Google Global Cache Beta

Installation and Operations Guide

Revised: June 23, 2011

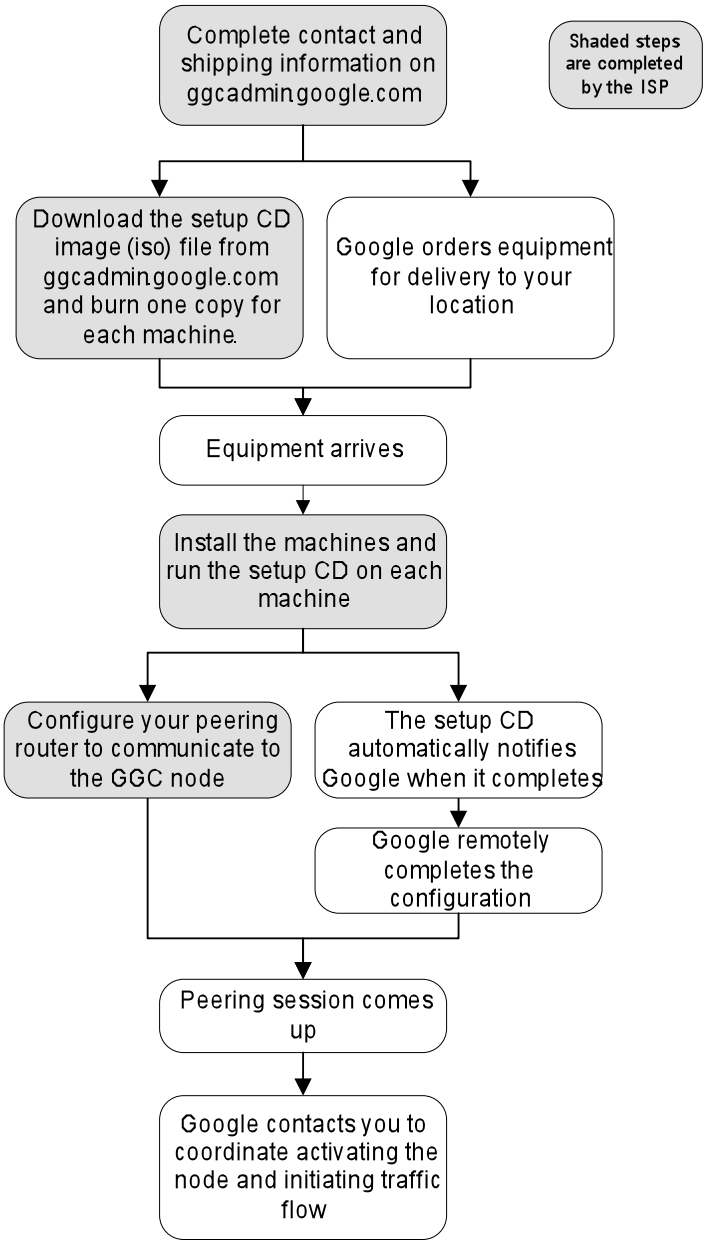
Contents

Summary

This document describes how to install a Google Global Cache (GGC) node. Please see the *Google Global Cache Application Note and Technical Overview* for general information on the GGC node.

Installation and Commissioning Process Overview

The following diagram shows the high level steps involved in the deployment and commissioning of a Google Global Cache node. The shaded boxes are steps that are completed by the cache host. This document will provide additional detail required to complete each step.



Hardware Installation

You will need

- Rack mount installation kit and instructions (included in server box)
- Phillips (crosshead) screwdriver
- Help to lift servers into position
- Copper Ethernet cables, 2 per server
- Dual AC power feeds (see table below).
- Dual C13-C14 power cords (provided with each server)

# Servers	Rack Space	Nominal Power	Peak Power	Amps @ 110 VAC	Amps @ 220VAC
4	8RU	1400W	1600W	15A	8A
8	16RU	2800W	3200W	30A	15A

Procedure

1. Install the servers in the rack according to the included instructions.
2. Connect the network switch to the port labeled Gb1 on each server. Do not connect the Gb2 port at this time.
3. Connect power, but do not turn the system on yet.

Both power supplies must be connected. It is strongly recommended that you connect each power supply to an independent power feed (ie: A and B power). However, both can be connected to the same circuit to protect from a power supply failure.

Switch Configuration

You will need

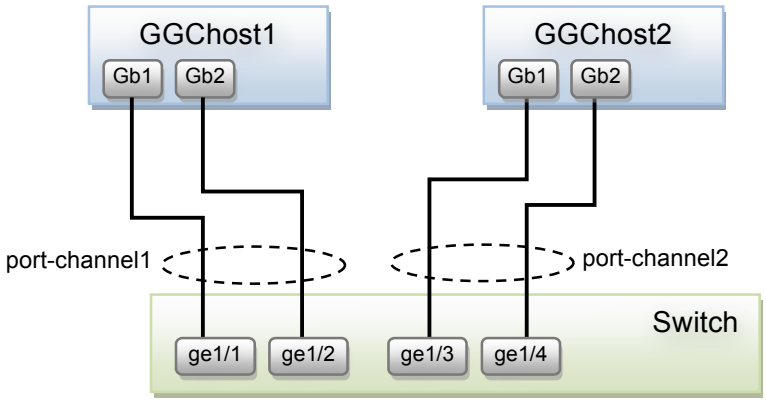
- Administrative access to the Ethernet switch connected to the GGC servers
- Switch port numbers connected to the GGC servers

Procedure

Please refer to your switch's documentation for the specific commands to configure the Ethernet ports facing the GGC machines as follows:

- 1000Mbps full duplex
    - Set to auto-negotiate
- Link Aggregation Control Protocol (LACP)
    - Passive mode
    - Load balanced on source/destination layer 3 information
    - Switch/Layer 2 mode
- All machines in the GGC node must be in a single layer 2 broadcast domain

Note: for the initial installation, LACP will be enabled, but only a single interface on each server will be connected to the switch. After the software installation step (below) is complete, the Gb2 interfaces will be connected and will automatically aggregate with the Gb1 interfaces.



Switch Configuration Examples

The following example is for illustration purposes only. Your configuration may vary. Please contact us if you require additional support.

Cisco Switch Configuration Fragment

```
!
interface GigabitEthernet1/1
 description GGChost1-Gb1
 switchport mode access
 channel-protocol lacp
 channel-group 1 mode passive
!
interface GigabitEthernet1/2
 description GGChost1-Gb2
 switchport mode access
 channel-protocol lacp
 channel-group 1 mode passive
!
interface Port-channel1
 description GGChost1
 switchport
 switchport mode access
!
interface GigabitEthernet1/3
 description GGChost2-Gb1
 switchport mode access
 channel-protocol lacp
 channel-group 2 mode passive
!
interface GigabitEthernet1/4
 description GGChost2-Gb2
 switchport mode access
 channel-protocol lacp
 channel-group 2 mode passive
```

```
!
interface Port-channel2
 description GGChost2
 switchport
 switchport mode access
end
```

IP Addressing

The GGC node requires a dedicated layer 3 subnet. The size of the subnet is set forth in the table below. Each server has a management IP address statically assigned to the Ethernet interface (or bonded set of interfaces in the case of LACP) as well as a number of virtual IP addresses (VIPs). User traffic is served from the VIPs. In the event of a server failure, other servers in the node pick up the failed server's VIPs.

# Servers	Subnet Size	Subnet Mask	Number of host IPs
4	/26	255.255.255.192	62
8	/26	255.255.255.192	62

Rules for the assignment of IP addresses with the GGC subnet are described below. The general guidelines are:

1. Assign the subnet gateway to the first address in the subnet.
2. If required, use the next addresses for HSRP or GLBP.
3. Assign the 4th IP address in the subnet to the Gb1 interface of the first server.
4. The server addresses must be contiguous. Do not assign or reserve addresses for the Gb2 interfaces as these will be aggregated with the Gb1 interfaces via LACP.
5. The 16th address in the subnet is reserved for the first Virtual IP. This address is used as the source for the BGP peering session with your network.

Address Number	Use	8 Server Node Example
0	Subnet address	10.10.10.64/26
1	Gateway set on cache servers	10.10.10.65
2,3	HSRP/GLBP gateways (optional)	Unused
4	First GGC server (configured manually during setup)	10.10.10.68
5	Second GGC server (configured manually during setup)	10.10.10.69
…continue for each server	Last GGC server (configured manually during setup)	10.10.10.75
(Last server+1) to 15	Reserved	10.10.10.76 – 10.10.10.79
16	First virtual IP – Used for BGP peer (configured remotely by Google)	10.10.10.80
16 to (last address - 1)	Additional virtual IPs (configured remotely by Google)	10.10.10.81 – 10.10.10.126
Last address in subnet	Broadcast	10.10.10.127

Server Naming / Reverse DNS

Please configure reverse DNS entries for all servers' IP addresses to **cache.google.com**

Proxies and Filters

No transparent proxies or filters may be placed in the path of communications between the GGC Node and Google's back-end servers.

Running the Setup CD

This section describes the steps to install the initial setup software on the machine. After completing this step the machine will automatically signal to Google to remotely begin the next step in the process.

You will need

- Copper Ethernet cable (straight through)
- Laptop with a 10/100/1000 copper Ethernet port and a web browser
- GGC Setup CD burned from the ISO image available on the GGCadmin portal (Installation goes fastest if you have one CD for each GGC server)
- IP information provided to Google in the GGCadmin portal
- Labels to mark the IP address on each server
- (optional) If desired, a local monitor can be used to observe the boot process.

Procedure

For each server:

1. Turn on the server from the power switch.
2. Insert the setup CD into drive.
3. Restart server from the power switch. To shut down, press and hold the power button and release when the system has been powered off.
    a. Please note, you should only force power down before the software has been installed. When software has been installed, please shut down safely by pressing the power button once, release immediately, and wait for the server to power off safely.
4. Power up the server once it has powered down successfully. The server will boot, perform its power-on self tests, and boot from the image. **This process will take 10-15 minutes.**
5. Connect the laptop's Ethernet port to the server's Gb2 port as shown below.

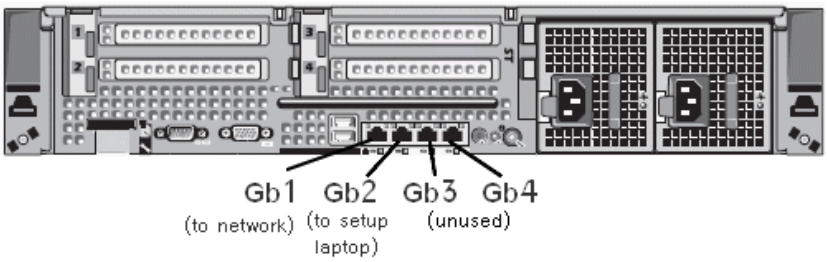


Figure 1: Server Connections During Setup

a. After the server has booted, the laptop should get an IP address from the DHCP server running on the Gb2 interface. If it does not, please release and renew the laptop IP address.

b. The DHCP server will only be active during the setup process. After setup is complete, the server will boot from the hard drive without starting any DHCP service.

6. Start a web browser on your laptop and navigate to http://**192.168.1.1**

   a. You should see the **Google Global Cache Setup** screen (Figure 2 below).

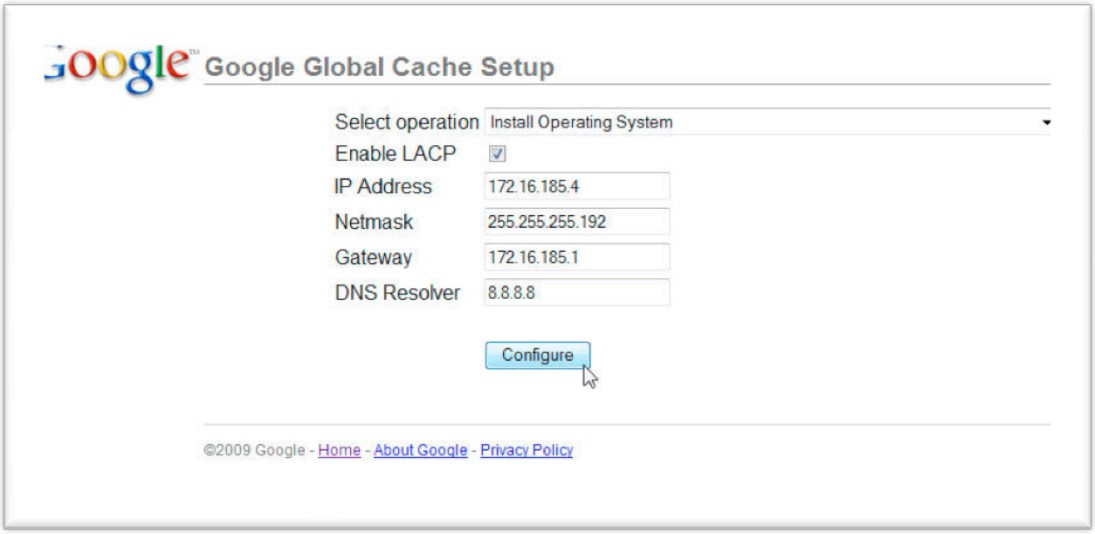


Figure 2: Google Global Cache Setup Page (in laptop browser)

7. Use the web interface to select the operation **'Install Operating System'**

   a. Be sure **'Enable LACP'** is checked unless you have been asked to uncheck it.

8. Follow the onscreen instructions for entering the IP information. This must match the information provided to Google on the GGCadmin portal.

   a. Please assign the first address(es) in the subnet to the gateway, followed by the first server IP address, then the second, etc... Server IP addresses must be contiguous. (See **IP Addressing** section above)

   b. **'DNS resolver'** must be set to 8.8.8.8. This is only used during this phase of the setup process.

c. The server will attempt to reach Google's network over the Gb1 port. If this connectivity fails, please validate the IP information, DNS server, and connectivity from Gb1 to www.google.com.

9. Label each server with the IP address you have assigned to it.
10. Upon validation of the IP information and connectivity, the server will begin final disk configuration.
    a. This step will take approximately 20-30 minutes. The "Overall install progress" is an approximation and may linger at various points. Please be patient and allow it to finish.
11. When the process is complete, the CD will eject and the server will reboot.
    a. If the setup process encounters an error after network connectivity is established, it will automatically report the error to Google for investigation. If this happens, please leave the server running with the CD in the drive for remote diagnostics.
12. If LACP was selected, after a successful installation, connect each server's Gb2 port to the same network switch as the Gb1 port. They should automatically form a LACP bundle. Gb3 and Gb4, if equipped on your servers, are not used.

Once the field setup process is complete, the setup CD will automatically report the configuration to Google so that the installation can be remotely completed and the node brought on-line.

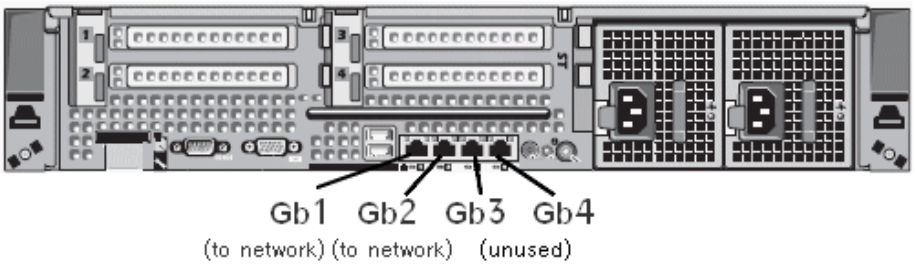
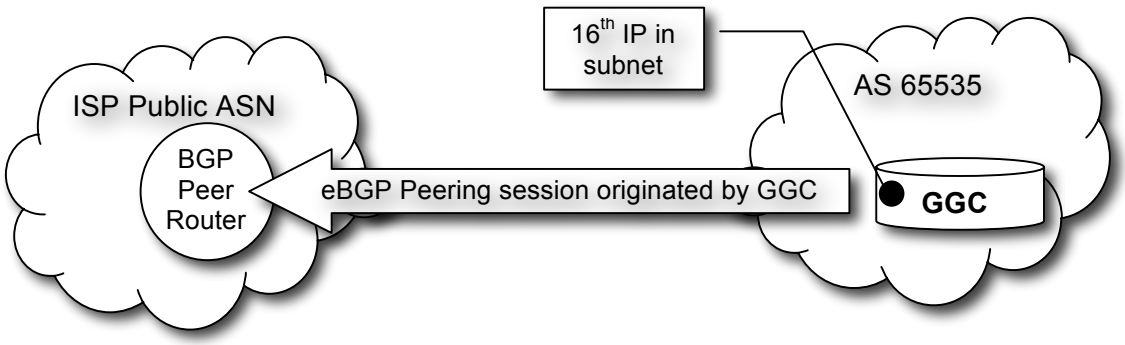


Figure 3: Server connections after setup

BGP Configuration



You will need

- IP address of the BGP peer router
- Administrative access to the BGP peer router

Note: only a single session is permitted to each GGC node. Redundancy is not required as an interruption this session will not impact traffic flow at the node.

Procedure

1. Your end of the session will be the router specified in the GGC admin portal
    - Use your public ASN (as provided in the GGCadmin portal) for your end of the eBGP session.
    - BGP multihop is permitted
2. The GGC end of the session will be the **16th usable IP address in the GGC subnet**.
    - This is a virtual address which will be configured after Google completes the remote installation.
    - The GGC ASN is always **65535**.
3. The session should be configured in passive mode. The connection is always initiated by the GGC end.
    - The session will not come up until Google completes the next step of the installation.
4. Do not configure monitoring on the BGP session
    - The GGC system does not interpret an interruption of the BGP feed as a loss of the node. The node will continue to serve based on the most recent valid feed received until the session is restored. Google will monitor the availability of the node and automatically shift traffic away in the event of an outage. Normal management activity may briefly interrupt the session at any time.
5. For configuration simplicity, MD5 signatures are not recommended. However, MD5 can be supported if required.

What to Announce Over the Peering Session

Google Global Cache uses BGP only as a mechanism to communicate the list of users that should be served from a node. It is not used for routing or to determine if the cache is online. An interruption to the BGP session has no effect on the cache.

User and Resolver Prefixes

In order for requests from a user to be served from the cache, **both** the IP address of the **user** and the IP address of the **DNS resolver they are using** must be advertised to the cache. If either of these IPs are not in the advertisement, the request will be served outside your network.

- DNS resolver IPs are used to map requests to the cache node.
- User IPs are used to build an access control list on the node itself.
- GGC will ignore any /32 prefixes

Peers and Downstream ASNs

Prefixes from other ASNs can be mapped to the cache, providing the following conditions are met:

- Both the DNS resolver and user prefixes must be advertised in order to both map and serve those users from the cache.
- If the other AS transits an AS with a peering relationship with Google, their traffic will not be mapped to the cache. If an exception is required, please contact ggc@google.com.
- Do not send the full Internet routing table to the GGC node. Only send the prefixes that should be served from the node.

Private Addresses

Only public IP addresses can be advertised to the node. Private addresses (ie: RFC 1918) will be ignored.

IPv6

IPv6 is not supported by GGC at this time.

Multiple Cache Nodes

There are two configuration options available when multiple cache nodes are deployed in the same network. These are described below. Please indicate your configuration preference to the GGC support team (ggc@google.com).

Users load balanced over multiple cache nodes

- If traffic can be load balanced across multiple cache nodes, send the same BGP advertisements to all nodes.
- Failure of one node will send traffic to the remaining node(s). If the load exceeds the capacity of the remaining node(s), it will overflow to caches on Google's network.

Users directed to a specific cache node

In some cases, network topology dictates that it is best to serve specific sets of users from specific cache nodes.

- Advertise the user and resolver prefixes to the cache node they should prefer.
- If a DNS resolver serves users from multiple nodes, advertise that resolver to the cache node that will serve the majority of the users. Advertise the user prefixes to the specific nodes where they should be served. Users will be redirected to the desired node.
    - For the best user experience, it is strongly recommended to provide dedicated DNS resolvers for each cache node.
- Do not advertise the same prefix to both nodes as the system will not be able

to determine the preferred location.
    - Overlapping prefixes are accepted. In this case, the more specific advertisement will determine the preferred node.
- Failover is configured at the node level. In the event of a failure, all prefixes advertised to the failed node will be served from the designated backup node or from Google.
    - Please indicate your failover preference to the GGC support team.

BGP Peer Configuration Examples

The following examples are for illustration purposes only. Your configuration may vary. Please contact us if you require additional support.

Cisco Option 1: Prefix list based route filtering

```
neighbor <IP address of GGC> remote-as 65535
neighbor <IP address of GGC> transport connection-mode passive
neighbor <IP address of GGC> prefix-list deny-any in
neighbor <IP address of GGC> prefix-list GGC-OUT out

ip prefix-list deny-any deny 0.0.0.0/0 le 32

ip prefix-list GGC-OUT permit <x.y.z/24>
ip prefix-list GGC-OUT permit <a.b.c/24>
```

Cisco Option 2: AS-PATH based route filtering

```
neighbor <IP address of GGC> remote-as 65535
neighbor <IP address of GGC> transport connection-mode passive
neighbor <IP address of GGC> filter-list 1 in
neighbor <IP address of GGC> filter-list 2 out

ip as-path access-list 1 deny .*

ip as-path access-list 2 permit _100_
ip as-path access-list 2 permit _200$
ip as-path access-list 2 permit ^300$
```

Juniper Option 1: Prefix based policy

```
neighbor <IP address of GGC> {
            description "GGC";
            import no-routes;
            export export-filter;
            peer-as 65535;
            passive;
}

policy-statement no-routes {
        term default {
               then reject;
            }
}
```

```
    policy-statement export-filter {
          term allow-routes {
              from {
               route-filter a.b.c.d/xy orlonger;
              }
           then accept;
              }
    }
```

Juniper Option 2: AS-PATH based policy

```
    neighbor <IP address of GGC> {
                description "GGC";
                import no-routes;
                export export-filter;
                peer-as 65535;
                passive;
    }

    policy-statement no-routes {
           term default {
                   then reject;
                }
    }

    policy-statement export-filter {
              term allow-routes {
              from {
                     from as-path-group GGC;
              }
              then accept;
              }
    }

    as-path-group GGC {
       as-path AS-PATH-NAME-1 "^100.*";
       as-path AS-PATH-NAME-2 "^200.*";
    }
```

Operations and Troubleshooting

Shutdown and Traffic Drain

In the event you need to shut the node down for scheduled maintenance of your data center or network, we ask that you inform us ahead of time so that we know the downtime is expected. You may still receive automated alert messages generated by our monitoring systems. If the outage is temporary and expected, these can safely be ignored.

You can perform a graceful traffic drain by shutting down the Ethernet interfaces facing all but one of the servers. When the Google monitoring system detects only a single server is reachable, the DNS system will stop handing out the node's IP addresses. Due to DNS

response caching, it can take up to 30 minutes for user traffic to fully drain away from the node. However, the one remaining server will redirect excess load, so no users will be denied service while the drain is taking effect.

If you need to power down a GGC machine, you can initiate a graceful shutdown by pressing the server's power button once. There will be a several second pause before the system shuts down.

To restore traffic to the node, simply re-enable all switch interfaces or restore power to all machines. Once the monitoring system detects more than one machine is reachable, the DNS system will once again begin handing out the node's IP addresses.

Hardware Monitoring and Repair

Google's monitoring system will remotely identify hardware failures. Your technical contact will be notified if we require any local assistance, troubleshooting, or RMA coordination. If you believe that hardware is not operating properly, please contact us at ggc@google.com.

Please keep the technical contacts section of the ggcadmin.google.com portal current. Our support team will rely on this information in the event we need to contact you.

Local Monitoring

While no monitoring is required, some local monitoring can be helpful. However, it is important to understand the following considerations:

- It can be helpful to monitor the availability and performance of the path between the GGC node subnet and Google's network. A sample host for video content origin is v1.cache1.googlevideo.com.
- If you monitor egress traffic from the node, bear in mind that traffic at the node will be impacted by youtube.com maintenance and availability.
- Binary and configuration changes are regularly pushed to machines in the node in a rolling fashion. If you are monitoring egress per machine, you will see occasional interruptions of service during the associated restarts. The GGC software ensures that the load for the machine under service is spread around the node during these events.
- The BGP session to the node is different from typical peering sessions. It is not used to establish the availability of the node. Brief interruptions of the session are normal and will not impact user traffic. If you are monitoring this session, you should not consider it an actionable alert unless the session is down for longer than an hour.

Playing a Test Video

Using Firefox, select and play a popular video from the www.youtube.com homepage. [Note: The base web pages of www.youtube.com may not be served from the cache. These host names will typically **not** resolve to the GGC node.] Observe the status bar at the bottom left of the page. You should see a message such as "*Transferring data from*

<*server host name*>". The actual server host name will vary depending on the content selected, but it will typically be similar to v1.lscache1.c.youtube.com.

Using nslookup or a similar tool, resolve the server host name

```
$ nslookup v1.lscache1.c.youtube.com
Server:         <your name server>
Address:        <your name server IP>

Non-authoritative answer:
Name:   v1.lscache1.l.google.com
Address: <server IP address>
```

If <server IP address> is in the subnet allocated to the GGC node, the video is playing from the cache. If this is not the case, see below.

Videos Not Playing From the Cache

There are several possible reasons why a video may not play from the cache.

The user's DNS resolver is not in the BGP feed to the GGC node.

The GGC system uses DNS to send the request to the node. The DNS request from the user will go to your resolver, which will then come to Google's authoritative resolvers. If this resolver's IP address is in a prefix that is being advertised to the node, the IP address returned to your DNS resolver (and then to the user) should be from the GGC node subnet. To determine the resolver Google is seeing, execute the following command from your test client:

```
nslookup -q=txt o-o.myaddr.l.google.com
```

This is a special host name that will return the IP address of the DNS resolver as seen by Google. You should see a response similar to:

```
Non-authoritative answer:
o-o.myaddr.l.google.com.google.com
text = "<IP address>"
```

Confirm that <IP address> is in the BGP feed to the GGC node. A common error is using a test resolver that relays requests through another resolver not in the BGP feed.

The mapping file is updated periodically, so if the address was added to the feed within the last 24 hours, please contact ggc@google.com to confirm that the change has been pushed to our production servers.

The client's IP is not in the BGP feed to the GGC node

If the DNS server is properly mapped, but the request is still not playing from the cache, it is possible that the BGP feed does not include the test client's IP address. If this is the case, the cache will get the request and then redirect it to a cache outside your network. Verify that the test client's IP address is in the feed and has been there for at least 1 hour.

The cache is overloaded and overflowing

If the same video plays from the cache sometimes, but not every time, the cache may be overflowing. As the cache reaches its configured serving capacity, it will begin redirecting requests to external caches. The service capacity of the cache is based on a combination of several factors:

- The number of servers in the node
- The number of interfaces connected on each server
- The amount of bandwidth provisioned between the cache and your network (reported on the GGCadmin portal)
- Manually configured limits

You can determine if the cache is overflowing by reviewing the Traffic graph on the GGCadmin portal (see below). If you suspect that the node should not be overflowing at the current traffic level, it is possible that an out of date limit is configured. Contact ggc@google.com to confirm that the capacity is set correctly.

The video is not popular enough to be in the cache

The cache will store the most popular videos your users are requesting. There is an admission mechanism that can prevent a video from being cached on the first play. If this is the case, a second playback should come from the cache.

Node Status in the GGCadmin Portal

During the installation process, the Status tab in the GGCadmin portal will provide information on the fulfillment, shipping, and turn-up status of the node. After the node is activated, the Status tab will provide several graphs that can be useful for evaluating node performance.

Traffic and Global Traffic Graphs

The **Traffic** graph shows ingress and egress traffic. Ingress traffic is cache fill coming from Google's origin servers. Egress is traffic from the cache sent towards the users. The ratio between the two will show you the cache's efficiency. If the egress line appears flat at the peaks, it is possible that there is not enough cache capacity and traffic is overflowing. Please contact ggc@google.com to find out what can be done.

The **Global Traffic** graph shows the global egress traffic from the Google Global Cache network. This graph is useful in determining if a traffic interruption at your node was part of a larger, global event. If so, then the most common explanation is an error or maintenance on the YouTube.com site, which can instantly reduce global demand from the caches.

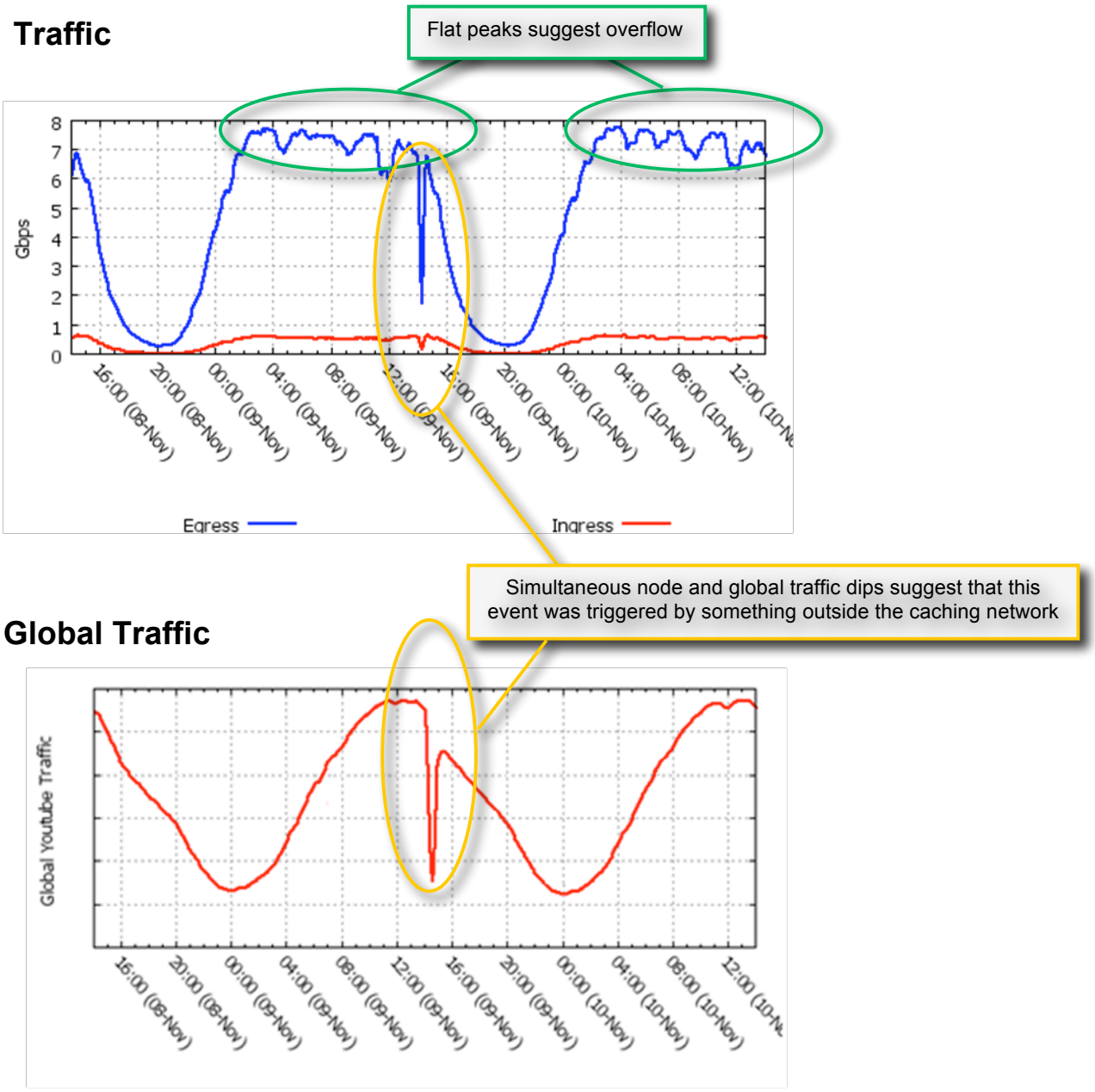


Figure 4: GGCAdmin Traffic Status Graphs

Packet Loss

The **Packet Loss** graph shows transmit packet loss (measured by retransmissions) from the node. If you are observing slow video playbacks or significant rebuffering, this graph can tell you if the node is having difficulty reaching your users. High packet loss is most often caused by congestion or faults in the access network between the cache and the users. The issue can sometimes be traced to a network bottleneck, such as improper load balancing across aggregated links, a faulty circuit, or a malfunctioning interface.