Coyote Point Systems White Paper

Accelerate, Secure, and Customize Your Web Services:

Load Balancing for Application Server Administrators

http://www.coyotepoint.com

The idea of load balancing is well defined in the IT world: A network device accepts traffic on behalf of a group of servers, and distributes that traffic according to load balancing algorithms and the availability of the services that the servers provide. From network administrators to server administrators to application developers, this is a generally well understood concept.

The implementation of load balancing, however, is another matter. There are often many questions regarding how load balancing is deployed, how the servers are configured, and how the overall network architecture may need to change to accommodate load balancing appliances.

The good news is that deploying a load balancer needn't be perplexing or difficult. In fact, installing a Coyote Point Equalizer™ load balancer into an existing web server infrastructure can easily be done with minimal changes to your existing configuration. This document outlines how a fairly common web server installation can be outfitted with an Equalizer in five easy steps to provide load balancing with minimal changes to your network architecture using a simple “drop-in” deployment strategy. The five steps are:

- Preparation
- Configure Equalizers on the network
- Configure Virtual Clusters
- Configure Server Gateways
- Changeover DNS

And best of all, **you don't need to be a networking guru to install an Equalizer**. If you understand networking from a server perspective, then you've got the knowledge necessary to drop a Coyote Point Equalizer into a network and configure a fully load balanced environment. The following section explains how Equalizer can easily fit into your existing network.

The Network

The Coyote Point Equalizer series of load balancers are flexible, and can be implemented in a number of different network configurations, depending on your infrastructure needs. One such network implementation, known as single-network configuration, is a “drop-in” implementation.

The drop-in implementation has several advantages that make Equalizer particularly simple to implement. These advantages include:

- No need for additional subnets or physical networks
- The servers do not need to have their IP addresses changed
- Only one small change is needed on the servers to fully implement load balancing
- Works without changing existing network infrastructure
- Cut over is seamless -- does not interrupt site traffic, even if connections go to the old IP

To show how this drop-in can easily be added into an existing infrastructure, take as an example a very common installation illustrated here in Diagram 1

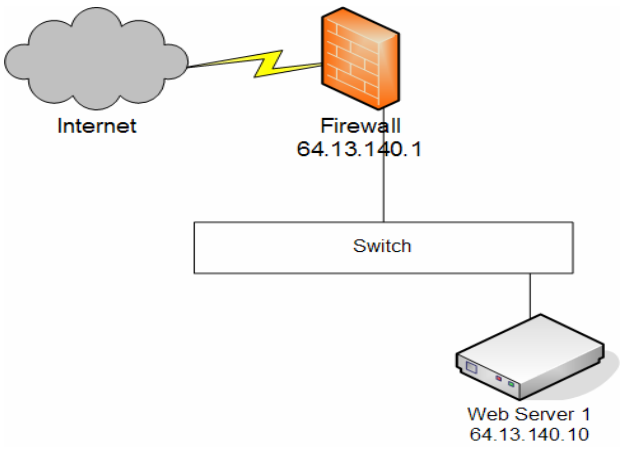

Diagram 1: Simple Web Serving Scenario

This is a very simple single-web server configuration: The domain name points in DNS to the web server with the IP address 64.13.140.10. The firewall, located at 64.13.140.1, acts as the default gateway for the single web server.

If your business depends on this web site, this is not an ideal situation. There's no redundancy in case the web server were to suffer a failure, and expanding the capacity would require either upgrading memory/processors for the system, or replacing it entirely with a more powerful system.

With a drop-in configuration, Equalizers sit on only one network, on one subnet, the same network and subnet that the web servers sit on. You don't need to add additional networks, change the IP addresses of your servers, or add any extra networking gear. The servers will still be accessible the same way they were before a load balancer was implemented.

Two Coyote Point Equalizer load balancers in a redundant configuration (Diagram 2) will be added, as well as two more web servers, bringing the number of web servers to three.

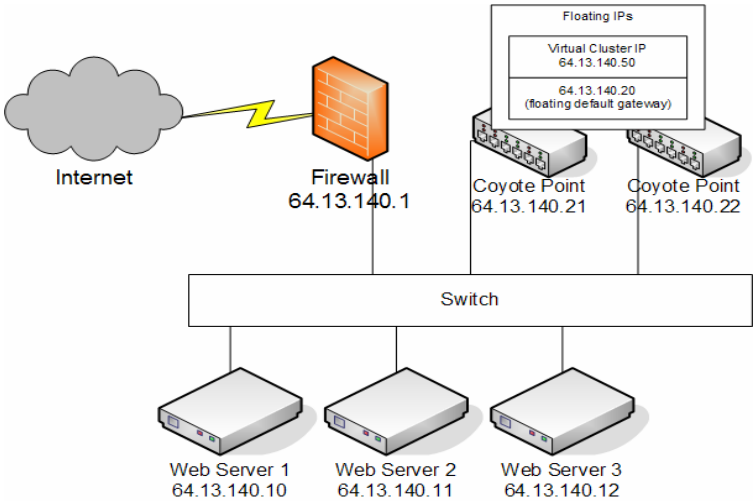

Diagram 2: New Load Balanced Environment

While the servers can still be individually accessed, all web traffic will be directed to a separate IP address, called a Virtual Cluster. The Virtual Cluster will accept traffic and distribute it to the available web servers. If the active Equalizer were to go off-line, the Virtual Cluster IP would automatically switch to the standby Equalizer.

Equalizer will also perform health checks on the three servers, ensuring that they are capable of serving up traffic. If one web server goes down, Equalizer stops sending traffic to that server and routes traffic to the remaining active servers. Once the server comes back up, Equalizer resumes sending traffic to it.

Each Equalizer has an individual IP address, which is used for management. In addition, both Equalizers share a floating IP address (the failover address). Like the Virtual Cluster address, the floating address exists only on the active Equalizer. This floating IP also serves as the default gateway for the web servers.

While the servers change their default gateway to the floating IP address, both Equalizers have their default gateway set to the firewall; the effect is that the outbound gateway for the entire configuration is still the firewall. The web servers have inbound and outbound Internet access just as they did before Equalizer was installed, and are limited only by the firewall's security profile. Equalizers need to be the default gateway to ensure traffic passes through Equalizer in both inbound and outbound directions.

Inbound traffic will be changed to use the Virtual Cluster IP Address instead of the server address previously used (64.13.140.10). This change will be made in DNS once installation is complete so that there is no interruption of service.

With this explanation of the network, the actual process of installation can begin.

Step 1: Preparation

There are a few preparatory steps that you'll want to take before implementation to ensure a successful deployment. First, you'll need four additional IP addresses on your network if you're running Equalizers in redundant mode, or two IP addresses if you're running a single Equalizer in stand-alone mode.

You'll also want to change the TTL (Time To Live) on your domain name (or names) to zero. This will make the cut-over from the single web server to the load balancer quicker. Your DNS provider (whoever shows up in a WHOIS for the domain, typically your ISP) should be able to accommodate this request.

Step 2: Get Equalizer on the network

The first step in getting Equalizer on the network is to set up the physical connection. You'll use Equalizer's Internal network interface ports (Diagram 3), and will disregard the External network port. One port from each Equalizer must be plugged into the same switch or hub infrastructure that the firewall (or upstream router) is plugged into. If you are using two Equalizers in a redundant configuration, also connect one internal port on one Equalizer to an internal port on the other Equalizer.

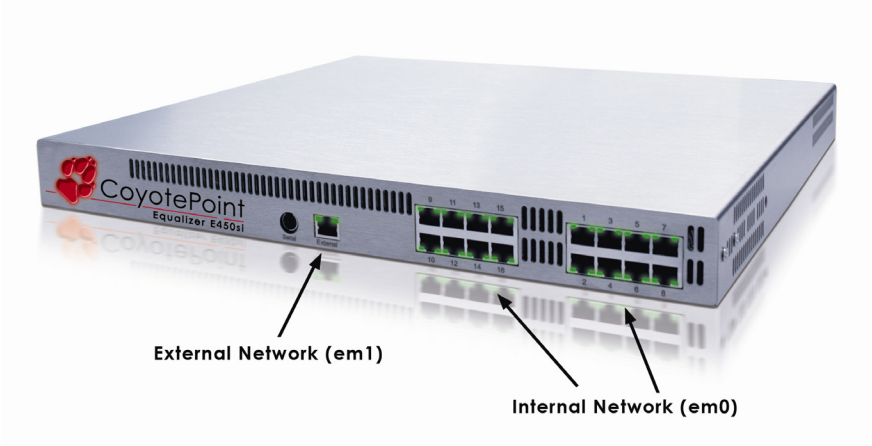

Diagram 3: Port Layout of an Equalizer

The servers can either be plugged directly into the Internal network ports, or they can plug into the same switch or hub infrastructure that each Equalizer plugs into.

Power up each Equalizer and give them IP addresses and hostnames. The IP addresses for eq1 and eq2 are assigned to the internal network cards on the Equalizers using the character line interface through Equalizer's serial port. Initial configuration of the box is done using the included serial cable and a serial terminal; or, a terminal emulator application (such as HyperTerminal, which is included with most versions of Microsoft Windows) can also be used. The installation manual describes other freely available terminal emulator options, and contains the information needed to make a connection to Equalizer.

Below is the IP scheme for this particular configuration. Again, three IP addresses (Table 1) will be needed in a redundant scenario.

Equalizer Hostname	*IP Address*
eq1	64.13.140.21
eq2	64.13.140.22
Floating	64.13.140.20

Table 1: IP Addressing for Equalizers

Once you assign IP addresses to the Equalizers, you should be able to reach them (using the **ping** command, for example) from other systems on the same subnet. You can now finish configuring Equalizer via a web browser over secure HTTP. If Equalizer's IP is 64.13.140.21, the URL would be https://64.13.140.21. This brings up Equalizer's Administration Interface. While the Administrative Interface works with any Java-enabled browser, we recommend using FireFox 2, Internet Explorer 6, or Internet Explorer 7.

Step 3: Configure a Cluster

The next step is to configure an Equalizer Cluster. The Cluster is what accepts connections on behalf of the servers. Eventually, the site's DNS entry will point to the Virtual Cluster IP.

From the main screen, you'll see an **Add** menu at the top. Scrolling over the word "Add" will bring up a sub-menu, with the option for Virtual Cluster (Diagram 4).

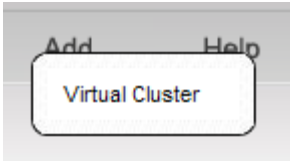

Diagram 4: Add Virtual Cluster

Click on this, and you'll be brought to the add cluster configuration screen (Diagram 5). There are many options, but the options that you're primarily interested in are the ones below.

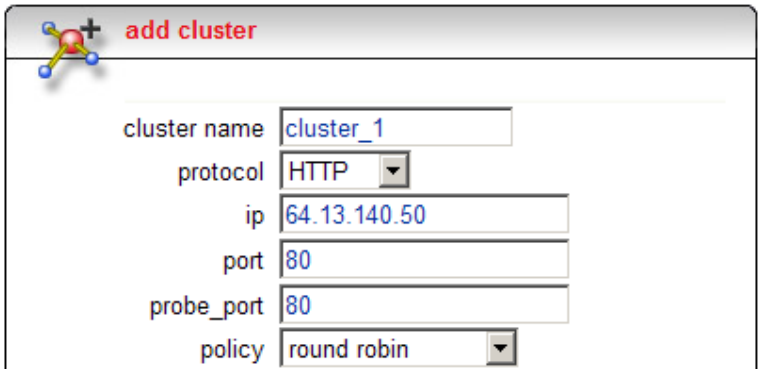

Diagram 5: Virtual Cluster Properties

In this example, the virtual cluster will be given the name **cluster_1**, and will be configured for the HTTP protocol on port 80 at the IP address shown. We'll leave the other cluster parameters set to their default values, and discuss a few of the more important ones briefly before we configure the servers in the cluster.

The default cluster load balancing policy is **round robin**, which distributes incoming requests to each server in the cluster one at a time, then loops back to the beginning of the list of servers. Other available load balancing policies include **static weight**, **adaptive**, **fastest response**, **least connections**, and **server agent**.

One important option in the virtual cluster configuration can be found by scrolling down into the **flags** section (Diagram 6). This option is called **spoof**, and this is what preserves the IP addresses of the client requests. Without it, all the servers in the cluster would see all inbound connections as coming from Equalizer's IP address. This is enabled by default, so you don't need to change anything.

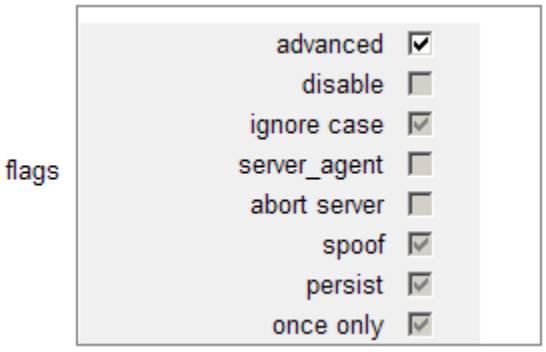

Diagram 6: Virtual Cluster Flags

Another important option configured by default is **persist**. When a Virtual Cluster is configured, persistence through active cookies is setup by default. This will keep a client tied to a specific server for the duration of their session. This is typically a requirement for interactive web sites. Even if it's not a requirement, persistence is generally benign, and will not adversely affect a site that doesn't require persistence.

Click on the **commit** button at the bottom of the screen to continue.

Add Servers to the Cluster

Now that the virtual cluster is configured, it's time to add the web servers to the cluster. In the cluster settings, mouse-over the **menu** link, which opens a menu with the **Add Server** option (Diagram 7).

Diagram 7: Add Server to Virtual Cluster

This will bring up the add server configuration screen. Give the server a name, add the IP address and the port, and click **commit** to add the server to the cluster (Diagram 8).

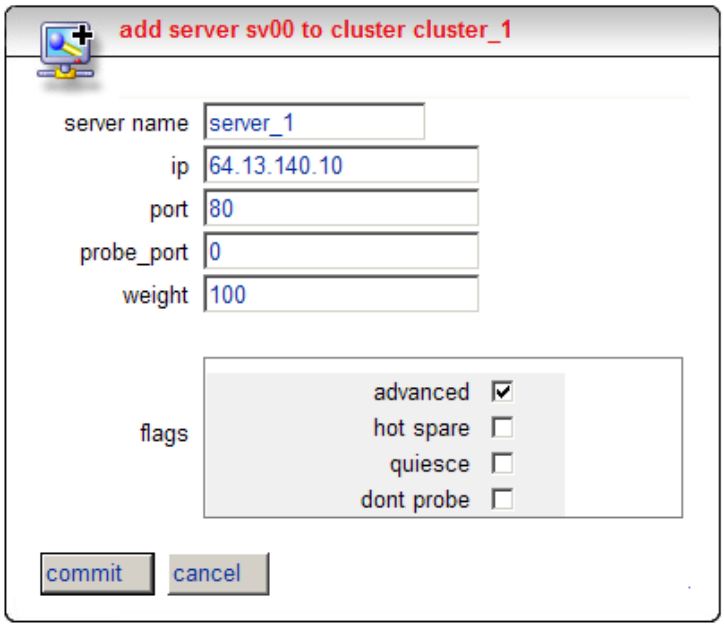

Diagram 8: Server Options

Repeat this with all your web servers.

Step 4: Configure Default Gateways on Servers and Test

There is only one configuration change that must be made on the server systems, and that's the default gateway. In Microsoft Windows, this setting can be found under the TCP/IP settings of a given network connection's control panel (Diagram 9).

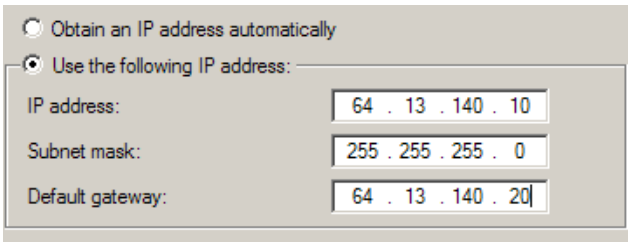

Diagram 9: Default Gateway Settings

The default gateway for all servers that are load balanced by Equalizer needs to be Equalizer's floating gateway IP address. All of your other settings, including the web server configuration, can remain unchanged. Equalizer will work seamlessly with this configuration.

Test Your Cluster

Now that your servers are configured, go back to the Equalizer Administration Interface and click on **View > Cluster Summary** in the main menu. A screen like the following is displayed:

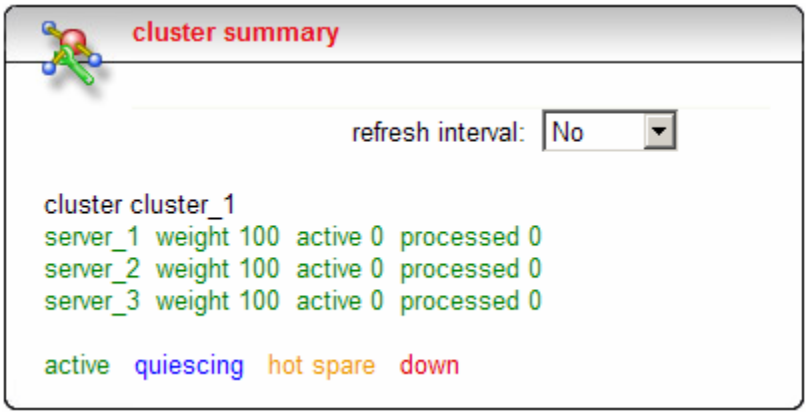

Diagram 10: Cluster Summary Screen

Click on the **refresh interval** drop down box, and select **1 min.** to refresh the numbers on the screen every minute.

Using one or more clients that are not on the same subnet as your web servers, open a browser and type the Virtual Cluster IP address into the location bar and you'll see the web site coming up on each client. You should see the numbers in the screen above changing as you hit the website from different client machines, as shown below.

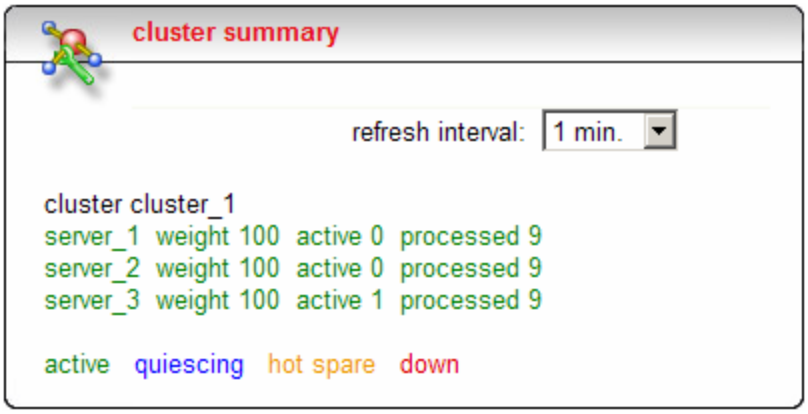

Diagram 11: Cluster Summary Screen during testing

(**Note:** if you don't have multiple clients available, then clear your browser's cookies after every web site access so you in effect open a new connection to the cluster every time you access it from the same client.)

Step 5: DNS Cut-over

With all the setup and configuration complete and tested, the site is now ready to switch over with a DNS cut-over.

Contact your DNS provider (again, typically your ISP) and have them switch the DNS for your site from the old IP (directly accessing the first web server) to the IP address of the Virtual Cluster on Equalizer. Also have them set the TTL for your DNS entry to 0.

Because DNS can (and will) be cached despite TTL, the effect will not be immediate for all clients. However, since both the old IP address and the Virtual Cluster are both serving up the website, this will not affect your overall site availability. Over a period of 24-48 hours, traffic will migrate to the Virtual Cluster without service interruption.

Drop In Simplicity

This solution can expand to many more web servers, Virtual Clusters, and serve multiple websites while providing advanced traffic management features such as persistence, health checking, and failover. In addition, global load balancing can be included with the Coyote Point Envoy solution, which provides failover and load distribution to multiple geographical locations.

With Coyote Point's Equalizer drop-in solution, load balancing needn't be difficult to implement or disruptive to your existing infrastructure. **With this easy to understand, easy to implement solution, there is no reason not to add the benefits of load balancing to your infrastructure today**.