



Overload Control

Revised: May 14, 2012, OL-25016-02

Overload Control Processes

Overload is a switch condition that exists when system resources cannot handle system tasks. Increases in call traffic or messages indirectly related to call traffic usually cause overload. The overload control processes are listed in [Table C-1](#).

The Overload Control feature supports the Cisco BTS 10200 Softswitch Call Agent (CA) and Feature Server (FS). Overload Control detects, controls, and manages overload from all types of networks (SIP, SS7, ISDN, MGCP, H.323).

Table C-1 **Overload Control Processes**

Overload Control Phase	Actions
1. Automatic detection and handling	Measures and compares factors to threshold values. Determines system congestion and machine congestion level (MCL). Detects Cisco BTS 10200 machine congestion conditions in 5 levels: none, mild, moderate, severe, emergency. Automatically reduces overload as described.
2. Reporting	Affects the following switch areas: <ul style="list-style-type: none">• Alarms• Logs• Billing• Measurements



Note

The monitoring of the CPU load of critical processes is *not* supported.

Detecting Overload

In the detection phase of Overload Control, any one of three factors can have the highest MCL. This value dictates the MCL for the entire system. The three factors are

- Critical process queue lengths—The `olm.cfg` configuration file has critical queue lengths for Cisco BTS 10200 processes like BCM, MGA, SGA, SIA, ISA, and H3A. You can define multiple (32 factors total) critical queues for any Cisco BTS 10200 process. The Cisco BTS 10200 monitors the usage proportion of each critical IPC queue.
- IPC buffer pool usage—Cisco BTS 10200 monitors the proportion of available buffers in the IPC buffers pool. This reflects MCL: the higher the usage, the greater the congestion.

Cisco BTS 10200 detects its own MCL in five levels:

- MCL0—No congestion and no need for any abatement.
- MCL1—Mild congestion. Call rejection starts as configured in `olm.cfg`.
- MCL2—Moderate congestion. Call rejection increases as configured in `olm.cfg`.
- MCL3—Severe congestion. Call rejection increases still more as configured in `olm.cfg`.
- MCL4—Emergency congestion. Cisco BTS 10200 rejects all calls including emergency calls.

Computing MCL

The Cisco BTS 10200 computes factor levels by calculating averages for each factor. The rate of sampling (number of slots) can be configured per factor (3–10 slots). The MCL is set according to a factor level. In [Table C-2](#) thresholds are set to 50, 70, 90, and 95 percent.

Table C-2 MCL Thresholds

Onset /abatement thresholds	Factor Level	MCL
—	0–49	MC0
<code>level_1_threshold = 50</code>	50–69	MC1
<code>level_2_threshold = 70</code>	70–89	MC2
<code>level_3_threshold = 90</code>	90–94	MC3
<code>level_4_threshold = 95</code>	95–100	MC4

Reducing Overload

When MCL exceeds MCL0, overload control reduces MCL as follows:

- Selectively reject new calls by the signaling adapters—A percentage of calls and messages are rejected at the current MCL level, based on olm.cfg. Emergency calls are not rejected at MCL 1–3, but all calls, including emergency calls, are rejected at MCL4.
- Tell the network to stop sending traffic—This starts when the Cisco BTS 10200 is mildly congested (at MCL1) and continues through all higher MCL levels until the overload condition abates to MC0. This action can only be applied to the following types of networks:
 - SS7 sends Automatic Control Level (ACL) parameter in ISUP release messages.
 - H.323 sends Resource Availability Indicator (RAI) message.
 - SIP sends 500 or 503 with a retry.
- CA stops sending triggers to POTS FS—When the FS is congested the following occurs:
 - FS notifies CA once of its congested status.
 - CA sends only emergency triggers to FS, as it manages FS’s congestion abatement.

Slowing Overload Reduction

Sudden abatement reduction might cause MCL to rapidly increase again. To counteract MCL “bouncing,” MCL reduces one MCL level at a time, regardless of how low computed MCL becomes. This permits the system MCL to reduce gracefully over a number of intervals.

Overload Implementation and Configuration

This section discussed the following items:

- [Configuring Emergency Call Handling](#)
- [Signal Adapter Call Rejection](#)
- [Configuring the SIP Response Code](#)
- [SIP Message Handling](#)
- [H.323 Message Handling](#)
- [SS7 Automatic Control Parameter](#)



Note

These tasks include examples of CLI commands that illustrate how to provision the specific feature. For a complete list of all CLI tables and tokens, refer to the [Cisco BTS 10200 Softswitch CLI Database](#).

Configuring Emergency Call Handling

The Cisco BTS 10200 checks the

- Called-party number for all incoming calls against the EMERGENCY-NUMBER-LIST
- Calling party category (CPC) in ISUP calls

If the Cisco BTS 10200 determines it is an emergency call and the MCL is 1, 2, or 3, the Cisco BTS 10200 gives it priority and does not reject the call. If the MCL is 4, The Cisco BTS 10200 rejects all calls, including emergency calls.

To add a number to the EMERGENCY-NUMBER-LIST, enter a command similar to the following:

```
add emergency-number-list digit_string=911;
```

```
Reply: Success: at 2006-02-28 09:48:40 by btsadmin
MNT add successful
Transaction 934823299797597704 was processed.
```

To display the EMERGENCY-NUMBER-LIST, enter:

```
show emergency-number-list;
```

```
DIGIT_STRING=911
```

```
Reply: Success: at 2006-02-28 09:48:45 by btsadmin
Entry 1 of 1 returned.
```

To delete a number from the EMERGENCY-NUMBER-LIST, enter a command similar to the following:

```
delete emergency-number-list digit_string=911;
```

```
Reply: Success: at 2006-02-28 09:52:20 by btsadmin
MNT delete successful
Transaction 934823480106794504 was processed.
```

Signal Adapter Call Rejection

The OLM process provides functionality by which adapters can call to see if a particular call or event should be rejected. This functionality determines whether a call or message should be accepted or rejected. The calling signaling adapter provides a message/call number (or allow a default value to be used by OLM), and the percentage of calls/messages to be rejected at the current MCL level (or allow a default value to be used by OLM).

The first parameter is an integer containing the message/call number value to be used for the call rejection calculation.

The second parameter is an integer containing the percentage of calls to be rejected.

The third parameter can be set to either `OLM_API_CALL_TYPE_ORDINARY` or `OLM_API_CALL_TYPE_EMERGENCY`. When set to `OLM_API_CALL_TYPE_ORDINARY`, the function always returns `FALSE` if the system MCL is at level 4. When set to `OLM_API_CALL_TYPE_EMERGENCY`, the function always returns true ignoring the other parameters unless the system MCL is at level 4 in which case it again returns false. Per the selection of parameter 4, emergency calls are not normally subject to any rejection at MCL 0 – 3, but all calls including emergency calls are rejected at MCL4.

Based on these parameters, the function then returns either false - do not accept the message/call/session, or true - accept the message/call/session. In the case of MCL4, the function always returns false.

SS7 (SGA) Implementation of Call Rejection

An SS7 (SGA) implementation provides defaults for the first two parameters, so OLM provides both the random seed for the percentage of rejections and the actual reject rates. In addition, the called-party number for all incoming calls is checked against the `EMERGENCY-NUMBER-LIST` table to determine whether the call is an emergency call. Also in ISUP implementations, the calling party category is checked to see if it is an emergency line. If the incoming call is an emergency call (that is, called-party number is found in the table or CPC is emergency), it is given priority and not rejected outright in case of MC1, 2, and 3. In case of MC4, all calls are to be rejected.

H323 Implementation of Call Rejection

For an H.323 implementation, the default behavior is to use for OLM reject percentage values (H.323 calls with `OLM_API_USE_DEFAULT_PERCENTAGE`). However, H.323 also provides the option to use command-line arguments to over-ride the default OLM percent values. In that case, H.323 passes the value. For the message number, H.323 always uses `OLM_API_USE_DEFAULT_MSG_NUM`.

SIA (SIP) Implementation of Call Rejection

An SIA (SIP) implementation provides the actual message number and the rejection percentage. This is because SIP not only deals with calls (invites), but with other messages such as register, subscribe, notify, and options messages. The messages may or may not have any call context.

ISA Implementation of Call Rejection

An ISA implementation provides defaults for the first two parameters, so OLM provides a random number for the percentage of rejections and the actual reject rates. In addition, the called-party number for all incoming calls is checked against the EMERGENCY-NUMBER-LIST table to determine whether the call is an emergency call. If the incoming call is an emergency call (that is, called-party number is found in the table), it is given priority and not rejected outright in case of MC1, 2, and 3. In case of MC4, all calls are to be rejected.

Configuring the SIP Response Code

When rejecting a SIP message during overload, the Cisco BTS 10200 can use either of the following:

- 500 Server Internal Error
- 503 Service Unavailable

Use the following command. The default value is 503.

```
add ca_config type=SIA-OC-REJECTION-RESP; datatype=integer; value=500;
```

SIP Message Handling

When processing an incoming SIP call, the Cisco BTS 10200 looks at the MCL of the CA. It uses the following factors to decide whether to accept or reject the message:

- SIP Message Type
- Call type (normal or emergency)
- Configured rejection percentage
- Current MCL status

SIP Message Types

This section provides information on the SIP message types.

Message Rejection: Invite

If it is overloaded, the Cisco BTS 10200 rejects a percentage of incoming invite messages. The percentage rejected is based on sia.cfg. Only new invite messages are checked for acceptance. Reinvite messages are always accepted.

Message Rejection: Register

If it is overloaded, the Cisco BTS 10200 rejects a configured percentage of register messages.

Message Rejection: Refer

If it is overloaded, the Cisco BTS 10200 rejects a percentage of incoming refer messages.

Message Rejection: Subscribe

If it is overloaded, the Cisco BTS 10200 rejects a configured percentage of out-of-dialog subscribe messages. The Cisco BTS 10200 also rejects subscribe messages without call contexts. The Cisco BTS 10200 does not reject subscribe messages received in an invite dialog.

Message Rejection: Options

If it is overloaded, the Cisco BTS 10200 rejects options messages. There is no configuration required; all options messages are rejected between MCL1 and MCL4.

Message: Unsolicited Notify Repression

If it is overloaded, the Cisco BTS 10200 does not send unsolicited notify messages (MWI requests) to endpoints. However, even if it is overloaded, the Cisco BTS 10200 receives and processes unsolicited notify requests.

UDP Messages

The Cisco BTS 10200 drops messages like stun if they are less than the configured size. This applies to UDP messages.

Message Rejection Logic

When the Cisco BTS 10200 rejects an incoming SIP call it responds with 500 or 503. Use the CLI to set the response code.

The Cisco BTS 10200 includes a Retry-After header in its response. The value (in seconds) in this header notifies the endpoint that the Cisco BTS 10200 does not receive further requests for the specified time. For example, “Retry-After: 5” means the endpoint should send the next request to the Cisco BTS 10200 only after 5 seconds has passed.

H.323 Message Handling

This section provides information on the H.323 message handling.

Call Rejection—System MCL

When a complete H.225 setup message is received, an application-provided call-back function is used to check whether the call should be rejected immediately or accepted based on the MCL state. The OLM determines whether the call should be rejected based on its default call reject percentages and the system-MCL. The H.323 process checks whether the call is an emergency call before releasing it. If it's an emergency call, it is accepted.

Call Rejection—IPC Queue

For incoming TCP-based calls, the H.323 process checks the IPC queue for congestion. If the queue is congested, the H.323 process responds to all valid setup message with an H.225 ReleaseComplete with a CauseCode=42, and CallCapacity information with CallsAvailable=0, to indicate congestion. The TCP connection is released immediately.

The H.323 process checks whether the call is an emergency call before releasing it. If it's an emergency call, it attempts one more time to post the call to the worker thread.

Congestion on Peer Gateway

A terminating peer H.323 Gateway (or IP-IP Gateway) can indicate congestion by sending a ReleaseComplete with CauseCode=42 or with CallCapacity data with CallsAvailable=0. If either of these two indications is received at the Cisco BTS 10200, the H.323 process sets the acl_set field in the TRUNK-GRP table associated with the H.323 gateway and starts a timer. The timer length is read from the PEER-GW-OVERLOAD-TIMER field of the H323-TG-PROFILE table configured for the TRUNK-GRP. For the length of the timer, BCM checks and attempts to route calls to an alternate destination.

Reporting Call Capacity

For every incoming call, the H.323 process reports call capacity information to the gatekeeper in the ARQ and DRQ messages and to peer H.323 gateways in the release complete message. In addition, the H.323 process reports call capacity information to the gatekeeper in the RRQ and RAI messages.

Each H.323 instance reports the call capacity as it relates to its available resources. The maximum call capacity is calculated based on the smaller of the provisioned MAX-VOIP-CALLS or the number of available shared memory call-data blocks for the Cisco BTS 10200 H.323 gateway. The current call capacity is the current number of active calls on the H.323 gateway.

Report Alternate Endpoints

The Cisco BTS 10200 virtual H.323 gateway includes provisioned alternate endpoint data. The alternate endpoint data is provisioned through the CLI in the H323-GW table and consists of a TSAP address for each endpoint. The H.323 process is updated by a database trigger whenever this table is changed. Additionally, a full-weight RRQ message including the updated alternate endpoint information is sent to the gatekeeper when the H323-GW table is changed.

Sending RAI to Gatekeeper

A resource availability indicator (RAI) message is sent to the gatekeeper when there is system-wide or H.323 IPC thread congestion. The sending of the RAI message is triggered when there is a system MCL or IPC queue MCL state change.

When a system moves from a noncongested to a congested state, an RAI message is sent with the almost out of resources parameter set to true. When a system transitions from a congested to no-congested state, an RAI message is sent with the almost out of resources parameter set to false. Additionally, the call capacity information is included in the RAI message.

SS7 Automatic Control Parameter

If the current Machine Congestion Level (MCL) is greater than MC0, the Cisco BTS 10200 system includes an automatic congestion level (ACL) parameter in every ISUP release message it sends to linked switches. Receiving the ACL induces the linked switches to reduce the call traffic offered to the system.

Some switches understand up to three levels of congestion indication. Others only understand two. Some ignore the ACL altogether depending on the ISUP variant. The attribute MAX ACL indicates the mapping of the current MCL to the ACL value in the release message, as stated in [Table C-3](#).

Table C-3 MCL to ACL Mapping

MAX ACL	Current MCL	ACL Value in Release Message
0	0, 1, 2, 3, 4	Not Present
2	0	Not Present
	1	1
	2	2
	3	2
	4	2
3	0	Not Present
	1	1
	2	2
	3	3
	4	3

The MAX_ACL value is hard coded in MDL per variant. For example, if a particular ISUP supports a maximum ACL value of 2, then at run time if the current MCL is 3, then according to [Table C-3](#), an ACL value of 2 is sent in all the release messages to this switch.

The SS7 (SGA) process monitors the MCL and includes the ACL in the release messages whenever the MCL is greater than MC0.

Operating

This section explains how to perform the following tasks:

- [Viewing MCL](#)
- [Measurements](#)

It also explains how this feature affects the measurements operational area.

Viewing MCL

To display the MCL, enter a command similar to the following:

```
status machine-congestion-level platform_id=CA146;

MACHINE CONGESTION LEVEL ON CALL AGENT CA146 IS... ->

ADMIN MCL -> NO_CONGESTION(0)
COMPUTED MCL -> NO_CONGESTION(0)
EFFECTIVE MCL -> NO_CONGESTION(0)
FEATURE SERVER CONGESTION ->
FSAIN205 IS NOT CONGESTED
FSPTC235 IS NOT CONGESTED
REASON -> ADM executed successfully
RESULT -> ADM configure result in success

Reply : Success: at 2006-02-28 09:54:27 by btsadmin
```

If platform_id is the FS, for example, FSPTC235, the output shows MCL. If platform_id is the CA, for example, CA146, the output includes congestion status of FSs as seen by the CA. Without this parameter the command displays the MCL of all platforms on the system.

Setting the Minimum System MCL



Warning

Manually setting minimum MCL means call processing is affected exactly as it would be if MCL were set at that level due to actual system overload/congestion. Use it for test purposes only.

To set the minimum system MCL, enter a command similar to the following:

```
control machine-congestion-level platform_id=CA146, mcl=2;

MACHINE CONGESTION LEVEL ON CALL AGENT CA146 IS... ->

ADMIN MCL -> NO_CONGESTION(0)
COMPUTED MCL -> NO_CONGESTION(0)
EFFECTIVE MCL -> NO_CONGESTION(0)
FEATURE SERVER CONGESTION ->
FSAIN205 IS NOT CONGESTED
FSPTC235 IS NOT CONGESTED
REASON -> ADM executed successfully
RESULT -> ADM configure result in success

Reply: Success: at 2006-02-28 09:54:27 by btsadmin
```

Measurements

These tables list new, modified, or deleted measurements.



Note

See the [Using BTS Measurements](#) chapter of the *Cisco BTS 10200 Operations and Maintenance Guide, Release 6.0.4* for a complete list of all traffic measurements.

Call Processing Measurements

Table C-4 lists the new call processing measurements provided to support this feature.

Table C-4 Call Processing Measurements Used by Overload Control

Measurement	Description
CALLP_OLM_OFFERED	The total number of calls offered to OLM
CALLP_OLM_ACCEPT	The total number of calls accepted by OLM
CALLP_OLM_REJECT	The total number of calls rejected by OLM
CALLP_OLM_ACCEPT_MCL0	Calls accepted by OLM at MCL0
CALLP_OLM_ACCEPT_MCL1	Calls accepted by OLM at MCL1
CALLP_OLM_ACCEPT_MCL2	Calls accepted by OLM at MCL2
CALLP_OLM_ACCEPT_MCL3	Calls accepted by OLM at MCL3
CALLP_OLM_REJECT_MCL1	Calls rejected by OLM at MCL1
CALLP_OLM_REJECT_MCL2	Calls rejected by OLM at MCL2
CALLP_OLM_REJECT_MCL3	Calls rejected by OLM at MCL3
CALLP_OLM_REJECT_MCL4	Calls rejected by OLM at MCL4
CALLP_OLM_REJECT_EMERGENCY	Emergency calls rejected at MCL4
CALLP_OLM_MCL1_COUNT	Total number of MCL1 occurrences
CALLP_OLM_MCL2_COUNT	Total number of MCL2 occurrences
CALLP_OLM_MCL3_COUNT	Total number of MCL3 occurrences
CALLP_OLM_MCL4_COUNT	Total number of MCL4 occurrences
CALLP_OLM_ISUP_MSG_DUMPED	Number of ISUP messages dumped at MCL4 by layer 3/4 interface (MIM) due to system overload.

Service Interaction Manager Measurements

Table C-5 lists the new Service Interaction Manager measurements provided to support this feature.

Table C-5 Service Interaction Manager Measurements Used by Overload Control

Measurement	Description
SIM_OC_TRIG_FILTERED	The number of triggers dropped when the FS is overloaded (a single counter is used by SIM, which tracks the trigger filtering for all the FS). SIM updates this counter every time it filters a trigger due to congestion on a FS.
SIM_OC_EMG_TRIG_FORCED	The number of emergency triggers (that is, TRIGGER_911) forced when the FS is overloaded. A single counter is used by SIM which tracks number of emergency triggers forced for all the FS. SIM updates this counter every time it forces an emergency trigger (TRIGGER_911) to FS.
SIM_OC_TRIG_FORCED	The number of triggers forced when the FS is overloaded. A single counter is used by SIM which tracks the number of forced triggers for all the FSs. SIM updates this counter every time it forces a trigger.

Traffic Measurements Monitor Counters

Table C-6 lists the new Traffic Measurements Monitor (TMM) measurements provided to support this feature.

Table C-6 TMM Timers Used by Overload Control

Measurement	Description
SIA_OC_RX_INVITE_REJECT	The total number of incoming invite messages rejected by SIA due to overload.
SIA_OC_RX_REGISTER_REJECT	The total number of incoming register messages rejected by SIA due to overload.
SIA_OC_RX_REFERER_REJECT	The total number incoming refer messages rejected by SIP due to overload.
SIA_OC_RX_SUBSCRIBE_REJECT	The total number of incoming subscribe messages rejected.
SIA_OC_RX_UN SOL_NOTIFY_SUPP	The total number of unsolicited notification requests suppressed without being sent to the endpoints.
SIA_OC_RX_OPTIONS_REJECT	The total number of incoming options messages rejected by SIA due to overload.

Miscellaneous Measurements

Table C-7 lists additional measurements added to support Overload Control.

Table C-7 *Miscellaneous Measurements Used by Overload Control*

Timer	Description
ISUP_CONG_CALL_REJECTED	The congestion-rejected calls on a per trunk group basis. This is implemented for SGA.
POTS_OC_DP_RECEIVED	The number of Detection Points (DPs) reported during periods of congestion. This is being pegged by the FS.
H323_OC_SETUP_REJECTED	The total number of incoming H.225 Setup messages rejected by the Cisco BTS 10200 due to overload.
MEAS_ISA_OC_SETUP_REJECTED	The number of ISDN calls rejected due to system overload.
MEAS_MGA_OC_CALL_REJECTED	The number of MGCP calls rejected due to system overload.

Troubleshooting

This section lists the [Events and Alarms](#) added to support this feature.

Events and Alarms

The FS sends an alarm when:

- MCL changes.
- An individual critical factor reaches its threshold.

The CA sends an Informational alarm when:

- It receives a congested notification.
- It receives an abatement notification from an FS.

Informational alarms are sent at fixed 25 percent increments. A configurable parameter, `info_alarm_step_size`, is added to each factor defined in `olm.cfg`. Ensure that the value allows sufficient warning. The default for `info_alarm_step_size` is 5, giving factor informational alarms at 5, 10, and 15 percent, and so on.

Congestion Status—Maintenance (112)

The Congestion Status alarm (major) indicates that MCL changes have occurred, “System MCL Level.” The System MCL Level is the effective MCL or the greater of either the computed MCL or the administrative MCL.

When a new Maintenance (112) alarm appears, the old Maintenance (112) alarm is cleared. When the system MCL falls to 0, the alarm is cleared.

The alarm is dampened using the `alarm_damping_time` setting in `olm.cfg`. The value of `alarm_damping_time` is the minimum amount of time that passes before the alarm is issued after a change has occurred.

For additional information, refer to the [“Maintenance \(112\)” section on page 7-61](#).

CPU Load of Critical Processes—Maintenance (113)

The CPU Load of Critical Processes alarm (info) indicates that the MCL from the CPU utilization factor crossed a multiple of the `info_alarm_step_size`. This alarm appears for every crossing of the `info_alarm_step_size`, but it must pass the next higher or lower level before it is issued again.

For additional information, refer to the [“Maintenance \(113\)” section on page 7-62](#).

Queue Length of Critical Processes—Maintenance (114)

The Queue Length of Critical Processes alarm (info) indicates that the MCL for defined critical process queue length factors crossed a multiple of the `info_alarm_step_size`. This alarm is issued for every crossing of the `info_alarm_step_size`, but it must pass the next higher or lower level before it is issued again.

For additional information, refer to the [“Maintenance \(114\)” section on page 7-62](#).

IPC Buffer Usage Level—Maintenance (115)

The IPC Buffer Usage Level alarm (info) indicates that the MCL for IPC buffer usage factor crossed a multiple of the `info_alarm_step_size`. This alarm appears for every crossing of the `info_alarm_step_size`, but it must pass the next higher or lower level before it is issued again.

For additional information, refer to the [“Maintenance \(115\)” section on page 7-63](#).

CA Reports the Congestion Level of FS—Maintenance (116)

CA Reports the Congestion Level of FS alarm (info) shows CA received a congestion or abatement notification from an FS.

For additional information, refer to the [“Maintenance \(116\)” section on page 7-63](#).

Logs

Use the INFO logs to get differing levels of information about the alarms:

- INFO1—Are included with each alarm
- INFO3—Print factor feature controlled by `olm.cfg` which shows system overview
- INFO4—Have extra detail
- INFO5—Shows exact details of the factor MCL computations

