H.323 Gatekeeper-Based Load Balancing

First Published: November 17, 2006
Last Updated: November 17, 2006

The H.323 Gatekeeper-Based Load Balancing feature replaces random gateway selection with a selection process based on the capacity of the gateway (or endpoint). With this feature enabled, gateways with the capacity to support more calls than other gateways are routed more calls. Additionally, when the gateways in a zone are out of resources, calls are distributed randomly among all the gateways, and only not to gateways configured with priority 0.

Contents

Finding Feature Information in This Module

Your Cisco IOS software release may not support all of the features documented in this module. To reach links to specific feature documentation in this module and to see a list of the releases in which each feature is supported, use the "Feature Information for H.323 Gatekeeper-Based Load Balancing" section on page 13.

Finding Support Information for Platforms and Cisco IOS and Catalyst OS Software Images

Use Cisco Feature Navigator to find information about platform support and Cisco IOS and Catalyst OS software image support. To access Cisco Feature Navigator, go to http://www.cisco.com/go/cfn. An account on Cisco.com is not required.

Contents

Prerequisites for H.323 Gatekeeper-Based Load Balancing

- This feature is supported on the following platforms Cisco AS5*xxx*, 3*xxx*/2*xxx*,17*xx*, and 7200.
- Gateways that report call capacity (voice or H.323) of the trunks dynamically in RAS messages are preferred. If the gateways do not support this functionality, voice call-capacity must be statically configured in the Cisco gatekeeper. Load balancing can be achieved only if the gatekeeper receives periodic capacity updates from the gateway, or when the capacity of the gateway changes.
- It is assumed that each call will consume only one bearer channel.
- Capacity reporting from gateway in IRR, ARQ, and DRQ messages.
- Tech-prefix, carrier-based, and trunk group-based routing are supported.

Restrictions for H.323 Gatekeeper-Based Load Balancing

- Dynamic changes in capacities are not available to gatekeeper, which may affect load balancing. For instance, the gatekeeper may not know the exact utilization of capacity, when the DSO's are:
  - busied or shutdown
  - non-RAS calls
  - mixed protocols (SIP) are used in the gateway.
- Voice interfaces, such as Foreign Exchange Stations (FXS) or Ephone do not support load balancing.
- Load balancing depends on the capacity reporting from the gateways, which can be delayed. Therefore load balancing is achieved over a period of time, and not on a per-call basis.
- Load balancing is not supported for E.164 registrations.
- Configuration of **endpoint max-calls** command is not supported in conjunction with this feature.
- IP-to-IP gateway selection should be explicitly specified through invia and outvia configurations in the gatekeeper.
- If a gateway does not report call capacity, it will not be selected for termination when the load balancing is configured. Gateways that do not dynamically report the call capacity; configure the call capacity with the **endpoint max-calls h323id** command in the gatekeeper.

Note This configuration supports only voice call capacity. H.323 IP-to-IP call capacity is not supported.

- The gatekeeper does not reject calls when the gateway exceeds maximum call capacity. Call rejection, when needed, is managed by the gateway.
- The images used in the gatekeeper cluster should support load balancing for deterministic working. When only primary endpoints advertise their capacities periodically to other gatekeeper clusters, load balancing may be inaccurate.

Information About H.323 Gatekeeper-Based Load Balancing

This feature provides capacity-based gateway selection, and random gateway selection from all the gateways when an overload condition exists (except for gateways with priority 0). Load is calculated based on the number of calls being handled and the maximum call capacity of the gateway. The gateway with the least load is returned as the primary endpoint by the gatekeeper in ACF. Load balancing is extended to all types of call routing such as tech-prefix, carrier-based, and IP-to-IP gateway selection. Load balancing would be achieved over a period of time and should not be calculated on per-call basis.

To configure H.323 Gatekeeper-Based Load Balancing, you should understand the following concepts.

- Tech-Prefix Based Routing
- IP-to-IP Gateway Selection
- Carrier-Based Routing

Tech-Prefix Based Routing

Tech-prefix based routing is used for routing calls based solely on the dialed prefix.

IP-to-IP Gateway Selection

The gatekeeper selects IP-to-IP gateway in a round-robin fashion.

Carrier-Based Routing

Carrier-based routing allows the gatekeeper to route calls based on the destination carrier. The destination carrier can be received from GKTMP server, from the ARQ in a Cisco IOS IP-to-IP gateway, or non-Cisco and third-party devices. The gatekeeper sorts the gateway based on the maximum available calls capacity (absolute) for each carrier. An algorithm returns maximum available gateway as the primary, and the rest of the gateways are reported as alternate endpoints in the order of the available capacity. With the exception of barring the endpoints marked as out-of-resources, the gatekeeper doesn't apply percentage based load-balancing based on the capacities.

Note In carrier based routing with multiple destination carriers, if all the endpoints in the first carrier are out of resources, one endpoint from that carrier is returned as the primary endpoint based on the least load. The endpoints from other carriers are alternates.

How to Enable H.323 Gatekeeper-Based Load Balancing

The H.323 Gatekeeper-Based Load Balancing feature monitors the current capacities of the gateway when multiple gateways are registered to the same zone and have the routing capability to handle the same call.

This section describes the following tasks:

- Enabling Capacity Based Gateway Selection (required)
- Verifying H.323 Gatekeeper-Based Load Balancing (optional)

Enabling Capacity Based Gateway Selection

Enabling this feature distributes gateways based on capacity utilization. To enable the selection of the gateway based on the relative size of the gateway, perform the following steps.

SUMMARY STEPS

1. **enable**
2. **configure terminal**
3. **gatekeeper**
4. **call-routing hunt-scheme percentage-capacity-util**
5. **timer cluster-element resource-update** *seconds*
6. **exit**
7. **end**

DETAILED STEPS

	Command or Action	Purpose
Step 1	**enable** **Example:** `Router> enable`	Enables privileged EXEC mode. • Enter your password if prompted.
Step 2	**configure terminal** **Example:** `Router# configure terminal`	Enters global configuration mode.
Step 3	**gatekeeper** **Example:** `Router(config)# gatekeeper`	Enters gatekeeper configuration mode.
Step 4	**call-routing hunt-scheme percentage-capacity-util** **Example:** `Router(config-gk)# call-routing hunt-scheme percentage-capacity-util`	Selects a terminating gateway based on capacity, even when an overload condition exists.
Step 5	**timer cluster-element resource-update** *seconds* **Example:** `Router(config-gk)# timer cluster-element resource-update 4`	Dynamically sends capacity updates to an alternate gatekeeper in a cluster scenario.

	Command or Action	Purpose
Step 6	**exit** **Example:** `Router(config-gk)# exit`	Exits gatekeeper configuration mode.
Step 7	**end** **Example:** `Router(config)# end`	Exits global configuration mode.

Examples

```
Router> enable
Router# configure terminal
Router(config)# gatekeeper
Router(config-gk)# call-routing hunt-scheme percentage-capacity-util
Router(config-gk)# timer cluster-element resource-update 4
Router(config-gk)# exit
Router(config)# end
```

Verifying H.323 Gatekeeper-Based Load Balancing

Use the following **show** commands on both originating and terminating gateways to verify the implementation of the H.323 Gatekeeper-Based Load Balancing feature.

show gatekeeper status

In the following sample output the Hunt Scheme: Percentage Capacity Util line indicates that the load balancing feature is enabled:

```
Router# show gatekeeper status

    Gatekeeper State: UP
    Load Balancing:   DISABLED
    Flow Control:     DISABLED
    Zone Name:        GKB
    Accounting:       DISABLED
    Endpoint Throttling:        DISABLED
    Security:         DISABLED
    Maximum Remote Bandwidth:              unlimited
    Current Remote Bandwidth:              0 kbps
    Current Remote Bandwidth (w/ Alt GKs): 0 kbps
    Hunt Scheme: Percentage Capacity Util
```

show gatekeeper circuits

The following example shows sample debug output for maximum and utilized circuit capacity information for each trunk:

```
Router# show gatekeeper circuits


                    CIRCUIT INFORMATION
                    ===================
Circuit       Endpoint   Max Calls Avail Calls Resources     Zone
-------       --------   --------- ----------- ---------     ----
XYZ
              Total Endpoints: 2
              TGW-1
                                                  23         16          Available
              TGW-2
                                                  46         31          Available
```

Troubleshooting Tips

- Use the **debug gatekeeper dns 10** command to verify domain name server information found on the gatekeeper.

  To display all information related to the load balancing, the **debug gatekeeper dns 10** command must be enabled before the **debug gatekeeper endpoint 10** command.
- Use the **debug gatekeeper endpoint 10** command to verify that endpoint capacities are updated.
- Use the **debug gatekeeper gup asn1** command in a cluster scenario to verify that the new Gatekeeper Update Protocol (GUP) asn1 resourceUpdateIndication is used for updating the capacity of trunks or carries between the gatekeepers in the gatekeeper cluster.
- Use the **debug gatekeeper main 10** command to verify that the routing information used by the gatekeeper.
- Use the **debug h225 asn1** command to display the GUP events or Abstract Syntax Notation 1 (ASN.1) details.

Note For examples of **show** and **debug** command output and details on interpreting the output, see the following resources:

- *Cisco IOS Debug Command Reference,* Release 12.4T
- *Cisco IOS Voice Troubleshooting and Monitoring Guide*
- *Troubleshooting and Debugging VoIP Call Basics*
- *Voice Multiservice Debug Lookup*

Configuration Examples for H.323 Gatekeeper-Based Load Balancing

This section provides the following configuration examples:

H.323 Gatekeeper-Based Load Balancing: Example

```
gatekeeper
zone local GK1 cisco.com 10.13.32.184
zone cluster local BGL-CL1 BGL-GK1
element GK2 10.13.32.182 1719
element GK3 10.13.32.183 1719
!
zone prefix GK1 9362* gw-priority 8 5850-GW4 GW1-5350
zone prefix GK1 9362* gw-priority 2 GW2-3745 GW3-3845
zone prefix GK1 9845* gw-priority 6 5850-GW4
zone prefix GK1 9845* gw-priority 4 GW3-3845 GW2-3745 GW1-5350
accounting
timer cluster-element announce 10
timer cluster-element resource-update 10
no shutdown
call-routing hunt-scheme percentage-capacity-util
server registration-port 1111
```

Gateway Configuration for Trunk Groups: Example

The following sample ouitput shows trunk group configuration for Analog, ISDN, and CAS voice interfaces on the gateway.

```
.
.
.
 trunk group  tg
 carrier-id ABC

voice-port 2/1/0
 trunk-group tg

interface Serial2:23
 trunk-group tg

controller T1 1/0
ds0-group 1 timeslots 1-10 type e&m-fgd
 cas-custom 1
 trunk-group tg
.
.
.
```

Additional References

The following sections provide references related to the H.323 Gatekeeper Based Load-Balancing feature.

Related Documents

Related Topic	Document Title
General trunk and carrier-based routing	• *VoIP Gatekeeper Trunk and Carrier Based Routing Enhancements* http://www.cisco.com/en/US/docs/ios/12_2t/12_2t11/feature/guide/ftgkrenb.html • *VoIP Gateway Trunk and Carrier Based Routing Enhancements* http://www.cisco.com/en/US/docs/ios/12_2t/12_2t11/feature/guide/ftgwrepg.html

Standards

Standard	Title
No new or modified standards are supported by this feature, and support for existing standards has not been modified by this feature.	—

MIBs

MIB	MIBs Link
No new or modified MIBs are supported by this feature, and support for existing MIBs has not been modified by this feature.	To locate and download MIBs for selected platforms, Cisco IOS releases, and feature sets, use Cisco MIB Locator found at the following URL: http://www.cisco.com/go/mibs

RFCs

RFC	Title
No new RFCs are supported by this feature.	—

Technical Assistance

Description	Link
The Cisco Technical Support & Documentation website contains thousands of pages of searchable technical content, including links to products, technologies, solutions, technical tips, tools, and technical documentation. Registered Cisco.com users can log in from this page to access even more content.	http://www.cisco.com/techsupport

Command Reference

This section documents new commands only.

call-routing hunt-scheme

To enable capacity-based load-balancing, use the **call-routing hunt-scheme** command in gatekeeper configuration mode. To disable this function, use the **no** form of this command.

call-routing hunt-scheme percentage-capacity-util

no call-routing hunt-scheme

Syntax Description

percentage-capacity-util	Selects the gateway with lowest percentage capacity utilized.

Command Default

This command is disabled.

Command Modes

Gatekeeper configuration

Command History

Release	Modification
12.4(11)T	This command was introduced.

Usage Guidelines

Use the **call-routing hunt-scheme** command to turn on load balancing based on the capacity of the gateway and verify that gateway capacity reporting is enabled.

Examples

The following example shows how to enable capacity-based load balancing:

```
Router(gk-config)# call-routing hunt-scheme percentage-capacity-util
```

Related Commands

Command	Description
timer cluster-element	Sets the length of time between capacity messages to the local gatekeeper.

timer cluster-element

To configure the length of time between dynamic capacity messages to the local gatekeeper, use the **timer cluster-element** command in gatekeeper configuration mode. To stop sending dynamic updates, use the **no** form of this command.

timer cluster-element {**announce** | **resource-update**} *seconds*

no timer cluster-element

Syntax Description

announce	Configures the lengh of time between announcement messages to the gatekeepers in the local cluster.
resource-update	Configures the lengh of time between resource update messages to gatekeepers in the local cluster.
seconds	Number of seconds between resource updates sent to the gatekeeper. The valid range is 1 to 60. There is no default value.

Command Default

Disabled by default.

Command Modes

Gatekeeper configuration

Command History

Release	Modification
12.4(11)T	This command was introduced.

Usage Guidelines

Use the **timer cluster-element** command to manage the length of time between resource updates and time between announcement messages sent to the gatekeeper.

Examples

The following example shows the time between resource update messages to gatekeepers in local cluster being set to 20 seconds:

```
Router(config-gk)# timer cluster-element resource-update 20
```

Related Commands

Command	Description
call-routing hunt-scheme	Enables capacity-based load-balancing.

Feature Information for H.323 Gatekeeper-Based Load Balancing

Table 35 lists the features in this module and provides links to specific configuration information. Only features that were introduced or modified in Cisco IOS Release 12.4(Pi5) or a later release appear in the table.

For information on a feature in this technology that is not documented here, see the "Cisco IOS H.323 Feature Roadmap."

Not all commands may be available in your Cisco IOS software release. For release information about a specific command, see the command reference documentation.

Use Cisco Feature Navigator to find information about platform support and software image support. Cisco Feature Navigator enables you to determine which Cisco IOS and Catalyst OS software images support a specific software release, feature set, or platform. To access Cisco Feature Navigator, go to http://www.cisco.com/go/cfn. An account on Cisco.com is not required.

Note Table 35 lists only the Cisco IOS software release that introduced support for a given feature in a given Cisco IOS software release train. Unless noted otherwise, subsequent releases of that Cisco IOS software release train also support that feature.

Table 35 Feature Information for H.323 Gatekeeper-Based Load Balancing

Feature Name	Releases	Feature Information
H.323 Gatekeeper-Based Load Balancing	12.4(11)T	H.323 Gatekeeper-Based Load Balancing feature balances calls based on the capacities of the gateway distribution of the calls when the gateways in a zone are considered to be in an overload condition. The following commands were introduced or modified by this feature: **call-routing hunt-scheme and percentage and timer cluster-element**.

Glossary

GK—Gatekeeper.

GUP—Gatekeeper Update Protocol

GW—Gateway.

IRR—Information request.

GKTMP—Gatekeeper Transaction Message Protocol.

RAI—ResourcesAvailableIndicate.

ARQ—Admission request.

LRQ—Location request.

CME—Cisco Call Manager Express.

IPIPGW—IP-to-IP call functionality on VoIP gateways.

Note See *Internetworking Terms and Acronyms* for terms not included in this glossary.