failure or collapse which in turn causes load increase on the other nodes.

The SMF measures different resources and defines the load based on those measurements. Also, the SMF updates the NRF about the load. Currently, the SMF applies overload protection on inbound messages. The external nodes throttle towards the SMF to come out of a congestion when overload protection is applied on the inbound interface (SBA Interface).

NOTE: The scope of this feature is only on overload due to inbound requests on SBA interface.

How it Works

The SMF protects inbound requests from overloading at Endpoint and Application levels.

- **Endpoint Level** – The protection is based on the HTTP request method without taking the message type into account.
- **Application Level** – The protection is based on the message type.

Message Priority

The SMF applies the overload protection on the incoming request messages after evaluating the resources' availability to process the request and the message priority. The high priority messages get the lower preference to throttle, and low-priority messages get higher preference. An overloaded NF applies the message prioritization schemes on the incoming messages during an overloaded condition. In such conditions, the NF excludes the messages of the highest priority from the overload protection mechanism.

Once you configure message priority, SMF starts classifying the messages based on their priority. This configuration is optional. If you chose not to use this configuration, SMF applies the overload protection technique without considering the message priority.

Overload Protection at Endpoint

For endpoints, the SMF offers overload protection at both the endpoint and client levels. The SMF defines the overload threshold limits for the inbound request messages. Based on the threshold range, the SMF can reject the inbound request messages. The SMF sends back an HTTP response with the configured status to the request initiator.

The following are the overload threshold limits defined in the SMF:

- **Low** – When this threshold is met, only the POST method (with generic URI contributing to resource allocation) is rejected.
- **High** – All messages are rejected with the configured (reject) statuses when this threshold is met.
- **Critical** – All messages are rejected with the configured (reject) statuses when this threshold is met.

Configuring Overload Protection

This section describes the configuration procedures involved in configuring the overload protection for inbound request messages.

Configuring Overload Protection at Endpoint Level

Use the following configuration to configure overload protection at endpoint level.

```
configure
   endpoint sbi
     overload-control threshold threshold_limit threshold_range action action_status
 action_code range
      commit
      end
```

NOTES:

- **overload-control** – Specify the overload control at endpoint level.
- **threshold** – Specify the threshold limit and range.
- *threshold_limit* –Specify the threshold limit. *threshold_limit* must be one of the following:
  - *low* – Specify the low threshold limit for overload protection.
  - *high* – Specify the high threshold limit for overload protection.
  - *critical*– Specify the critical threshold limit for overload protection.
- *thershold_range* – Specify the threshold range. *threshold_range* must be an integer in the range of 10 – 100000.
- **action** – Specify the action to be taken for the threshold limit.
- *action_status* – Specify the action for the threshold limit. *action_status* must be:
  - **reject** – Rejects the inbound messages if the specified threshold range is met.
- *action_code* – Specify the action status code. *action_code* must be:
  - **reject-code** – Specify the reject status code.
- *range* – Specify the range of the action code. *range* must be an integer in the range of 100 – 600.

The following is an example configuration:

```
overload-control threshold low 500 action reject reject-code 501
overload-control threshold critical 10000 action reject reject-code 329
```

Configuring Overload Protection at Client Level

Use the following configuration to configure overload protection at client level.

```
configure
   endpoint sbi
      overload-control client threshold threshold_limit threshold_range action
action_status action_code range
      commit
      end
```

NOTES:

- **overload-control client** – Specify the overload control at client level.
- **threshold** – Specify the threshold limit and range.
- *threshold_limit* –Specify the threshold limit. *threshold_limit* must be one of the following:
  - *low* – Specify the low threshold limit for overload protection.
  - *high* – Specify the high threshold limit for overload protection.
  - *critical*– Specify the critical threshold limit for overload protection.
- *thershold_range* – Specify the threshold range. *thershold_range* must be an integer in the range of 10 – 100000.
- **action** – Specify the action to be taken for the threshold limit.
- *action_status* – Specify the action for the threshold limit. *action_status* must be:
  - **reject** – Rejects the inbound messages if the specified threshold range is met.
- *action_code* – Specify the action status code. *action_code* must be:
  - **reject-code** – Specify the reject status code.
- *range* – Specify the range of the action code. *range* must be an integer in the range of 100 – 600.

The following is an example configuration:

```
overload-control client threshold low 50 action reject reject-code 329
overload-control client threshold critical 20000 action reject reject-code
501
```

Verifying the Overload Protection Configuration

Use the **show running-config** command to view the overload protection configuration in the SMF Ops Center. The following is a sample output of the **show running-config** command.

```
[cluster1/data] example# show running-config
endpoint sbi
 overload-control threshold low 5000 action reject reject-code 555
 overload-control threshold high 7000 action reject reject-code 329
 overload-control threshold critical 10000 action reject reject-code 503
 overload-control client threshold low 750 action reject reject-code 329
 overload-control client threshold high 500 action reject reject-code 329
 overload-control client threshold critical 1000 action reject reject-code 503
 interface n11
  overload-control threshold low 4000 action reject reject-code 555
  overload-control threshold high 6000 action reject reject-code 329
  overload-control threshold critical 7000 action reject reject-code 503
  overload-control client threshold low 500 action reject reject-code 329
  overload-control client threshold high 700 action reject reject-code 329
  overload-control client threshold critical 800 action reject reject-code 503
 exit
```

Configuring the Message Priority

Use the following configuration to configure message priority for the inbound request messages.

configure
  overload-control threshold *threshold_limit threshold_range* **action reject reject-code** *range* **exclude message-priority** *priority_value*
  end

NOTES:

- **overload-control** – Specify the overload control at endpoint level.
- **threshold** – Specify the threshold limit and range.
- *threshold_limit* – Specify the threshold limit. *threshold_limit* must be one of the following:
  - low – Specify the low threshold limit for overload protection.
  - high – Specify the high threshold limit for overload protection.
  - critical– Specify the critical threshold limit for overload protection.
- *thershold_range* – Specify the threshold range. *thershold_range* must be an integer in the range of 10–100000.
- **action** – Specify the action to be taken for the threshold limit.
- *action_status* – Specify the action for the threshold limit.*action_status* must be:
  - **reject** – Rejects the inbound messages if the specified threshold range is met.
- **exclude message-priority** – Excludes the messages from the overload protection mechanism depending on the assigned priority.
- *priority_value* – Specifies the priority value.

The following is an example configuration:

overload-control threshold *low 1000* **action reject reject-code** *100* **exclude message-priority** *8*

overload-control threshold *high 2000* **action reject reject-code** *100* **exclude message-priority** *5*

If the priority value is 8, then the messages received with priority 8 or higher are not throttled. This applies even when the system threshold is lower than the priority value. The 3GPP defined message priority is 0–31 as per *3GPP TS 29.500, section 6.8.4*.

Monitoring and Troubleshooting

This section provides information regarding bulk statistics available to monitor and troubleshoot this feature.

Statistics

The following statistics are available in support of Overload Control.

Bulk Statistics	Statistics Type	Description
endpoint_overload_status	Gauge	Contains Endpoint-Name, Interface-Name and Overload-Level as labels. Once any level(low/high/critical) is hit, the gauge value will be set to 1. In normal condition the value is set to 0.
endpoint_client_overload_status	Gauge	Contains Endpoint-Name, Interface-Name, peer-host name and Overload-Level as labels. Once any level(low/high/critical) is hit, the gauge value will be set to 1. In normal condition the value is set to 0.
endpoint_pending_request	Gauge	Display current outstanding request for an endpoint. It contains Endpoint name and Interface Name as label.
endpoint_client_pending_request	Gauge	Display current outstanding request for a peer connected with an endpoint. It contains Endpoint name, Interface Name and peer host address connected to the endpoint as label.
endpoint_overload_exclude	Counter	Display the messages with their priority details that were excluded from the overload control mechanism. The metric is incremented for every message, which bypasses the overload control mechanism.