

Resonate **Global Dispatch™**

Multi-Site Service Level Control

***Technical White Paper Series
2004***



Copyright © 2001-2004 Resonate, Inc. 08/2004

Abstract

Intranets, extranets, and the Internet are the lifelines of modern enterprises as mission-critical business activities become more dependent on geographically dispersed information, services, and computing resources. Across the global corporate network or over the Internet, employees and customers expect fast, reliable service for electronic mail (email), electronic commerce (e-commerce) and a host of other essential IP-based applications. But as this distributed environment becomes more indispensable to business operations, its expanding complexities – from applications to geographies, networking infrastructures, server systems, and user bases – make its deployment and management increasingly difficult. Web masters, system/network administrators, and business managers struggle to match existing system resources with client requirements while trying to prepare for imminent and exponential increases in demand for global applications and services.

Resonate Global Dispatch™ is a WAN-based, multi-site, service level control solution that provides high availability and optimum performance for geographically distributed e-business applications. It enables multiple, dispersed points of presence (POPs) to act as a single, highly available system and directs user requests to the POP best able to maximize service levels. Additionally, it can enhance a user's experience by routing them to the POP with the content most relevant to their needs or it can direct intranet users to the POP closest to their location to save on WAN costs.

Table of Contents

<i>Resonate Global Dispatch Overview</i>	4
Introduction	4
Site Availability.	5
Response Time	5
Site Load.	5
Persistence.	5
Directed Traffic Tables	5
Secure Socket Layer	6
Multi-Homed Servers	6
<i>Intelligently Managing Global Traffic</i>	6
Global Dispatch Scheduler	6
Using Directed Traffic Table.	6
Using Persistence Table.	7
Quick Response to Requests.	9
Shadow Mode: Testing Configuration Before Deploying.	9
Integration with Authoritative DNS Server.	9
<i>Gathering Valuable Site Metrics</i>	11
Global Dispatch Agents	11
Load Computation	11
<i>Easy Configuration and Management</i>	13
Global Dispatch Manager	13
<i>Bringing it all Together</i>	15
<i>Global Service Level Control</i>	16

Resonate Global Dispatch Overview

Introduction

By gathering and using latency, load, and availability information, Global Dispatch software directs client requests to the physical site best suited to respond.

Global Dispatch combines multiple physical sites into a single (or more, if needed) virtual hostname such as `www.company.com` or `company.com` (figure 1). Clients use the virtual hostname to access resources that may be available across multiple physical sites. When a client's local DNS server makes an address resolution request for a virtual hostname, Global Dispatch responds with the IP address of the most appropriate physical site based on the latency and load information it receives from Global Dispatch agents installed at each site.

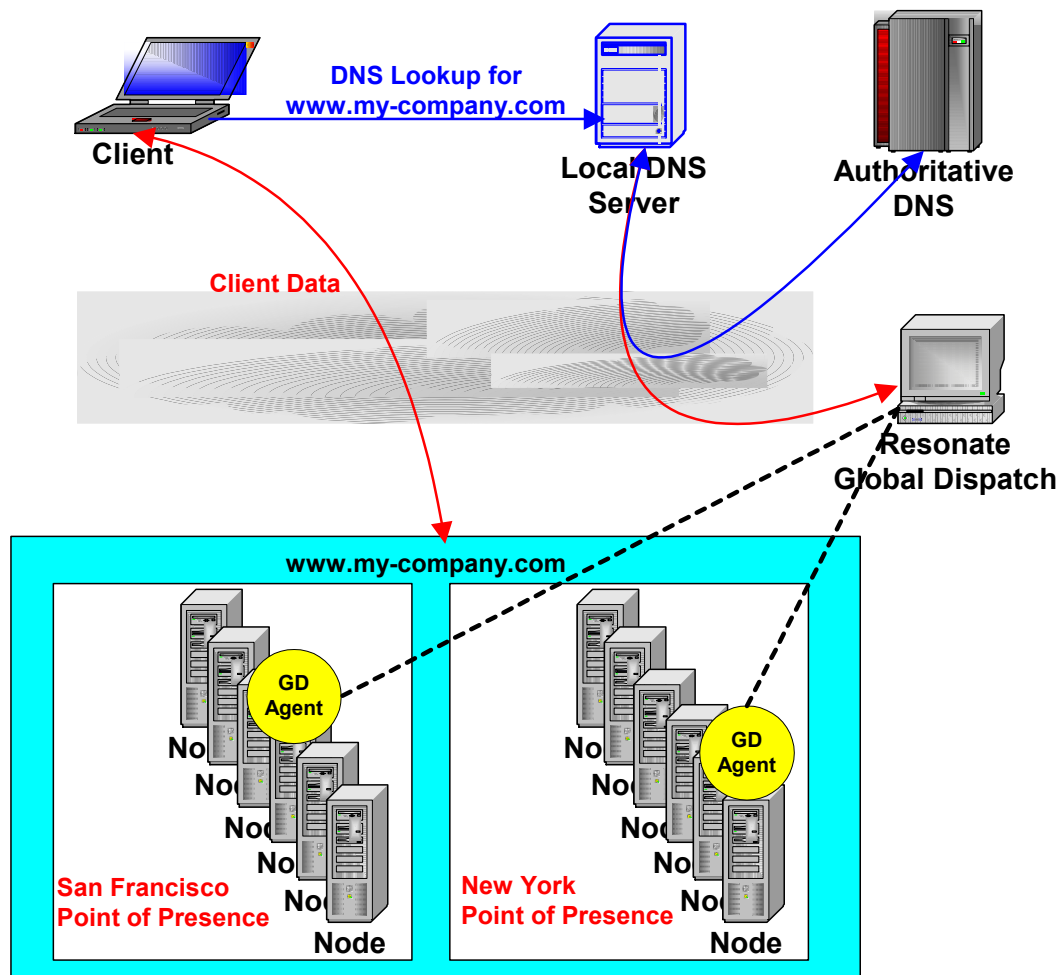


Figure 1 - Global Dispatch configuration.

Global Dispatch uses the following metrics and policies to determine the site best suited to respond to client requests:

Site Availability.

Global Dispatch actively monitors the availability of each site through the Global Dispatch agents and feedback from the Central Dispatch agents if Central Dispatch is used as the local traffic management solution. Requests are routed around failed sites and are only sent to available resources. In the event of a downed site, clients are rerouted until the site becomes available again, at which time it is automatically re-integrated into the list of available sites. For example, if a database is unavailable at a particular site but the site is still available, Global Dispatch software can be configured to automatically direct users to alternate sites where the required resources are available.

Response Time

Global Dispatch measures the network latency between the client's local DNS server and each site. Through triangulation between the client local DNS server, the Global Dispatch Scheduler and Global Dispatch Resonate Global Dispatch: Multi-Site Service Level Control 4 agent, the client is sent to the site with the fastest response time (i.e., the smallest latency).

Site Load.

Measures the load at each site and distributes incoming requests to the least-busy sites. Integration with Central Dispatch enables more precise load statistics to be captured, such as CPU load across the site and other custom-defined criteria allowing administrators to address specific load requirements needed in their environment.

Persistence.

For a designated period of time from the first request, all subsequent requests from the same client can be assigned to the same site that handled the initial request. This is called a "sticky session" and accommodates requests that generate and utilize state information (typically storing it in the server's database). With persistent session support, server databases need only be synchronized periodically, avoiding the high overhead and cost associated with real-time synchronization of content.

Directed Traffic Tables

Global Dispatch enables designated traffic to be directly routed to a specified site, avoiding the standard tests for the best-suited site. For example, a Web site administrator may want to designate a particular physical site to handle all requests from customers of a particular ISP, a particular region of the world, or a particular company.

The intelligent scheduling capabilities and an easy-to-use management console enable administrators to efficiently direct requests across multiple sites and also accommodate diverse

company requirements. Administrators gain control over geographically dispersed traffic to deliver the best experience to its users and to save costs by optimally allocating server resources and reducing traffic on private backbone WANs.

To effectively manage WAN traffic and server resources, Global Dispatch software integrates three major components: a Scheduler that receives and resolves client requests; agents that monitor and communicate physical POP status; and the Global Dispatch Manager console that allows configuration and gathering of site statistics. These three components integrate tightly to deliver a responsive system to maximize service levels for globally distributed traffic.

Secure Socket Layer

Beside the standard Connect and URL Connect, Global Dispatch supports SSL Connect, which test availability of a pop for a secure host.

Multi-Homed Servers

In the case of multi-home servers, Global dispatch can binds to all internet addresses available on this particular machine or just one interface specified by the user.

Intelligently Managing Global Traffic

Global Dispatch Scheduler

Global Dispatch Scheduler (item C in Figure 1) ensures that traffic is directed to the most available and appropriate site. The Global Dispatch Scheduler takes the client requests for the virtual hostname through the client's local DNS server and DNS hierarchy. It then directs the client to the most appropriate physical site by first determining the optimal site and then returning the IP address (or Virtual IP address) of that site.

A single IP address can be returned for the address resolution or multiple IP addresses can be returned for failover. The Scheduler provides the intelligence to route clients around failed sites or to direct requests based on predefined rules and policies. These predefined rules allow Global Dispatch to send clients to sites that are closest in proximity, least loaded or offering the best response time. Global Dispatch observes the following policies, in order, when directing clients to the most appropriate site (shown in Figure 2):

Using Directed Traffic Table.

If any directed traffic table entries have been put in place (e.g., all traffic from a designated IP address range mapped to a specific server), then these rules are first used to arrive at a target IP address.

Using Persistence Table.

The Scheduler then checks the persistence table at this point to determine if a "sticky session" was previously created that still applies to this request. The persistence tables can be synchronized across Schedulers at administrator-defined intervals to ensure session state is maintained.

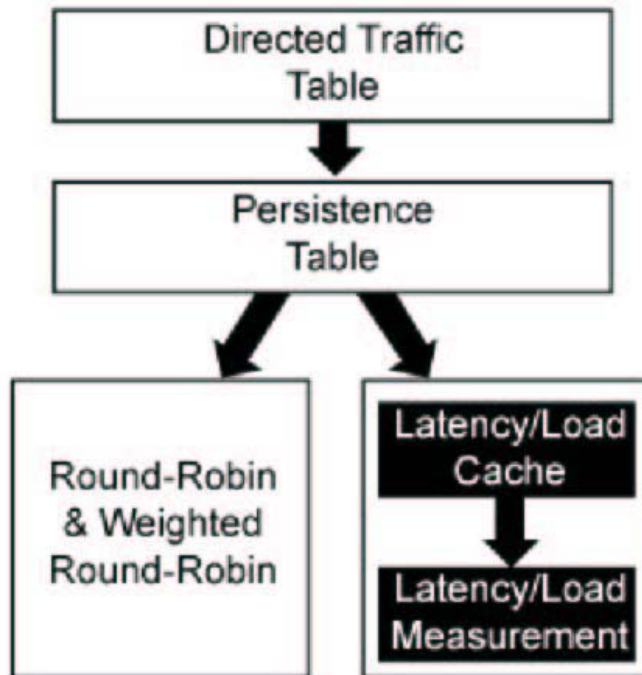


Figure 2 - Traffic Scheduling decision flow

Using Load Balancing Rules.

If neither persistence or directed traffic features have been enabled, the Scheduler chooses a physical POP based on the latency and load information it receives from the Global Dispatch agents installed at each POP. The following load balancing rules can be selected in Global Dispatch (shown in Figure 3):

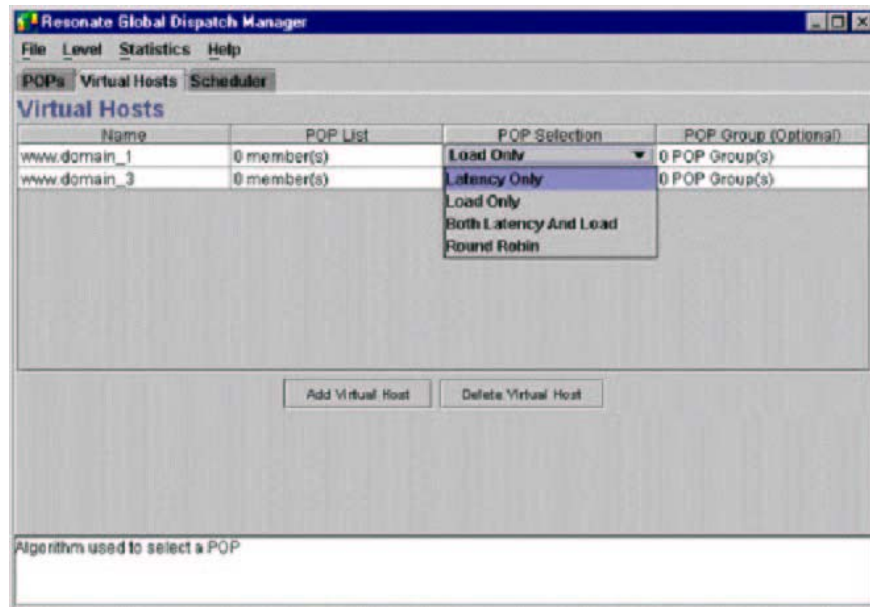


Figure 3 - Load balancing rules.

Basic Round-Robin: sic round-robin selection of sites based on all available sites.

Weighted Round-Robin. Round-robin selection of sites adjusted using the site weights to send proportionally more traffic to POPs with a higher weight value. This helps if some sites have a higher capacity and ability to respond to client requests.

Latency Only. Global Dispatch chooses the POP with the lowest latency value unless the POP is unavailable. To maximize performance, the Scheduler maintains a cache in RAM where it stores latency information for client DNS servers. When the Scheduler gets a request from a client, it first checks the cache for saved latency information about that particular DNS. The cache is automatically flushed at default or user-specified intervals. If the request results in a cache miss, then the Scheduler generates requests to the Global Dispatch agents for measuring the latencies between the agents and the requestor's local DNS.

Load Only. When using the load only method, Global Dispatch uses an algorithm that calculates a quota and the ability for a POP to service a request. The Scheduler assigns a quota to each POP participating in a virtual host site. This quota will be dynamically modified by the changes of the load – so the higher the quota, the better its ability to serve traffic. The quota will be higher if the site load is low and the quota will be lower if the site load is high.

Both Latency and Load. Global Dispatch sorts the POPs based on their latency value (in order of increasing latency). It will then run down this list and select a POP which has not yet filled it's quota of hits. This

mode will try to send users to the closest POP, as long as the load distribution is even.

Quick Response to Requests.

When latency measurements are chosen, Global Dispatch has the ability to take immediate action and send an IP address back to the local DNS server with roundrobin without having to waiting for the latency measurements to be returned. This mode of operation is especially useful to provide a quick response to client requests while still enabling latency measurements to be factored into subsequent decisions.

The Scheduler quickly returns an IP address back to the client's local DNS server with round-robin while latency measurements are performed in the background and stored in the cache for future requests.

Shadow Mode: Testing Configuration Before Deploying.

This mode enables administrators to test the Global Dispatch configuration before putting it into actual production. In this mode, Global Dispatch software responds to DNS queries using a simple round-robin scheduling method. Simultaneously, it logs the results of "what-if" calculations to show which POP would have been selected had certain rules been activated. Shadow Mode helps administrators better understand site traffic and routing patterns at initial site setup.

Additionally, it provides a means to safely test adjustments that may be warranted because of fluctuations in site activity.

Integration with Authoritative DNS Server.

The Scheduler can be used in conjunction with the authoritative DNS servers either as a delegate to the authoritative server or it can act as a proxy to the authoritative DNS server to resolve A-records (common hostname request for IP address).

When Global Dispatch is a delegate of the authoritative server (figure 4), hostnames such as www.domain.com, ftp.domain.com, are delegated as subdomains to Global Dispatch. The delegation redirects the client's local DNS server from the Authoritative DNS server to the Global Dispatch Scheduler for subdomains.

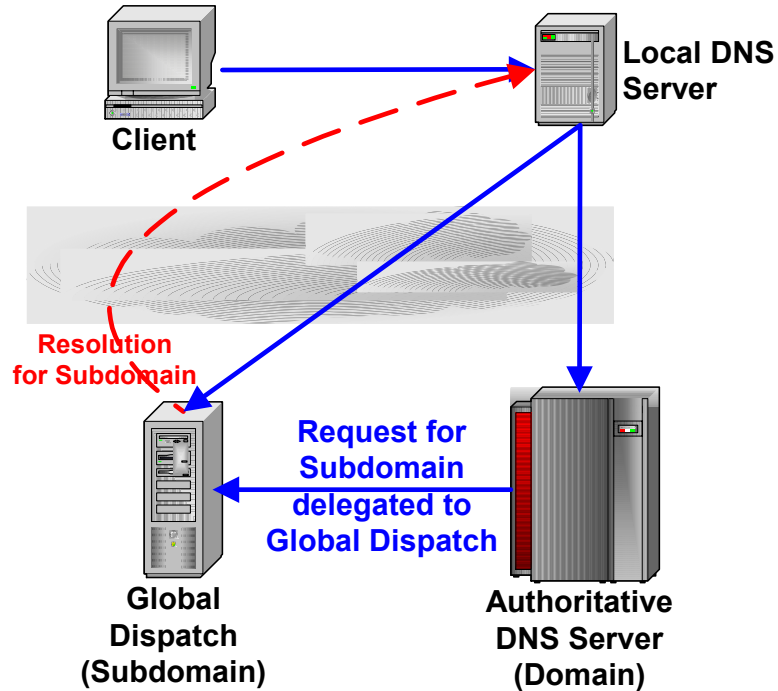


Figure 4 - Global Dispatch as a delegate to the Authoritative DNS Server

For instance, when `www.domain.com` is requested, the authoritative DNS server redirects, or delegates, `www.domain.com` to be resolved by the Global Dispatch Scheduler. The Scheduler then responds to the client's local DNS server with the appropriate IP address.

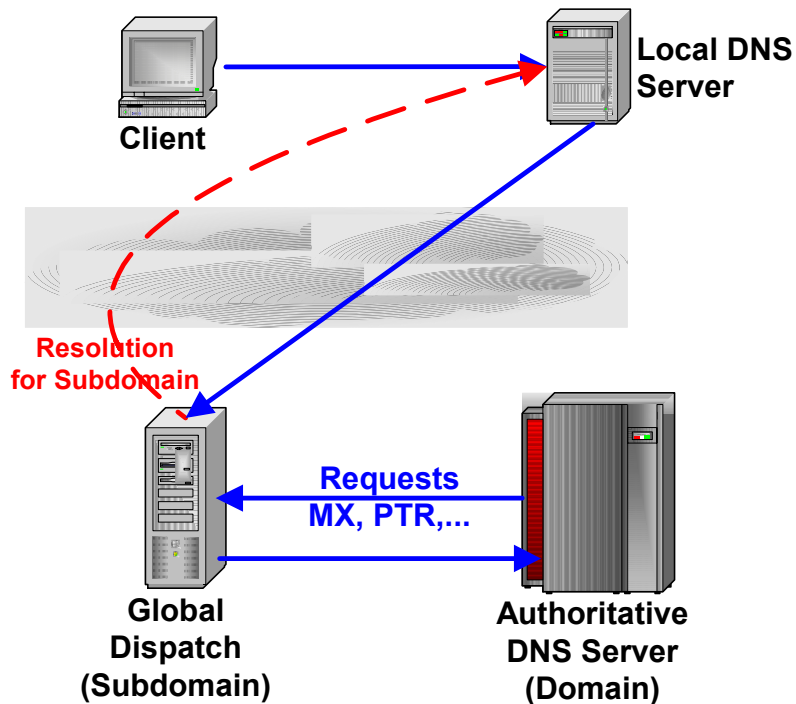


Figure 5 - Global Dispatch as a proxy to the authoritative DNS server

The authoritative DNS server will still handle the domain.com request, but will pass the sub.domain.com requests on to Global Dispatch. When the Scheduler acts as a proxy to the authoritative DNS server (figure 5), it can also handle requests for domain.com and it is treated as a peer to the authoritative DNS server.

It can resolve both domain.com requests and subdomain requests, such as www.domain.com. Any requests that cannot be resolved by Global Dispatch Scheduler in this mode will be requested from the authoritative DNS server and returned back to the Scheduler. This will include any non-A records.

The Scheduler is able to direct incoming requests from the client's local DNS server to the appropriate site based on predefined policies giving administrators ultimate flexibility in directing traffic and delivering predictable service levels. The Global Dispatch Scheduler, however, does not act alone. It gathers site information from the Global Dispatch agent to aid in its decision-making.

Gathering Valuable Site Metrics

Global Dispatch Agents

Global Dispatch Agents (item D in Figure 1) gather statistics from the physical sites to provide the Schedulers with site availability, load and response times to the client DNS servers. The agents are installed at each physical POP and are the Scheduler's point of contact for information about a POP. Agents can be loaded on any local server, which allows the Global Dispatch Scheduler to work with any local traffic management solution including Resonate Central Dispatch™ sites, hardware-based load balanced sites, and sites hosted on a single Web server. Each agent measures both the POP load and the latency between its POP and the client DNS server making the address resolution request.

The agent, independent of DNS requests, periodically communicates load information to the Scheduler; while latency information is measured and communicated when initiated by a DNS latency request. If the load measurements cannot be completed during successive tries, the Scheduler will mark the site as down and will direct traffic to other sites. In addition, a single POP can have multiple agents, each carrying a share of the triangulation work (latency measurement) for that POP. This configuration minimizes scheduling overhead in the largest Internet environments and also prevents agents from being a single point of failure.

Load Computation

Global Dispatch agents offer administrators tremendous flexibility in choosing the load calculation method that is most appropriate for

individual site environments. Agents determine site load in one of the following ways (shown in Figure 6):

- Connect.** Connects to the POP and measure how long it takes to establish the connection.
- URL-Connect.** Requests a specific file from the POP and measure how long it takes to begin receiving the file.
- Central Dispatch CPU Load or Open Connections.** Collects the average CPU load and open connections across a Central Dispatch cluster and reports the information back to the Scheduler.
- Custom Script.** Runs an administrator-defined script which determines the load. The ability to create a script to return load values is an extremely flexible method to collect load metrics from specialized applications and services. Administrators can develop custom scripts incorporating non-standard metrics, returning a load value ranging from 0 to 100, to influence Global Dispatch scheduling decisions.
- Load Server.** Connects to an administrator-defined load server process running on a remote host server to return a load value. The load server can run a server process to measure the load at the site and return a value back to the agent to influence the Global Dispatch scheduling decision.

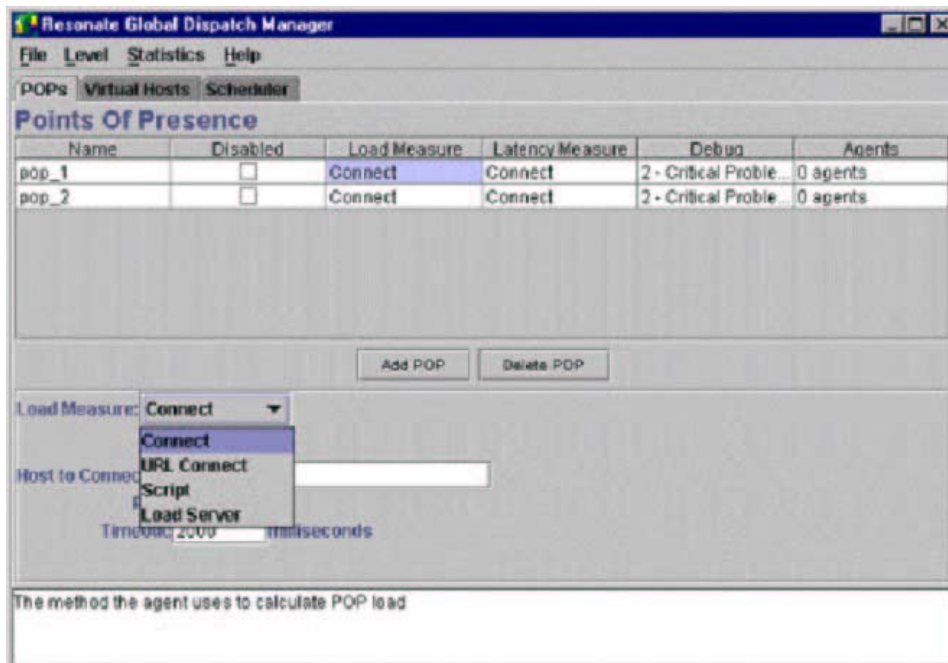


Figure 6 - Determining site load

Agents. Agents also provide the Scheduler with latency information by measuring the time it takes for them to connect to the client DNS server, and thereby supplying the response time between the client DNS and each Global Dispatch site. The agents use a variety of methods to gather latency between the physical site and the client DNS server. Each method can be used in different environments to deliver reliable latency measurement results (shown in Figure 7):

TCP Connect. With this method, the agent sends a TCP connection request to the client's local DNS server. The roundtrip time is measured as the latency.

ICMP Ping. Using the ICMP (Internet Control Message Protocol) method, the agent measures the time it takes to receive a response to a single ICMP request packet sent to the client DNS. Because only a single packet is exchanged, this method is faster and provides a more reliable measure of network latency (virtually eliminating CPU overhead)

Reverse DNS Lookup. With this option, the IP address of the requesting client DNS server is sent to the agents. The agents will then request a hostname lookup for that IP address from the client DNS. The time taken to get a response for this query will be measured as the latency. The Global Dispatch agents provide reliable and accurate site load and client response statistics that enable Global Dispatch to make intelligent decisions about where traffic should be directed. Using multiple agents provides failover and the ability to share the load across agents at large sites. Choosing these load and latency measurements can be easily accomplished in the Global Dispatch Manager.

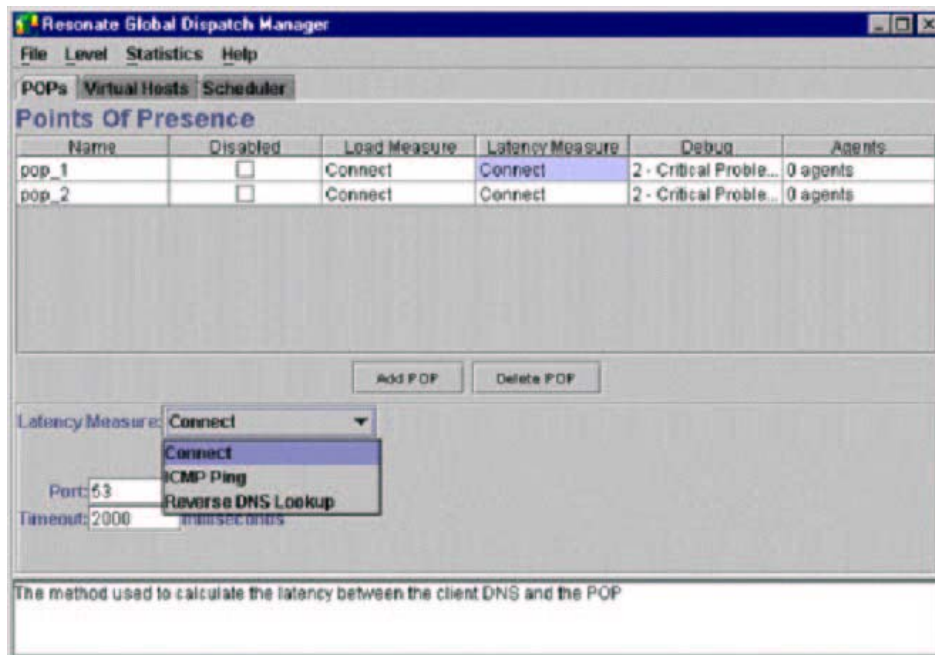


Figure 7 - Selecting latency measurement

Easy Configuration and Management

Global Dispatch Manager

Global Dispatch Manager (shown in figure 8) enables administrators to

quickly and easily set up, configure, and manage multiple, geographically distributed sites.

The Global Dispatch Manager is a graphical console that gives administrators a single view into their Global Dispatch sites and allows management from local or remote locations. This tool enables administrators to quickly bring up a redundant site with an easy-to-use interface. Once the Global Dispatch sites are up and running, statistics can be gathered through the Global Dispatch Manager to provide valuable usage information that helps administrators troubleshoot and fine-tune the site configuration.

The Global Dispatch Manager has been created to facilitate quick configuration of redundant sites through basic and advanced configuration screens. With error checking, pull down menus and simple check boxes, companies can ensure their multi-site configuration is created quickly and easily from the start. For additional flexibility, all configuration and commands in the Global Dispatch Manager can be executed through a command line interface.

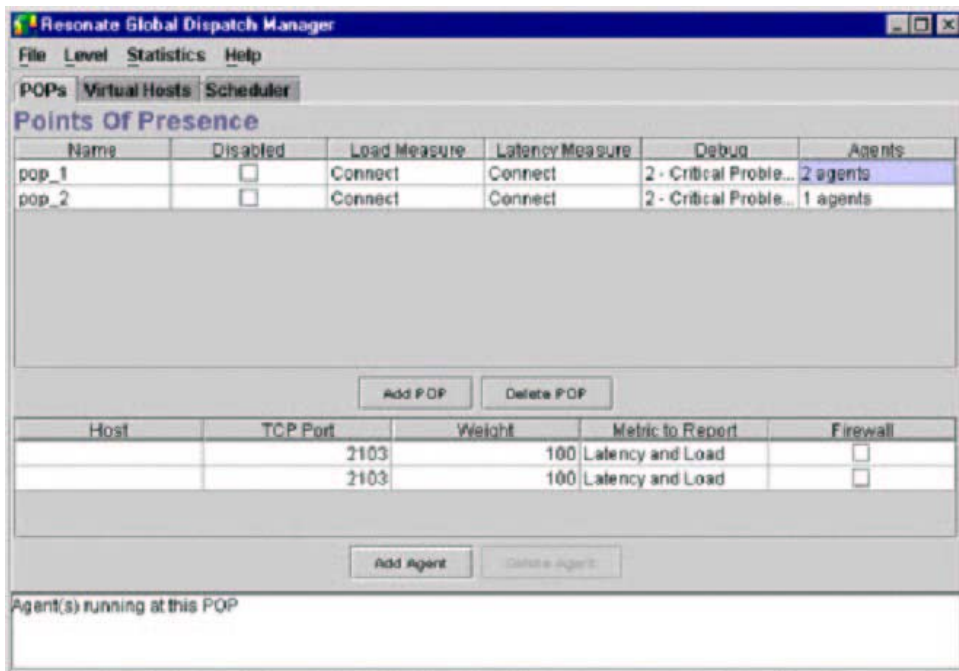


Figure 8 - Easy configuration with Global Dispatch Manager

The Global Dispatch Manager and command line utilities also provide site statistics to enable administrators to quickly troubleshoot problems and to proactively monitor traffic patterns. The Global Dispatch statistics program queries the Scheduler for information; it counts and refreshes statistics, and flushes the Scheduler cache as requested. To provide important system management information to administrators, Global Dispatch statistics describe the health of POPs, as well as traffic and system efficiency. Global Dispatch logs the results of all DNS requests including the source IP address of the DNS query, the physical POP chosen, and the IP address returned. The Global Dispatch Manager provides easy access to the following

available statistics (shown in Figure 9):

- Current size of the cache in bytes or number of entries.
- Number of hits and misses on the cache.
- Average time taken for Global Dispatch to respond to DNS requests.
- Number of requests for a particular host name.
- Number of times a particular POP was selected in response to a DNS request.
- Current status of all physical POPs.
- Number of times a latency measurement failed from each physical POP. Global Dispatch statistics can be continuously displayed to provide on-going site evaluation.

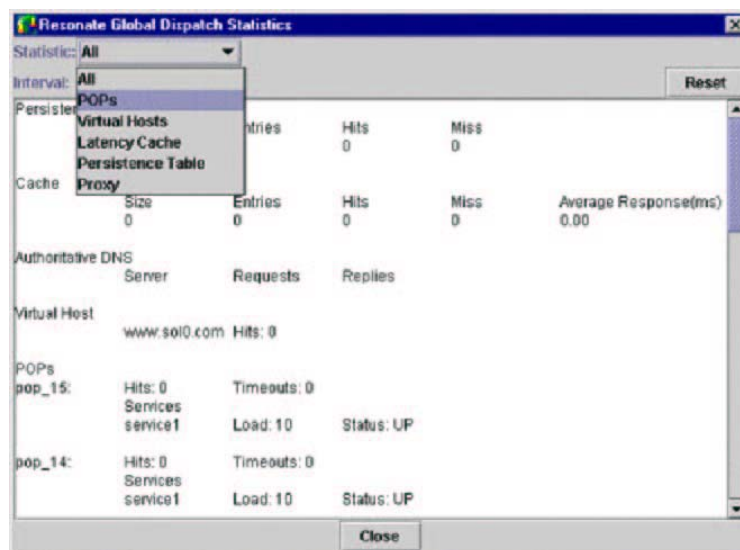


Figure 9 - Proactive management with Global Dispatch statistics

Bringing it all Together

All the Global Dispatch elements tie tightly together to provide the highest level of availability and performance to deliver the best user experience. When a client connects to a Web site, it must resolve the site name (e.g., www.resonate.com or resonate.com) to an IP address. This process is illustrated in Figure 10. The process steps include the following:

1. The client sends a request to its local DNS.
2. If the requested site information is not cached in the local DNS, the client's local DNS server sends a request to the requested site's authoritative DNS server (or Global Dispatch Scheduler if the Scheduler is acting as a proxy) through the DNS hierarchy. The authoritative DNS server examines the subdomain name requested and passes the IP address of the Global

Dispatch Scheduler back to the local DNS server. The local DNS server then communicates with the Global Dispatch Scheduler directly.

3. Global Dispatch determines if the scheduling information has been defined through the directed traffic table or persistence table. If these features are enabled, skip to 5.
4. The Global Dispatch Scheduler uses the load balancing policy configured (load, latency, latency and load, round-robin, or weighted round-robin) to determine the most appropriate site for the client. Load and latency metrics are captured from the Global Dispatch agents at each site and returned back to the Scheduler.
5. The IP address (or Virtual IP) of the appropriate POP is returned to the local DNS server. Multiple IP addresses are returned if this feature is enabled.
6. The local DNS server then passes this address (or addresses) back to the client.
7. Client traffic is then directed to the most available, appropriate POP.

Note: Global Dispatch also supports multiple virtual host names at each POP, allowing fine granularity in directing traffic to meet varying browser access patterns or unique server configurations.

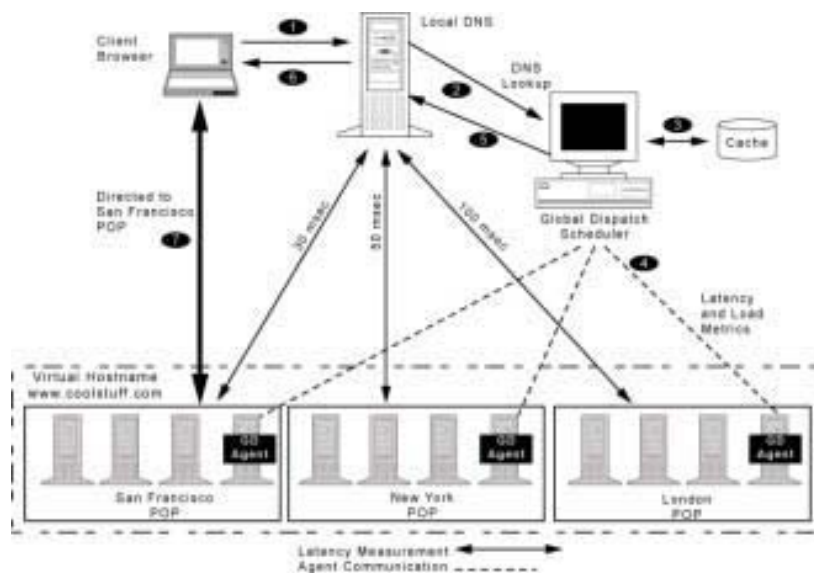


Figure 10 - Resolving a virtual hostname using latency and load.

Global Service Level Control

When a business is dependent on the successful access of data, a company can't afford an Internet site going down or slow response times. Resonate Global Dispatch provides a WAN-based, multi-site, service level control solution that provides high availability and

optimum performance for geographically distributed e-business applications. The Global Dispatch Scheduler enables intelligent management of globally distributed traffic to ensure traffic is always directed to the most available, responsive sites while the agents gather the valuable site metrics that help in the decision making process. And, lastly the Global Dispatch Manager provides a single interface to configure, monitor and proactively manage distributed sites. Resonate Global Dispatch is the only software solution that has been uniquely designed to provide high availability and scalability without compromising performance while still offering the greatest flexibility and control over geographically distributed traffic. Its flexible software solution layers onto existing infrastructure, including existing DNS and local server load balancing solutions, enabling companies to quickly and easily deploy a redundant site and ultimately deliver a great end user experience. Integrating with all of Resonate's products including Central Dispatch, Commander, and Enterprise Services Console, Resonate Global Dispatch ensures service levels can be easily controlled and maintained across multiple applications, services and geographies.

Additional Information

A variety of other product and technical information, including solutions documents and application notes, is also available. Visit the Resonate Web site at www.resonate.com or contact your Resonate sales representative to obtain further information.

Disclaimer

Resonate is a registered trademark, the Resonate logo, Keeping E-Business Open for Business, Resonate Central Dispatch, Resonate Global Dispatch, Resonate Commander, and Resonate Commander Solutions are trademarks of Resonate, Inc. All other trademarks are the property of their respective owners.
